Unraveling the Relationship between Education and Health: Genetic Controls, Heterogeneity across Sociodemographic Groups, and Variation across Biomarkers of Health Risk

Unraveling the Relationship between Education and Health: Genetic Controls, Heterogeneity across Sociodemographic Groups, and Variation across Biomarkers of Health Risk

A dissertation presented by

Meghan Zacher

to

The Department of Sociology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the Subject of

Sociology

Harvard University

Cambridge, Massachusetts

August 2019

Dissertation Advisor: Mary C. Waters Meghan Zacher

Unraveling the Relationship between Education and Health:
Genetic Controls, Heterogeneity across Sociodemographic Groups,
and Variation across Biomarkers of Health Risk

Abstract

Despite decades of research demonstrating better health among the higher educated, the causal effect of education on health is still debated. This is due in part to mixed evidence obtained in quasi-experimental work. These puzzling patterns could be explained by the influence of uncontrolled confounders in observational research, by effect heterogeneity across individuals or environments, or by variation in effects across manifestations of health. The empirical chapters of this dissertation draw motivation from these observations to further unravel the relationship between education and health among older adults in the United States.

First, I assess the utility of a novel control variable: a measure of genetic selection into education. Genetic selection is operationalized using a polygenic score (PGS) that predicts years of schooling based on many hundreds of thousands of genetic variants across the genome. Among European-ancestry respondents to the Health and Retirement Study (HRS) and the Wisconsin Longitudinal Study (WLS), I find that controlling for the PGS significantly attenuates the association between education and later health. The level of attenuation I observe is comparable to that obtained when controlling instead for measures of other known confounders, including family background and childhood health. Additional results suggest that the education PGS reflects more proximal confounders of the education-health link that may not be adequately controlled using survey measures alone. Crucially, however, the positive relationship between education and health is robust to this particular measure of genetic selection into years of schooling.

Next, I evaluate whether the association of education with health varies across sociodemographic groups defined by socioeconomic (SES) origin, race, and gender using data from the HRS. In so doing, I take a more complex intersectional perspective than has been used in prior work. This is important, as exposure to discrimination, which shapes opportunities to use resources in support of health, may depend on multiple sociodemographic characteristics simultaneously. Results underscore the importance of one intersection in particular: that between SES origin and race. In line with prior work, I find that the association of years of schooling with self-reported health is stronger for those from low-SES backgrounds; however, this is only the case among whites. Seen from the other angle, the association of education with self-reported health and mortality is weaker for blacks than for whites, but primarily among those from low-SES origins. For both self-reported health and mortality, I find the smallest gain in health per year of schooling among low-SES origin black men, the group with the highest risk of poor health and mortality overall.

In the final empirical chapter, I use data from the HRS to assess whether educational disparities in biomarkers of health risk vary across their distributions. Fundamental cause theory implies that such disparities will be largest where related resources can most successfully be leveraged to improve outcomes. For many biomarkers, this could be in the unhealthy tail of the distribution, where unequal access to and efficacy of medical interventions may exacerbate disparities. Consistent with this theory, I find that educational disparities in blood sugar and blood pressure are largest at their least healthy levels, precisely the points where impacts on subsequent morbidity and mortality are greatest. Meanwhile, high-density lipoprotein (HDL) or "good" cholesterol—a biomarker that is not regularly targeted by medication—does not display such a pattern. These results are not only of theoretical and substantive interest; they also provide methodological guidance for future work on biomarkers of health risk, which is timely given the recent proliferation of such measures in social science datasets.

Table of Contents

Acknowledgements

I am immensely grateful to the many people who enriched my graduate school experience and to those who enhanced the quality of this dissertation. First and foremost, my advisors—Mary Waters, Jason Beckfield, and Alexandra (Sasha) Killewald—have been constant sources of positive energy and encouraging feedback. Mary has often helped me see the big picture in a set of convoluted results; she has also extended numerous training and research opportunities my way over the years, for which I am very appreciative. Likewise, Jason helped me find my feet early on in graduate school by connecting me with potential collaborators; he has also provided much-needed insight into sociological perspectives on health. Last but not least, Sasha offered crucial feedback on countless drafts of this dissertation, resulting in substantial improvements with each iteration; she inspires me to be a better writer, a more rigorous analyst, and a more effective teacher. Collectively, my advisors' mentorship made this final, frenzied year of graduate school one of the most fulfilling.

I also acknowledge the Social Science Genetic Association Consortium, which has provided invaluable opportunities to learn, collaborate, and fund my research. In particular, I thank Daniel Benjamin, David Cesarini, Samantha Cherney, and Patrick Turley. Their instruction in social science genetic concepts and methods continues to pique my intellectual curiosity.

Next, I want to thank my graduate school cohort, from whom I have learned so much: Laura Adler, Kristina Brant, Robert Manduca, Roland Neil, Bo Yun Park, Nathan Robinson, Adam Travis, and Zach Wehrwein. Without them, problem sets would have been considerably less effective, not to mention markedly less enjoyable. I am also grateful to my dear friend and colleague, Robbee Wedow, whose suggestions have been extremely helpful in the writing of this dissertation and in navigating academia more generally.

Also deserving of recognition are the brilliant researchers who encouraged me to pursue a graduate degree in the first place. I owe a huge thank you to Carolyn Liebler and Rachel Magennis, whose offers of employment as a research assistant while an undergraduate at the University of Minnesota

allowed me to decline an unwanted internship many years ago. Thanks also to Hollie Nyseth Brehm and Joachim Savelsberg, who encouraged me to pursue independent research as an undergraduate while providing stellar examples of how to conduct engaged scholarship. Finally, I am deeply grateful to the Centre for Behavioural Research in Cancer (CBRC) in Melbourne, Australia, where I worked for several years prior to graduate school. Working at CBRC showed me that the impact of research does not necessarily end with publication; it has the potential to spur real policy and societal change. Warranting special mention are Sarah Durkin, Linda Hayes, Michelle Scollo, and Melanie Wakefield, who have been fantastic mentors over the years.

Much thanks is due to the Department of Sociology's administrative staff, including (but not limited to) Nancy Branco, Dotty Lukas, Jessica Matteson, and Suzanne Ogungbadero, who ensured both my timely academic progress and material support. I am also appreciative of the funding provided by the Harvard Graduate School of Arts and Sciences.

Finally, I owe a special debt of gratitude to those outside of academia. I thank my family and friends for providing many much-needed opportunities for distraction. I thank my parents, Tod and Kathy, for their unwavering support, for always being ready to celebrate, and for generously gifting their frequent flyer miles. And lastly, I thank my partner, Liam, for so many things. Thank you for reminding me that life should be prioritized in the work-life equation; thank you for organizing our social calendar; thank you for showing me how to use my phone. Most of all, thank you for moving across the world five years ago so that we could have this experience together. The next adventure awaits!

Chapter 1. Introduction

The distribution of health within and between populations has been the focus of scientific enquiry for centuries. Hippocrates (c. 460–370 B.C.) was the first to argue that the origin of disease was not supernatural, but instead rational, depending in part on climate and behavior (Lloyd 1983). The systematic study of health disparities was refined in nineteenth-century Europe, when improvements in national population statistics enabled the study of variation by district or occupation (Krieger 2011). Building on these early observations, epidemiologists, demographers, and sociologists have theorized that health reflects the biological embodiment of both natural and social environments encountered over the life course (Krieger 2001, 2011). Segments of the population that occupy distinct physical, structural, or cultural contexts are therefore expected to experience different patterns of health and mortality.

This dissertation focuses on health disparities across a dimension of population segmentation that is highly salient in the United States today: educational attainment. Prior research shows that educational disparities in health exist across time and space (Cutler and Lleras-Muney 2008; Hummer and Lariscy 2011). The magnitude of these disparities are often staggering. At age 25, college graduates in the U.S. are expected to live five years longer than those with a high school diploma and more than 11 years longer than those who did not complete high school (Rostron et al. 2010). Moreover, educational disparities in health have widened in recent decades (Goesling 2007; Liu and Hummer 2008; Masters et al. 2012; Meara et al. 2008; Montez et al. 2011).

Some have therefore advocated investment in education as public health policy (Cohen and Syme 2013; Galea et al. 2011; Hahn and Truman 2015). Woolf et al. (2007), for example, estimates that eliminating educational disparities would have averted eight times more deaths between 1996 and 2002 than the medical advancements that occurred over that period. This line of thinking assumes that schooling causally affects subsequent health, with effects occurring through some combination of economic, behavioral, and structural mechanisms (Cutler and Lleras-Muney 2008; Freese and Lutfey 2011; Link and Phelan 1995; Mirowsky and Ross 2003).

The causal impact of educational attainment on health is difficult to demonstrate empirically, however. Many factors influence both schooling choices and later health and thus could confound causal estimates in observational research (Cutler and Lleras-Muney 2008). Quasi-experimental studies—those that net out confounding bias using natural experiments such as policy changes or discordant education between twins—return mixed results (Galama et al. 2018; Grossman 2015). However, since their estimates apply to varying non-representative subsets of the population, mixed results may indicate not a null effect of education on health, but rather heterogeneity in effects across population subgroups. More generally, the effect of education is likely to vary across dimensions of health depending on disease etiology and the ability to prevent or treat adverse outcomes (Phelan et al. 2010).

Thus, there remains a need for research that (1) better measures and controls for potential confounders of the education-health link, (2) studies heterogeneity in the effect of education on health across populations or subgroups, and (3) investigates the effect of education on diverse manifestations of health. The three empirical chapters of this dissertation contribute to these gaps in the literature using data from the Health and Retirement Study (2016; RAND 2016), a representative survey of older U.S. adults that is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and conducted by the University of Michigan. Below, I briefly describe the prior work on education and health that motivates this dissertation. I also introduce the empirical chapters to follow, each of which aims to further unravel the relationship between education and health.

SOCIOLOGICAL PERSPECTIVES ON EDUCATIONAL ATTAINMENT AND ITS RELATIONSHIP WITH HEALTH

Weber (1978[1922]) highlighted that, as a key contributor to both social class and status, education is likely to have major implications for subsequent life chances. Indeed, contemporary work shows that more education is associated with higher earnings, higher occupational status and prestige, lower unemployment, and even a lower divorce rate (Hout 2012). As noted earlier, there is also a robust relationship between education and health, such that more educated people are, on average, healthier

(Cutler and Lleras-Muney 2008; Hummer and Lariscy 2011).

Social scientists have attempted to determine what characteristics select people into schooling trajectories for decades. The status attainment model emphasized childhood social class, as those whose parents have higher levels of schooling are likely to attain high levels themselves (Blau and Duncan 1967; Pfeffer and Hertel 2015). The Wisconsin Model argued that academic performance, aspirations, and encouragement also matter (Sewell et al. 1969). Scholars have since identified childhood health (Case et al. 2005; Jackson 2009; Palloni 2006) and cognitive and non-cognitive skills and personality (Bowles and Gintis 1976; Farkas 2003; Lleras 2008) as additional factors selecting people into educational outcomes.

Because educational attainment is not randomly assigned, it is difficult to evaluate the extent to which the association of schooling with subsequent outcomes, including health, reflects a causal effect. Nonetheless, there are several reasons to suspect that the education-health link is due, at least in part, to the effects of schooling. Schooling may improve health by enhancing economic prospects, thereby minimizing financial stress and enabling the procurement of nutritious food, safe housing, quality medical care, and other goods and services that support wellbeing (Hout 2012; Hummer and Lariscy 2011; Link and Phelan 1995; Lynch 2006). Education may also instill the skills and norms that are required to recognize and enact healthy behaviors throughout the life course (Cockerham 2005; Mirowsky and Ross 2003). Finally, education propels people into structural positions that may support health passively; for example, a town council in a wealthy neighborhood may pass strong restrictions on pollution, benefitting all local residents regardless of personal involvement or interest (Freese and Lutfey 2011).

ROADMAP FOR THIS DISSERTATION

Each of the three empirical chapters of this dissertation addresses one of the major issues confronting research on the relationship between education and health. These issues include confounding in observational research, heterogeneity across contexts and individuals, and variation across dimensions and distributions of health. In the sections that follow, I discuss each of these issues in turn while introducing the empirical chapters that they motivate.

Confounding in observational research

As mentioned earlier, social scientists have identified several characteristics that select people into educational outcomes. These include family socioeconomic status, childhood health, and abilities, skills, and personality (Blau and Duncan 1967; Bowles and Gintis 1976; Case et al. 2005; Farkas 2003; Jackson 2009; Lleras 2008; Palloni 2006; Pfeffer and Hertel 2015; Sewell et al. 1969). Many epidemiological, demographic, and sociological studies attempting to describe the effect of education on health therefore control for measures of these concepts in a standard regression framework.

Such studies have consistently shown that the positive relationship between educational attainment and health remains when holding measures of potential confounders constant (Cutler and Lleras-Muney 2008; Grossman 2015). For example, after controlling for an extensive list of variables reflecting childhood circumstances and health, Montez and Hayward (2014) find a strong independent association between educational attainment and active life expectancy. Further, though some have suggested that intelligence may explain socioeconomic disparities in health (Gottfredson 2004), research controlling for measures of cognitive performance shows, again and again, that a gradient remains (Batty et al. 2006; Conti et al. 2010; Link et al. 2008; Schnittker 2005; Zheng 2017).

Critics argue, however, that even extensive lists of controls likely leave confounding influences unaccounted for in observational research. A key confounder may be entirely unmeasured in the data at hand and thus cannot be controlled. When a dataset does contain some measure of a potential confounder, the resulting operationalization is often incomplete or inaccurate such that the confounding concept continues to bias causal estimates. Consider, for example, that family background is often measured using parental education; at best, measures of family income or parental occupation are also available, but often as static measures that do not identify instability in socioeconomic conditions across childhood. These variables may also fail to capture additional confounding influences related to duration and quality of family interactions. Even in the unlikely scenario in which all known confounders are measured perfectly, unknown confounders may continue to generate spurious results.

In sum, control variables feature heavily in sociological research on educational disparities in

health, and additional or better measures of potential confounders are always welcome. In Chapter 2 (*Polygenic Scores as Controls for Genetic Selection into Education in Models of Health*), I assess the utility of an innovative control variable that reflects genetic selection into schooling. Genetic selection is operationalized using a polygenic score (PGS) that predicts a person's years of schooling based on the estimated effects of many hundreds of thousands of genetic variants across their genome (Dudbridge 2013; Lee et al. 2018). PGSs may be useful control variables (Cesarini and Visscher 2017; Conley 2016; Freese 2018) if they are correlated with more proximal confounders of the education-health link—family background, childhood health, abilities and skills or personality traits—especially those that are unmeasured, measured poorly, or that remain unknown.

Heterogeneity across contexts and individuals

Quasi-experimental research gets around confounding by isolating exogenous variation in education. Such studies provide mixed support for a causal effect of education on health and mortality (Galama et al. 2018; Grossman 2015). In a well-known instrumental variable (IV) study, for example, Lleras-Muney (2005) finds that one additional year of schooling induced by changes in compulsory schooling laws in the U.S. reduced 10-year mortality rates by over six percentage points (see also, Fletcher 2015). Similar studies from European countries, however, return null results (Albouy and Lequien 2009; Braakmann 2011; Clark and Royer 2013; Johnston et al. 2015). In another type of quasi-experimental study, researchers compare the health outcomes of identical twins with discordant education (Kohler et al. 2011). As with IV studies, results vary (Amin et al. 2015; Behrman et al. 2011; Fujiwara and Kawachi 2009; Lundborg 2013; Lundborg et al. 2016; Madsen et al. 2010).

Importantly, effects estimated in such studies are based on non-representative subsets of the population. In studies instrumenting education with changes in compulsory schooling laws, estimated effects reflect those who were compelled to attain more education than they would have without the policy change. And, twin study estimates reflect effects among twins with different levels of education, or only about one-third of all twin pairs (Boardman and Fletcher 2015; Lundborg 2013). Thus, evidence

against a causal effect of education on health in quasi-experimental research does not necessarily confirm there is no effect overall. Instead, inconclusive findings may reflect effect heterogeneity. As such, there is now a push to examine not *whether* education affects health, but *under what contexts and conditions and for which groups of individuals* this effect is largest and smallest (Montez and Friedman 2015).

In Chapter 3 (*The Association of Education with Health and Mortality by Socioeconomic Origin, Race, and Gender)*, I focus on heterogeneity in the relationship between education and health across sociodemographic groups. In so doing, I evaluate two theories purporting that, while education helps people accrue resources that are important for health, some individuals have useful resources to draw on whether or not they achieve high levels of education (Ross and Mirowsky 2006, 2010). The theory of resource substitution holds that people with the fewest alternative resources are thus likely to benefit most from the resources accrued through education. Alternatively, under resource multiplication, those with access to alternative resources stand the most to gain from education.

Prior work has evaluated support for these theories by comparing the association of education with health across sociodemographic groups defined in terms of socioeconomic origin (Ross and Mirowsky 2011) or gender (Ross and Mirowsky 2006, 2010; Ross et al. 2012). I contend that a more complex intersectional perspective could shed light on additional sources of heterogeneity, as both access to alternative resources and the ability to use one's education in support of health can be impacted by discrimination, exposure to which may depend on joint group membership. Thus, I evaluate whether the relationship of education with health differs jointly by socioeconomic origin, race, and gender.

Variation across dimensions and distributions of health

The effects of education on health are likely to vary across health conditions. The theory of fundamental causation, for example, implies that health disparities by socioeconomic status (SES)—of which education is a key component—depend on sociohistorical factors including extant knowledge regarding the ability to prevent, treat, or cure disease (Link and Phelan 1995; Phelan et al. 2010). Empirical work supports this hypothesis, as SES disparities in health are greatest when prevention and

treatment strategies exist but are not universally accessible (Chang and Lauderdale 2009; Clouston et al. 2016; Glied and Lleras-Muney 2008; Masters et al. 2015; Phelan et al. 2004; Phelan and Link 2005; Tehranifar et al. 2009). It is only for these health conditions that resources associated with SES can be marshaled to improve outcomes.

The measures of health traditionally available to sociologists are generally limited to self-reports and mortality. Self-reports capture several important manifestations of health, including perceived overall health, difficulties with physical functioning, symptoms of suboptimal mental health, and past medical events. However, symptoms must cross some threshold of severity to be reflected in self-reports, making it difficult to study health in younger populations or to investigate the emergence of health issues across the life course. And, many self-reports are subjective. Not only is it difficult to know what factors lead people to perceive their health in a certain way, the considerations that go into perceptions of health may vary across the same population subgroups that are the subject of health disparities research. For example, self-rated overall health is more strongly related to mortality for those with higher education and income (Dowd and Zajacova 2007) and high-SES individuals judge their health more harshly, relative to objective measures, than those of low-SES (Dowd and Zajacova 2010).

Mortality also has limitations as a measure of health. All-cause mortality merges together many diverse conditions and events that lead to death, each of which may evolve through unique social and biological mechanisms. Cause-specific mortality also presents complications, requiring very large samples for sufficient statistical power.

In recent years, physical measurements and biological specimens have been collected of respondents to large social surveys in the U.S. and elsewhere. Using these data, biomarkers of health or health-related risk can be constructed (Harris and Schorpp 2018; McDade et al. 2007). Biomarkers include blood sugar, blood pressure, different forms of cholesterol, and more. Each of these new measures is objective as well as continuous, reflecting a range of health or health-related risk.

Chapter 4 (*Unconditional Quantile Regression and Educational Disparities in Biomarkers of Health Risk*) uses these novel measures to assess not only whether the association of education with

health varies across dimensions of health, but also whether the magnitude of educational disparities varies across their distributions. This chapter contributes to research attempting to map the shape of the relationship between education and health (Montez et al. 2012). It also tests an implication of fundamental cause theory (Link and Phelan 1995; Phelan et al. 2010) that suggests that disparities will be greatest at points in the distribution of health at which medical interventions may be triggered to improve outcomes. Finally, results have methodological implications for future research using biomarkers.

SUMMARY

In sum, each of the three empirical chapters of this dissertation addresses an issue that continues to complicate our understanding of the relationship between education and health. Chapter 2 assesses the utility of a new and innovative control for genetic selection into education in observational research. Chapter 3 studies heterogeneity in the association of education with health across sociodemographic groups using a more complex intersectional approach than has been taken in prior research. Finally, Chapter 4 assesses educational disparities in several objective and continuous biomarkers of health risk, with particular attention given to variation in disparities across the distributions of these measures.

Unraveling the relationship between education and health is an important and timely goal, as educational disparities are robust across time and space (Cutler and Lleras-Muney 2008; Hummer and Lariscy 2011), and in the U.S., appear to be widening (Goesling 2007; Liu and Hummer 2008; Masters et al. 2012; Meara et al. 2008; Montez et al. 2011). In Chapter 5 (*Conclusion*), I discuss the overarching contribution of this dissertation to the literature on educational disparities in health as well as emerging directions in this research area. In particular, I reflect on issues related to the incorporation of genetic data, the push to study heterogeneity, and the study of biomarkers of health-related risk.

Chapter 2: Polygenic Scores as Controls for Genetic Selection into Education in Models of Health

Despite the robust positive association between education and health, education's causal effect on health remains contested. Many factors, including family background, childhood health, and skills or personality, may influence both schooling choices and later health. Though the educational gradient in health persists after accounting for these confounders (Batty et al. 2006; Conti et al. 2010; Cutler and Lleras-Muney 2008; Link et al. 2008; Montez and Hayward 2014; Schnittker 2005; Zheng 2017), extant studies leave a unique aspect of selection uncontrolled: a person's DNA.

DNA influences both educational outcomes and health, albeit in complex ways that are only partially understood (Collins et al. 2003; Heath et al. 1985; Lee et al. 2018). One thing that is clear is that the effects of DNA on education and health are not independent: genetic effects on educational attainment are correlated with those on a variety of health outcomes and behaviors (Bulik-Sullivan et al. 2015; Wedow et al. 2018). Genetics may therefore confound the effect of education on health in observational research. In the current study, I examine the extent to which this is the case.

This question could not be answered empirically until recently, with the development of variables known as polygenic scores (PGSs) (Dudbridge 2013). PGSs sum up the estimated effects of a person's genetic variants on an outcome and therefore may be considered measures of genetic selection. I draw on a PGS that predicts educational attainment; it was constructed based on a sample of over 1.1 million people and explains around 11% of variation in years of schooling in representative samples of European ancestry individuals in the United States (Lee et al. 2018).

Genetic effects on social and behavioral outcomes like educational attainment do not occur solely within the body; they must also be driven or enabled by the environment. As such, they may evolve through the development of traits and behaviors that lead to self-selection into, environmental pressures to follow, or structural barriers precluding particular educational trajectories. To date, scholars have considered that the education PGS might correlate with or influence familial factors, childhood health,

cognitive and non-cognitive skills, and personality traits (Belsky et al. 2016; Belsky et al. 2018; Conley et al. 2015; Domingue et al. 2015), all of which are associated with educational outcomes (Blau and Duncan 1967; Case et al. 2005; Farkas 2003; Fletcher and Lehrer 2011; Jackson 2009; Lleras 2008; Palloni 2006; Pfeffer and Hertel 2015; Sewell et al. 1969).

If genetic effects on education are all environmentally mediated, why not instead control for characteristics relevant to the more proximal social processes selecting people into schooling? This is what prior research has attempted by controlling for long lists of potential confounders (Batty et al. 2006; Conti et al. 2010; Cutler and Lleras-Muney 2008; Link et al. 2008; Montez and Hayward 2014; Schnittker 2005; Zheng 2017). However, confounding constructs, such as family background, childhood health, and early skills or personality, are often incompletely measured, if measured at all, in the survey and administrative data that social scientists rely on. The education PGS may also correlate with traits we are not yet aware confound the association of education with health, traits that would otherwise be entirely uncontrolled.

It has therefore been suggested that PGSs have potential for reducing bias in observational research (Cesarini and Visscher 2017; Conley 2016; Freese 2018). To my knowledge, this possibility has not yet been tested for the association of education with health. Accordingly, I assess the extent to which controlling for the education PGS attenuates the estimated effect of education on several dimensions of health using two complementary U.S. datasets of older individuals of European ancestries.

EDUCATION AND HEALTH

As a key contributor to both social class and status (Weber 1978[1922]), educational attainment is expected to have major implications for subsequent life chances (Hout 2012; Torche 2011). The current study takes interest in the effect of education on health. Educational disparities in health are found across time periods and societies (Cutler and Lleras-Muney 2008; Hummer and Lariscy 2011). In the U.S., high school graduates are expected to live between six and seven years longer, and college graduates over 11 years longer, than those who did not finish high school (Rostron et al. 2010).

Education may affect health for several reasons (Cutler and Lleras-Muney 2008; Link and Phelan 1995). It promotes economic wellbeing (Hout 2012), which enables people to access health-promoting goods and services, including quality health care. Education may also impart the norms, knowledge, and ability to maintain a healthy lifestyle (Mirowsky and Ross 2003). And, those with higher education often find themselves embedded in structural positions that confer health advantages without purposive action (Freese and Lutfey 2011).

Nonetheless, studies utilizing natural experiments have produced mixed evidence for a causal effect of education on health and mortality (Galama et al. 2018; Grossman 2015). Several such studies instrument education using changes in compulsory schooling laws. For example, Lleras-Muney (2005) estimates that one additional year of schooling reduced 10-year mortality rates in the U.S. by over six percentage points. Similar research using other policy reforms, however, has returned null results (e.g., Albouy and Lequien 2009; Clark and Royer 2013). Such mixed results should, perhaps, be unsurprising, as estimates from instrumental variable designs are based on varying non-representative subsets of the population and reflect the effects of particular educational transitions that differ from study to study.

Thus, there remains a need to investigate the effect of education on health in population-representative data, which is typically survey-based and observational. Of course, as far back as Durkheim (1956[1922]), sociologists have noted that people are not selected into educational trajectories at random. A major challenge for observational work is therefore to identify, measure, and adjust for those aspects of selection that may confound the education-health link. The confounders most often considered include family background and socioeconomic status (SES), childhood health, and adolescent abilities or skills.

Each of these factors is associated with, and causally prior to, both educational and adult health outcomes. For example, family background influences educational outcomes (Blau and Duncan 1967; Pfeffer and Hertel 2015) and those from high-SES households have better health in adulthood, on average (Ben-Shlomo and Kuh 2002; Galobardes et al. 2004; Montez and Hayward 2011). Evidence also suggests that early life experiences of poor health limit educational attainment (Case et al. 2005; Fletcher and

Lehrer 2011; Jackson 2009; Palloni 2006) and are associated with worsened health later in the life course (Case et al. 2005; Haas 2007; Lam et al. 2019). Similarly, early life cognitive performance and non-cognitive skills or personality predict both improved educational outcomes (Farkas 2003; Lleras 2008; Sewell et al. 1969) and better subsequent health (Conti et al. 2010; Gottfredson and Deary 2004; Hauser and Palloni 2011). Some of the positive association between educational attainment and adult health could therefore be due not to a protective causal effect of schooling, but to the fact that socially advantaged, healthier, and smarter or more conscientious kids stay in school longer, on average, and are set up for better health starting in childhood.

When holding measures of these confounders constant, there remains an independent positive association of education with health (Batty et al. 2006; Conti et al. 2010; Cutler and Lleras-Muney 2008; Link et al. 2008; Montez and Hayward 2014; Schnittker 2005; Zheng 2017), suggesting a causal effect of education. However, measures available in the survey and administrative data researchers rely on may not adequately reflect confounding concepts. Frequently, a key confounder is entirely unmeasured in the data at hand and thus cannot be controlled. For example, many surveys contain no measures of cognitive ability and non-cognitive skills or personality, particularly in adolescence.

Even when datasets contain a measure of a potential confounder, the resulting operationalization is often incomplete. While most surveys inquire about parental education, for instance, parental occupation and income are often left out, or at best included as static measures. Moreover, people may report their parental education, occupation, and income inaccurately. Similar issues exist with respect to childhood health, which is often measured using a single subjective and retrospective assessment. And, when early abilities or skills are measured, they are typically based on test scores, grades, or teacher assessments, which may fail to capture relevant talents and behaviors. Some of these measures are also subjective and may be influenced by teacher bias (Downey and Pribesh 2004; Riegle-Crumb and Humphries 2012).

Confounding concepts may continue to bias causal estimates when associated measures are controlled but operationalized with error. And, additional confounders likely exist that researchers remain

unaware of. As I describe below, controlling for education PGSs may therefore reduce remaining bias in education's estimated effect on health in observational work.

GENETIC SELECTION INTO EDUCATION

Studies of twins suggest that up to 40% of variation in educational attainment is traceable to genetics (Branigan et al. 2013; Heath et al. 1985). Until recently, this was virtually all that was known about genetic selection into education. The effects of specific genetic variants, and the mechanisms that drove them, were a black box.

Genome-wide association studies (GWAS) estimate relationships between an outcome of interest and millions of genetic variants, or segments of DNA (McCarthy et al. 2008; Visscher et al. 2017). They demonstrate that traits like education are influenced by a large number of genetic variants, each of which has a very small effect (Chabris et al. 2015; Rietveld et al. 2014). For example, the most recent GWAS of educational attainment found 1,271 genetic variants were significantly associated with years of schooling (Lee et al. 2018). Among the variants demonstrating significant associations, the median effect size was small, predicting a difference of just 1.7 weeks of schooling.

For outcomes like education, it therefore makes sense to aggregate genetic effects from across the genome rather than focusing on the effects of individual genetic variants. Polygenic scores (PGSs) do just that, predicting a trait by adding up the effects—as estimated in an independent GWAS—of many hundreds of thousands of a person's genetic variants (Dudbridge 2013). A PGS constructed using the most recent GWAS of education explains 11% of variation in years of schooling in representative U.S. datasets of European-ancestry individuals (Lee et al. 2018). By comparison, maternal education—one of the strongest social predictors of educational outcomes—explains 14% (Lee et al. 2018).[1]

I refer to the education PGS as a measure of genetic selection into education. This is not to imply

[1] These figures represent incremental R^2s. They are calculated by comparing the predictive power of models controlling only for sex, year of birth, their interaction, and measures of genetic ancestry to those that add either the education PGS or maternal education, respectively. The incremental R^2 obtained by adding the PGS to a model that already controls for sex, year of birth, their interaction, measures of genetic ancestry, and maternal education is smaller but nonetheless substantial, at around 7% (Lee et al. 2018).

that the effects that comprise the PGS evolve through genetic mechanisms alone. On the contrary, genetic effects on non-biological outcomes like education cannot possibly evolve through purely biological mechanisms; they must be socially mediated and exist only insofar as they relate to institutional processes or norms that structure access to such outcomes (Jencks 1980). For example, in a society that forbids women to attend school, having two X-chromosomes would completely determine the educational trajectories of half of the population. However, this genetic effect would exist not for biological reasons, but rather because of the social environment.

Genetic effects on education may thus occur due to correlation with or the development of characteristics that lead to self-selection into, environmental pressures to follow, or structural barriers that preclude particular educational trajectories. Scholars have considered, for example, whether the education PGS reflects social advantage, childhood health, and early skills or personality traits. Research suggests that it does.

Specifically, the education PGS is modestly positively correlated with family SES[2] and thus its relationship with educational outcomes arises partially due to factors related to family background (Belsky et al. 2016; Belsky et al. 2018; Conley et al. 2015; Domingue et al. 2015). The education PGS is also positively correlated with cognitive performance (e.g., test scores) in adolescence, which is the single largest mediator of the relationship between the education PGS and years of schooling (Belsky et al. 2016; Domingue et al. 2015). Non-cognitive skills and related personality traits—self-control, sociability, and openness to experience—have likewise been shown to mediate this relationship (Belsky et al. 2016; Okbay et al. 2016; see also, Krapohl et al. 2014). And, while prior research has not found support for the notion that childhood health mediates the association between the education PGS and years of schooling (Belsky et al. 2016; see also, Krapohl et al. 2014), further investigation is warranted as results may depend on the measures of childhood health employed.

[2] Why does this correlation exist? DNA affects an individual's education and their subsequent socioeconomic outcomes. By extension, we can assume that an individual's parents' DNA affected the socioeconomic environment in which that individual grew up. The process of reproduction also guarantees that parents pass their DNA down to their offspring. This leads to positive correlations between a person's education PGS, their parents' PGSs, and their family SES.

In sum, the education PGS is correlated with several characteristics, including family background, early life abilities and skills or personality, and (maybe) childhood health. These characteristics are all known to select people into schooling, as described earlier (Blau and Duncan 1967; Case et al. 2005; Farkas 2003; Fletcher and Lehrer 2011; Jackson 2009; Lleras 2008; Palloni 2006; Pfeffer and Hertel 2015; Sewell et al. 1969). Moreover, nearly half of the relationship between the education PGS and years of schooling remains unexplained by these measures (Belsky et al. 2016). Thus, the PGS may be associated with myriad additional selection-related characteristics, perhaps including some that are rarely studied or controlled in observational research.

The education PGS as a control variable

The processes described in the previous section are potentially consequential for estimating the effect of education on health. Measured characteristics have so far been insufficient to account for all the mediating pathways linking the education PGS to educational attainment. In that sense, the education PGS is currently akin to a propensity score for which we do not know what measures of selection were included in the model.

This may be its strength as a control variable, as measured traits may likewise be insufficient to control for all the confounding pathways linking educational attainment to subsequent health. Selection into social environments—including years of schooling—is notoriously difficult to explain with measured characteristics alone. The education PGS may be well placed to attenuate the confounding that remains, as it may be correlated with or influence confounding concepts that are measured poorly, entirely unmeasured, or not yet recognized as confounders. In fact, some have argued that DNA directly or indirectly influences virtually all traits to some extent (Turkheimer 2000). Because of this, scholars have reasoned that PGSs may be valuable as control variables in observational research settings (Cesarini and Visscher 2017; Conley 2016; Freese 2018).

Roadmap for the current study

The statistical models I estimate demonstrate the utility of the education PGS as a control variable when investigating the effect of education on health under various circumstances. Initial models assume that nothing is known about a respondent's family background, childhood health, or adolescent abilities and skills, and thus only demographic confounders can be controlled using survey measures. The next set of models reflects the common case in which survey measures of family background and health in childhood exist and can be controlled, but no indicators of adolescent abilities or skills are available. The final models correspond to the oft-celebrated scenario in which adolescent cognitive performance is also measured and can be controlled.

I expect the proportional attenuation of education's effect on health following the inclusion of the education PGS to decline across these sets of models, as the more proximal confounders that the PGS may reflect are controlled. Nonetheless, I expect the education PGS to reduce the association of education with health to some extent in all models. Even in the best-case scenario, the education PGS may continue to reflect and thus adjust for poorly measured and unknown confounders of the education-health link.

METHODS

Data

I use two complementary datasets: the Wisconsin Longitudinal Study (WLS) and the Health and Retirement Study (HRS). Both are surveys of older U.S. adults. As described below, the WLS is limited in terms of geographic scope and range of educational outcomes, while the HRS is nationally representative. However, the WLS contains more detailed measures from adolescence. It also includes a sibling sample, allowing for analyses using family fixed effects.

Wisconsin Longitudinal Study (WLS)

The WLS began in 1957 when Wisconsin students in their final (senior) year of high school were surveyed about their plans for the future (Herd et al. 2014). In 1964, one-third of those who went on to

graduate were enrolled in the WLS (n = 10,317). This sample is roughly representative of non-Hispanic white American men and women with at least a high school education[3] born between 1938 and 1940. In 1993, the WLS expanded to include a randomly selected sibling from each respondent. Respondents were most recently surveyed in 2010-2012 when original graduates were around 70 years old.

The WLS began collecting saliva samples for genotyping in 2007. Genetic data is now available for roughly 9,000 graduates and siblings,[4] and the PGS I use has been constructed for 8,509 European-ancestry respondents (Okbay et al. 2018a). I further drop 141 siblings who did not complete high school, as they would not have been eligible for the original WLS sample. Of those remaining, 6,713 responded to both the in-person and mail-in 2010-2012 surveys and are thus eligible for the main analytic sample. The final sample includes the 6,018 respondents who had non-missing information across the 11 measures of health required to construct the dependent variables, which are described below.

Health and Retirement Study (HRS)

The HRS is a panel survey of U.S. households (RAND 2016). It began in 1992 with a sample of U.S. adults born between 1931 and 1941 and their spouses. Additional birth cohorts have since been added to the sample, and respondents have been followed up biennially. Since 1998, the HRS has surveyed a nationally representative sample of the U.S. population over age 50 and their spouses. Over the years, over 30,000 individuals have contributed data to the HRS.

Saliva samples were collected for genotyping beginning in 2006,[5] and the PGSs I use have been constructed for 8,652 European-ancestry respondents (Okbay et al. 2018b). I restrict this sample to those

[3] In the late 1950s, approximately 75% of Wisconsin students graduated from high school (Hauser and Willis 2005).

[4] Saliva was first collected for genotyping in 2007 and 2008 by mail; in 2010, additional samples were collected during home interviews. Genotyping was performed at the Center for Inherited Disease Research (CIDR) using the Illumina HumanOmniExpress beadchip array, which genotypes 713,014 SNPs. For further information, see the quality control report (Wisconsin Longitudinal Study 2016).

[5] DNA samples were first collected from HRS respondents during in-home interviews in 2006, at which time a random subsample of households was asked to participate. In 2008, all remaining households were asked to provide saliva samples for genotyping. Genotyping was performed by the Center for Inherited Disease (CIDR) using the Illumina HumanOmni2.5 beadchip array, which genotypes roughly 2.5 million SNPs. For further information, see Crimmins et al. (2013, 2015) and quality control reports (Health and Retirement Study 2012a, 2013a).

who responded to at least one HRS survey after providing saliva for genotyping (through 2012), those whose biomarkers were assessed, those ages 50 and over (the population that is technically eligible for HRS sampling), and those born in the U.S. This leaves 7,986 respondents eligible for analysis. A total of 7,726 of these respondents had non-missing information across the 11 measures of health used to construct summary measures in at least one wave and thus are included in analyses. Respondents could contribute up to two observations to the analysis, depending on biomarker availability. There are thus 12,629 HRS observations included in this study.

Measures

Summary measures of health

Both the WLS and HRS assess many aspects of health, including self-reported symptoms, diagnoses, and physical and biological measurements. Rather than choose and rely on just a few, I combine measures found in both the WLS and HRS using principal components factor analysis (PCFA). Put simply, PCFA analyzes correlations between a set of measures to identify the unique dimensions or factors across which the total variance is spread. Variables can then be constructed for the most important factors using a weighted linear combination of the original measures, where the weighting is determined using the structure of correlations in the data.

I draw on the 11 measures of health that are found in both the WLS and HRS datasets.[6] These include a five-category measure of self-rated health, presence of basic (physical) limitations, presence of instrumental (cognitive) limitations, self-reported memory, and number of depressive symptoms. I also include indicators of whether the respondent has ever been diagnosed with high blood pressure, heart

[6] Self-rated health is a five-category variable indicating whether the respondent's health is poor, fair, good, very good, or excellent. Presence of any basic (physical) limitations is a binary indicator of whether the respondent has difficulty with any of the following basic physical activities: walking across a room, getting into or out of a chair, getting into or out of bed, bathing, dressing, eating/feeding, and toileting. Presence of any instrumental limitations is a binary indicator of whether the respondent has difficulty with any of the following instrumental activities: using a map, making phone calls, taking medication, and managing money. Self-reported memory is measured using the Health Utilities Index (HUI) in the WLS, while in the HRS, a simple five-category variable is used. Depressive symptoms are operationalized using versions of the Center for Epidemiologic Studies Depression score. In the HRS, I create three categories of total cholesterol based on Fletcher (2017). Categories indicate whether the respondent's total cholesterol is under 200 milligrams per deciliter [mg/dL]), under 240 mg/dL, or 240 mg/dL or higher.

disease, stroke, and diabetes or high blood sugar. In the WLS, I use a variable indicating whether the respondent was ever diagnosed with high cholesterol; in the HRS, classifications based on total measured cholesterol, as collected by trained interviewers, are used instead. Finally, I include body mass index (BMI) category, as designated by the Centers for Disease Control (2016).

A total of 6,018 WLS respondents and 12,629 HRS observations from 7,729 unique respondents had non-missing information for each measure of health and are therefore included in the PCFAs, which are conducted separately in each dataset. First, the original measures are standardized (mean = 0, standard deviation [SD] = 1) across included WLS or HRS respondents. As is standard, I retained PCFA-constructed factors with eigenvalues above 1 and rotated results obliquely to allow factors to be correlated. Results are presented in Appendix Table A2.1.

In both the WLS and HRS, this process returns three factors, which are remarkably similar across datasets. The first indicates general physical and mental or cognitive health; its highest loading variables include self-reported health, presence of basic (physical) and instrumental (cognitive) limitations, self-reported memory, and depressive symptoms. In the WLS, it explains 24.4% of the variance across the 11 measures of health included in the PCFA, and in the HRS, 26.4%. The second factor represents cardiovascular conditions (high blood pressure, cholesterol, heart disease, stroke) and explains 20.7% and 20.5% of the total variance in the WLS and HRS datasets, respectively. The third and final factor indicates metabolic health (diabetes or high blood sugar, BMI), explaining 18.4% of the variance across measures in the WLS and 15.9% in the HRS.

I use these factors to construct analogous summary measures of health, which are all standardized (mean = 0, SD = 1) and scaled so that higher values indicate better health. Correlations between these summary measures and original health variables are presented in Appendix Table A2.2.[7]

[7] Robustness checks (not shown) demonstrate that the use of original survey measures returns results that are substantively the same as those obtained when using PCFA-constructed measures. Use of the PCFA-constructed measures, however, allows for a more parsimonious analysis.

Educational attainment

In both the WLS and HRS, I operationalize educational attainment with self-reported years of completed schooling. All WLS respondents graduated from high school so the minimum level of educational attainment is 12 years. The survey top-codes years of schooling at 20. Among HRS respondents, I collapse those who reported very low levels of schooling into a single category (5 years or less), and schooling is top-coded at 17 years.

Note that this measure is not identical to the measure of educational attainment employed in the study from which the education PGS is derived. As described below, Lee et al. (2018) harmonized measures of education collected by separate surveys in diverse educational systems; they did so by mapping to internationally comparable years of schooling. Because my primary interest is in the effect of education on health—rather than the effect of the PGS itself—this is not a major concern. It is an issue I return to in the Discussion, however, as measurement in the GWAS stage has implications for the predictive power of resulting PGSs and therefore their utility as control variables.

Polygenic scores (PGSs) for education and measures of genetic ancestry

I operationalize genetic selection into education using a PGS (Dudbridge 2013). PGSs predict an outcome by aggregating the effects of hundreds of thousands or even millions of single-nucleotide polymorphisms (SNPs), the type of genetic variant that is responsible for most genetic differences between humans. At each SNP, a person possesses two alleles, each of which may be considered either a risk allele or a non-risk allele. A person's genotype at a particular SNP is simply the number of risk alleles (0, 1, or 2) that are found at that genetic site. A PGS is the weighted sum of a person's genotypes, where the weight of a particular SNP is the effect of an additional risk allele at that SNP on the outcome of interest. Equation 1 provides a standard formula for individual i's PGS (PGS_i),

$$PGS_i = \sum_{j=1}^{J} x_{ij} w_j, \qquad \text{Equation 1}$$

where x_{ij} is the genotype of individual i at SNP j (it may equal 0, 1, or 2), and w_j indicates the weight or effect size of an additional risk allele at genetic variant j, as estimated in an independent GWAS (McCarthy et al. 2008; Visscher et al. 2017).

In the GWAS, researchers estimate a separate regression for each genetic variant they wish to study. Specifically, they regress their outcome of interest on the number of risk alleles (0, 1, or 2) found at a particular SNP, controlling for a limited number of covariates such as sex, year of birth or age, and genetic ancestry. Genetic ancestry is operationalized using principal components of the genetic data.[8] Once genetic effects are estimated, researchers further adjust for confounding due to linkage disequilibrium (LD), a term used to describe the fact that genetic variants are not inherited independently; genotypes at genetic variants located close to each other on the genome tend to be correlated. It is too computationally intensive to control for genotype at all SNPs within the regression models themselves; thus, the adjustment is made on the back end by specifying the structure of LD in the research sample or a reference population. This ensures that the genetic effects are not double-counted when calculating PGSs.

The PGSs I employ are constructed using the most recent GWAS of educational attainment (Lee et al. 2018), which was conducted in a sample of 1,131,881 people drawn from 71 separate datasets. Measures of educational attainment were harmonized across datasets into internationally equivalent years of schooling using International Standard Classification of Education (ISCED) guidelines. As mentioned earlier, this GWAS found 1,271 SNPs that were significantly associated with years of schooling, and PGSs constructed from the GWAS explain around 11% of the variation in years of schooling in independent U.S. samples after adjusting for year of birth, sex, their interactions, and ancestry.

Graciously, the authors of this GWAS have constructed education PGSs for use in the WLS

[8] Ancestry and race are related but distinct concepts. Race is a set of categories that differ across time and space depending on which human differences are viewed as salient (Fujimura et al. 2014). Ancestry is instead a multidimensional, continuous concept reflecting the history of human migration and reproduction.

(Okbay et al. 2018a) and HRS (Okbay et al. 2018b) datasets.[9] After making sample restrictions (as described earlier), I standardize these PGSs (mean = 0, SD = 1) within the respective analytic samples. Additional information regarding PGS construction is provided in Appendix Text A2.1.

Note that PGSs were not constructed for respondents of non-European ancestries (Lee et al. 2018). As described in Appendix 6 of Conley and Fletcher 2017 (see also, Martin et al. 2017), due to the history of human migration and subsequent isolation, ancestral groups differ in terms of genetic variability and the structure of correlations between inherited genotypes (LD). This complicates several critical processes: genotyping, imputing variants that are not directly genotyped, and adjusting estimated effects for confounding.[10] Genetic effects may also be differentially mediated or moderated by environmental factors, including racialized experiences. Most genetic analyses therefore restrict ancestral heterogeneity to European ancestry groups.

Statistical geneticists still worry that subtle variation due to ancestry could confound effects of the PGS. This is not a major issue in the current study, as I am not interested in the effects of the education PGS, per se. However, measures of genetic ancestry may be useful control variables in and of themselves. Ancestry groups may cluster in particular regions, cities, or neighborhoods due to past or present social networks. If characteristics of these places (e.g., labor market, climate) influence residents' educational outcomes and health, a spurious association between education and health could ensue. Therefore, I control for genetic ancestry, which is operationalized using the top ten principal components (PCs) of the

[9] The WLS and HRS datasets were omitted from the GWAS sample for the purpose of PGS construction (Okbay et al. 2018a; Okbay et al. 2018b).

[10] Common genotyping chips are optimized to measure a sample of genetic variants that differ between Europeans. Chips may therefore fail to directly genotype important sources of genetic variability in non-Europeans. The accuracy with which non-genotyped variants are imputed is also likely to vary across ancestries. It is likely to be particularly inaccurate among those with substantial African ancestry, given the greater genetic variance among African populations. This is due to the genetic bottleneck that occurred when around 2,000 individuals migrated out of Africa 100,000 years ago: genetic variability is reduced in the descendants of the migrants compared to the descendants of those who remained in Africa. Because of the greater genetic variability in African populations, more genetic variants need to be directly genotyped to obtain the same imputation accuracy; this is not, however, done in practice. Relatedly, a larger number of genotyped individuals is needed to infer the LD structure in African populations, which is used to impute and to adjust genetic effects for confounding; unfortunately, genetic samples from African populations are, if anything, more limited than those from European and other ancestral groups. For further information, see Appendix 6 of Conley and Fletcher (2017).

genetic data, which are the ten largest dimensions of genetic variation in the sample. The PCs are provided with the PGSs (Okbay et al. 2018a; Okbay et al. 2018b).

In Appendix Table A2.3, I show that controlling for the top ten PCs on their own barely attenuates the estimated effects of education on health. Thus, in the main analysis, I incorporate controls for genetic ancestry alongside the education PGS and stress that the vast majority of the observed attenuation is due to the education PGS rather than the ancestry PCs.

Control variables

The statistical models I estimate add sets of control variables sequentially. The idea is that the utility of the education PGS as a control may decline when the more proximal confounding characteristics it reflects are controlled directly. Initial models control for demographic variables only, in order to reflect a research scenario in which little is known about other proximal confounders of the education-health link. Intermediary models add controls for family background characteristics and childhood health, and final models control for a measure of ability in adolescence that I refer to as cognitive performance.

Demographics characteristics: In both datasets, I control for demographic characteristics including sex, year of birth, age, a squared term for age, the interaction between year of birth and age, and the interaction between year of birth and age-squared.

Family background: Measures of family background reflect family SES, structure, and place. In both the WLS and HRS, maternal and paternal years of schooling are expressed as six-category variables (less than eight years; eight; nine to eleven; twelve; 13-15; and 16 or more years). In the WLS, I also use paternal occupation (farming; unskilled; skilled; white collar; professional; and not in labor force) and a continuous measure of family income in 1957, when graduate respondents were about to complete high school. Both of these variables are derived from tax data by the WLS. No such data are available in the HRS; instead, I use a measure of perceived SES in childhood (poor; average; or well-off) and an indicator of whether the respondent's father was consistently employed throughout childhood versus unemployed for several months or absent/deceased.

In the WLS, family structure is operationalized with a variable indicating whether the respondent lived with both parents through age 16 and the number of siblings they report having. In the HRS, paternal presence is subsumed in the measure of paternal employment.

Publicly available WLS data does not report place of birth. Instead, I use a categorical variable indicating the population of the graduate's hometown in 1957. In the HRS, I group region of birth into four categories (Northeast, Midwest, South, or West).

Childhood health: In both the WLS and HRS, respondents were asked to reflect on their health as children and report it as poor, fair, good, very good, or excellent. I control for this measure as a five-category variable.

In the WLS, respondents were also asked whether they had each of 11 different health conditions as a child. Conditions include asthma, bronchitis, diphtheria, ear infections, hepatitis, meningitis, mononucleosis, pneumonia, polio, tonsillitis, and whooping cough. I create a measure of childhood health by adding one for each condition they had, top-coding the variable at five conditions or more (only 0.5% of respondents fall into this category). I further construct a variable in the WLS indicating whether the respondent experienced extended activity limitations because of their health sometime before the age of 16: they may have missed school for a month or more, been confined to their home or bed for a month or more, or been restricted from physical activity for three months or more due to a health condition.

Cognitive performance in adolescence: Cognitive performance in adolescence is operationalized in the WLS using the respondent's centile rank on the Henmon-Nelson test of mental ability (Henmon et al. 1957). At the time when original WLS respondents and their siblings attended school, the Henmon-Nelson test was administered in all Wisconsin high schools; it was a multiple-choice test consisting of 90 questions assessing verbal and quantitative ability. The test was taken when respondents were in their freshman or junior year of high school. Centile rank is based on all test-takers nationwide in a given year; it adjusts for the age at which the respondent took the test.

Unfortunately, no measure of adolescent cognitive performance is available in the HRS.

Analysis

I begin by estimating missing values of independent variables among otherwise-eligible respondents using chained imputations across 20 iterations.

I then proceed with the regression analyses, in which I assess the extent to which controlling for the education PGS attenuates the association of educational attainment with health. I do so separately in the WLS and HRS datasets for each of the three summary measures of health. All models employ ordinary least squares (OLS) linear regression and cluster standard errors at the family (WLS) or household (HRS) level. In the WLS, a single observation is taken from each respondent, *i*, while in the HRS, respondents can contribute up to two observations to the analysis; observations within individuals are indexed with *t*. Recall that summary measures of health are standardized (mean = 0, SD = 1) within the WLS and HRS samples. Thus, coefficients on years of schooling indicate the expected SD change in the relevant summary measure of health for a year increase in educational attainment, holding covariates constant.

The series of models I estimate reveal the extent to which the PGS confounds the effects of education on health when holding constant sets of more proximal, known confounders of the education-health relationship. Model 1A estimates the association of years of schooling (*EduYrs*) with health (*Health*) controlling for demographic characteristics, indicated by the vector Dem. Model 1B adds the education *PGS* and a vector (PC) including the top ten principal components of the genetic data (i.e., genetic ancestry).

$$Health_{it} = \alpha_0 + \alpha_1 EduYrs_i + \theta^T \mathrm{Dem_{it}} + \varepsilon_{it} \qquad \text{Model 1A}$$

$$Health_{it} = \beta_0 + \beta_1 EduYrs_i + \beta_2 PGS_i + \delta^T \mathrm{Dem_{it}} + \rho^T \mathrm{PC_i} + \epsilon_{it} \qquad \text{Model 1B}$$

I use seemingly unrelated estimation to test whether the coefficient on *EduYrs* differs significantly between Models 1A and 1B. Seemingly unrelated estimation calculates the covariance

matrices for the models of interest jointly. It can be used to conduct tests of significance between models that rely on overlapping data, such as the nested models I evaluate here. I estimate p-values separately for each of the 20 multiply imputed datasets, and I report the averages. I also calculate the percent attenuation of the effect of education on health with the introduction of the education PGS using Equation 2.

$$100 * \frac{(\hat{\alpha}_1 - \hat{\beta}_1)}{\hat{\alpha}_1} \qquad \text{Equation 2}$$

The next two models incorporate, in addition to demographic characteristics, controls for measures of family background (Fam) and childhood health (ChHlth). As with Models 1A and 1B, Model 2A estimates the association of health with years of schooling controlling for these measures only; Model 2B adds the control for the education PGS and measures of genetic ancestry.

$$Health_{it} = \alpha_0 + \alpha_1 EduYrs_i + \theta^T \mathrm{Dem_{it}} + \gamma^T \mathrm{Fam_i} + \tau^T \mathrm{ChHlth_i} + \varepsilon_{it} \qquad \text{Model 2A}$$

$$Health_{it} = \beta_0 + \beta_1 EduYrs_i + \beta_2 PGS_i + \delta^T \mathrm{Dem_{it}} + \sigma^T \mathrm{Fam_i} + \lambda^T \mathrm{ChHlth_i} + \rho^T \mathrm{PC_i} + \epsilon_{it} \qquad \text{Model 2B}$$

Finally, Models 3A and 3B parallel those above while adding a control for yet another potential confounder of the effect of education on health: cognitive performance in adolescence (*CogPerf*). These models are estimated in the WLS sample only, as no measure of cognitive performance in early life is available in the HRS.

$$Health_{it} = \alpha_0 + \alpha_1 EduYrs_i + \theta^T \mathrm{Dem_{it}} + \gamma^T \mathrm{Fam_i} + \tau^T \mathrm{ChHlth_i} + \mu CogPerf_i + \varepsilon_{it} \qquad \text{Model 3A}$$

$$Health_{it} = \beta_0 + \beta_1 EduYrs_i + \beta_2 PGS_i + \delta^T \mathrm{Dem_{it}} + \sigma^T \mathrm{Fam_i} + \lambda^T \mathrm{ChHlth_i} + \pi CogPerf_i + \rho^T \mathrm{PC_i} + \epsilon_{it} \quad \text{Model 3B}$$

Family fixed effects

I also estimate models using family fixed effects in a sub-sample of n = 2,232 siblings from n = 1,114 families in the WLS. These models effectively hold constant all stable family characteristics by investigating differences in health between siblings with differing levels of education. This analysis is useful as the measures of family background I control for in the main analysis are likely incomplete. The PGS could therefore function as a useful control variable in part because it reflects remaining uncontrolled social advantage. But if the education PGS attenuates the association of education with health within families (i.e., between siblings), it suggests that the score is not just reflecting family-level confounders of the education-health link. It must also effectively net out the confounding influences of childhood health, abilities and skills or personality, and potentially many other factors.

RESULTS

Table 2.1 presents key descriptive statistics. Summary measures of health are standardized across included respondents within datasets (mean = 0, SD = 1); all are scaled so that higher values indicate better health. In both datasets, measures tend to be skewed towards poorer health, with the median falling above the mean. For example, the median of physical and mental health is 0.3 in both the WLS and the HRS. Further, in the WLS, physical and mental health ranges from 6.5 SDs below the mean to 1.3 SDs above; in the HRS, it ranges from 4.7 SDs below to 1.7 SDs above.

The education PGS is also standardized across included respondents within datasets; it is relatively symmetrically distributed, ranging from -3.5 to 4.0 in the WLS and from -3.7 to 3.9 in the HRS. Average education in the WLS is 14.0 years, higher than the mean in the HRS (13.2 years). This is because the WLS sample includes only high school graduates while 12.7% of HRS respondents did not graduate high school.

Table 2.1. Descriptive statistics

	WLS		HRS[a]	
	Mean (SD) or %	**Min, Max**	**Mean (SD) or %**	**Min, Max**
Summary measures of health				
Physical & mental health	0.00 (1.0)	-6.5, 1.3	0.00 (1.0)	-4.7, 1.7
Cardiovascular health	0.00 (1.0)	-3.5, 2.7	0.00 (1.0)	-3.6, 2.5
Metabolic health	0.00 (1.0)	-4.1, 2.7	0.00 (1.0)	-3.5, 2.5
Key independent variables				
Education PGS	0.00 (1.0)	-3.5, 4.0	0.00 (1.0)	-3.7, 3.9
Years of education	14.01 (2.4)	12, 20	13.17 (2.5)	5, 17
Highest degree attained				
Less than high school	0.0%	-	12.7%	-
High school or GED	49.9%	-	38.8%	-
Associate's or some college	16.9%	-	24.1%	-
Bachelor's or higher	33.2%	-	24.4%	-
Demographic controls				
Female	52.9%	-	58.1%	-
Year of birth	1939.6 (4.0)	1920, 1960	1937.9 (10.0)	1905, 1961
Age	70.51 (4.0)	47, 92	69.99 (9.8)	50, 101
N possible observations	6,018		12,629 (7,726 respondents)	
N complete observations [b]	4,322		10,076 (6,053 respondents)	

[a] In the HRS, descriptive statistics for measures of health and age represent 12,629 observations (up to two per respondent). All other figures represent means across respondents.

[b] Information is missing primarily for control variables measuring family background, childhood health, and cognitive performance in adolescence, descriptive statistics for which can be found in Appendix Table A2.4.

Though WLS respondents were born between 1920 and 1960, over half were born in a single year (52.1% in 1939), reflecting the fact that a single graduating cohort was originally selected for sample inclusion. Just 1.3% of the sample was born before 1930 and 3.4% were born in 1950 or later. Average age at the time of the survey is 70.5 years. HRS respondents are comparable in terms of average birth year (1938), though the spread around the mean is more substantial. One in five respondents was born before 1930 (21.8%) and 15.1% were born in or after 1950. Average age at the time of the survey is similar to that in the WLS (70.0 years).

Descriptive statistics for additional control variables are provided in Appendix Table A2.4.

The education PGS as a control variable

Controlling for demographic characteristics only

In Table 2.2, I present results from models of the three summary measures of health on years of education. Results from Models 1A and 1B, which control for demographic characteristics, are also presented in Figure 2.1. Put briefly, in Models 1A and 1B, education is positively and significantly related to all three summary measures of health in both datasets. However, in all cases, the estimated effect of a year of schooling is attenuated when controlling for the education PGS in Model 1B. In both datasets, the proportional attenuation is smallest for physical and mental health and largest for cardiovascular health.

In Model 1A, each year of schooling predicts a 0.064-SD ($p < .001$) increase in physical and mental health in the WLS and a 0.105-SD ($p < .001$) increase in the HRS. When holding the education PGS constant (Model 1B), the magnitudes of these effects decline modestly, to 0.061 ($p < .001$) and 0.094 ($p < .001$), respectively. Thus, the inclusion of the education PGS as a control reduces the estimated effect of education on physical and mental health by 4.8% in the WLS and by 10.5% in the HRS.

Parallel models of cardiovascular health show that controlling for the education PGS attenuates the estimated effect of years of schooling by 21.1% in the WLS and by 26.1% in the HRS. Specifically, a year of schooling is associated with a 0.031-SD ($p < .001$) increase in cardiovascular health in the WLS and a 0.032-SD increase in the HRS, holding only demographics constant. When controlling for the

Table 2.2. Effect of a year of education on standardized summary measures of health, before and after controlling for the education PGS

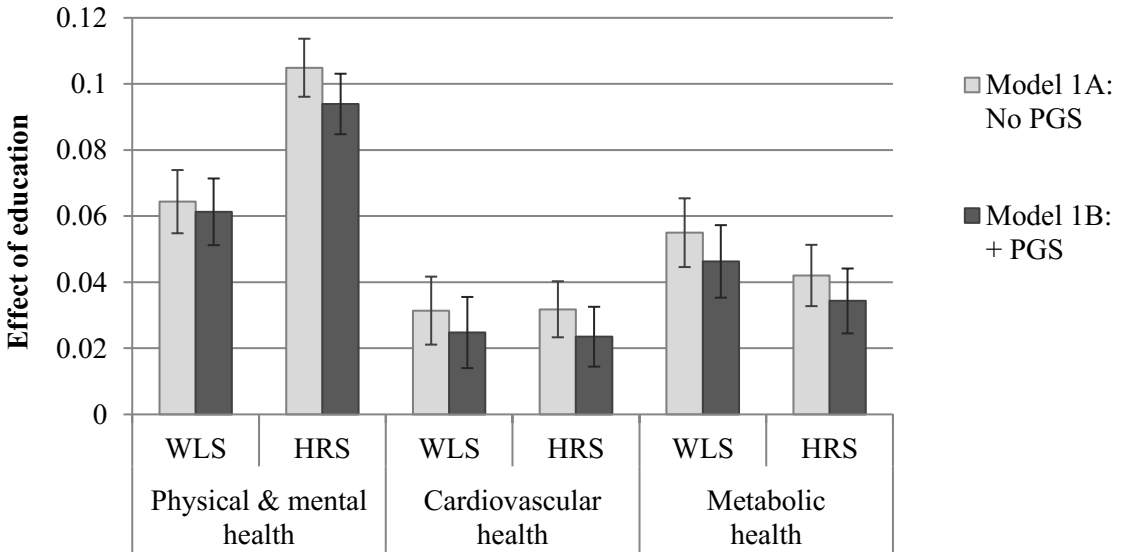
	Model 1		**Model 2**		**Model 3**	
	1A: No PGS	**1B:** + PGS	**2A:** No PGS	**2B:** + PGS	**3A:** No PGS	**3B:** + PGS
Dependent variable	$\hat{\alpha}_{\text{EduYrs}}$ (SE)	$\hat{\beta}_{\text{EduYrs}}$ (SE)	$\hat{\alpha}_{\text{EduYrs}}$ (SE)	$\hat{\beta}_{\text{EduYrs}}$ (SE)	$\hat{\alpha}_{\text{EduYrs}}$ (SE)	$\hat{\beta}_{\text{EduYrs}}$ (SE)
WLS						
Physical & mental health	0.064 ***	0.061 ***	0.058 ***	0.055 ***	0.047 ***	0.045 ***
	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)
Cardiovascular health	0.031 ***	0.025 ***	0.028 ***	0.022 ***	0.030 ***	0.025 ***
	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)
Metabolic health	0.055 ***	0.046 ***	0.048 ***	0.040 ***	0.048 ***	0.043 ***
	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
HRS						
Physical & mental health	0.105 ***	0.094 ***	0.084 ***	0.076 ***	-	-
	(0.004)	(0.005)	(0.005)	(0.005)		
Cardiovascular health	0.032 ***	0.024 ***	0.019 ***	0.012 *	-	-
	(0.004)	(0.005)	(0.005)	(0.005)		
Metabolic health	0.042 ***	0.034 ***	0.030 ***	0.024 ***	-	-
	(0.005)	(0.005)	(0.005)	(0.005)		

Notes: Model 1 controls for demographic characteristics; Model 2 adds controls for family background and childhood health; and Model 3 adds a control for cognitive performance in adolescence. Models 1B, 2B, and 3B also control for the education PGS and the top ten principal components of the genetic data (measures of genetic ancestry). Standard errors are adjusted for clustering within families (WLS) or households (HRS).

PGS = Polygenic score; SE = Standard error; WLS = Wisconsin Longitudinal Study; HRS = Health and Retirement Study; WLS post-imputation n = 6,018; HRS post-imputation n = 12,629 observations from 7,726 respondents.

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; † $p < 0.1$ (two-tailed test)

Figure 2.1. Effect of a year of education on standardized summary measures of health, before (Model 1A) and after (Model 1B) controlling for the education PGS



Notes: Effects are estimated in Models 1A (light grey bars) and 1B (dark grey bars), as shown in Table 2.2. Both models control for demographic characteristics; Model 1B also controls for the education PGS and the top ten principal components of the genetic data (measures of genetic ancestry). 95% confidence intervals are shown; standard errors are clustered at the family (WLS) or household (HRS) level. PGS = Polygenic score; WLS = Wisconsin Longitudinal Study; HRS = Health and Retirement Study.

education PGS as well, estimates are reduced to 0.025 ($p < .001$) and 0.024 ($p < .001$), respectively.

Finally, a year of schooling is expected to improve metabolic health by 0.055 SDs ($p < .001$) in the WLS and by 0.042 SDs ($p < .001$) in the HRS, holding demographics constant. Incorporating the education PGS as a control, respective estimates decline by 15.8% and 18.3% to 0.046 ($p < .001$) and 0.034 ($p < .001$).

In all cases, the PGS's utility as a control variable is at least marginally statistically significant. The difference in the estimated effect of education on physical and mental health with and without controlling for the education PGS is marginally significant at $p = .072$ on average across the 20 imputed datasets in the WLS; it is significant at $p < .001$ in the HRS. For cardiovascular and metabolic health, the difference is highly significant in both datasets ($p < .001$).

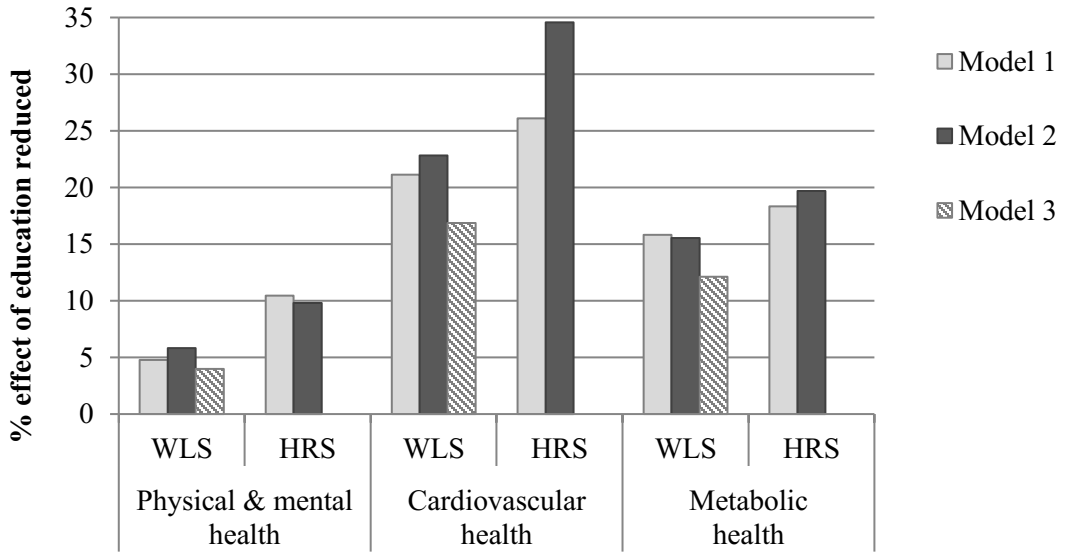
That said, controlling for the education PGS certainly does not upend decades of previous research. A substantial and statistically significant positive association of education with each measure of health remains when controlling for the education PGS.

Controlling for additional confounders

Subsequent models investigate whether the education PGS continues to attenuate the estimated effect of education on health when measures of more proximal confounders are controlled directly. I expect the percent attenuation to decline as additional proximal confounders are incorporated into the model. Figure 2.2 presents the percent reductions in the estimated effects of education on health in each set of models.

Models 2A and 2B reflect a research scenario in which some measures of family background and childhood health, but not adolescent ability or skills, can be controlled. The estimated effect of education on each measure of health is smaller in Model 2A compared to Model 1A, due to the introduction of family background and childhood health controls (Table 2.2). However, as shown in Figure 2.2, controlling for the education PGS reduces the estimated effect of education on measures of health to a similar extent in Model 2B versus 2A as in Model 1B versus 1A. Thus, the utility of the education PGS as

Figure 2.2. Percent reduction in the effect of a year of education on standardized summary measures of health when controlling for the education PGS, by model



Notes: Percent reductions are calculated by comparing estimates from Models 1A and 1B (light grey bars), Models 2A and 2B (dark grey bars), and Models 3A and 3B (striped bars), as shown in Table 2.2. Model 1 controls for demographic characteristics; Model 2 adds controls for family background and childhood health; and Model 3 adds a control for cognitive performance in adolescence. Models 1B, 2B, and 3B also control for the education PGS and the top ten principal components of the genetic data (measures of genetic ancestry). PGS = Polygenic score; WLS = Wisconsin Longitudinal Study; HRS = Health and Retirement Study.

a control variable appears about equal in models that control for demographics only versus those that also control for measures of family background and childhood health.

For example, a year increase in education is expected to improve physical and mental health by 0.058 SDs ($p < .001$) in the WLS and by 0.084 SDs ($p < .001$) in the HRS, holding demographics, family background, and childhood health constant (Model 2A). When controlling for the education PGS as well, estimated effects decline to 0.055 SDs ($p < .001$) and 0.076 SDs ($p < .001$) (Model 2B). As with the previous set of models, the estimated effect of education on physical and mental health is significantly smaller in the model controlling for the education PGS than the model omitting the PGS as a control (WLS $p < .001$, HRS $p = .027$). However, the levels of attenuation in the effect of education on physical and mental health—5.8% in the WLS and 9.8% in the HRS—are similar to those obtained previously. The same is true for cardiovascular (WLS: 22.8%; HRS: 34.6%; both $p < .001$) and metabolic health (WLS: 15.5%; HRS: 19.7%; both $p < .001$).

Models 3A and 3B add a control for cognitive performance in adolescence, another potential confounder of the education-health link. These models are estimated in the WLS only, no such measure is available in the HRS. As expected, controlling for cognitive performance in Model 3A attenuates the association of education with physical and mental health, compared to Model 2A. A year of schooling is now associated with a 0.047-SD ($p < .001$) improvement in physical and mental health. When the PGS is added to the model, the estimated effect of education declines 4.0% to 0.045 SDs ($p < .001$). However, the difference in effects estimated in Models 3A and 3B is not statistically significant ($p = .125$).

Adding the control for cognitive performance does not attenuate the association of educational attainment with cardiovascular or metabolic health. This suggests that measured cognitive ability does not explain the relationship between education and these forms of health. The education PGS, however, continues to attenuate estimated effects of education significantly, however, by 16.9% for cardiovascular health ($p < .001$) and by 12.1% for metabolic health ($p < .001$).

For each measure of health, the attenuation of education's estimated effect following the introduction of the education PGS is smallest in WLS models that control for the most comprehensive set

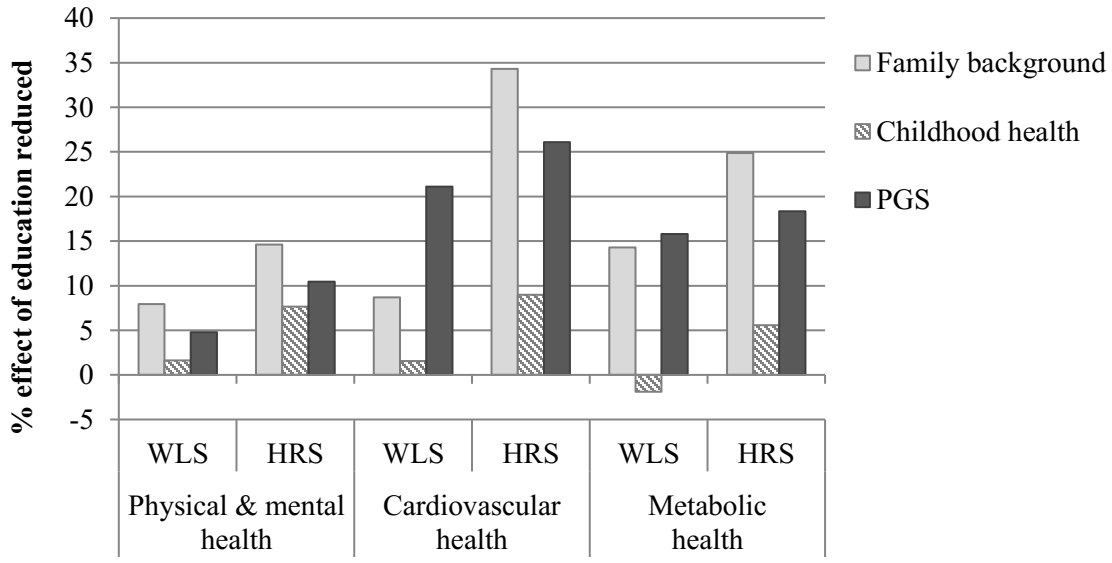
of more proximal potential confounders: demographics, family background, childhood health, and cognitive performance in adolescence. Differences in attenuation across models are modest. For physical and mental health, adding the education PGS results in a proportional attenuation of 4.0% in the most comprehensive model versus 5.8% in the model preceding it; corresponding figures for cardiovascular health are 16.9% versus 22.8% and for metabolic health 12.1% versus 15.5%. However, these results provide suggestive evidence that the education PGS is particularly useful as a control variable when certain more proximal confounders of the education-health relationship are unmeasured or uncontrolled. In particular, the education PGS may function as a partial control for cognitive performance. This should not be surprising, as a sizeable portion of genetic selection into educational outcomes is explained by measures of cognitive performance in adolescence (Belsky et al. 2016; Domingue et al. 2015).

Comparing the education PGS to more proximal confounders

While controlling for the education PGS results in significantly smaller estimates of education's effects on health, the scale of this attenuation needs to be contextualized. How does the proportional attenuation obtained when controlling for the education PGS compare to that obtained when instead controlling for the more proximal confounders that social scientists typically focus on? Figure 2.3 provides the answer to this question, presenting the percent attenuation obtained by adding family background measures (light grey bars), measures of childhood health (dark grey bars), or the education PGS (striped bars) to models that control for demographic characteristics only.

Adding measures of family background to models that otherwise control only for demographic characteristics reduces the association between education and health by between 8% and 34%, depending on the dataset and measure of health considered. Measures of childhood health attenuate associations by 9% at most. And, adding the education PGS to models that otherwise hold only demographics constant reduces associations of education with health by between 5% and 26%. The education PGS's value as a control variable is therefore more or less on par with measures of family background and childhood health, which are currently among the most recognized confounders of the education-health relationship.

Figure 2.3. Percent reduction in the effect of a year of education on standardized summary measures of health when controlling for the education PGS compared to that obtained when controlling for measures of family background or childhood health



Notes: Percent reductions are calculated by comparing estimates from models controlling for demographic characteristics only to those also controlling for: family background (light grey bars); childhood health (dark grey bars); or the education PGS and the top ten principal components of the genetic data (measures of genetic ancestry) (striped bars). PGS = Polygenic score; WLS = Wisconsin Longitudinal Study; HRS = Health and Retirement Study.

Family fixed effects

While Models 2A/2B and 3A/3B above control for measured family background characteristics, the sets of variables used likely omit important aspects of the respondent's childhood environment. To more completely control for family background, I use family fixed effects with 2,232 siblings from 1,114 families in the WLS.

First, as shown in Models 1A and 1B in Table 2.3 and in Figure 2.4, a year of schooling is associated with significantly better health across all three dimensions in family fixed effect models that control for demographic characteristics, both before and after incorporating the education PGS. Effect sizes are also similar to those presented above. The same is true for models that also control for childhood health (Models 2A and 2B) and adolescent cognitive performance (Models 3A and 3B). In this respect, results concur with the results from the main analysis presented above. Controlling for the education PGS does not fundamentally change what we know about the relationship between education and health: more education predicts better health even when genetic selection is accounted for.

Second, Figure 2.5 shows that controlling for the education PGS attenuates the association of education with two of the three dimensions of health between siblings: physical and mental health and metabolic health. The education PGS reduces the estimated effect of education on physical and mental health by 11.2% in models controlling for fixed effects and demographics only (light grey bars); by 11.0% in models also controlling for childhood health (dark grey bars); and by 8.8% in models incorporating a control for cognitive performance in adolescence (striped bars). These levels of attenuation are not statistically significant. In parallel models of metabolic health, estimated effects are attenuated significantly after the introduction of the education PGS, by 24.7% in models controlling for fixed effects and demographics only ($p = .028$); by 24.3% when adding childhood health ($p = .024$); and by 21.1% when controlling for adolescent cognitive performance ($p = .017$).

Notice that, as in the main analysis, the percent attenuation observed for these measures of health following the introduction of the education PGS to the model is weakest in the most comprehensive models: those that control for cognitive performance in adolescence. Again, this suggests that the

Table 2.3. Effect of a year of education on standardized summary measures of health, before and after controlling for the education PGS, using family fixed effects in the WLS

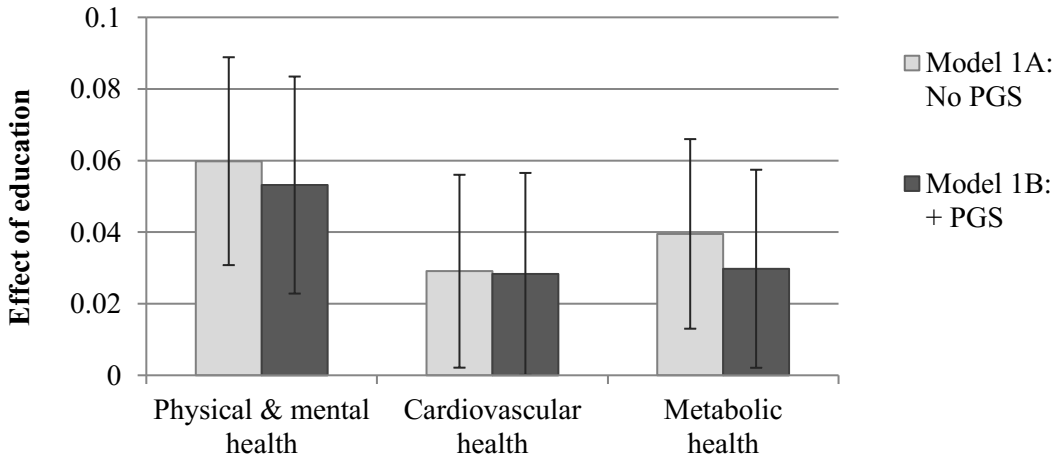
	Model 1		**Model 2**		**Model 3**	
	1A: No PGS	**1B:** + PGS	**2A:** No PGS	**2B:** + PGS	**3A:** No PGS	**3B:** + PGS
Dependent variable	$\hat{\alpha}_{\text{EduYrs}}$ (SE)	$\hat{\beta}_{\text{EduYrs}}$ (SE)	$\hat{\alpha}_{\text{EduYrs}}$ (SE)	$\hat{\beta}_{\text{EduYrs}}$ (SE)	$\hat{\alpha}_{\text{EduYrs}}$ (SE)	$\hat{\beta}_{\text{EduYrs}}$ (SE)
Physical & mental health	0.060 *** (0.015)	0.053 ** (0.015)	0.060 *** (0.015)	0.054 *** (0.015)	0.050 ** (0.015)	0.046 ** (0.016)
Cardiovascular health	0.029 * (0.014)	0.028 * (0.014)	0.029 * (0.014)	0.028 * (0.014)	0.029 * (0.014)	0.028 † (0.015)
Metabolic health	0.040 ** (0.014)	0.030 * (0.014)	0.040 ** (0.014)	0.031 * (0.014)	0.046 ** (0.014)	0.036 * (0.015)

Notes: Model 1 controls for demographic characteristics; Model 2 adds controls for family background and childhood health; and Model 3 adds a control for cognitive performance in adolescence. Models 1B, 2B, and 3B also control for the education PGS and the top ten principal components of the genetic data (measures of genetic ancestry). Standard errors are adjusted for clustering within families.

PGS = Polygenic score; SE = Standard error; WLS = Wisconsin Longitudinal Study; Post-imputation n = 2,232 individuals from 1,114 families.

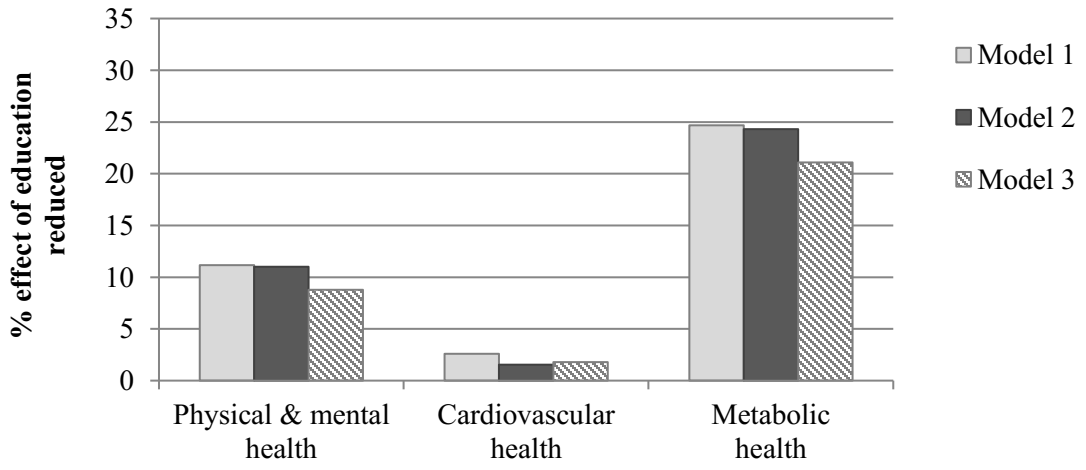
*** p < 0.001; ** p < 0.01; * p < 0.05; † p < 0.1 (two-tailed test)

Figure 2.4. Effect of a year of education on standardized summary measures of health, before and after controlling for the education PGS, using family fixed effects in the WLS



Notes: Effects are estimated in Models 1A (light grey bars) and 1B (dark grey bars), as shown in Table 2.3. Both models control for demographic characteristics and family fixed effects; Model 1B also controls for the education PGS and the top ten principal components of the genetic data (measures of genetic ancestry). 95% confidence intervals are shown. PGS = Polygenic score; WLS = Wisconsin Longitudinal Study.

Figure 2.5. Percent reduction in the effect of a year of education on standardized summary measures of health when controlling for the education PGS, using family fixed effects in the WLS



Notes: Percent reductions are calculated by comparing estimates from Models 1A and 1B (light grey bars), Models 2A and 2B (dark grey bars), and Models 3A and 3B (striped bars), as shown in Table 2.3. Model 1 controls for demographic characteristics and family fixed effects; Model 2 adds controls for childhood health; and Model 3 adds a control for cognitive performance in adolescence. Models 1B, 2B, and 3B also control for the education PGS and the top ten principal components of the genetic data (measures of genetic ancestry). PGS = Polygenic score; WLS = Wisconsin Longitudinal Study.

education PGS acts as a partial control for this construct. Also, note that the magnitudes of attenuation observed in the family fixed effect models are greater than those obtained in the main analysis. I return to this finding in the Discussion.

For cardiovascular health, the education PGS does not appear to play any role as a control variable within families. While this finding could be a result of limited variation in cardiovascular health between siblings, additional models indicate that roughly 85% of variation in cardiovascular health exists within, rather than between, families. Instead, it could be that the education PGS in part reflects aspects of one's family or childhood environment that are constant between siblings but that are associated with cardiovascular health when considering unrelated individuals. More generally, the confounders of education's effect on cardiovascular health are likely to differ from those relevant to other dimensions.

DISCUSSION

This paper set out to estimate the extent to which genetic selection into schooling confounds the estimated effect of education on later health in observational research. This is an important question, as scholars continue to debate the existence of a causal effect of education in part because schooling choices have long been recognized as non-random (Blau and Duncan 1967; Bourdieu 1984; Durkheim 1956[1922]; Parsons 1961; Sewell et al. 1969) and because selection factors are notoriously difficult to control (Cutler and Lleras-Muney 2008). Measures of genetic selection might correlate with or influence personal characteristics that, for whatever reason, cannot be controlled directly. Controlling for genetic selection may therefore reduce bias in the estimated effect of education on health in observational research (Cesarini and Visscher 2017; Conley 2016; Freese 2018).

I operationalize genetic selection using powerful PGSs (Dudbridge 2013; Lee et al. 2018) that predict a person's years of schooling by summing up the estimated effects of many genetic variants across the genome. My results are remarkably robust across two U.S. datasets of older individuals of European ancestries. Key findings are italicized below.

First and foremost among them is that *controlling for the education PGS does not upend the*

established robust and positive relationships between education and several key dimensions of health. In both the HRS and WLS datasets, there remains an educational disparity in health after adjusting for the PGS. This is true of all summary measures studied, including physical and mental, cardiovascular, and metabolic health.

Nonetheless, *the education PGS appears to have potential as a control variable*. The association between education and each measure of health is attenuated significantly when controlling for the PGS, even when holding constant demographic characteristics, family background, and childhood health, and (at least for some health outcomes) cognitive performance in adolescence. Further, while the attenuation of the estimated effect of education on health with the incorporation of the education PGS as a control may be modest, it is on par with measures of family background and childhood health, which are currently among the most recognized confounders of the education-health relationship.

Further, *the proportional attenuation of the relationship between education and each measure of health with the addition of the PGS is somewhat stronger when adolescent cognitive performance is omitted as a control*. This is consistent with the idea that the education PGS is a useful control variable primarily when more proximal confounders of education's effect on health—such as early abilities and skills—are unmeasured or measured with error. By implication, controlling for PGSs could also limit bias induced by characteristics that we remain unaware confound the education-health link, as DNA may influence all traits to some extent (Turkheimer 2000).

The education PGS does not, however, appear to act as a partial control for unmeasured aspects of family background and childhood health, as the proportional attenuation obtained when adding the PGS to models that omit measures of these confounders is no greater than that obtained when they are included. This result is consistent with prior research showing that childhood health does not mediate the relationship between the education PGS and educational outcomes (Belsky et al. 2016). That said, measures of childhood health employed here are less than ideal, as they are limited in scope and recalled many years later. Future research could assess whether education PGSs may pick up on and thus partially control for more objective measures of childhood health, such as biomarkers, which reflect variation in

health or disease risk even before symptoms are expressed.

Models utilizing family fixed effects generally support results from the main analysis, *though results differ in two ways*. *First, the education PGS reduces the magnitude of the estimated effect of education on physical and mental health and on metabolic health to a larger extent within families than in the main analysis of the full sample*. This is consistent with prior research, as Domingue et al. (2015) finds that the association of the education PGS with years of schooling is greater between siblings than it is among unrelated individuals. They suggest that holding highly influential family and environmental factors constant accentuates the impact of the more subtle characteristics that tend to distinguish siblings, such as specialized skills, interests, and personality traits—all characteristics that the education PGS may be correlated with (see also, Boardman and Fletcher 2015; Krapohl et al. 2014) and which might also influence later health. Future research could utilize data with measures of these traits to assess whether this is the case.

Second, for cardiovascular health, the education PGS does not appear to play any role as a control variable within families, though it did significantly attenuate the estimated effect of education in the main analysis. This finding does not appear to be a result of limited variation in cardiovascular health between siblings. It could be that the education PGS in part reflects aspects of one's family or childhood environment that are constant between siblings but that are associated with cardiovascular health among unrelated individuals. In any case, my results echo prior work that suggests that cardiovascular and metabolic health is informed by complex processes (e.g., Gaydosh et al. 2018) that remain only partially understood. Further, these findings illustrate the importance of breaking health down into its unique dimensions to better understand the social and biological etiology of its specific forms.

The current study has several limitations. First, while the PGSs I use are calculated based on one of the largest genetic association studies to date, they measure genetic selection imperfectly. The measure of education utilized in Lee et al. (2018) is rough, as it necessitated harmonizing information from many different datasets; while necessary, this likely increases error in estimated genetic effects and resulting PGSs. Indeed, while the PGS based on Lee et al. (2018) explains roughly 11% of variation in years of

schooling in representative samples of European-ancestry U.S. adults, Rietveld and colleagues (2013) suggest that PGSs based on population parameters of genetic effects could explain around 20%. Moreover, PGSs reflect additive genetic effects only; estimates from twin studies suggest that, including non-additive effects, DNA could explain roughly 40% of variation in educational outcomes (Branigan et al. 2013; Heath et al. 1985). Random measurement error in the education PGS likely renders my estimates of attenuation conservative. As newer and more powerful education PGSs become available in future years, they are likely to become more effective as control variables.

A second and related issue is that, though demographic and environmental characteristics are likely to modify the effects of genetic variants on outcomes, current PGSs reflect only average genetic effects. Specifically, they reflect average effects in the GWAS sample, which is often comprised of data from medical case-control studies, opt-in biobanks (e.g., UK Biobank), and databases from genomic testing companies (e.g., 23andMe). GWAS samples may therefore be of higher SES than the general population and may be self-selected for interest in medical or genealogy research. These factors raise important concerns about the portability of PGSs across population subgroups (Mostafavi et al. 2019). In the context of the current study, interacting the education PGS with independent variables may reduce additional confounding due to moderated genetic effects.

Third, statistical geneticists worry that small differences in genetic ancestry could bias the effect of the education PGS on schooling outcomes. This is not a major concern in the context of this study, as I am not interested in the effect of the PGS on outcomes, per se. Nonetheless, I follow methods of best practice by limiting my analyses to European-ancestry individuals and by controlling for the top 10 principal components of the genetic data (i.e., the 10 key dimensions of genetic ancestry).

While methodologically motivated, the restriction of the sample in this study and others like it to European-ancestry individuals is ethically questionable. Any beneficial implications of such studies may not be applicable to ethnically diverse populations, and, from a purely intellectual standpoint, current studies privilege the experiences of those of European descent. In the case of the current article, however, there is reason for optimism. Even in a sample of European ancestry adults—those for whom PGSs are

currently most predictive of outcomes (Lee et al. 2018; Martin et al. 2017)—controlling for the education PGS appears to neither drastically alter nor eliminate positive associations of education with health.

I utilize summary measures of three major dimensions of health in order to demonstrate general proof-of-concept regarding the utility of PGSs as control variables; this strategy also allows me to compare the robustness of results across two datasets collecting different measures of health. Nonetheless, future research could employ alternative measures of health and wellbeing that are commonly used in social science research. Also, I am unable to investigate patterns when additional potential confounders—including additional aspects of childhood health and non-cognitive skills or personality in adolescence—are controlled. However, this "limitation" also underscores one of my key arguments: that such confounders are often unavailable to control.

Despite the limitations described above, my results are remarkably robust across two U.S. datasets that differ in terms of geographic representation and range of age and educational outcomes. In sum, PGSs for educational attainment—measures of genetic selection—are likely to be useful control variables in studies of the education-health link. This study thus provides some of the first empirical evidence of the utility of PGSs as control variables in observational research. Nonetheless, it does not appear that genetic information will force a major reconsideration of the relationship between education and health: this relationship is positive, pervasive, and robust to confounders, including measures of genetic selection.

Chapter 3: The Association of Education with Health and Mortality by Socioeconomic Origin, Race, and Gender

The positive relationship between education and health is one of the most robust in the social sciences, and the health advantage experienced by educated individuals is thought to be due in part to a causal effect of schooling (Cutler and Lleras-Muney 2008; Hummer and Lariscy 2011). Schooling may affect health by promoting access to flexible resources—economic, cognitive, social, and structural—that support wellbeing throughout the life course (Freese and Lutfey 2011; Link and Phelan 1995; Mirowsky and Ross 2003). Theories of resource substitution and multiplication, outlined by Ross and Mirowsky (2006, 2010), posit that the effects of education on health will differ across the sociodemographic groups that also structure access to these resources, such as socioeconomic (SES) origin, race, and gender.

Resource substitution exists if education and the resources it generates stand in for resources that would otherwise be absent. By extension, the largest effect of education on health should be found among disadvantaged sociodemographic groups. Alternatively, under resource multiplication, education has the greatest impact on subsequent health for those with the most alternative resources to build on.

Prior research suggests patterns of resource substitution by SES origin (Ross and Mirowsky 2011). This result is often taken to suggest that education ameliorates the deleterious effects of a disadvantaged childhood on later health. However, results by race are more consistent with resource multiplication, with greater health benefits accruing to whites than to blacks in the United States (Farmer and Ferraro 2005; Holmes and Zajacova 2014; Shuey and Willson 2008). Thus, education's equalizing effects do not appear to extend to health disparities by race.

These contradictory findings may reflect the fact that the effect of education on health depends on the ability to successfully deploy education-related resources for health promotion and to avoid added stressors that accompany schooling itself or the high-SES spaces it affords access to (Masters et al. 2015). As long as the achievements of some sociodemographic groups are rewarded less generously with health-relevant resources than others, the effects of those achievements on subsequent wellbeing are likely to be

muted. The effect of education on health may therefore also vary across the intersection of multiple sociodemographic characteristics, as joint group membership may affect exposure to discrimination and other potential stressors (Borrell et al. 2006; Colen et al. 2018; Pew Research Center 2016).

In the current study, I therefore take a more complex intersectional approach than has been used in prior work. Specifically, rather than assessing heterogeneity in the association of education with health by SES origin *or* by race *or* by gender, I evaluate heterogeneity by SES origin, race, and gender simultaneously. I do so using self-reported health and mortality data from non-Hispanic whites and blacks in the Health and Retirement Study, a survey of older U.S. adults.

Without considering the unique experiences of individuals at the intersections of these sociodemographic categories, the theory of resource substitution would expect the smallest effect of education on health to obtain among white men from high-SES origins: those with the most alternative resources stemming from their SES origin, their race, and their gender. Alternatively, resource multiplication would predict the smallest effect to be found among black women from low-SES backgrounds. Instead, I find the smallest association of education with both self-reported health and mortality among low-SES origin black men.

Further, I show that, while higher education is associated with reduced disparities in self-reported health by SES origin for whites, it does not have the same substitutive, equalizing effect on health among blacks. Moreover, racial disparities in both self-reported health and mortality are exacerbated among the highly educated, particularly among those from low-SES origins. Education is not a remedy for all background-based and demographic disparities in health.

PATTERNS OF RESOURCE SUBSTITUTION AND MULTIPLICATION IN PRIOR WORK

First outlined by Ross and Mirowsky (2006, 2010), the theories of resource substitution and multiplication posit that the effects of education on health differ across sociodemographic groups depending on whether members have alternative access to health-enhancing resources. Resource substitution exists if the effect of education on health is strongest for those with the fewest resources

stemming from background or status. Education may enable those from disadvantaged groups to accrue the same health-promoting resources that more advantaged individuals can access in other ways. Or, education may be a distinct resource that can be used in lieu of advantages stemming from family background or demographic status. Ross and Mirowsky (2006) argue that "the less there is of one resource, the more important another will be" (Pp.1400).

Contrasted with the theory of resource substitution is resource multiplication, which exists if education affects health most for those who are otherwise advantaged. Such a pattern by SES origin may develop if, for example, family contacts are instrumental for gaining entry into the most health-enhancing careers, neighborhoods, or social groups, particularly for the highly educated. By race or by gender, a pattern of resource multiplication may exist if high educational achievement is rewarded more generously in the labor market for whites or for men, respectively, assuming that labor market success improves subsequent health (Link and Phelan 1995; Lynch 2006).

SES origin

Much of the existing research on resource substitution and multiplication focuses on family SES, as those from low-SES origins have worse health in adulthood than those from more advantaged backgrounds, even when holding attained SES constant (Ben-Shlomo and Kuh 2002; Galobardes et al. 2004; Montez and Hayward 2011). It is therefore encouraging that the largest effects of education on health appear to accrue to those from socioeconomically disadvantaged families (Andersson and Vaughan 2017; Bauldry 2015; Luo and Waite 2005; Ross and Mirowsky 2011; Schaan 2014; Schafer et al. 2013).[11] Consistent with resource substitution, these studies show that health disparities by SES origin are greatest among less-educated adults; among those with high education, virtually no disparity exists.

[11] A few studies have returned null or inconsistent interaction effects; these studies have assessed mortality (Hayward and Gorman 2004), active life expectancy (Montez and Hayward 2014), and diagnosed chronic conditions (Nandi et al. 2012). Just one study to my knowledge has supported resource multiplication by SES origins. Bauldry (2014) finds that, among young adults, the association of college with self-reported health is strongest for those most likely to attend. However, because Bauldry predicts probability of college attendance using not only measures of SES origin but also of academic performance and aspirations for the future, results are not entirely comparable with the other cited literature.

The effects of education on health may be largest for those from low-SES backgrounds because their subsequent life chances rely primarily on their own skills, behaviors, and achievements, which can be improved through education (Hout 2012; Link and Phelan 1995; Mirowsky and Ross 2003). Meanwhile, health-promoting skills and norms may be modeled in the home from an early age for those from high-SES backgrounds, making schooling less critical (Hernández-Alava and Popli 2017; Laureau 2003; Mollborn and Lawrence 2018). Those from high-SES origins may also succeed economically regardless of their own education (Hout 1984, 1988; Torche 2011) by drawing on family for information and contacts, or through direct receipt of financial and other assistance (Flaster 2018; Henretta et al. 2012). To the extent that economic outcomes influence subsequent health (Link and Phelan 1995; Lynch 2006), this would produce a greater effect of education on health among those from low-SES origin.

Race

To my knowledge, no research has drawn on theories of resource substitution and multiplication to describe heterogeneity in effects of education on health by race. Nonetheless, patterns similar to those observed by SES origin—declining disparities with increasing education—might be expected, given the correlation between race and SES and the fact that non-Hispanic whites evince better health and lower mortality than blacks. (Williams et al. 2010). Quite the reverse, the health advantages experienced by whites actually increase with education (Farmer and Ferraro 2005; Holmes and Zajacova 2014; Masters et al. 2012; Shuey and Willson 2008; Zajacova and Hummer 2009). Evidence thus points to resource multiplication by race, whereby the protective effects of education on health accrue mostly to whites.

This may be because discrimination limits the effects of education on health among U.S. blacks. Discrimination in the labor market restricts socioeconomic attainment, and evidence suggests it is most consequential for highly educated blacks (Tomaskovic-Devey et al. 2005). Insofar as economic outcomes affect health (Link and Phelan 1995; Lynch 2006), this should dampen the effect of education on health among blacks. Similarly, interpersonal discrimination may reduce average wellbeing among blacks by inducing health-harming stress (Goosby et al. 2018; Lewis et al. 2015; Williams and Mohammed 2009).

It may also offset education's positive effect on health, as blacks with higher levels of education are more likely to report experiencing discrimination (Borrell et al. 2006; Colen et al. 2018; Hochschild 1995; Pew Research Center 2016), perhaps because they are more often in spaces frequented by whites.

The reduced association of education with health among blacks may also stem from early life experiences in segregated neighborhoods (Sampson and Winter 2016; Williams and Collins 2001). The quality of segregated black schools has historically been lower than others, which could result in reduced returns to health (Frisvold and Golberstein 2011). Similarly, heightened exposure to neighborhood-related environmental toxins or stress in childhood could irreversibly impact health across the life course (Galobardes et al. 2004; Shonkoff et al. 2009), weakening the effect of education on later outcomes.

Another potential explanation is that competing obligations grow with education for blacks more so than for whites. In general, blacks in the U.S. are more likely than whites to lend to extended kin (Gerstel 2011; O'Brien 2012) and to live in multigenerational or extended households (Cross 2018; Perkins 2019). These patterns may be most dramatic among the highly educated if the likelihood of having a network member in need shrinks sharply with education for whites but not for blacks.

Gender

Patterns by gender are more complex than those by SES origin or race, as they occasionally support the resource substitution perspective and other times resource multiplication. Subjective aspects of health follow patterns of resource substitution by sex, as the relationships between education and self-reported health, self-reported physical limitations, and depressive symptoms are stronger for women than for men (Ross et al. 2012; Ross and Mirowsky 2006, 2010). Ross and Mirowsky (2006) find that this is explained in part by the fact that education affords opportunities for creative expression at work more for women than for men, where creative work predicts improved mental health.

Meanwhile, the negative association of education with mortality is steeper for men, largely due to deaths stemming from behavioral risks (Ross et al. 2012). Ross and colleagues (2012) reason that education might improve these outcomes more for men because men have "decision-making latitude"

(Pp.1180), while women have historically been subject to strict, overarching social controls. Skills generated through education—such as self-regulation and direction—may thus factor into consequential behavioral choices (e.g., smoking) more for men than for women. Prior research supports this idea (Denney et al. 2010; Olson et al. 2017).

Notably, women report poorer health and disablement due to nonfatal conditions than men, while men have higher rates of life-threatening chronic illnesses and higher mortality rates (Case and Paxson 2005; Crimmins et al. 2011). Thus, in some sense, both sets of patterns are consistent with a resource substitution perspective: on health outcomes for which men are at a disadvantage, they benefit more from education, while the reverse is true on health outcomes for which women are at a disadvantage.

THE INTERSECTION OF SES ORIGIN, RACE, AND GENDER

Considering SES origin, race, and gender simultaneously, the theory of resource substitution would expect the smallest effect of education on health to be found among high-SES origin white males—those at an advantage in terms of their family background, their race, and their gender. The largest effect would be seen among those with the fewest alternative resources: black females from disadvantaged families. Resource multiplication would expect the reverse.

I expect less simplistic patterns. The effect of education on health depends not only on access to resources, but also on one's ability to deploy resources effectively for health promotion and to avoid added stressors that accompany schooling itself or the high-SES spaces it affords access to (Masters et al. 2015). As described below, these factors may vary across the intersection of groups defined by SES origin, race, and gender. A more complex intersectional perspective may therefore shed light on additional sources of heterogeneity in effects of education on health.

The intersectional approach emerged out of black feminist scholarship, which argued that the experiences of black women could not be explained by the additive effects of blackness and being female (Collins 1990; Crenshaw 1989). Instead, the effects of these characteristics were augmented by their interaction. Drawing on the intercategorical approach to intersectionalism (Choo and Ferree 2010; McCall

2005), I assess how inequalities are structured by interactions between memberships in various groups.

Prior work suggests that joint membership in sociodemographic groups defined by SES origin, race, and/or gender may structure effects of education on health due to differential exposure to discrimination and other stressors. As described in earlier sections, discrimination in the labor market and in social settings may offset the effect of education on health for blacks compared to whites (Borrell et al. 2006; Colen et al. 2018; Goosby et al. 2018; Hochschild 1995; Lewis et al. 2015; Pew Research Center 2016; Williams and Mohammed 2009). Here, I contend that the regularity of discriminatory and other stressful experiences—and the resulting effects of education on health—may vary by not only by race, but also by SES origin and/or gender.

For example, Borrell et al. (2006) finds that black men are more likely than black women to report experiencing racial discrimination in several realms of life, including getting a job and at work. Black men are also more likely than black women to report being seen as suspicious or being inexplicably stopped by police (Pew Research Center 2016). Recall that research also suggests discrimination increases with education for blacks (but presumably not for whites) (Borrell et al. 2006; Colen et al. 2018; Hochschild 1995; Pew Research Center 2016; Tomaskovic-Devey et al. 2005) and that discrimination worsens health (Goosby et al. 2018; Lewis et al. 2015; Williams and Mohammed 2009). If the racial gap in exposure to discrimination grows with education more for men than for women, we might expect the difference in the effect of education on health by race to also be most dramatic among men.

Similarly, discrimination on the basis of SES origin may occur more frequently for blacks than for whites. Indeed, Colen et al. (2018) find that upward socioeconomic mobility predicts reduced discrimination for whites, but enhanced discrimination for blacks. Discrimination is known to occur on the basis of black-identifiable voices (Fischer and Massey 2004) and names (Bertrand and Mullainathan 2004), both of which may be more prevalent among blacks from low-SES origins compared to high-SES origin blacks and whites of either background (Gaddis 2017).

The concept of John Henryism (James 1994; James et al. 1987) and the related theory of skin-deep resilience (Brody et al. 2013; Chen et al. 2015; Miller et al. 2015) also suggest that the effect of

education on health is moderated by both SES origin and race. These perspectives imply that the constant striving required to achieve high-SES status—or to achieve a high level of education—induces health-harming stress for those from disadvantaged backgrounds regardless of race (Stephens et al. 2012a, 2012b). However, this stress may be particularly problematic for low-SES origin racial minorities, who must adjust not only to an unfamiliar class environment, but also to a heightened threat of discrimination that persists in the high-SES spaces that education subsequently affords access to (Borrell et al. 2006; Colen et al. 2018; Goosby et al. 2018; Hochschild 1995; Lewis et al. 2015; Pew Research Center 2016; Williams and Mohammed 2009).

Consistent with this idea, Gaydosh et al. (2018) find that physical health—specifically, metabolic syndrome—is predicted by the interaction between childhood disadvantage, race, and education among young adults in the U.S. A college education predicts better health among whites and among blacks from relatively advantaged backgrounds. However, among blacks from less advantaged origins, college education actually predicts poorer health. In the language of resource substitution and multiplication, this suggests that the larger racial disparities in health seen at high levels of education (Farmer and Ferraro 2005; Holmes and Zajacova 2014; Wilson and Shuey 2008)—patterns of resource multiplication by race—may be particularly evident among those from low-SES backgrounds. Likewise, it suggests that patterns of resource substitution by SES origin—larger positive effects of education on health for those from low-SES backgrounds (Ross and Mirowsky 2011)—will not be found among blacks.

In sum, while prior work on resource substitution and multiplication has tended to focus on single sociodemographic characteristics, effects of education on health may vary across the intersection of SES origin, race, and gender. In particular, the effects of education on health may vary across the intersection of race and gender and/or by race and SES origin. Heterogeneity at the intersection of the SES origin, race, *and* gender remains understudied, providing additional motivation for taking a more complex intersectional perspective, as I do here.

METHODS

Data

I use data from the Health and Retirement Study (HRS), a panel survey of U.S. households that is funded by the National Institute on Aging and conducted through the University of Michigan (RAND 2016). The HRS began in 1992 with a sample of U.S. adults born between 1931 and 1941 and their spouses. Additional birth cohorts were later added to the sample, and respondents have been re-interviewed biennially. Since 1998, the HRS has surveyed a nationally representative sample of the U.S. population over age 50 and their spouses. I draw data from survey years 1998 through 2014; I begin in 1998 rather than 1992 as this is the first year respondents were asked to report their health in childhood, a key covariate.

I restrict the sample of self-reported health to observations taken from respondents ages 50 to 64; likewise, hazard models of mortality are limited to those who were between ages 50 and 64 in their first eligible survey year. The lower bound of this range is the age at which people are eligible for HRS sampling. I set an upper bound due to research showing that the association of education with health morphs over the life course. Educational disparities in mortality, for example, decline among the elderly due to selective attrition and to natural aging processes that render resources associated with educational attainment less relevant for health (Willson et al. 2007).[12] In addition, I limit the sample to U.S.-born non-Hispanic white and non-Hispanic black respondents, as sample sizes are too small to generate precise estimates of education's association with health among Hispanics and the contents and implications of education for health may vary by nativity (Acevedo-Garcia et al. 2007; Kimbro et al. 2008).

There are 16,439 respondents meeting these criteria and eligible for both the SRH and mortality analysis. Nearly 10% of respondents (n = 1,626) are missing data on a key covariate.[13] I multiply impute missing independent variables across 10 imputations and use the imputed data in analysis.

[12] Results for those ages 65 and above are available on request.

[13] The vast majority is missing information on maternal education (n = 1,535). This is in part because many respondents were not asked their parental years of schooling at all; early HRS surveys asked only whether each parent had attained fewer than eight years of schooling versus eight or more years.

The analysis of self-reported health pools 57,587 observations from the 16,439 respondents between survey years 1998 and 2014. Each respondent may therefore contribute between one and nine observations to the analysis of self-reported health (average = 3.5). The mortality analysis includes the same 16,439 respondents and tracks mortality outcomes through 2014.

Measures

Health

Self-reported health (SRH) correlates strongly with objective measures, like the presence of diagnosed conditions and timing of subsequent mortality (Idler and Benyamini 1997; Jylhä 2011). Moreover, when reporting SRH, people may consider the severity, number, and duration of any unpleasant physical or psychological symptoms they experience. SRH was ascertained using the following question: "Would you say your health is excellent, very good, good, fair, or poor?" I dichotomize SRH as follows: (0) Poor or fair health; (1) Good, very good, or excellent health.[14]

Mortality data is taken from the National Death Index through 2011, the latest year for which it was linked with the HRS sample. From 2012 through 2014, mortality data is obtained through exit interviews with respondents' family members or spouses. In the hazard models I estimate, mortality is operationalized with a binary indicator of whether the respondent was deceased at a particular age.

Education

In the main analysis, education is measured using years of formal schooling. The HRS top-codes years of education at 17 or more, and I collapse those who reported very low levels of schooling into a single category. The resulting variable ranges from five to 17. I center this variable at its mean across respondents; I do this so that when years of schooling is interacted with other characteristics, coefficients on those other characteristics can be interpreted as effects when holding education at its mean.

[14] In robustness tests (not shown), I replicate the main analysis using alternative models and measures of SRH. Results are substantively the same as those presented here.

In robustness tests, I operationalize education using highest degree attained: (1) Less than high school; (2) High school or GED; (3) Associate's degree or some college; (4) Bachelor's degree or higher.

Socioeconomic origin, race, and gender

I use a binary measure of maternal education to operationalize SES origin, as maternal education is associated with child and adult health as well as educational attainment (Case et al. 2005; Currie and Moretti 2003; Gage et al. 2013). I dichotomize maternal education as follows: (0) Less than twelve years of education; (1) twelve years of education or more.[15]

I then construct an eight-category nominal variable indicating SES origin (maternal education twelve years or more versus less than twelve years), race (non-Hispanic white or non-Hispanic black), and gender (male or female) to use in analyses.

Control variables

I control for a person's age, age-squared, year of birth, and interactions between age and year of birth and between age-squared and year of birth. I also control for childhood health, since research suggests that poor health in early life both deters educational attainment and affects later health (Haas 2007; Jackson 2009; Palloni 2006). Childhood health is assessed with a simple question asking respondents to consider their health through age 16 and rate it as excellent, very good, good, fair, or poor.

Analysis

Self-reported health

To analyze SRH, I estimate a series of logistic regression models. Model 1 does not take an intersectional approach; instead, its goal is to establish that average health disparities by education, SES

[15] Paternal education more often goes unreported than maternal education. Among otherwise-eligible respondents, 9.3% are missing maternal years of education versus 16.3% paternal education. Missingness on paternal education is also more strongly linked to race: 7.3% of whites and 15.6% of blacks in the sample are missing maternal education, compared to 11.4% and 32.0% for paternal education, respectively.

origin, race, and gender in this sample are consistent with those identified in previous research (e.g., Cutler and Lleras-Muney 2008; Montez and Hayward 2011; Williams et al. 2010; Case and Paxson 2005). I anticipate that, all else constant, years of schooling will be associated with higher odds of good SRH, while low-SES origin, black race, and female sex will be associated with lower odds.

Model 1 therefore regresses good or better SRH (*GoodSRH*, indexed by person, *i*, and time, *t*) on years of education (*EduYrs*) and binary indicators of low-SES origin (*LowSES*), black race (*Black*), and female sex (*Female*). Control variables (indicated by the vector X) include age at interview, age-squared, year of birth, the interaction between age and year of birth, the interaction between age-squared and year of birth, and health in childhood. To correct for non-independence of observations from the same respondent or household, I cluster standard errors at the household level. Models of SRH include 57,587 observations from 16,439 respondents.

$$\text{logit}(GoodSRH_{it}) = \beta_0 + \beta_1 EduYrs_i + \beta_2 LowSES_i + \beta_3 Black_i + \beta_4 Female_i + \tau^T \mathrm{X}_{\mathrm{it}} \qquad \text{Model 1}$$

Model 2 builds on Model 1 by incorporating a limited intersectional perspective, one that is comparable to that employed in prior research. Specifically, Model 2 adds interactions between years of education and the three binary sociodemographic variables. Based on prior work (e.g., Farmer and Ferraro 2005; Holmes and Zajacova 2014; Ross and Mirowsky 2006, 2011; Ross et al. 2012), I expect to find a greater increase in odds of good SRH per year of schooling for those from low-SES origins and for women—evidence of resource substitution by SES origin and gender. Thus I expect odds ratios associated with β_5 and β_7 to be greater than 1. I anticipate a smaller improvement in odds of good health per year of schooling for black respondents—indicative of resource multiplication by race—and thus for the odds ratio associated with β_6 to be less than 1.

$$\text{logit}(GoodSRH_{it}) = \beta_0 + \beta_1 EduYrs_i + \beta_2 LowSES_i + \beta_3 Black_i + \beta_4 Female_i + \beta_5 EduYrs_i LowSES_i + \beta_6 EduYrs_i Black_i + \beta_7 EduYrs_i Female_i + \tau^T \mathrm{X}_{it} \quad \text{Model 2}$$

The approach used in Model 2 assumes that any heterogeneity in the effect of education on health across one sociodemographic characteristic is constant regardless of a person's other characteristics. Taking a more complex intersectional perspective, Model 3 considers membership in all three sociodemographic groups simultaneously by regressing good or better SRH on education, an eight-category variable indicating SES origin, race, *and* gender, and their interaction.[16] While I specify high-SES origin white males as the reference group here, I also calculate and present odds ratios on years of education for the seven other sociodemographic groups.

$$\begin{aligned}\text{logit}(GoodSRH_{it}) = {} & \beta_0 + \beta_1 EduYrs_i \\ & + \beta_2 LowSES\_Wh\_Male_i \\ & + \beta_3 HighSES\_Wh\_Fem_i + \beta_4 LowSES\_Wh\_Fem_i \\ & + \beta_5 HighSES\_Bl\_Male_i + \beta_6 LowSES\_Bl\_Male_i \\ & + \beta_7 HighSES\_Bl\_Fem_i + \beta_8 LowSES\_Bl\_Fem_i \\ & + \beta_9 EduYrs_i LowSES\_Wh\_Male_i \\ & + \beta_{10} EduYrs_i HighSES\_Wh\_Fem_i + \beta_{11} EduYrs_i LowSES\_Wh\_Fem_i \\ & + \beta_{12} EduYrs_i HighSES\_Bl\_Male_i + \beta_{13} EduYrs_i LowSES\_Bl\_Male_i \\ & + \beta_{14} EduYrs_i HighSES\_Bl\_Fem_i + \beta_{15} EduYrs_i LowSES\_Bl\_Fem_i \\ & + \tau^T \mathrm{X}_{it}\end{aligned} \quad \text{Model 3}$$

[16] It would be mathematically equivalent to estimate a model incorporating a four-way interaction between educational attainment, SES origins, race, and gender along with all lower-level interactions and main effects. Such a model would draw primary attention to the statistical significance of higher-order interaction terms. Some have termed this approach "intersectionality as testable explanation" (Evans et al. 2018). Where relevant, I present results from such models below.

Differences in odds ratios across groups may not align with differences in effects on probabilities (Long and Freese 2014; Long and Mustillo 2018; Mood 2010). Thus, I also estimate and present predicted probabilities of good SRH by years of education for each sociodemographic group. To do so, I hold age at its mean across observations (58 years) and other covariates at their means across respondents.

Finally, to assess whether patterns hold at different points in the distribution of education, I estimate a model that is identical to Model 3 but that instead uses highest degree attained to operationalize education (Model 3b).

Mortality

To study all-cause mortality, I estimate a series of discrete-time hazard models with a logit link function. These models use data from the same 16,439 respondents as the analysis of SRH. Time is indexed with year of age, starting with the age at which the respondent was first eligible for the analytic sample and ending with the respondent's age at the end of 2014. Respondents who died over this period are eliminated from risk set after the age at which they are recorded deceased.

The mortality analysis follows a series of models that are analogous to Models 1-3 described above. Specifically, Model 1 regresses an indicator of mortality status (0 = Not deceased; 1 = Deceased) on education and indicators of low-SES origin, black race, and female sex. Model 2 adds interactions between education and each of the three sociodemographic variables. Model 3 incorporates a more complex intersectional angle by interacting education with the eight-category variable indicating joint membership in groups defined by SES origin, race, and gender. Finally, Model 3b is identical to Model 3 except that education is operationalized using highest degree attained rather than years of schooling.

As with the analysis of SRH, in addition to age (which indexes time), each model controls for age-squared, year of birth, interactions between age, age-squared, and year of birth, and childhood health. I cluster standard errors at the household level. To ease interpretation, I present odds ratios by group and estimate the probability of mortality at age 58, holding covariates at their respondent-level means.

RESULTS

Table 3.1 presents unweighted descriptive statistics. Of the 57,587 observations of SRH, three in four report being in good or better health (76.6%). And, of the 16,439 respondents included in the analysis, 15.2% were deceased by 2014. On average, respondents attained 13.13 years of education. Just 14.4% did not complete high school; half earned a high school diploma or GED (55.6%); 6.8% earned an Associate's degree or attended some college; and 23.2% attained a Bachelor's degree or higher. Around half of respondents' mothers attained twelve or more years of education (57.8%); these respondents are considered to be from high-SES origins. Three in four are non-Hispanic white (75.9%), 24.1% are non-Hispanic black, and just over half are female (56.5%). Across all observations in the analysis of SRH and at entry into the mortality analysis, respondents range in age from 50 to 64 years.

Self-reported health

Table 3.2, Model 1 presents results from a logistic regression of good SRH on years of education and binary indicators of the three focal demographic characteristics. All else constant, a year of education increases odds of good health by 24.9% ($p < .001$). As expected, those from low-SES origins have lower odds of good health than those from socioeconomically advantaged backgrounds (odds ratio [OR] = 0.747, $p < .001$) and black respondents have worse odds than whites (OR = 0.576, $p < .001$). Unexpectedly, odds of good health do not differ significantly by gender (OR = 1.048, $p = .154$).

Model 2 adds a limited intersectional angle, interacting years of education with the three focal sociodemographic variables. The association of education with odds of good health is greater for those from low-SES origins (interaction OR = 1.045, $p = .009$). Further, odds of good SRH increase with years of schooling significantly less for blacks than for whites (interaction OR = 0.934, $p < .001$) and marginally more for women than for men (interaction OR = 1.029, $p = .057$). Results by SES origin and potentially gender thus support a pattern of resource substitution, while those for race follow resource multiplication. Each of these results is consistent with prior research (e.g., Farmer and Ferraro 2005; Holmes and Zajacova 2014; Ross and Mirowsky 2006, 2011; Ross et al. 2012).

Table 3.1. Descriptive statistics

Variable	N	%
Total observations for self-reported health	**57,587**	**100**
Observations with complete data [c]	52,299	
Self-reported health		
Poor or fair	13,482	23.4
Good, very good, or excellent	44,105	76.6
Age[a]	Mean (*SD*): 57.97 (*3.9*)	
Total respondents	**16,439**	**100**
Respondents with complete data [c]	14,813	
Mortality by end of 2014	2,494	15.2
Years of education[b]	Mean (*SD*): 13.12 (*2.5*)	
Highest degree attained		
< High school or GED	2,359	14.4
High school diploma or GED	9,137	55.6
Associate's degree or some college	1,123	6.8
Bachelor's degree or higher	3,820	23.2
Sociodemographic characteristics		
Low-SES origin (Maternal ed. <12 years)	6,285	42.2
High-SES origin (Maternal ed. 12+ years)	8,619	57.8
White (non-Hispanic)	12,479	75.9
Black (non-Hispanic)	3,960	24.1
Male	7,158	43.5
Female	9,281	56.5
High-SES white males	3,383	22.7
Low-SES white males	1,742	11.7
High-SES white females	3,807	25.5
Low-SES white females	2,631	17.7
High-SES black males	598	4.0
Low-SES black males	686	4.6
High-SES black females	831	5.6
Low-SES black females	1,226	8.2
Year of birth[b]	Mean (*SD*): 1946.48 (*8.2*)	
Childhood health		
Poor	231	1.4
Fair	790	4.8
Good	2,413	14.8
Very good	4,062	24.9
Excellent	8,830	54.1

[a] Unless noted otherwise, this variable is centered at the observation-level mean in analyses.

[b] Unless noted otherwise, this variable is centered at the respondent-level mean in analyses.

[c] Maternal education was missing for n = 1,535 respondents and childhood health was missing for n = 113 respondents. Imputed values are used in analyses.

Table 3.2. Odds ratios (OR) from logistic regression models of good self-reported health

	Model 1	**Model 2**	**Model 3**
	OR (SE)	OR (SE)	OR (SE)
Education (Years)	1.249 (0.010) ***	1.228 (0.017) ***	1.221 (0.023) ***
Low-SES origin (Ref: High)	0.747 (0.029) ***	0.768 (0.030) ***	-
Black (Ref: White)	0.576 (0.023) ***	0.546 (0.022) ***	-
Female (Ref: Male)	1.048 (0.035)	1.068 (0.036) †	-
Sociodemographic group			(p < 0.001)
High-SES white males	-	-	Ref.
Low-SES white males	-	-	0.781 (0.055) ***
High-SES white females	-	-	1.113 (0.063) †
Low-SES white females	-	-	0.831 (0.052) **
High-SES black males	-	-	0.577 (0.051) ***
Low-SES black males	-	-	0.414 (0.036) ***
High-SES black females	-	-	0.541 (0.044) ***
Low-SES black females	-	-	0.457 (0.032) ***
Education X Low-SES origin	-	1.045 (0.018) **	-
Education X Black	-	0.934 (0.015) ***	-
Education X Female	-	1.029 (0.015) †	-
Education X Sociodemographic group			(p = 0.002)
High-SES white males	-	-	Ref.
Low-SES white males	-	-	1.066 (0.029) *
High-SES white females	-	-	1.015 (0.028)
Low-SES white females	-	-	1.090 (0.033) **
High-SES black males	-	-	0.991 (0.040)
Low-SES black males	-	-	0.949 (0.030) †
High-SES black females	-	-	1.002 (0.040)
Low-SES black females	-	-	1.008 (0.029)
N observations (N respondents)	57,587 (16,439)	57,587 (16,439)	57,587 (16,439)
Pseudo R^2	0.1063	0.1074	0.1078

Notes: Years of education is centered at the mean across respondents (13.12 years). Models control for age, age-squared, year of birth, the interaction between age and year of birth, the interaction between age-squared and year of birth, and childhood health. Standard errors are adjusted for clustering at the household level.

*** p < .001; ** p < .01; * p < .05; † p < .1 (two-tailed test)

Ref. = Reference group; SE = Standard error; SES = Socioeconomic origin

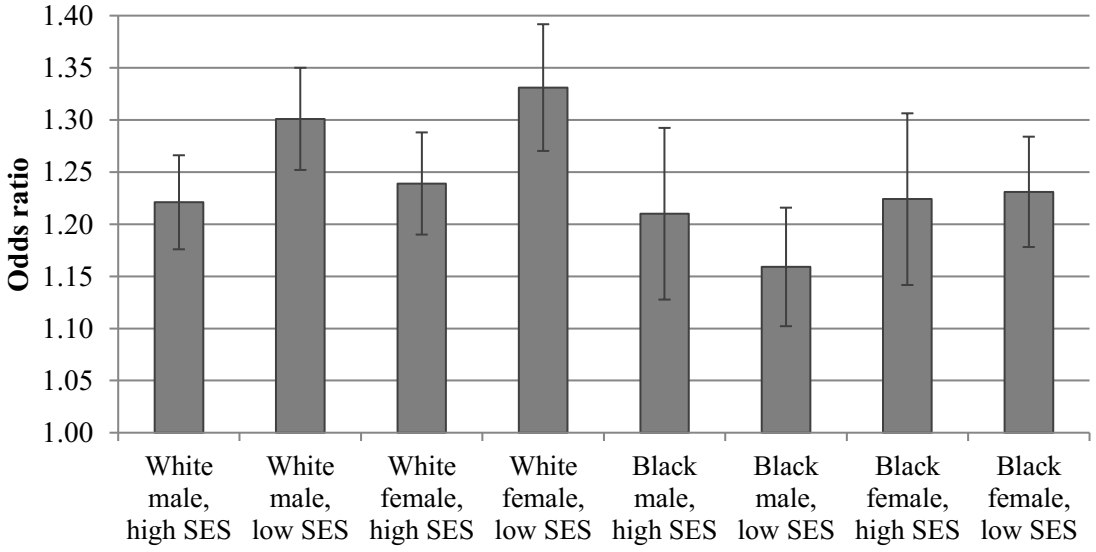
Taking a more complex intersectional approach, Model 3 interacts years of education with the eight-category indicator of sociodemographic group. Figure 3.1 presents the estimated effect of education by sociodemographic group, demonstrating that additional schooling is associated with elevated odds of good SRH at a high level of statistical significance ($p < .001$) for all (see also, Appendix Table A3.1). As expected given the results of Model 2, the terms involved in the interaction between education and sociodemographic group are jointly significant ($p = .002$), indicating that the association of schooling with SRH differs across groups.

Neither the most status- and background-advantaged group (high-SES origin white men) nor the least (low-SES origin black females) demonstrates the greatest gain in odds of good SRH per year of schooling. In fact, odds of good SRH increase a similar amount per year of schooling for high-SES white men (22.1%) and low-SES black women (23.1%). Instead, the association is strongest for white women from low-SES origins, among whom an additional year of schooling increases odds of good SRH by 33.1%. Meanwhile, low-SES origin black men experience the smallest gain in odds of good SRH per year of schooling, at just 15.9%.

While patterns of heterogeneity thus do not uniformly support theories of resource substitution or multiplication, support for both theories is evident in subsets of respondents. Consistent with prior research and the results of Model 2, there is evidence in Model 3 that the largest effects of education on health accrue to those from disadvantaged backgrounds, suggestive of resource substitution by SES origin. However, this is only the case for whites.

Among white men, odds of good SRH increase by 30.1% per year of schooling for those from low-SES backgrounds compared to a much smaller 22.1% for those from high-SES origins (difference $p = .021$). Results are similar for white women ($OR_{low\text{-}SES} = 1.331$; $OR_{high\text{-}SES} = 1.239$; difference $p = .023$). Among black respondents, there is no sign of resource substitution by SES origin, as those from low-SES origins do not appear to experience greater health returns to schooling than those from high-SES origins among either men ($OR_{low\text{-}SES} = 1.159$; $OR_{high\text{-}SES} = 1.210$; difference $p = .326$) or women ($OR_{low\text{-}SES} = 1.231$; $OR_{high\text{-}SES} = 1.224$; difference $p = .891$). A model regressing good SRH on all possible three-way

Figure 3.1. Effects of a one-year increase in education on odds of good self-reported health, by sociodemographic group



Notes: Group-specific odds ratios are calculated using Model 3 in Table 3.2. See Appendix Table A3.1 for further information. Error bars represent 95% confidence intervals. SES = Socioeconomic origin.

interactions between the four focal variables and lower level terms confirms that this pattern, as the interaction between education, SES origin, and race, is statistically significant ($p = .033$).[17]

Viewed from a different angle, these results indicate resource multiplication by race—such that the largest effects of education on odds of good health accrue to whites—but only among those from low-SES origins.[18] Among low-SES origin males, whites' odds of good SRH increase by 30.1% per year of education compared with 15.9% among blacks (difference $p < .001$). This is also found among females ($OR_{whites} = 1.331$; $OR_{blacks} = 1.231$; difference $p = .014$). Among those from high-SES origins, however, white and black men ($OR_{white} = 1.221$; $OR_{black} = 1.210$; difference $p = .820$) and women ($OR_{white} = 1.239$; $OR_{black} = 1.224$; difference $p = .744$) experience roughly equivalent gains per year of schooling.

No other higher-level terms estimated in the model of SRH on all possible three-way interactions between education, SES origin, race, and gender is statistically significant. The four-way interaction between these variables, estimated in another model, is also insignificant ($p = .561$). Still, it is worth noting that in Model 3, the largest difference in the estimated effects of education on health by gender is found between low-SES origin black men and women. Specifically, odds of good health increase by 23.1% for each year of education for low-SES origin black women but by only 15.9% for low-SES origin black men, a marginally significant difference ($p = .069$).

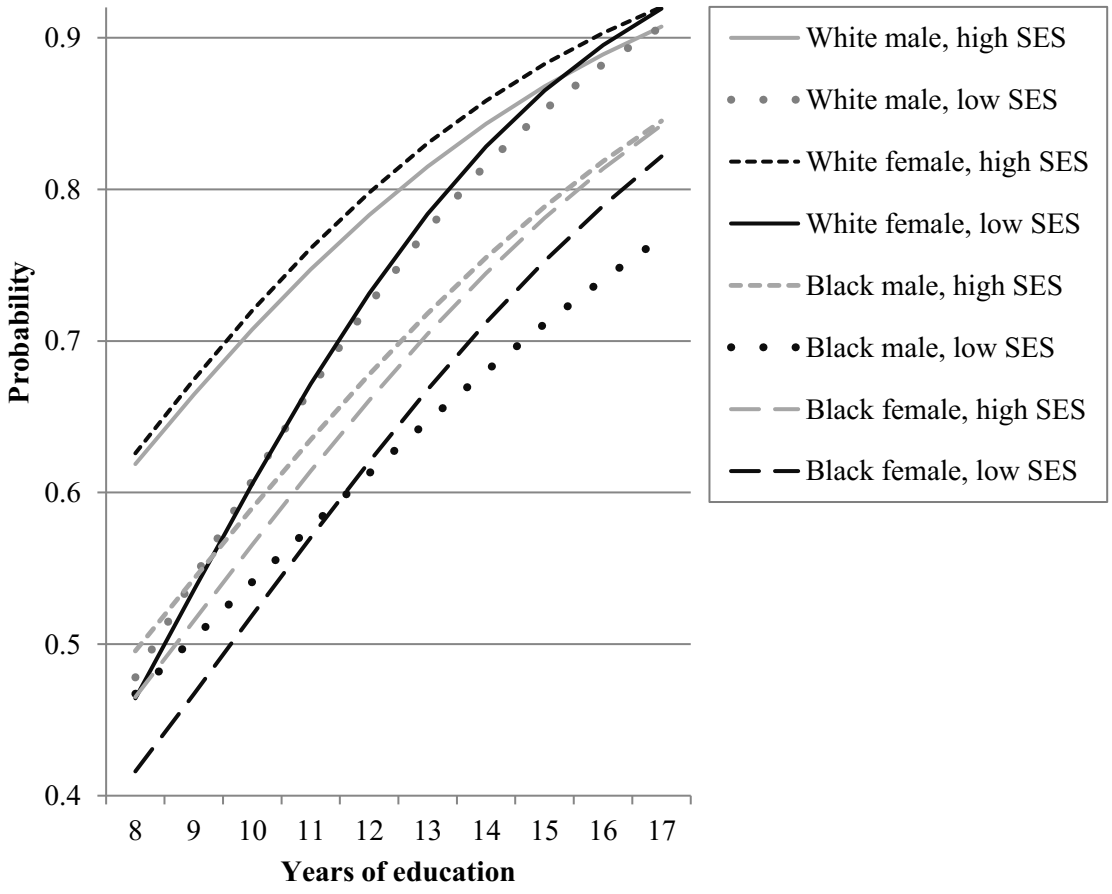
Predicted probabilities

Figure 3.2 presents predicted probabilities of being in good SRH by years of schooling and sociodemographic group. Probabilities are calculated based on Model 3 in Table 3.2, holding age at 58 years and covariates at their respondent-level means. As indicated by the regression results, the probability of good SRH increases with education for all groups. Probabilities range from 0.42 to 0.63 at

[17] These findings are also evident when estimating Model 2 separately by race. The interaction between education and low-SES origins indicates resource substitution among whites (interaction OR = 1.071, $p = .001$) but not blacks (interaction OR = 0.977, $p = .461$).

[18] These findings are also evident when estimating Model 2 separately by SES origin. The interaction between education and black race indicates resource multiplication among those from low-SES origins (OR = 0.904, $p < .001$) but not those from high-SES origins (OR = 0.996, $p = .883$).

Figure 3.2. Predicted probability of good self-reported health, by sociodemographic group and years of education



Notes: Predicted probabilities are calculated using Model 3 in Table 3.2. Age is held at 58 years and covariates are held at their respondent-level means. SES = Socioeconomic origin.

low levels (eight years) of schooling and from 0.77 to 0.92 at high levels (17 years or more). Thus, unlike what might be predicted by resource substitution, variation in probabilities across groups does not shrink substantially at high levels of education, nor does it grow, as would be expected based on resource multiplication.

Also consistent with the regression results, disparities in health by SES origin decline with education for whites. Among white men with eight years of schooling, 62% and 48% of those from high- and low-SES origins are expected to report good SRH, respectively. Among those with 17 years or more, however, there is virtually no difference by SES origin, with a predicted 91% of both high- and low-SES origin white men reporting good SRH. Similarly, good SRH is predicted for 63% of high-SES origin white women with eight years of education compared with 46% of those from low-SES backgrounds. Among those with at least 17 years of schooling, no disparity exists (high- and low-SES: 92%).

The predicted probabilities suggest that the lack of such a pattern among blacks—that is, no resource substitution by SES origin—may have something to do with the fact that health disparities by SES origin are smaller for blacks than for whites, especially at low levels of schooling. For example, among black males with eight years of schooling, 50% of those from high-SES origins are expected to have good SRH, compared with 47% of those from less advantaged backgrounds; the corresponding figures for white men are, as noted above, 62% and 48%. While advantaged family background appears to protect whites with low education from poor health in middle age, it may not do so for blacks.

These results have implications for health disparities by race. Specifically, among those from high-SES origins, the probability of good SRH increases at a similar rate for whites and blacks and for men and women. As a result, racial disparities in the probability of good SRH change very little with education. Meanwhile, among those from low-SES origins, results are indicative of resource multiplication by race. That is, whites evince greater gains in the probability of good SRH with education than do blacks, and consequently, racial health disparities grow with education. For example, at eight years of schooling, there is very little difference in probabilities of good SRH by race among low-SES origin men (whites: 48%; blacks: 47%). At 17 years or more, however, 91% of low-SES origin white men

are expected to report good SRH, much more than the 77% of low-SES origin black men. This pattern is present but less pronounced among women.

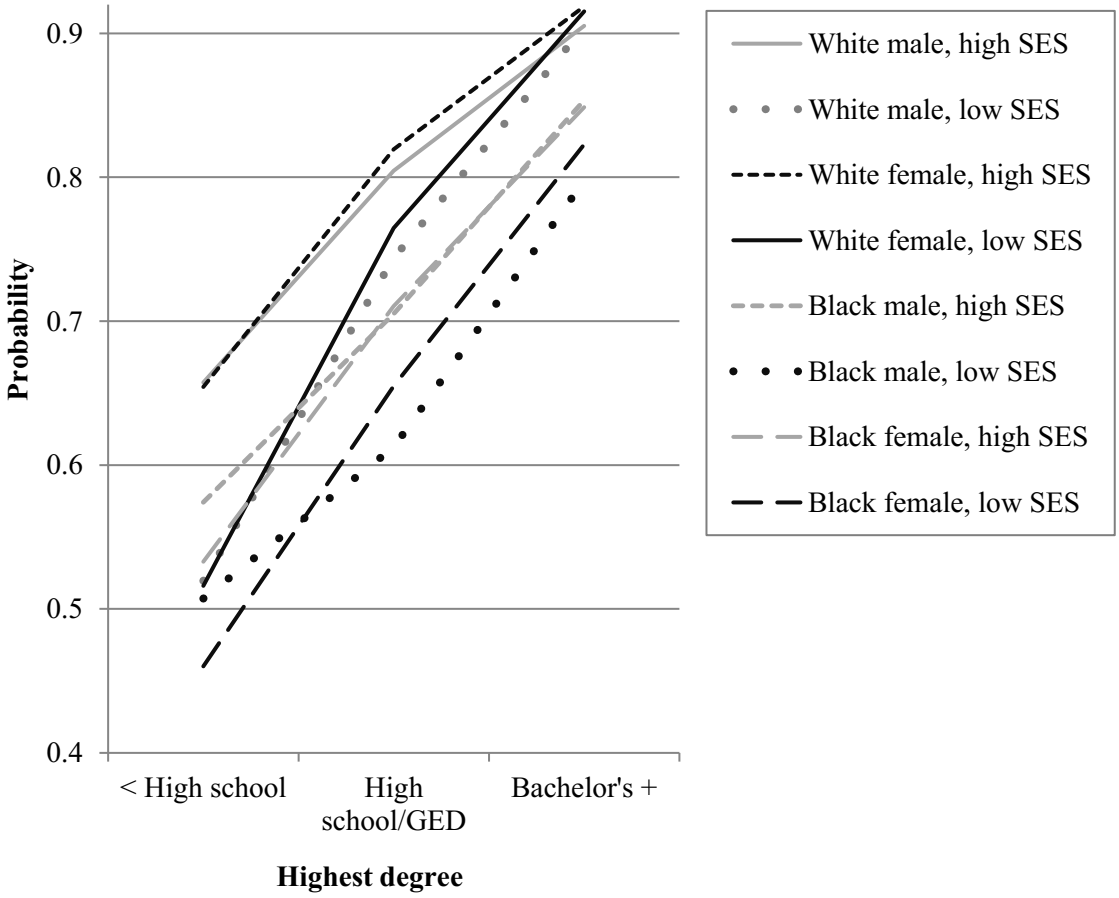
Patterns by gender are remarkably similar for most groups, with at most small disparities at all levels of education. The two groups for which this is not the case are low-SES origin black men and women. At low levels of schooling, low-SES black women are less likely to report good health than low-SES origin black men. However, this disparity narrows and then reverses with education.

Highest degree attained

Probabilities of good SRH are plotted by highest degree attained in Figure 3.3; they are calculated based on Model 3b, results of which are presented in Appendix Table A3.2 and Appendix Figure A3.1. Overall, results are consistent with those above. Probabilities of being in good health increase with degree attainment for all groups. Among whites, disparities by SES origin shrink with each degree attained; no such pattern appears for blacks. Meanwhile, among those from low-SES origins, racial disparities in predicted probabilities grow with degree attainment, while this pattern is not evident for those from advantaged SES backgrounds.

Results further suggest that effects of education on health are more heterogeneous across groups when considering the transition between less than high school and a high school diploma or GED, compared to that between high school and college graduation. The difference in the predicted probability of good SRH between high school graduates and non-graduates is much smaller for low-SES origin black men than other sociodemographic groups, while the effect of attaining a Bachelor's degree or higher is not substantially different for low-SES origin black men compared to other groups. Insofar as education moderates the magnitude of disparities in probabilities of good SRH, it appears to be due largely to differences in returns to moderate levels of schooling across sociodemographic groups.

Figure 3.3. Predicted probability of good self-reported health, by sociodemographic group and highest degree attained



Notes: Predicted probabilities are calculated using Model 3b in Appendix Table A3.2. Age is held at 58 years and covariates are held at their respondent-level means. SES = Socioeconomic origin.

Mortality

In Table 3.3, Model 1, I present results from a hazard model of mortality on years of education and binary indicators of sociodemographic characteristics. Holding sociodemographic characteristics constant, each year of schooling is associated with 10.3% lower odds of mortality ($p < .001$). Those from low-SES origins do not experience significantly higher odds of mortality compared with those from more advantaged backgrounds, though the point estimate falls in the anticipated direction (OR = 1.077, $p = .115$). As expected, blacks have higher odds of mortality than whites (OR = 1.377, $p < .001$) and women have lower odds than men (OR = 0.645, $p < .001$).

Model 2 interacts years of education with the focal sociodemographic variables. Unlike the results for SRH, the association of education with odds of mortality does not differ by SES origin (interaction OR = 1.016, $p = .416$). This finding is consistent with previous studies of mortality and active life expectancy that have also failed to detect an interaction between personal and parental years of schooling (Hayward and Gorman 2004; Montez and Hayward 2011). The association of education with mortality does differ by race, such that blacks experience a smaller reduction in odds of mortality per year of schooling than do whites (interaction OR = 1.067, $p < .001$). Finally, women experience significantly greater reductions in odds of mortality for each year of education attained (interaction OR = 0.925, $p < .001$). Thus, as in the SRH analysis, education appears to substitute for disadvantages stemming from female gender while exacerbating those associated with minority race.

In Model 3, I interact years of education with the eight-category indicator of sociodemographic group. The association of education with odds of mortality differs significantly across groups (interaction $p < .001$), as expected given the results of Model 2. As shown in Figure 3.4 and Appendix Table A3.3, odds ratios are below one and statistically significant for most groups. As in the analysis of SRH, low-SES origin white women experience the greatest effect of education on odds of mortality (OR = 0.826, $p < .001$), while odds of mortality decline the least—and in fact, do not decline significantly—for low-SES origin black men (OR = 0.979, $p = .345$).

Table 3.3. Odds ratios (ORs) from logistic regression models of mortality

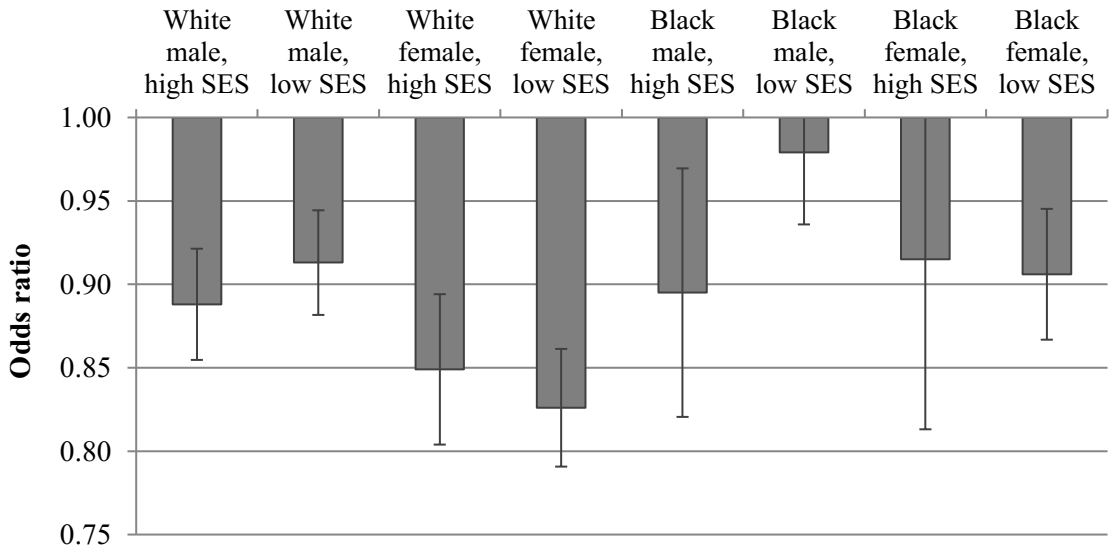
	Model 1	**Model 2**	**Model 3**
	OR (SE)	OR (SE)	OR (SE)
Education	0.897 (0.007) ***	0.896 (0.014) ***	0.888 (0.017) ***
Low SES origin (Ref: High)	1.077 (0.051)	1.069 (0.051)	-
Black (Ref: White)	1.377 (0.068) ***	1.531 (0.082) ***	-
Female (Ref: Male)	0.645 (0.026) ***	0.589 (0.026) ***	-
Sociodemographic group			(p < 0.001)
High-SES white males	-	-	Ref.
Low-SES white males	-	-	0.987 (0.069)
High-SES white females	-	-	0.535 (0.039) ***
Low-SES white females	-	-	0.596 (0.045) ***
High-SES black males	-	-	1.313 (0.162) *
Low-SES black males	-	-	1.613 (0.156) ***
High-SES black females	-	-	0.869 (0.112)
Low-SES black females	-	-	0.948 (0.085)
Education X Low SES origin	-	1.016 (0.019)	-
Education X Black	-	1.067 (0.018) ***	-
Education X Female	-	0.925 (0.015) ***	-
Education X Sociodemographic group			(p < 0.001)
High-SES white males	-	-	Ref.
Low-SES white males	-	-	1.022 (0.018)
High-SES white females	-	-	0.958 (0.023) †
Low-SES white females	-	-	0.956 (0.020) *
High-SES black males	-	-	0.995 (0.038)
Low-SES black males	-	-	1.044 (0.024) †
High-SES black females	-	-	1.000 (0.050)
Low-SES black females	-	-	1.010 (0.024)
N observations (N respondents)	191,685 (16,439)	191,685 (16,439)	191,685 (16,439)
Pseudo R^2	0.0447	0.0461	0.0465

Notes: Years of education is centered at the mean across respondents (13.12 years). Models control for age, age-squared, year of birth, the interaction between age and year of birth, the interaction between age-squared and year of birth, and childhood health. Standard errors are adjusted for clustering at the household level.

*** p < .001; ** p < .01; * p < .05; † p < .1 (two-tailed test)

Ref. = Reference group; SE = Standard error; SES = Socioeconomic origin

Figure 3.4. Effects (odds ratios) of a one-year increase in education on odds of mortality, by sociodemographic group



Notes: Group-specific odds ratios are calculated using Model 3 in Table 3.3. See Appendix Table A3.3 for further information. Error bars represent 95% confidence intervals. SES = Socioeconomic origin.

Here, however, the protective effect of education on odds of mortality is no greater for whites from low-SES origins than for those of higher-SES backgrounds. Thus, while Model 2 shows that there is no support for resource substitution by SES origin overall, Model 3 confirms that resource substitution by SES origin does not exist even among whites when considering mortality in middle and older age.

Results regarding race are more consistent with those for SRH, though the three-way interaction between education and binary indicators of SES origin and race is not statistically significant ($p = .305$) in a model of mortality on all possible three-way interactions between education and binary measures of SES origin, race, and gender. Still, there is suggestive evidence that the pattern of resource multiplication by race is primarily present for those of low-SES origin.[19] As noted above, the association of education with odds of mortality is smallest (and insignificant) among low-SES origin black men (OR = 0.979, $p = .345$); it is significantly larger in magnitude for white men from low-SES backgrounds (OR = 0.913, $p < .001$) (difference $p = .013$). Likewise, low-SES origin black women evince significantly smaller reductions in odds of mortality per year of schooling (OR = 0.906, $p < .001$) than white women (OR = 0.826, $p < .001$) (difference $p = .003$). In contrast, among those of high-SES origin, white (OR = 0.888, $p < .001$) and black (OR = 0.895, $p = .009$) men experience a similar reduction in odds of mortality per year of schooling (difference $p = .864$), though white women (OR = 0.849, $p < .001$) do experience a marginally greater effect than black women (OR = 0.915, $p = .121$) (difference $p = .058$).

Finally, while results from Model 2 suggest that women experience greater reductions in odds of mortality per year of schooling than men, Model 3 suggests that this is primarily among those from low-SES origins. Low-SES origin white women experience a 17.4% decline in odds of mortality per year of education, compared with 8.7% among men. Among low-SES origin blacks, women evince a 9.4% decline in odds of mortality per year of schooling, compared to an insignificant 2.1% among men. This pattern is marginally statistically significant, as confirmed in a model including all possible three-way interactions between the four focal variables (education and binary indicators of SES origin, race, and

[19] These findings are also evident when estimating Model 2 separately by SES origin. The interaction between education and black race indicates resource multiplication among those from low-SES origins (OR = 1.081, $p < .001$) but not those from high-SES origins (OR = 1.021, $p = .552$).

gender); the interaction between education, SES origin, and gender has a p-value of .080.[20]

No other three-way interaction is statistically significant, nor is the four-way interaction between education and binary indicators of SES origin, race, and gender, which was estimated in a separate model ($p = .622$).

Predicted probabilities

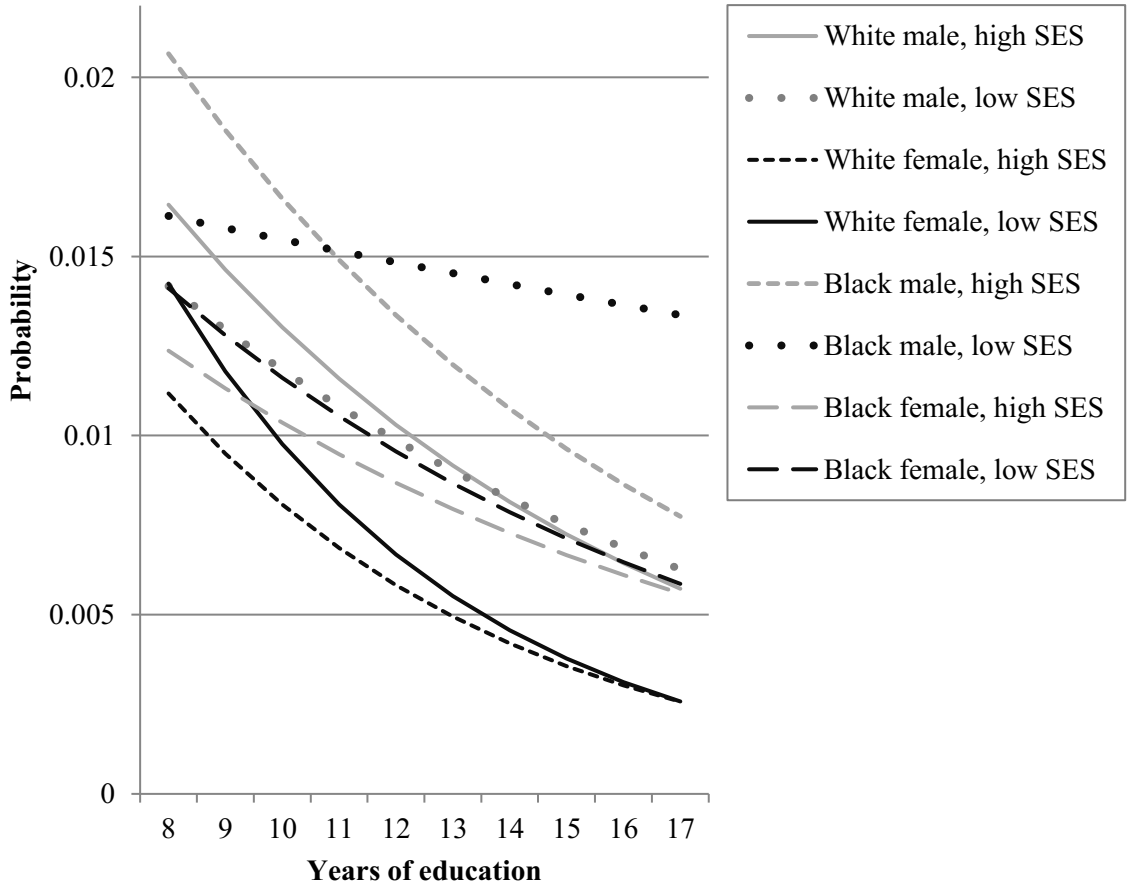
Predicted probabilities of mortality at age 58, constructed using results of Model 3 in Table 3.3, are presented by years of schooling and sociodemographic group in Figure 3.5. The probability of mortality declines with education for all groups. But as with the analysis of SRH and unlike what might be predicted by theories of resource substitution and multiplication, variation in probabilities across groups does not decline nor grow substantially at high levels of education.

At low levels of schooling, black males from high-SES origins have the highest probability of death (e.g., 0.021 at eight years of schooling). Mortality rates for other groups range from about 0.011 to 0.016 at eight years of schooling. The most striking pattern across years of schooling is seen for low-SES origin black men, among whom the probability of mortality declines very little with education in comparison to the other groups. Among those with 17 or more years of schooling, low-SES origin black males have a probability of death that is nearly double that of the next-highest group (0.0134 versus 0.0077). This finding parallels the SRH analysis, where the probability of good health improved with education least for low-SES origin black men.

Also consistent with the SRH analysis, disparities in the probability of mortality by race increase with education among those from low-SES origins. The probability of death among those with eight years of schooling is 0.014 for low-SES origin white men compared with 0.016 for low-SES origin black men; at 17 or more years of schooling, probabilities among low-SES origin black men (0.013) are more than

[20] These findings are also evident when estimating Model 2 separately by SES origin. The interaction between education and female gender suggests a pattern of resource substitution for those from low-SES origins (interaction OR = 0.913, $p < .001$) but not those from high-SES origins (interaction OR = 0.968, $p = .264$).

Figure 3.5. Predicted probability of mortality, by sociodemographic group and years of education



Notes: Predicted probabilities are calculated using Model 3 in Table 3.3. Age is held at 58 years and covariates are held at their respondent-level means. SES = Socioeconomic origin.

double that of low-SES origin white men (0.006). Among women from disadvantaged backgrounds, a similar pattern by race exists, with equivalent probabilities by race at eight years of schooling (both 0.014) but disparate probabilities at 17 years of more (0.003 among whites and 0.006 among blacks). Racial disparities in the probability of mortality do not grow at all among high-SES origin men and grow only slightly among high-SES origin women.

Highest degree attained

Results of Model 3b, which operationalizes education using highest degree attained, can be found in Appendix Table A3.4 and Appendix Figures A3.2 and A3.3. Results align with those presented above. However, as with SRH, the reduced effect of education on mortality for low-SES origin black men compared with other groups is more pronounced when considering the impact of a high school education or GED versus than when considering the effect of a Bachelor's degree. Likewise, among low-SES origin women, racial disparities in the probability of mortality are relatively small among those with less than a high school education and increase among those with a high school diploma or GED—in line with the theory of resource multiplication by race. However, the magnitude of the racial disparity declines among those with a Bachelor's degree or higher. In sum, differences in returns to moderate levels of schooling across sociodemographic groups appear to drive the patterns reported here.

DISCUSSION

This article draws on theories of resource substitution and multiplication, which purport that the effect of education on health will differ across the sociodemographic groups that structure access to health-promoting resources (Ross and Mirowsky 2006, 2010). Consistent with the logic behind both of these theories, I find that, while education is positively associated with health for all, the magnitude of this association differs significantly across groups defined by SES origin, race, and gender. This finding is robust across analyses of self-reported health and mortality.

However, the complex intersectional perspective I take shows that support for theories of

resource substitution and multiplication does not apply uniformly across the population. While prior work finds evidence of resource substitution by SES origin—larger effects of education on health for those from disadvantaged families (Andersson and Vaughan 2017; Bauldry 2015; Luo and Waite 2005; Ross and Mirowsky 2011; Schaan 2014; Schafer et al. 2013)—I find that this pattern is only evident among whites. Among blacks, the association of education with health is similar regardless of SES origin.

Importantly, this finding applies to self-reported health but not to mortality. This is in line with prior work. While a sizeable body of research finds stronger associations of education with self-reported health, physical functioning, and depressive symptoms among those from disadvantaged SES origins (Andersson and Vaughan 2017; Bauldry 2015; Luo and Waite 2005; Ross and Mirowsky 2011; Schaan 2014), studies of mortality and severe medical diagnoses have failed to find such a pattern (Hayward and Gorman 2004; Montez and Hayward 2014; Nandi et al. 2012). Perhaps this is because behaviors that are clearly linked to mortality are correlated with education to a similar extent regardless of SES background. Research on educational disparities in smoking supports this idea (Andersson and Maralani 2015).

Among black respondents, I find no evidence of resource substitution by SES origin. However, it is worth highlighting that disparities in self-reported health by SES origin are small for blacks no matter their level of education. That is, while high-SES origin protects less educated whites from poor health, they do not appear to do the same for blacks. This finding aligns with prior research on health (Walsemann et al. 2016) as well as studies showing, for example, that parental wealth is less predictive of offspring wealth for blacks than whites (Pfeffer and Killewald 2015) and that household stability is less important for educational outcomes among black than white children (Perkins 2019).

Relatedly, my results show that education is less strongly associated with self-reported health and mortality for blacks than whites, thus providing evidence of resource multiplication by race. This resonates with existing research (Farmer and Ferraro 2005; Holmes and Zajacova 2014; Hummer and Lariscy 2011; Masters et al. 2012; Shuey and Willson 2008; Zajacova and Hummer 2009) while adding an additional piece to the puzzle. Specifically, resource multiplication by race is evident primarily among those from low-SES backgrounds.

Though I cannot test the mechanisms underlying these patterns directly, the theory of skin-deep resilience (Brody et al. 2013; Chen et al. 2015; Miller et al. 2015) and the related concept of John Henryism (James et al. 1987; James 1994) suggest that results may be driven in part by the additional stressors that accompany upward mobility for racial minorities compared to whites from low-SES origins. When pursuing higher education, low-SES origin racial minorities must strive in the face of not only an unfamiliar class environment (Stephens 2012a, 2012b), but also a heightened threat of health-harming discrimination (Borrell et al. 2006; Colen et al. 2018; Goosby et al. 2018; Hochschild 1995; Lewis et al. 2015; Pew Research Center 2016; Williams and Mohammed 2009). Stress due to discrimination could uniquely compromise the health of highly-educated blacks from low-SES origins, producing the smaller positive effects of education on health for low-SES blacks compared with low-SES whites seen here and elsewhere (Gaydosh et al. 2018).

Labor market discrimination could also explain the depressed associations of education with health among low-SES blacks. Prior research shows that employers discriminate against black applicants using names (Bertrand and Mullainathan 2004) and accents (Fischer and Massey 2004). These signals of minority group membership may be more common among blacks from low-SES origins than among blacks from relatively advantaged backgrounds (Gaddis 2017), and they are unlikely to be found among whites regardless of socioeconomic background. Discrimination may thus prevent low-SES origin blacks from using their education for upward mobility, weakening the effects of education on health.

Notably, I find that improvements in health associated with education are smallest not for the groups predicted by theories of resource substitution or multiplication—high-SES origin white men or low-SES origin black women, respectively—but instead for low-SES origin black men. This may be because discrimination is particularly pervasive for highly educated black men from low-SES origins, offsetting the positive effects of education on later health. Consider, for example, that discrimination is more common among black men than women (Borrell et al. 2006; Pew Research Center 2016) and, as noted above, among upwardly mobile blacks. In any case, this result is discouraging, given that low-SES origin black men are also the group most likely to experience poor health and mortality in general.

On a more positive note, results suggest that, while low-SES origin black men gain less from a high school education than other groups in terms of both self-reported health and mortality, there is less heterogeneity across groups in the effect of earning a Bachelor's degree. Moreover, while Gaydosh and colleagues (2018) find that the probability of presenting with metabolic syndrome (a multifaceted condition involving high blood pressure, high blood sugar, and obesity) increases with college education among low-SES origin blacks, I find that education is not detrimental to the health of any group—it invariably predicts better overall health.

While results by gender are less straightforward than those by SES origin or race, nowhere is there evidence that men accrue greater health returns to schooling than women. Where the association of schooling with health does differ by gender, it is stronger for women. This pattern is indicative of resource substitution, as female gender is associated with resource disadvantage in terms of income, status, power, and authority (Ross and Mirowsky 2006).

This result differs somewhat from prior research. While Ross and colleagues (2012) find that the association of education with self-reported health is stronger for women (a pattern that I reproduce here), they find that the association of education with mortality is stronger for men. Differences in our results may be driven by the age compositions of our samples: Ross and colleagues' (2012) sample is diverse with respect to age, including those ages 25 and up, while the current analysis is limited to those of middle and older age. Indeed, Ross et al. (2012) find that the heightened association of education with mortality for men is driven by causes of death that disproportionately afflict younger adults (e.g., accidents, violence, smoking-related diseases). Note also that Zajacova and Hummer (2009) show that there are no systematic gender differences in educational effects on mortality among older Americans, besides stronger returns to the highest levels of schooling among white men.

A few limitations of the current study bear mentioning. First, while the sample used here is nationally representative of the U.S. population ages 50 to 64, the subsample of non-Hispanic blacks is relatively small. Thus, estimates are less precise among blacks than among whites, particularly among blacks from high-SES origins. Future research using a larger sample of non-Hispanic blacks and

extending the analysis to other ethnic groups (e.g., Hispanics) is warranted.

Second, I analyze two common measures of general health: self-reported health and mortality. It may be that the patterns I explore vary depending on the etiology of the health condition in question. Studying heterogeneity in the association of education with a diverse range of health conditions may suggest mechanisms producing the patterns described here.

Further, the theories I evaluate propose that the association of education with health may vary depending on the value of health-promoting resources a person has access to outside of their education. Following previous research (and the theories themselves), I do not measure these resources directly. Instead, they are proxied using key sociodemographic characteristics across which they are structured.

While this strategy is useful to evaluate support for these theories, it also leaves unanswered questions regarding which resources matter and why. This is particularly troubling given the mixed support for resource substitution and resource multiplication. Why might education substitute for resources lacking due to disadvantaged SES origin (among whites), but not for those that are absent for those of minority race (among those from low-SES origins)? A few potential explanations are provided above. Namely, it may be that access to resources is not all that matters. So too does the ability to successfully deploy those resources for health promotion and to limit stress that may accompany higher education (Masters et al. 2015), both of which may be weakened by discrimination. Future research would therefore do well to study resources and mechanisms directly.

Scholars have come to refer to education as a great equalizer, as disparities in economic outcomes by SES origin are smallest among college graduates (Hout 1984, 1988; Torche 2011). The theory of resource substitution likewise predicts that education equalizes health outcomes across sociodemographic groups (Ross and Mirowsky 2006, 2010). While prior research generated substantial support for this theory with respect to SES origin, my results suggest that this optimism must be tempered. Education does not appear to eliminate health disparities by race and in fact may exacerbate them, particularly among those from disadvantaged backgrounds. More generally, disparities in the probabilities of good health and mortality differ substantially even among the highly educated. Echoing others (Masters et al.

2015; Montez and Friedman 2015), my results underscore the need to situate education within the contours of broader society. As long as the achievements of some sociodemographic groups are rewarded less generously than others, the effects of those achievements on health are likely to be muted.

Chapter 4: Unconditional Quantile Regression and Educational Disparities in Biomarkers of Health Risk

Decades of research documents that, on average, those with more education have better health (Cutler and Lleras-Muney 2008; Hummer and Lariscy 2011). Recent studies have refined our understanding of the association between education and health by, for example, studying its optimal functional form (Montez et al. 2012) and assessing whether it is modified by sociodemographic characteristics (Zajacova and Hummer 2009) or by features of the institutional environment (Montez et al. 2017a; Montez et al. 2019). However, the shape of the relationship between education and health remains incompletely mapped, as prior work has focused on educational disparities at single points in the distribution of health.

Studies of binary measures in effect estimate educational disparities in the probability of negative health events, such as mortality (Masters et al. 2012; Miech et al. 2011). When health is operationalized using continuous measures, models generally estimate differences in conditional means (Brummett et al. 2011; Gruenewald et al. 2009). Very little research sheds light on education's association with health across the illness-wellness continuum. Are educational disparities of the same magnitude at all points in the distribution of health?

The answer to this question may rely on key observations from fundamental cause theory (Link and Phelan 1995; Phelan et al. 2010). High socioeconomic status (SES)—of which education is a key component—is associated with material resources, social connections, knowledge, and skills (Link and Phelan 1995; Mirowsky and Ross 2003). The utility of these and other (Freese and Lutfey 2011) resources for health promotion is not guaranteed, however, as they can only be marshaled when preventive methods or treatment options exist. Indeed, prior research finds larger SES disparities in health conditions that are amenable to prevention or intervention (Chang and Lauderdale 2009; Glied and Lleras-Muney 2008; Masters et al. 2015; Phelan et al. 2004; Phelan and Link 2005; Tehranifar et al. 2009). When there are no known methods to prevent or treat a health problem, or when such methods are

implemented universally (Clouston et al. 2016), resources associated with SES are of limited use.

The same logic should apply to educational disparities across the distribution of health or related risk. Disparities are likely to be largest at points in the distribution at which education-linked resources can be leveraged to improve outcomes most effectively. When medical treatments exist, this is likely to be in the unhealthy tail of the distribution, as medical interventions are typically only triggered—and therefore become relevant for health disparities—when some elevated threshold of risk is crossed.

Using data from the Health and Retirement Study, a survey of older adults in the United States, I study variation in the magnitude of educational disparities across unconditional quantiles of blood sugar, blood pressure, and "good" and "bad" cholesterol. I expect educational disparities in blood sugar, blood pressure, and bad cholesterol to be greatest at their least healthy levels, given that medical intervention on these measures—access to and efficacy of which may vary by education—is only initiated once the clinical threshold of risk is surpassed. I do not expect the same result to obtain for good cholesterol, which is not regularly targeted by medication. Further, while the biomarkers I study are all influenced by diet, physical activity, and tobacco use (American Diabetes Association [ADA] 2018; Grundy et al. 2018; Whelton et al. 2018), I do not expect lifestyle to fully explain the observed variation in disparities across their distributions.

In addition to further describing the shape of the relationship between education and health and testing an implication of fundamental cause theory, this study thus also hints at mechanisms. Specifically, it posits that differential access to and experience within the health care system underlies educational disparities in biomarkers of health risk, particularly at unhealthy levels. Identifying the mechanisms producing disparities at unhealthy points in the distribution of blood sugar, blood pressure, and cholesterol should be of high priority, as this is precisely where change is most strongly associated with risk of subsequent morbidity and mortality (Coutinho et al. 1999; Lewington et al. 2002; Navarese et al. 2018; Stamler et al. 2000).

The current study also has methodological implications for future research on biomarkers of health risk, which are now available in many social surveys (Harris and Schorpp 2018; McDade et al.

2007). Studying conditional means may mask disparities that exist primarily or even only at the least healthy biomarker levels. Yet disparities at levels of risk below clinical thresholds should not be ignored, as they may also have meaningful consequences for subsequent health (Brunner et al. 2006; Coutinho et al. 1999; Lewington et al. 2002; Rapsomaniki et al. 2014; Selvin et al. 2010; Vasan et al. 2001).

INSIGHTS FROM FUNDAMENTAL CAUSE THEORY

Link and Phelan's (1995; Phelan et al. 2010) fundamental cause theory urges scholars to consider what structural conditions put people "at risk of risks" to their health. They identify SES as one such condition, as it affects various aspects of health through numerous mechanisms. Perhaps the most obvious mechanisms operate at the individual level, as SES comprises multiple flexible resources—money, social connections, knowledge and skills, power and prestige—that can be deployed to prevent poor health or ameliorate it should it occur. Mirowsky and Ross (2003) likewise argue that education facilitates learned effectiveness and along with it the motivation and skills required to construct a healthy lifestyle and adapt to new health needs.

Freese and Lutfey (2011) elaborate on structural mechanisms that link SES with health without purposive action or health-directed agency on the part of the individual. Those of high-SES may benefit from spillover effects of the actions of others in their network. SES-related variation in habitus (Bourdieu 1984) may drive disparities in health-related behaviors (Cockerham 2005; Cutler and Lleras-Muney 2010; Pampel et al. 2010). And, status or prestige can benefit those of high-SES without their own purposive action because it influences perceptions by those in the healthcare system and other institutions and therefore affects experience (Spencer and Grace 2016). In sum, SES is associated with both individual and structural resources that can support health.

Despite SES being a fundamental cause of health, disparities are not evident for all health outcomes. This is to be expected given the mechanisms thought to link SES with health. When preventive methods and treatments do not exist, SES-related resources cannot be marshaled—actively or passively—to support health. Thus, SES disparities are larger for health outcomes that are preventable or treatable

(Chang and Lauderdale 2009; Glied and Lleras-Muney 2008; Masters et al. 2015; Phelan et al. 2004; Phelan and Link 2005; Tehranifar et al. 2009), especially those for which prevention and treatment is not yet universal (Clouston et al. 2016).

EDUCATION AND THE PROMOTION OF HEALTH ACROSS ITS DISTRIBUTION

The current study expands on observations from fundamental cause theory. I posit that, for aspects of health that are amenable to prevention or treatment, education is likely to be more important at points in the distribution where a wider variety of resources can promote positive outcomes, or where such resources are particularly effective. Put in descriptive terms, I ask whether educational disparities vary in magnitude across the distribution of health or related risk. I expect that they do, such that disparities are larger at unhealthy levels.

There is little prior research on this topic. Studies find that education is associated with both an increased likelihood of good self-reported health (SRH) and a reduced likelihood of bad SRH (Hardy et al. 2014; Mackenbach et al. 1994; Reile and Leinsalu 2013). These results are sensitive to the choice of reference group—what is considered medium health—so the magnitude of education's association with good and bad SRH cannot be readily compared. Moreover, SRH is subjective. While people reflect on diagnosed health conditions and symptoms when reporting being in poor health, a wider range of factors appears to underlie reports of good health (Mackenbach et al. 1994; Martinez-Sanchez and Regidor 2002; Reile and Leinsalu 2013; Shields and Shooshtari 2001; Shooshtari et al. 2007). Considerations may also differ systematically across sociodemographic groups (Chaparro et al. 2019; Dowd and Zajacova 2007, 2010). Put simply, SRH is not a suitable measure of health for investigating the question at hand, in part because it is not continuous and in part because it is subjective.

Barcellos et al. (2019) instead study variation in the effect of education on later health using continuous, objective outcomes. Exploiting changes in the minimum school-leaving age in the United Kingdom, the authors find that reductions in body size induced by additional schooling were largest at unhealthy points in its distribution. Meanwhile, education actually increased blood pressure, but only at

low (healthy) levels. Given its regression discontinuity design, Barcellos and colleagues' estimates are not representative of the population; they reflect individuals who would have attained a very low level of schooling in the absence of the reform. Still, they provide evidence that long-term causal effects of education differ across the distribution of health for some subset of the population.

Moreover, Barcellos et al. (2019) highlight the practical significance of distributional variation in the relationship between education and health. Education was found to have no effect on rates of hypertension, as education's impact on blood pressure occurred only within its healthy range. Meanwhile, the concentration of education's effects in the unhealthy tail of body size is important not only because it resulted in reduced rates of obesity, but also because the relationship of body size to morbidity and mortality is nonlinear such that reductions at its highest levels are likely to have the greatest impact on subsequent health (Aune et al. 2016). The same may be true of the biomarkers studied here, including blood sugar (Coutinho et al. 1999), blood pressure (Lewington et al. 2002), and cholesterol (Navarese et al. 2018; Stamler et al. 2000).

Descriptive research on educational disparities across the range of biomarkers is therefore likely to return methodological implications for future work, which is timely given the recent proliferation of such measures in social science datasets (Harris and Schorpp 2018; McDade et al. 2007). Common analytic strategies, including the analysis of conditional means, might mask disparities in biomarkers that exist only or primarily in particular segments of their distributions. That said, if educational disparities exist across the range of biomarkers, studies that dichotomize measures following clinical guidelines will overlook unequal outcomes at moderate levels, which also have consequences for subsequent health (Brunner et al. 2006; Coutinho et al. 1999; Lewington et al. 2002; Rapsomaniki et al. 2014; Selvin et al. 2010; Vasan et al. 2001).

Educational attainment and the prevention and treatment of poor health

Following key tenets of fundamental cause theory (Link and Phelan 1995; Phelan et al. 2010), I anticipate educational disparities in some biomarkers of health risk to be greatest at their least healthy

levels. I do not expect this distributional variation to be explained by lifestyle factors alone. Of course, more educated people tend to follow healthier diets, engage in more frequent physical activity, and avoid tobacco use (Cutler and Lleras-Muney 2010; Pampel et al. 2010). And, these behaviors affect health: healthy diet, exercise, and non-smoking are known to reduce risk of heart disease, stroke, kidney disease, Type 2 diabetes, lung disease, various cancers, and joint pain (ADA 2018; Grundy et al. 2018; Whelton et al. 2018). As a result, such behaviors partially mediate the relationship between education and health (Cutler and Lleras-Muney 2008; Brummett et al. 2011).

However, I anticipate these lifestyle factors to mediate educational disparities to a similar extent across the distribution of health. Lifestyle not only affects the likelihood of experiencing severe health events; it is also associated with reductions in indicators of health-related risk even at low and modest levels (Bottai et al. 2014; Moon et al. 2017). As a result, I do not expect distributional variation in biomarker disparities to be eliminated when holding constant proxies for health lifestyle.

While lifestyle may link education with health at various points in its distribution, additional education-related resources are relevant primarily when a health threat emerges or when some threshold of risk is crossed. Risk factors may be discovered during routine medical examinations or when seeking attention for concerning symptoms. Resources held by the highly educated—private medical insurance, money to pay out-of-pocket expenses, generous workplace leave policies, norms regarding healthcare utilization—may lead to earlier assessment and treatment (Ayanian et al. 2000; Ayanian et al. 2003).

Once a threat to health is identified, SES-related resources are likely to affect both treatment decisions and likelihood of successful adherence to medical advice. Treatment options vary in cost and complexity, and the most expensive, complicated regimens are often the most effective. Thus, economic resources and memory are required to achieve optimal control of many conditions (Lutfey and Freese 2005; Osterberg and Blaschke 2005).

Less educated individuals may also struggle to adhere to medical advice regarding medication, health monitoring, and behavior change due to constraints imposed by their job, household, or social network (Lutfey and Freese 2005; Osterberg and Blaschke 2005). Consider that taking medications at the

same time every day—especially when multiple doses are required—may be difficult for people whose work schedule changes from week to week or whose job does not permit break time, work conditions that are more common among the less educated. The status that accompanies high education may also affect treatment and improve outcomes, as physicians may instruct low-SES patients to follow basic medication regimens regardless of their willingness or ability to try alternatives (Lutfey and Freese 2005; Spencer and Grace 2016; van Ryn and Burke 2000; van Ryn et al. 2006).

In sum, I expect to find larger educational disparities in several biomarkers of health risk at the least healthy points in their distributions. This has major implications for socioeconomic disparities in health, as differences at the least healthy levels of the biomarkers I study are likely to have a greater impact on morbidity and mortality than those at moderate and lower levels (Coutinho et al. 1999; Lewington et al. 2002; Navarese et al. 2018; Stamler et al. 2000). I do not expect patterns to be explained fully by measures of health lifestyle, which may affect biomarkers to a similar extent across their distributions. Instead, distributional variation in disparities may be driven by differential access to and efficacy of medical intervention by education level, as medical interventions become relevant for health and health disparities only after some elevated level of risk is surpassed.

Put differently, in the absence of medical interventions, educational disparities in biomarkers might exist due to lifestyle factors; however, disparities may be relatively uniform in magnitude across the biomarkers' distributions. Medical interventions might exacerbate the magnitude of disparities primarily in their unhealthy tails. There, differences in diagnosis, treatment assignment, and treatment success drive additional gaps in outcomes by pulling the unhealthy tail in among those with more education-linked resources. This distributional pattern should therefore be found primarily for biomarkers that are subject to medical intervention (Chang and Lauderdale 2009; Glied and Lleras-Muney 2008; Masters et al. 2015; Phelan et al. 2004; Phelan and Link 2005; Tehranifar et al. 2009). Hypotheses regarding the specific biomarkers I study are derived below.

Biomarkers of health risk

Chronic diseases are now extremely common among U.S. adults, affecting six in ten (Buttorff et al. 2017). Risk of chronic disease and related morbidity and mortality is heightened among those with unhealthy biomarker profiles (ADA 2018; Grundy et al. 2018; Whelton et al. 2018). Thus, I study biomarkers of health risk, including blood sugar, blood pressure, and both "good" and "bad" cholesterol. I assess whether educational disparities in these biomarkers vary across their distributions, and I test whether indicators of health lifestyle explain the observed variation. I briefly discuss each measure here, ending with hypotheses.

Blood sugar: High blood sugar slowly damages the eyes, kidneys, nerves, heart, and blood vessels (ADA 2018). It often exhibits few outward signs until major complications develop, making early detection critical for health promotion. High blood sugar is characteristic of diabetes, which is estimated to affect 12% of American adults (Centers for Disease Control [CDC] 2017). Unfortunately, 3% of Americans, or nearly one in four with diabetes, do not know they have it (CDC 2017).

The prevalence of undiagnosed diabetes is higher among those with less education. The CDC (2017) estimates that 4% of American adults who did not complete high school have diabetes but do not know it; comparable figures for high school graduates and those with more than a high school education are 3% and 2%, respectively. Similarly, while 28% of those without a high school diploma have blood sugar levels indicative of pre-diabetes but do not know it, just 25% of high school graduates and 20% of those with more than a high school education have unreported pre-diabetes.

If, due to lack of diagnosis, those with low education are less likely manage their blood sugar successfully, disparities at the highest levels of blood sugar are likely to be greater than those at low or moderate levels. Even among those who receive a diabetes diagnosis, treatment success might vary by education, as the best method for controlling blood sugar takes trial and error to find and may ultimately be both complex and expensive, leading to different treatment assignment and/or poorer medication adherence among those with less education (Lutfey and Freese 2005; Spencer and Grace 2016).

Blood pressure: High blood pressure or hypertension is a risk factor for heart attack and stroke

(Whelton et al. 2018). The most recent guidelines classify nearly half (46%) of American adults as hypertensive (Whelton et al. 2018). And, as is the case with high blood sugar, high blood pressure is more common among U.S. adults of low-SES and education (Gillespie and Hurvitz 2013; National Center for Health Statistics 2014).

Early detection and treatment is critical for managing the impact of high blood pressure on cardiovascular outcomes. When high blood pressure is discovered, medication is typically prescribed to bring it down to healthy levels, and consistent monitoring is recommended (Whelton et al. 2018). Thus, like blood sugar, I expect educational disparities in blood pressure to be largest at its least healthy levels.

The mechanisms underlying this pattern may differ from those described for blood sugar. Prevalence of undiagnosed hypertension is no higher among those of low education (Ayanian et al. 2003; Paulose-Ram et al. 2017). Instead, differential treatment plans or rates of adherence to medical advice may produce larger disparities at high levels of blood pressure. Research shows, for example, that among those with diagnosed hypertension, likelihood of successful control varies by SES, with those of higher SES achieving acceptable levels of blood pressure more often than those of low SES (Gillespie and Hurvitz 2013; National Center for Health Statistics 2014).

Cholesterol: High total cholesterol is estimated to affect 12.4% of U.S. adults (Carroll et al. 2017). Total cholesterol is comprised of both bad (non-high density lipoprotein, or non-HDL) and good (HDL) forms of the compound. Non-HDL cholesterol, particularly low-density lipoproteins (LDL), can build up in arteries; it is therefore associated with cardiovascular disease (Grundy et al. 2018).

Chang and Lauderdale (2009) show that, in the early 2000s, average LDL cholesterol declined with income, a finding that is attributed to greater uptake of medical innovations (statins and related medications) by those of high-SES. While statin use has diffused widely since the early 2000s, it remains underutilized among some groups, including the uninsured and racial and ethnic minorities (Salami et al. 2017), who may have lower average levels of education. Thus, I again expect greater educational disparities in bad (non-HDL) cholesterol at higher levels.

HDL is known as the good form of cholesterol because it assists in the removal of bad cholesterol

from the body; higher levels have traditionally been viewed as desirable. HDL cholesterol is not, however, a widespread target of medical intervention, as the causal effect of circulating levels of HDL on cardiovascular outcomes is debated (Rader and Hovingh 2014; Toth et al. 2013). HDL cholesterol is thus influenced primarily by lifestyle factors including diet and exercise; it may also be raised inadvertently by drugs targeting bad forms of cholesterol. It is therefore a useful negative case for the current study.

Summary: The research outlined above leads me to the following hypotheses:

- For blood sugar, blood pressure, and bad (non-HDL) cholesterol, educational disparities will be greatest at unhealthy levels. As outlined in the previous section, I do not expect that variation in disparities across the distributions of these biomarkers will be explained by mediating lifestyle factors.
- While education will be associated with higher good (HDL) cholesterol on average, this association will not change across the range of HDL. In particular, I do not predict a greater association of education with unhealthy levels of HDL.

METHODS

Data

I draw data from the Health and Retirement Study (HRS), a panel survey of U.S. households funded by the National Institute on Aging (RAND 2016). The HRS began in 1992 with a sample of U.S. adults born between 1931 and 1941 and their spouses; additional birth cohorts have since been added to the sample, and respondents have been followed up biennially. Since 1998, the HRS has surveyed a nationally representative sample of the U.S. population over age 50 and their spouses. Over 30,000 individuals have contributed data to the HRS.

The HRS began collecting physical measurements, including blood pressure and dried blood spots for biomarker assessment, in 2006 (Crimmins et al. 2013, 2015). At that time, half of HRS households were invited to provide an extended in-home interview; the other half was invited to do so in 2008. New households and those that had participated in 2006 were invited to do so again in 2010, and in

2012, the same was done for those who participated in 2008. Thus, physical measurements were assessed up to two times per respondent, either in 2006 and 2010 or in 2008 and 2012.

Eligibility for the current analysis requires that the respondent was included in the physical measurement sub-sample at some point between 2006 and 2012 and that they were age 50 or above at the time. There are 27,748 observations from 18,902 respondents matching these criteria. Of these, 23,908 observations from 17,370 respondents have non-missing information across the four biomarkers I study. My regression analyses include the 20,927 observations from 15,077 respondents with no missing data across the covariates I use.[21]

Measures

Biomarkers of health risk

The first biomarker I study is hemoglobin-A1C (A1C) (ADA 2018; Hanas and John 2010), an indirect measure of average blood glucose or blood sugar over the course of several months. A1C values greater than or equal to 5.7% are indicative of pre-diabetes, while levels 6.5% or above justify a diabetes diagnosis. In diagnosed diabetics, A1C indicates how well treatment is working to keep blood sugar under control. Among non-diabetics, increasing A1C—even at modest levels—is associated with heightened risk of future diabetes, heart disease, stroke, and death (Brunner et al. 2006; Coutinho et al. 1999; Selvin et al. 2010).

Next, I assess systolic blood pressure (SBP), which is averaged across three trials.[22] The latest guidelines suggest that hypertension be diagnosed when SBP exceeds 130 millimeters of mercury (mmHg) (Whelton et al. 2018). Risk of adverse cardiovascular outcomes increase with blood pressure,

[21] By far the most common variable to be missing across otherwise-eligible respondents is maternal education (n = 2,186 observations or 73.3% of those missing data). In robustness checks (not shown), I estimate models with a limited set of control variables across all respondents with non-missing biomarker data; results are substantively the same as those from the complete case analysis presented here.

[22] I use SBP rather than diastolic blood pressure (DBP) because it is more closely related to subsequent cardiovascular events and is therefore given more weight by physicians when making diagnostic and prescriptive recommendations (Whelton et al. 2018). In additional analyses (not shown), I obtain comparable results using mean arterial pressure, which is a weighted average of SBP and DBP.

even at modest levels (Lewington et al. 2002; Rapsomaniki et al. 2014; Vasan et al. 2001).

Finally, I study two forms of cholesterol: non-high-density lipoprotein (non-HDL) and HDL. Non-HDL encompasses several forms of bad cholesterol that raise the risk of cardiovascular disease, including low-density lipoproteins (LDL) (Grundy et al. 2018; Stamler et al. 2000).[23] The targeted level of non-HDL cholesterol varies depending on an individual's risk of future cardiovascular events. Those with higher risk due to family history or comorbid conditions are recommended to receive more aggressive treatment targeting lower cholesterol than those without additional risk factors (Grundy et al. 2018). HDL or "good" cholesterol, on the other hand, is associated with improved cardiovascular outcomes (Rader and Hovingh 2014; Toth et al. 2013).

It is important to note that, for all four biomarkers studied, abnormal levels in the healthy tail of the distribution are associated with poor health outcomes (Allard-Ratick et al. 2019; Carson et al. 2010; Currie et al. 2010; Jacobs et al. 1992; Kang and Wang 2016; Selvin et al. 2010). It is unclear precisely what levels of blood pressure and non-HDL cholesterol are too low and what levels of HDL cholesterol are too high. For blood sugar, Carson et al. (2010) finds elevated mortality rates among non-diabetics with A1C levels below 4%, which affects only 0.1% of observations in the analytic sample used here. As described below, I focus on the 10^{th} to the 90^{th} percentiles of biomarkers rather than more extreme levels to avoid interpretational issues related to these nonlinearities.

For analysis, I standardize (mean = 0; standard deviation [SD] = 1) all four biomarkers across the n = 20,927 observations in the sample. Density plots are provided in Appendix Figure A4.1.

Education

I operationalize educational attainment with self-reported years of completed schooling. Very few people reported attending school for 5 years or less, so I collapse those who reported such levels into a single category. Years of schooling was top-coded by the HRS at 17 years.

[23] The HRS does not collect measures of LDL cholesterol. Non-HDL cholesterol is a useful alternative, as it can be calculated by subtracting HDL cholesterol from total cholesterol, neither of which requires fasting to collect.

Prior research demonstrates that years of schooling may not be the optimal way to operationalize education in research on health, as it may obscure nonlinearities (Montez et al. 2012). Thus, in robustness checks, I operationalize schooling instead using highest degree attained. In those analyses, respondents are divided into four categories: did not finish high school; completed high school or earned a GED; earned an Associate's degree or completed some college; and earned a Bachelor's degree or higher.

Covariates

Age and cohort: Because earlier birth cohorts obtained less education, on average, and are likely to be less healthy than younger, more educated cohorts, age and year of birth could confound the association of education with biomarkers of health risk, perhaps especially in the unhealthy tails of their distributions. Thus, in all models, I control for respondent's age at the completion of the relevant HRS survey, a squared term for age, year of birth, and interactions between age and year of birth and between age-squared and year of birth.

Sex, race, and socioeconomic and health background: In some models, I also control for demographic characteristics and background factors that could be associated both with educational attainment and with biomarker profiles, particularly at the extreme ends of their distributions. These include an eight-category indicator of sex and race (non-Hispanic white male; non-Hispanic white female; non-Hispanic black male; non-Hispanic black female; Hispanic male; Hispanic female; other race male; other race female) and measures of the respondent's socioeconomic background and childhood health.

To account for socioeconomic background, I use maternal education, which is expressed as a seven-category variable (less than eight years; eight years; more than eight years, no further details[24]; nine-eleven years; twelve years; 13-15 years; and 16 years or more). Another variable indicates the respondent's perceived socioeconomic status in childhood: poor, about average, or pretty well off. A dummy variable indicates whether the respondent's father was present and consistently employed

[24] Respondents answered questions about their parental education in their first HRS survey only, and early HRS surveys asked whether each parent had attained fewer than eight years of schooling versus eight or more years. Later HRS surveys instead recorded responses in years, ranging from 0 to 17 years or more.

throughout adolescence versus unemployed for several months or absent or deceased. Another variable indicates the respondent's region of birth (Northeast, Midwest, South, West, or outside the U.S.).

I also utilize a measure of childhood health, which is known to influence both educational outcomes and later health (Haas 2007; Jackson 2009; Palloni 2006) and which may be most strongly related to biomarkers in their unhealthy tails. Respondents were asked to reflect on their health as children and report it as poor, fair, good, very good, or excellent. I merge "poor" and "fair" into one category because very few respondents reported having had poor health in childhood. Childhood health is expressed as a nominal four-category variable in analyses.

Health lifestyle and behaviors: I also assess the extent to which body mass index (BMI) and smoking status mediate the association between education and biomarkers at different points in the distribution of health. BMI is a proxy for diet and physical activity, which the HRS does not measure comprehensively. BMI is calculated as weight in kilograms divided by squared height in centimeters. Where possible, I calculate BMI from objective measurements collected during extended in-home HRS interviews. I also control for BMI-squared, as the relationship between BMI and biomarkers may be nonlinear. Smoking status is expressed as a three-category variable (never, former, or current smoker).

Analysis

I use unconditional quantile regression (UQR) to assess variation in the association of education with biomarkers across their distributions. First described by Firpo et al. (2009), UQR estimates the association of independent variables with the unconditional distribution of the outcome (see also, Killewald and Bearak 2014). Controlling for covariates mitigates concerns regarding confounding and selection, but quantiles are not defined with respect to covariates as they are in a conditional regression framework.

Estimating a UQR model requires transforming the dependent variable of interest using the recentered influence function (RIF) in Equation 1,

$$RIF(Health;\ q_\tau, F_{Health}) = q_\tau + \frac{(\tau - \mathbf{1}\{Health \leq q_\tau\})}{f_{Health}(q_\tau)}, \qquad \text{Equation 1}$$

where q_τ is the value of the outcome variable, Health, at the quantile, τ, of interest, and F_{Health} is the cumulative distribution function of Health. $\mathbf{1}$ is a function that equals one if the value of Health is less than or equal to q_τ and 0 otherwise. Finally, $f_{Health}(q_\tau)$ is the density of Health at q_τ. Notice that the RIF transformation does not utilize the independent variables at all; the transformed dependent variables are functions of the distribution of health across the sample of interest and the chosen quantile. I transform each measure of health for each of nine quantiles, specifically, quantiles 0.10 (i.e., the 10th percentile) through 0.90 (i.e., the 90th percentile), in increments of 0.10. I do so using the command `rifreg` (Firpo et al. 2009) in Stata 15.0 (StataCorp 2017).

I then estimate ordinary least squares (OLS) linear regression on the RIF-transformed dependent variables for each quantile of interest. Quantile-specific coefficients estimated through this process indicate the marginal effect of a one-unit shift in the independent variable on the unconditional quantile of the dependent variable, holding covariates constant. I cluster standard errors at the household level. To account for the uncertainty involved in the estimation of the RIF, standard errors are bootstrapped across 200 repetitions. To test whether estimates differ significantly across quantiles, I use seemingly unrelated estimation.

As shown below, I estimate a series of UQR models for each measure of health. The dependent variable (*RIF*) and coefficients are quantile-specific and therefore are indexed with *q*. The focal independent variable, educational attainment, is expressed as *EduYrs*. Recall that respondents could contribute up to two observations to the data; time-varying measures are therefore indexed by person (*i*) and time (*t*).

Model 1 assesses whether total educational disparities in health-related risk differ across the distribution of each biomarker; this model is entirely descriptive, as it controls only for age, age-squared,

year of birth (*YOB*), and their interactions. I expect β_1—an indicator of educational disparities—to grow across quantiles of blood sugar, blood pressure, and non-HDL cholesterol, but not across quantiles of HDL cholesterol, which is not regularly targeted by medical intervention. Note that, for this model and others, I also estimate parallel equations using ordinary least squares (OLS) linear regression to obtain estimates of education's association with conditional means.

$$RIF_{Health_{it},q} = \beta_{0,q} + \beta_{1,q}EduYrs_i + \beta_{2,q}Age_{it} + \beta_{3,q}Age_{it}^2 + \beta_{4,q}YOB_i + \beta_{5,q}Age_{it}YOB_i + \beta_{6,q}Age_{it}^2YOB_i + \varepsilon_{it,q} \quad \text{Model 1}$$

In Model 2, I aim to eliminate the confounding influences of demographic characteristics and socioeconomic or health-related background. To do so, I incorporate controls for sex and race (SexRace) and maternal education, perceived socioeconomic status in childhood, an indicator of paternal presence and employment, region of birth, and self-reported health in childhood (Background).

$$RIF_{Health_{it},q} = \beta_{0,q} + \beta_{1,q}EduYrs_i + \beta_{2,q}Age_{it} + \beta_{3,q}Age_{it}^2 + \beta_{4,q}YOB_i + \beta_{5,q}Age_{it}YOB_i + \beta_{6,q}Age_{it}^2YOB_i + \tau_q^T\text{SexRace}_i + \delta_q^T\text{Background}_i + \varepsilon_{it,q} \quad \text{Model 2}$$

The final models aim not to reduce confounding, but rather to assess whether lifestyle factors explain distributional heterogeneity in the association of education with biomarkers of health risk. I do not anticipate that they will, as the mediating role of lifestyle is expected to be similar across the biomarkers' distributions. To test this hypothesis, Model 3 controls for BMI, BMI-squared, and smoking status, limited but influential indicators of health-related lifestyle and behaviors. Results support my hypothesis if the percent change in the coefficient on years of schooling (β_1) between Models 2 and 3 is similar across quantiles. The ratio of effects estimated on the 90th and 10th percentiles is therefore expected to be

more or less the same in Models 2 and 3.

$$RIF_{Health_{it},q} = \beta_{0,q} + \beta_{1,q}EduYrs_i + \beta_{2,q}Age_{it} + \beta_{3,q}Age_{it}^2 + \beta_{4,q}YOB_i \quad \text{Model 3}$$
$$+ \beta_{5,q}Age_{it}YOB_i + \beta_{6,q}Age_{it}^2YOB_i$$
$$+ \tau_q^T \text{SexRace}_i + \delta_q^T \text{Background}_i$$
$$+ \beta_{7,q}BMI_{it} + \beta_{8,q}BMI_{it}^2$$
$$+ \beta_{9,q}FormerSmoke_{it} + \beta_{10,q}CurrentSmoke_{it} + \varepsilon_{it,q}$$

Models 1-3 assume that the association of a year of schooling with health does not depend on the particular year of education under consideration. Though this assumption justifies a common and parsimonious modeling strategy, prior research has shown that the education-health relationship depends both on years of schooling and on the attainment of milestone educational credentials (Montez et al. 2012). Thus, I estimate Models 1b-3b, which operationalize education using highest degree attained rather than years of schooling, but which are otherwise analogous to the models described above.

RESULTS

Descriptive statistics for the analytic sample of 20,927 observations and 15,077 respondents are provided in Table 4.1. In analyses, the four biomarkers are standardized with mean = 0 and SD = 1, though Table 4.1 also provides information on biomarkers in their original units. Unhealthy biomarker profiles are common in this sample. The average level of blood sugar is 5.9%, which is considered pre-diabetic (ADA 2018); in fact, for more than half of observations, blood sugar is consistent with either a diabetes diagnosis (14.3%) or pre-diabetes (43.9%). Likewise, the average systolic blood pressure in the sample is 131.0mmHg, just above the level that current guidelines consider hypertensive (Whelton et al. 2018), which nearly half of observations (48.4%) surpass. The average non-HDL cholesterol is 144.1mg/dL. Finally, the average HDL cholesterol in the sample is 54.4mg/dL; 58.1% of observations fall

Table 4.1. Descriptive statistics

	Mean or %	(SD)	Min, Max
Observations (n = 20,927)			
Standardized biomarkers			
Blood sugar	0.00	1.00	-2.08, 8.01
Blood pressure	0.00	1.00	-3.00, 4.66
Non-HDL cholesterol	0.00	1.00	-2.91, 4.31
HDL cholesterol	0.00	1.00	-2.41, 4.18
Unstandardized biomarkers			
Blood sugar (%)	5.85	1.00	3.78, 13.84
Blood pressure (mmHg)	131.04	20.36	70.0, 226.0
Non-HDL cholesterol (mg/dL)	144.12	39.09	30.3, 312.6
HDL cholesterol (mg/dL)	54.40	16.01	15.8, 121.4
Age	67.34	10.36	50, 101
Respondents (n = 15,077)			
Years of education	12.86	2.90	5, 17
Highest degree			
Less than high school	17.3%		
High school diploma or GED	34.5%		
Associate's or some college	24.8%		
Bachelor's degree or higher	23.5%		
Year of birth	1942.41	11.23	1908, 1963
Sex and race/ethnicity			
Males, non-Hispanic white	30.3%		
Females, non-Hispanic white	39.8%		
Males, non-Hispanic black	5.9%		
Females, non-Hispanic black	9.6%		
Males, Hispanic	4.9%		
Females, Hispanic	6.8%		
Males, Other race	1.2%		
Females, Other race	1.6%		

Notes: Descriptive statistics for additional control variables are provided in Appendix Table A4.1.

below what is considered the ideal level of HDL (Grundy et al. 2018).

On average, respondents obtained 12.9 years of schooling. Less than one in five (17.3%) did not graduate high school; one in three earned a high school diploma or GED (34.5%); 24.8% earned an Associate's degree or attended some college; and 23.5% obtained a Bachelor's degree or higher. The average age across observations is 67.3 years; the range encompasses those ages 50 to 101.[25] Descriptive statistics for additional control variables can be found in Appendix Table A4.1.

Educational attainment and the distribution of biomarkers of health risk

Table 4.2 presents biomarker means across four levels of schooling (less than 12 years, 12 years, 13-15 years, and 16 years or more). As expected, average blood sugar and blood pressure decline with years of schooling, while average HDL cholesterol increases: higher education is predictive of reduced health risk. Patterns are less clear for non-HDL cholesterol.

In Figure 4.1, I present density plots of biomarkers for those with less than 12 years of schooling and for those with 16 years or more (see also, Appendix Table A4.2). Consistent with my hypothesis, the distributions of blood sugar and blood pressure do not simply move to the right—towards higher levels and poorer health—while keeping their shape intact for those with less schooling. Instead, among those with less than 12 years of schooling, the right tails of these distributions—the unhealthy tails—are stretched further to the right. This is not the case for non-HDL cholesterol, the distribution of which is virtually identical by education level. Further, the opposite pattern obtains for HDL cholesterol, for which the healthy tail of the distribution is pulled further outwards for those with higher education.

Blood sugar and blood pressure

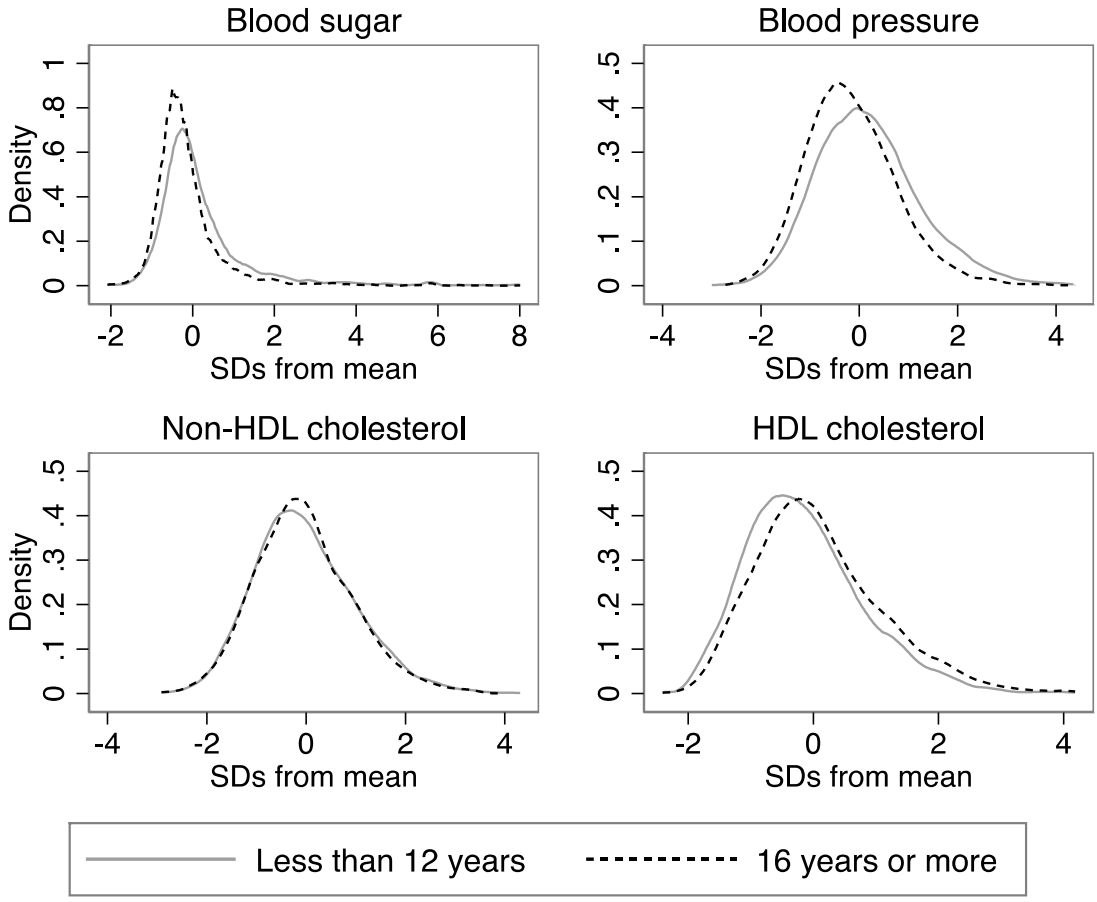
I begin by discussing results for blood sugar and blood pressure, the biomarkers for which results in Figure 4.1 (see also, Appendix Table A4.2) support my first hypothesis. That is, there appear to be

[25] Prior research shows that health disparities differ across age groups, with smaller differences in health by education among the elderly (Willson et al. 2007). In robustness checks (not shown), I assess results when including only those ages 50 to 69; results from these models are substantively the same as those presented here.

Table 4.2. Means of biomarkers by years of education

	Years of education			
	< 12	12	13-15	≥16
Standardized biomarkers				
Blood sugar	0.20	0.02	-0.05	-0.13
Blood pressure	0.19	0.05	-0.06	-0.15
Non-HDL cholesterol	-0.02	0.03	0.02	-0.04
HDL cholesterol	-0.13	-0.02	0.04	0.09
Unstandardized biomarkers				
Blood sugar (%)	6.06	5.88	5.80	5.72
Blood pressure (mmHg)	134.98	132.02	129.89	127.90
Non-HDL cholesterol (mg/dL)	143.18	145.26	144.92	142.53
HDL cholesterol (mg/dL)	52.25	54.03	55.01	55.91
Observations	N = 3,884	N = 6,886	N = 4,976	N = 5,181

Figure 4.1. Distributions of standardized biomarkers by years of education



Notes: Figure includes data from the n = 9,065 observations from respondents with fewer than 12 years of education or 16 years or more. Biomarkers are standardized across all n = 20,927 observations.

larger educational disparities in both blood sugar and blood pressure at unhealthy quantiles of their unconditional distributions.

Model 1 in Table 4.3 presents the estimated associations of a year of schooling with the conditional means of all biomarkers as well as with three points in their unconditional distributions (the 10^{th}, 50^{th}, and 90^{th} percentiles) when controlling for age, age-squared, year of birth, and interactions between age and year of birth and between age-squared and year of birth. Estimated effects of education on all nine deciles of each measure are provided in Figure 4.2. First, note that education is significantly associated with the conditional means as well as all deciles of blood sugar and blood pressure, such that higher education predicts lower (healthier) biomarkers. Schooling predicts reduced health-related risk at all points in the distributions of blood sugar and blood pressure.

Second, as hypothesized, the estimated effects of a year of schooling increase monotonically across quantiles, such that they are largest in the unhealthy tails. Variation in associations across quantiles is sizeable. A year of education predicts a relatively small 0.010 SD ($p < .001$) decline in the 10^{th} percentile of blood sugar. The effect on the 90^{th} percentile is nine times larger, with a year of schooling associated with a 0.091 SD reduction ($p < .001$). For blood pressure, the estimated effect of a year of schooling ranges from a 0.018 SD ($p < .001$) decline in the first decile to a 0.058 ($p < .001$) reduction in the last. In Figure 4.2, the 95% confidence intervals on effect estimates for the first and last deciles of both blood sugar and blood pressure do not overlap; the differences are statistically significant at $p < .001$.

In Model 2, I add controls for additional potential confounders, including demographic characteristics and socioeconomic and health background. While changes in effect sizes across the distribution of blood sugar become much less dramatic with the addition of these control variables, they do not disappear. The association of education with the 90^{th} percentile of blood sugar (-0.045, $p < .001$) is still over six times larger in magnitude than the association with its 10^{th} percentile (-0.007, $p = .004$). And, for blood pressure, the estimated effect of education on the last decile (-0.053, $p < .001$) remains nearly four times greater than that on its first (-0.014, $p < .001$). Again, the relationship of education with both blood pressure and blood sugar is significantly greater ($p < .001$) when considering the 90^{th} percentiles of

Table 4.3. Association of a year of education with standardized biomarkers across their distributions

	Mean	**Quantile .10**	**Quantile .50**	**Quantile .90**
	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)
Blood sugar				
Model 1: C/F Age, cohort [a]	-0.039 (0.003) ***	-0.010 (0.002) ***	-0.023 (0.002) ***	-0.091 (0.009) ***
Model 2: + Sex, race, background [b]	-0.019 (0.003) ***	-0.007 (0.002) **	-0.014 (0.002) ***	-0.045 (0.010) ***
Model 3: + BMI, smoking status [c]	-0.014 (0.003) ***	-0.005 (0.002) *	-0.010 (0.002) ***	-0.033 (0.010) **
Blood pressure				
Model 1: C/F Age, cohort [a]	-0.034 (0.003) ***	-0.018 (0.003) ***	-0.032 (0.003) ***	-0.058 (0.006) ***
Model 2: + Sex, race, background [b]	-0.031 (0.003) ***	-0.014 (0.003) ***	-0.029 (0.004) ***	-0.053 (0.007) ***
Model 3: + BMI, smoking status [c]	-0.026 (0.003) ***	-0.010 (0.003) **	-0.024 (0.004) ***	-0.047 (0.007) ***
Non-HDL cholesterol				
Model 1: C/F Age, cohort [a]	-0.006 (0.003) *	-0.004 (0.004)	-0.005 (0.003)	-0.011 (0.005) *
Model 2: + Sex, race, background [b]	-0.010 (0.003) **	-0.008 (0.004)	-0.008 (0.004) *	-0.010 (0.006)
Model 3: + BMI, smoking status [c]	-0.008 (0.003) **	-0.007 (0.004)	-0.006 (0.004)	-0.006 (0.006)
HDL cholesterol				
Model 1: C/F Age, cohort [a]	0.026 (0.003) ***	0.013 (0.003) ***	0.025 (0.003) ***	0.037 (0.006) ***
Model 2: + Sex, race, background [b]	0.028 (0.003) ***	0.009 (0.004) **	0.027 (0.003) ***	0.040 (0.006) ***
Model 3: + BMI, smoking status [c]	0.020 (0.003) ***	0.005 (0.004)	0.019 (0.003) ***	0.028 (0.007) ***

Notes: Standard errors are clustered at the household level and standard errors for quantile regression estimates are bootstrapped across 200 iterations.

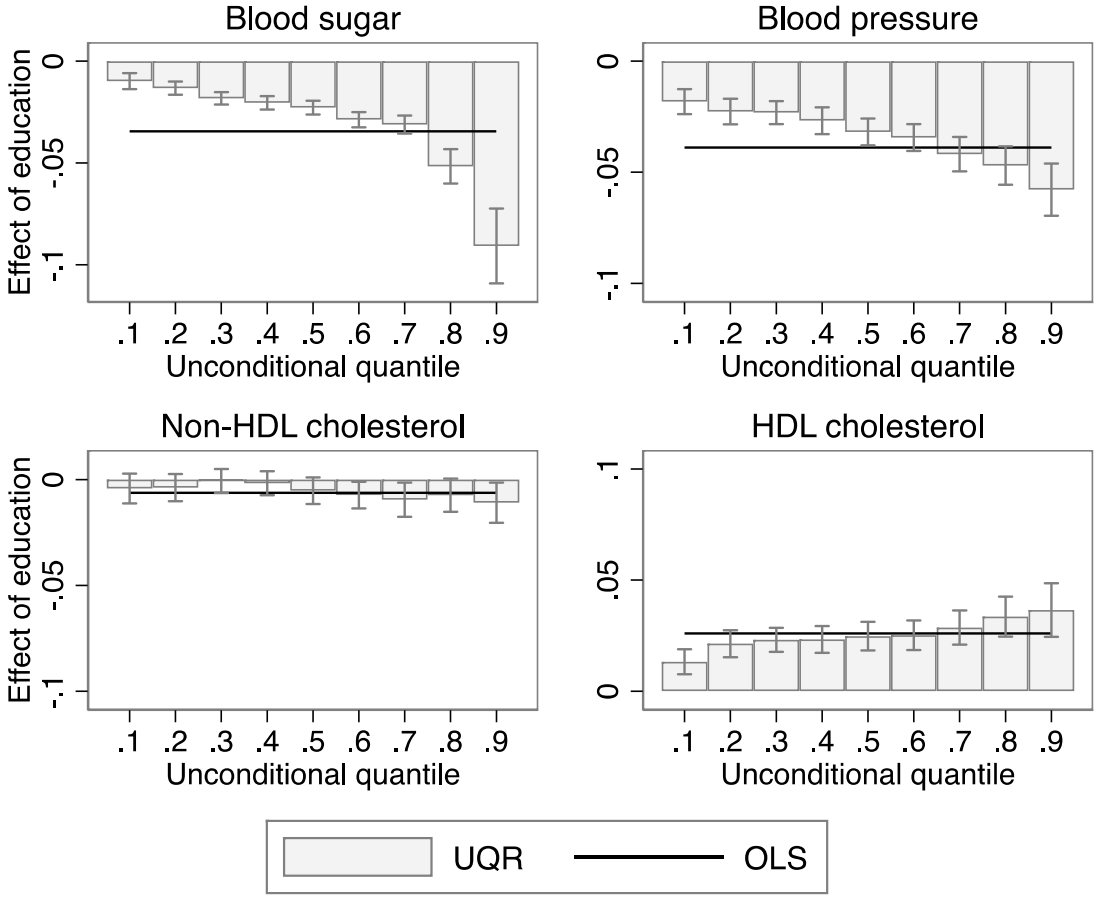
[a] Control variables include age, age-squared, year of birth, the interaction between age and year of birth, and the interaction between age-squared and year of birth.

[b] Control variables include those from Model 1 as well as sex and race/ethnicity, maternal education, perceived socioeconomic status in childhood, an indicator of paternal presence/employment, region of birth, and health in childhood.

[c] Control variables include those from Model 2 as well as body mass index (BMI), BMI-squared, and smoking status.

*** p < .001; ** p < .01; * p < .05 (two-tailed test); N = 20,927 observations from 15,077 respondents

Figure 4.2. Association of a year of education with standardized biomarkers across their distributions, from Model 1



Notes: Estimates are based on ordinary least squares regression (OLS) and unconditional quantile regression (UQR) models that control for age, age-squared, year of birth, the interaction between age and year of birth, and the interaction between age-squared and year of birth (Model 1 in Table 4.3). Biomarkers are standardized across the n = 20,927 observations. Standard errors are clustered at the household level and are bootstrapped across 200 iterations; 95% confidence intervals are shown.

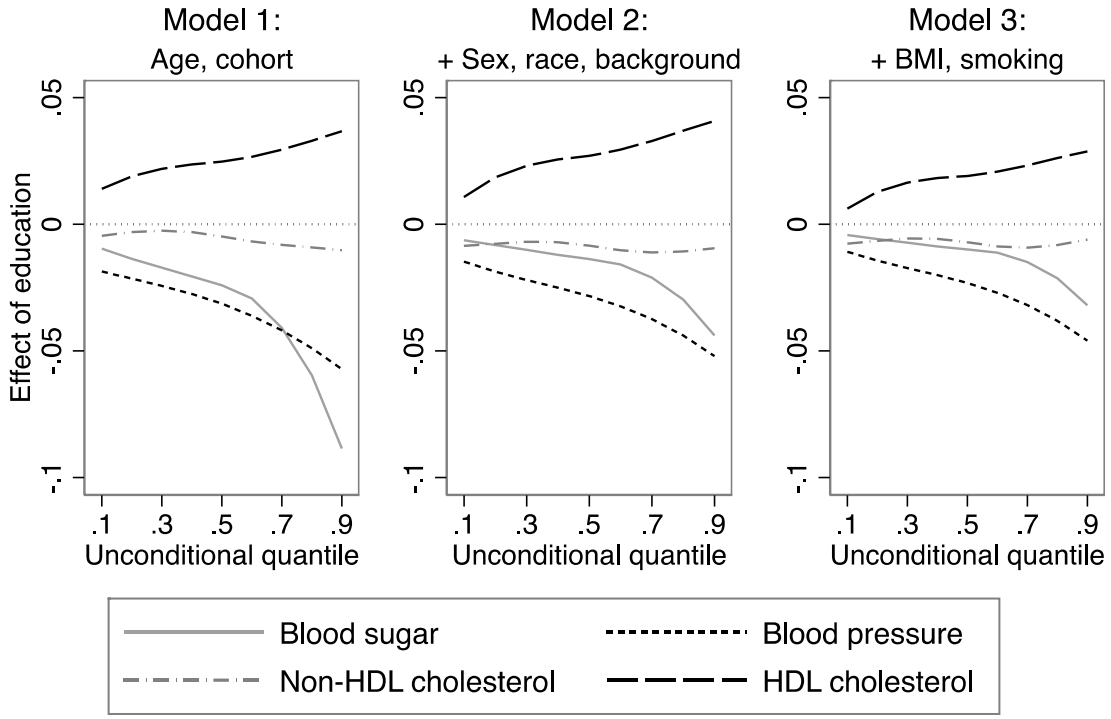
their unconditional distributions than the 10th percentiles.

Figure 4.3 presents the association of a year of schooling with each biomarker across its unconditional distribution as estimated in Models 1, 2, and 3 (see also, Table 4.3). BMI and smoking status—indicators of health lifestyle and behaviors—are controlled in Model 3 to assess the extent to which they mediate the association of education with health across quantiles. For blood sugar, the addition of these variables to the model reduces the magnitude of estimated educational disparities by a similar extent—between 26 and 33%—at all points in its distribution. For blood pressure, the mediating role of BMI and smoking is also pervasive and if anything, stronger in the healthy tail of the distribution than in the unhealthy tail. As a result, the ratio of effects estimated on the first and last deciles of blood sugar and blood pressure are not reduced in Model 3 compared with Model 2.

These results are in line with my expectations. The association of education with the 90th percentile of blood sugar (-0.033, $p = .001$) remains significantly larger in magnitude than its association with the 10th percentile (-0.005, $p = .034$) (difference $p = .005$). Similarly, a year of education is expected to reduce the 90th percentile of blood pressure by 0.047 SDs ($p < .001$), while it is expected to reduce the 10th percentile by just 0.010 SDs ($p = .003$) (difference $p < .001$).

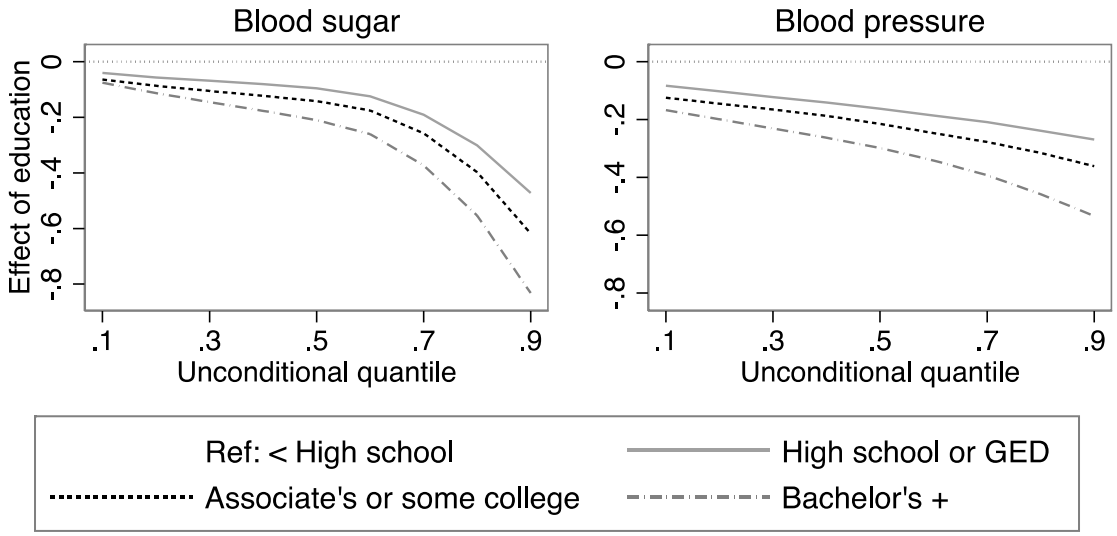
Results from models operationalizing education using highest degree attained are consistent with those presented above. As shown in Figure 4.4 and Appendix Table A4.3, Model 1b estimates that the 10th percentile of blood sugar among college graduates is 0.039 SDs ($p = .017$) lower than that among those who did not complete high school. At the 90th percentile, however, the disparity is estimated to be much larger, at nearly one-half of a standard deviation (-0.490, $p < .001$). These patterns are not explained by background characteristics or by lifestyle factors (Figure 4.5; Appendix Tables A4.4 and A4.5). Further, estimated disparities between those who did and did not complete high school (Appendix Figure A4.2) and between high school and college graduates (Appendix Figure A4.3) are greatest at higher quantiles of blood sugar. Similar patterns are found for blood pressure.

Figure 4.3. Association of a year of education with standardized biomarkers across their distributions, from Models 1, 2, and 3



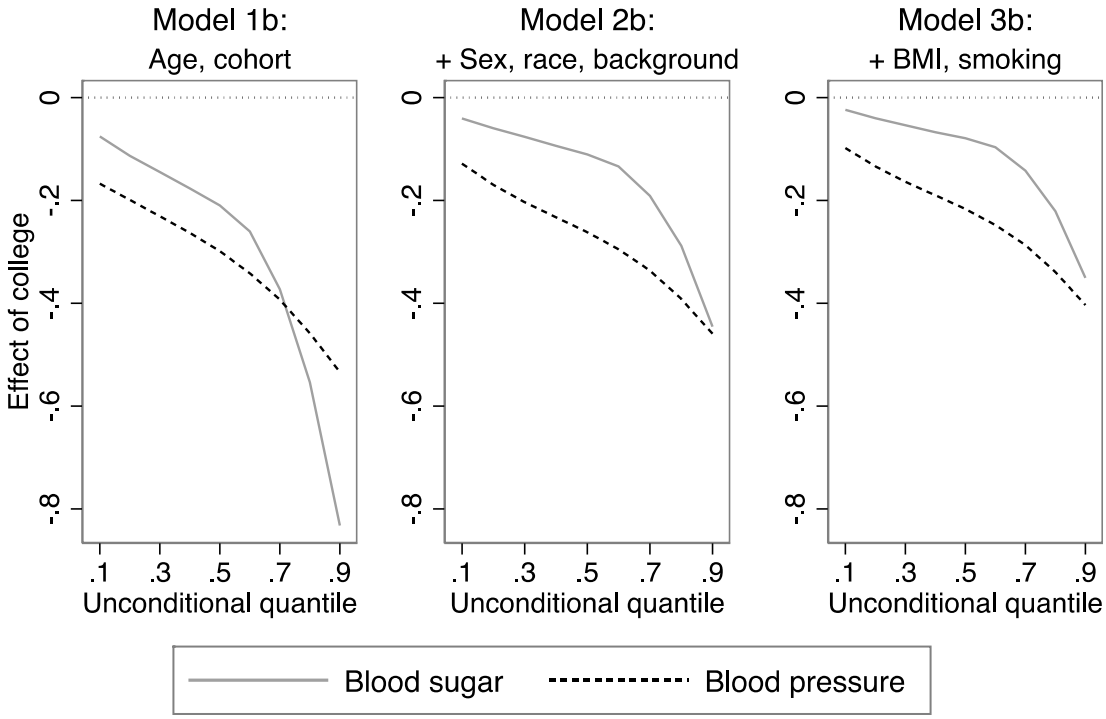
Notes: Estimates are based on unconditional quantile regression models, as shown in Table 4.3. Model 1 controls for age, age-squared, year of birth, the interaction between age and year of birth, and the interaction between age-squared and year of birth. Model 2 adds controls for sex and race/ethnicity, maternal education, perceived socioeconomic status in childhood, an indicator of paternal presence/employment, region of birth, and health in childhood. Model 3 adds controls for body mass index (BMI), BMI-squared, and smoking status. Lowess smoothing functions are applied. Biomarkers are standardized across the n = 20,927 observations.

Figure 4.4. Estimated disparities in standardized biomarkers across their distributions between those who did not complete high school (reference category) and those with higher degrees, from Model 1b



Notes: Estimates are based on unconditional quantile regression models that control for age, age-squared, year of birth, the interaction between age and year of birth, and the interaction between age-squared and year of birth, as shown in Model 1b, Appendix Table A4.3. Lowess smoothing functions are applied. Biomarkers are standardized across the n = 20,927 observations.

Figure 4.5. Estimated disparities in standardized biomarkers across their distributions between those who did not complete high school (reference category) and those with a Bachelor's degree or higher, from Models 1b, 2b, and 3b



Notes: Estimates are based on unconditional quantile regression models, as shown in Appendix Tables A4.3-A4.5. Model 1b controls for age, age-squared, year of birth, the interaction between age and year of birth, and the interaction between age-squared and year of birth. Model 2b adds controls for sex and race/ethnicity, maternal education, perceived socioeconomic status in childhood, an indicator of paternal presence/employment, region of birth, and health in childhood. Model 3b adds controls for body mass index (BMI), BMI-squared, and smoking status. Lowess smoothing functions are applied. Biomarkers are standardized across the n = 20,927 observations.

Cholesterol

While results for blood sugar and blood pressure support my hypothesis, those for non-HDL cholesterol do not. As Model 1 in Table 4.3 and Figure 4.2 show, point estimates for the association of education with non-HDL cholesterol are negative at all quantiles. They increase slightly in magnitude across quantiles (as per my hypothesis), but the difference between the estimated effect on the first and last deciles is not statistically significant ($p = .245$). I return to these results in the Discussion.

I expected to find a different pattern for HDL cholesterol than for the other biomarkers, as HDL cholesterol is not widely targeted by medication. Instead, it is influenced primarily by lifestyle factors such as diet and exercise, or may be raised inadvertently by drugs intended to lower LDL cholesterol. I therefore did not expect to find larger educational disparities at low, unhealthy quantiles of HDL.

Results shown in Model 1 of Table 4.3 as well as Figure 4.2 support this expectation, though surprisingly, an inverse pattern obtains. That is, the positive association of education with HDL cholesterol is greatest at its highest, healthiest levels. While background and lifestyle factors do appear to mediate some of the association between education and HDL (Table 4.3, Models 2 and 3; Figure 4.3), the variation across quantiles remains. In Model 3, a year of schooling is found to be associated with a non-significant 0.005 SD ($p = .195$) increase in the 10^{th} percentile of HDL cholesterol, a 0.019 SD ($p < .001$) increase in the median, and a 0.028 SD ($p < .001$) increase in the 90^{th} percentile. In all models, the difference in the effects of education estimated on the first and last deciles of HDL cholesterol is statistically significant ($p \leq .001$).

A model incorporating an interaction between sex and race demonstrates that this unexpected result is driven entirely by non-Hispanic white women (Appendix Figure A4.4). For all other groups, the relationship of education with HDL is small and roughly constant across the distribution of HDL. Thus, with the exception of white women, patterns for HDL align with my hypothesis.

As with the analyses of blood sugar and blood pressure, results for non-HDL and HDL cholesterol are substantively the same when operationalizing education using highest degree attained rather than years of schooling (Appendix Tables A4.3-A4.5).

DISCUSSION

This study was motivated by a gap in prior research: while there is clearly a positive association between education and average health and an inverse association between education and the probability of negative health events (Cutler and Lleras-Muney 2008; Hummer and Lariscy 2011), it remained unknown whether educational disparities varied across the distribution of health. Using data from U.S. adults ages 50 and above, I analyzed four biomarkers that are commonly used in clinical evaluations of health-related risk, three of which are commonly targeted by medication.

For two such biomarkers—blood sugar and blood pressure—I find significant variation in educational disparities across quantiles, with the largest disparities in the unhealthy tails of their distributions. The most educated people with the best biomarker profiles are actually not much better off than the healthiest people who completed very little schooling. However, educated individuals with the worst biomarker profiles are substantially healthier than their less educated peers.

This result is consistent with my hypothesis, which draws on fundamental cause theory (Link and Phelan 1995; Phelan et al. 2010) and related research. Specifically, educational disparities are largest for health conditions that are amenable to prevention or intervention (Chang and Lauderdale 2009; Glied and Lleras-Muney 2008; Masters et al. 2015; Phelan et al. 2004; Phelan and Link 2005; Tehranifar et al. 2009). This is thought to be because it is only for these aspects of health that resources can help to secure better outcomes. Analogously, I posited that educational disparities would be largest at points in the distribution of health where education-related resources can most effectively be mobilized to improve wellbeing. This is likely to be in the unhealthy tail of health, where thresholds that trigger medical intervention have been crossed.

Put differently, in the absence of medical interventions, educational disparities in biomarkers might be relatively uniform in magnitude across their distributions. Medical interventions are likely to exacerbate the magnitude of disparities primarily at the least healthy levels, where differences in diagnosis, treatment assignment, and treatment success drive additional gaps in outcomes. Specifically, these factors are likely to rein in unhealthy biomarker tails among those with more education and

resulting health-relevant resources.

My results for blood pressure differ from those obtained in a recent paper that instruments educational attainment with changes to the minimum school-leaving age in the United Kingdom (Barcellos et al. 2019). Barcellos and colleagues (2019) find that if anything, additional education spurred by the policy change resulted in increased blood pressure, with effects concentrated at low, healthy levels. Our contradictory findings may be explained by differences in study design and the population of inference. Barcellos and colleagues' (2019) estimates reflect those whose educational attainment was affected by the policy change, presumably those with a relatively low level of schooling. The current study instead offers a descriptive overview, showing how educational disparities in blood pressure vary across its distribution in a representative sample of older U.S. adults.

The results I obtain for non-HDL or "bad" cholesterol are not consistent with my hypothesis; instead, I find little variation in educational disparities across quantiles of non-HDL cholesterol. This might be because statins and other cholesterol-lowering drugs target just one form of non-HDL cholesterol: low-density lipoproteins (LDL). The other forms of cholesterol that comprise non-HDL may not be as easily manipulated by medication as LDL. Evidence suggests, for example, that the percentage of people who manage to reach their LDL cholesterol target is higher than the percentage that achieves their non-HDL goal (Virani et al. 2011). Variation across quantiles could be muted in the current study if educated people do not have as much of an advantage in terms of non-HDL control as they would for LDL control.

In addition, the result for non-HDL cholesterol may stem from the way cholesterol targets are set. For those with diagnosed diabetes or hypertension, the goal of treatment is generally to reduce corresponding biomarkers to levels within or slightly above what is considered the normal range (ADA 2018; Whelton et al. 2018). But for cholesterol, the target level is often even lower, particularly for those with a heightened risk of cardiovascular disease due to family history or comorbid conditions. Thus, it is precisely for those with the highest risk that non-HDL cholesterol, when properly controlled, is likely to be lowest. If those with low education are more likely to be considered at high risk of cardiovascular

disease due to family history or comorbid conditions than more educated people with the same initial cholesterol level, their treatment will be more aggressive, on average. This could reduce variation in educational disparities across quantiles of non-HDL cholesterol.

HDL or "good" cholesterol differs from the other three biomarkers I assess. HDL cholesterol is associated with improved cardiovascular outcomes, though because its causal role is debated, it is not widely targeted by medication (Rader and Horvingh 2014; Toth et al. 2013). Thus, I did not expect to find substantial variation in educational disparities across the distribution of HDL. Consistent with my expectations, I did not find larger educational disparities at unhealthy (low) levels of HDL. Instead, I found larger disparities at healthy (high) levels of HDL. Follow-up analyses reveal that this result is driven entirely by non-Hispanic white women. For all other groups, patterns align with my hypothesis, showing no systematic variation in educational disparities across quantiles of HDL cholesterol. Future research should assess why this is the case.

HDL cholesterol is not a perfect example of a negative case, in part because it may inadvertently be raised by medication intended to lower non-HDL cholesterol. However, it is challenging to identify an alternative—a biological measurement that shows substantial variation in the population, that is known to affect health, and yet that cannot be influenced with medication at all. There are no such measures in the dataset I use, though as biomarkers become increasingly available in social surveys, their range may expand to include more appropriate options.

The primary limitation of the current study is that I cannot investigate the main hypothesized mechanism for variation in educational disparities across quantiles of these biomarkers directly. That is, I cannot estimate distributional variation in biomarkers in the counterfactual scenario in which no medical intervention was possible. As a result, findings can be interpreted as consistent with the theory I present, but they cannot rule out alternative explanations.

In particular, though I expect that differences in diagnosis, treatment assignment, and treatment success are responsible for variation in educational disparities across the range of blood sugar and blood pressure, I cannot test this idea directly. The HRS does collect information on diabetes and hypertension

diagnosis. However, these measures are unsuitable for mediation analyses. Those who had high blood sugar or blood pressure at some earlier point in time are more likely to have been diagnosed with the corresponding health condition and to have less healthy subsequent biomarker profiles. Diagnoses may therefore spuriously appear to mediate the association between education and biomarkers, particularly at unhealthy points in their distributions. Fine-grained longitudinal data would be needed to confidently assess the roles of these potential mediators. This would be a fruitful area for future research.

My results do show that, as expected, rough proxies of health lifestyle and behaviors—BMI and smoking status—are not enough to explain the larger educational disparities at unhealthy points in the distributions of blood sugar and blood pressure. In fact, controlling for these measures does not attenuate the observed variation across quantiles at all. Future research could attend to a more comprehensive set of potentially influential measures, such as diet and exercise at different points in the life course.

The patterns I uncover for blood sugar and blood pressure are discouraging, as the points at which educational disparities are largest are also the points at which the association with subsequent morbidity and mortality is likely to be greatest (Coutinho et al. 1999; Lewington et al. 2002). Guidelines for clinical intervention are established precisely for this reason: to reduce the particularly strong risk of negative health outcomes associated with biomarkers in excess of some threshold. At the same time, a key finding of research on fundamental cause theory is that medical knowledge and technologies also drive socioeconomic disparities in health outcomes (Chang and Lauderdale 2009; Clouston et al. 2016; Glied and Lleras-Muney 2008; Masters et al. 2015; Phelan et al. 2004; Phelan and Link 2005; Tehranifar et al. 2009). Likewise, the current study suggests that, by exacerbating disparities in the unhealthy tails of these biomarkers' distributions, unequal access to or efficacy of medical intervention may aggravate inequalities in resulting health crises and death.

The lack of attention in prior research to distributional heterogeneity in the association of education with health is perhaps unsurprising, as until recently there were few objective, continuous measures of health available to study. The recent collection of biomarkers in social surveys makes this research possible. In addition to being both continuous and objective, biomarkers reflect health risk that

may emerge before physiological symptoms do, and thus are often the first line of defense against chronic disease for clinicians (Harris and Schorpp 2018; McDade et al. 2007). The current study highlights the utility of biomarkers for more completely mapping the contours of educational disparities in health.

It also provides practical guidance for future biomarker research. Results underscore that the analysis of conditional means might mask disparities in biomarkers that exist primarily in particular segments of their distributions. That said, studies dichotomizing biomarkers following clinical guidelines risk overlooking unequal outcomes at moderate levels, which also have consequences for subsequent health (Brunner et al. 2006; Coutinho et al. 1999; Lewington et al. 2002; Rapsomaniki et al. 2014; Selvin et al. 2010; Vasan et al. 2001).

Chapter 5: Conclusion

This dissertation builds on decades, if not centuries, of scientific research documenting disparities in health within and between populations (Krieger 2001, 2011). I focused on educational disparities in health among older adults in the United States, which prior research demonstrates are large and growing (Goesling 2007; Liu and Hummer 2008; Masters et al. 2012; Meara et al. 2008; Montez et al. 2011). For example, estimates indicate that at age 25, college graduates can expect to live more than five years longer than those with a high school degree and at least 11 years longer than those who did not complete high school (Rostron et al. 2010).

In this respect, the results of this dissertation are consistent with prior research. In all three empirical chapters, I found that those with more education evinced better health across numerous measures. And yet, mounting evidence—including the results presented here—suggests that the relationship between education and health is more complex than descriptive average disparities might imply.

EXECUTIVE SUMMARIES AND EMERGING DIRECTIONS

The empirical chapters of this dissertation drew motivation from a puzzle presented by the conflicting results of prior research. While research based on observational data tells a consistent story—one of better health among the higher-educated—quasi-experimental research has often failed to produce evidence of a causal effect of education on health (Galama et al. 2018; Grossman 2015; Montez and Friedman 2015). This is surprising at first glance, as increased educational attainment may influence later health through a combination of economic, behavioral or normative, and structural mechanisms (Freese and Lutfey 2011; Link and Phelan 1995; Mirowsky and Ross 2003).

Differences in results from observational and quasi-experimental work may arise due to confounders that are not adequately controlled in observational work, resulting in upwardly biased estimates of education's effect on health. The first empirical chapter of this dissertation (*Chapter 2:*

Polygenic Scores as Controls for Genetic Selection into Education in Models of Health) thus investigated the utility of an innovative control for selection into years of schooling in studies of health. Specifically, it investigated a control for genetic selection.

First and foremost among my findings is that controlling for genetic selection into schooling does not upend the established positive relationship between education and health. This result is robust across two datasets and three unique dimensions of health. Nonetheless, the measure of genetic selection used appears to have potential as a control variable that is on par with measures of family background and childhood health, currently among the most recognized confounders of the education-health relationship.

I operationalized genetic selection into schooling using a polygenic score (PGS), which predicts years of schooling based on the estimated effects of many hundreds of thousands of genetic variants across the genome (Dudbridge 2013; Lee et al. 2018). I posited that this PGS may be a useful control variable in research on education and health because it correlates with more proximal confounders of the education-health link, such as family background, abilities, skills, and personality traits (Belsky et al. 2016; Belsky et al. 2018; Conley et al. 2015; Domingue et al. 2015; Okbay et al. 2016), that may be poorly measured. It may also be correlated with additional confounders that remain unknown. My results are consistent with this idea, as the proportional attenuation in the estimated effect of education on health obtained when controlling for the PGS is greatest when a key confounder—cognitive performance in adolescence—is omitted from the model.

While the results of this chapter lend support to arguments for the incorporation of PGSs in social science research (Cesarini and Visscher 2017; Conley 2016; Conley and Fletcher 2017; Freese 2018), this emerging research space is not without controversy. In particular, the education PGS that I use has stirred both excitement and debate. Freese (2018) illustrates this tension well, stating, "To be clear: nobody is saying genes *determine* educational attainment. Nobody is even saying genetic information *predicts* educational attainment all that well. But standard sociological variables do not predict educational attainment that well either" (Pp.525).

Several uses for PGSs in the social sciences have been suggested. First, as advanced in this

dissertation and suggested elsewhere (Cesarini and Visscher 2017; Conley 2016; Freese 2018), PGSs may be used to control for selection. This use is relevant even to those who are disinterested in genetic effects, as controlling for PGSs may reduce confounding bias and improve estimate precision by shrinking residual error.

Second, PGSs may be used to study processes of cumulative (dis)advantage (Freese 2018). We know that a person's DNA is more or less set from birth, and PGSs show us that DNA ends up being correlated with a variety of important outcomes much later in life. The sociological puzzle, then, is why. In this sense, DNA is similar to the family characteristics that have held the interest of sociologists for decades (Blau and Duncan 1967; Pfeffer and Hertel 2015). I might ask, for example, when in the life course the relationship between the education PGS and health emerges and what environmental factors moderate its change over time.

Third, PGSs can be used to study gene-by-environment interactions (GxE), that is, whether the effects of environmental conditions on outcomes vary by genetics or vice versa. One GxE study, for example, indicates that genetic predisposition to smoke is now a stronger predictor of smoking status than it was for earlier birth cohorts, who were not subject to the strict tobacco control policies and norms of today (Domingue et al. 2016; see also, Wedow et al. 2018). Extending this idea to the topic of education and health, future research could assess, for example, whether the effect of the education PGS on later health outcomes differs across times or places that are characterized by different educational policies (e.g., mandatory minimum school-leaving ages) or different health-related institutions and technologies (e.g., insurance schemes, sanitation, immunization).

These uses of PGSs are not inherently controversial. The controversy over PGSs—and the education PGS in particular—relates to their interpretation. PGSs can be thought of as propensity scores for which we do not know what variables were included in the prediction model. Studies trying to "back out" these "variables" have shown that the education PGS is correlated with cognitive performance, but not perfectly so; it has also been shown to be related to non-cognitive skills and personality traits, and still, about half of the relationship between the PGS and educational outcomes remains unexplained

(Belsky et al. 2016; Belsky et al. 2018; Conley et al. 2015; Domingue et al. 2015; Okbay et al. 2016). Lots of mediating characteristics and behaviors likely shape the relationship between the education PGS and educational outcomes, and we do not yet fully understand this process. In my view, sociologists have a role to play in future research identifying these mechanisms. Sociologists can also help to ensure that the interdisciplinary scholars working in this space do not slip into simplistic and inaccurate descriptions of the education PGS as a measure of genetic intelligence.

The interpretation of PGSs is likely also to affect whether and how they are used out in the world. PGSs were designed in part to advance precision medicine by identifying people with elevated health risk or indicating optimal treatments. Corollary applications of PGSs for social or behavioral outcomes—precision education or precision policy, for example—may be neither feasible nor useful. The issue confronting schooling-related applications, for example, is not necessarily that students requiring additional help cannot be identified, but rather that resources are insufficient to do much about it. Moreover, allocating educational resources or determining policy based on DNA may have major adverse consequences that far outweigh potential benefits (Conley and Fletcher 2017; Duster 2003).

Another pressing issue is that those of non-European ancestries are regularly omitted from genome-wide and PGS analyses, including Chapter 2 of this dissertation. Any resulting beneficial applications may thus be less effective for them (Martin et al. 2017). From a purely intellectual standpoint, current studies privilege the experiences of those of European descent.

Despite these issues, genetic data is becoming more widely available than ever before. People are going to use this data no matter what, and a sociological voice at the table is likely to bolster discussion of the ways in which the environment mediates and moderates genetic effects as well as the unintended consequences of genetic research for inequality in outcomes and intellectual representation. The first three uses of genetic data described above are, in my assessment, consistent with the sociological mission, such that related research is likely to advance knowledge of social processes more so than it will tell us about genetics, per se.

The second empirical chapter of this dissertation (*Chapter 3: The Association of Education with*

Health and Mortality by Socioeconomic Origin, Race, and Gender) was motivated by another possible explanation for the inconsistent results in observational and quasi-experimental work on education and health. It may be that effects of education on health differ across individuals and environments, resulting in discrepancies between the average effects estimated in observational work and local average effects—effects present in a particular non-random subset of the population—in quasi-experimental work. Building on the emerging literature on heterogeneity (Montez and Friedman 2015), this chapter took a complex intersectional perspective and evaluated the association of education with health across groups defined by socioeconomic (SES) origin, race, and gender.

My results underscore the importance of one intersection in particular: that between SES origin and race. As in prior work (Ross and Mirowsky 2011), I find that the association of years of schooling with self-reported health is stronger for those from low-SES backgrounds. This finding supports Ross and Mirowsky's (2006, 2010) theory of resource substitution by SES origin, as one's own education and the resources it generates appear to make up for those lacking due to background or status. However, my results add an additional piece to the puzzle: this pattern is apparent among whites, but not among blacks.

Seen from the other angle, the association of education with self-reported health and mortality is weaker for blacks than for whites, a finding that is again consistent with prior work (Farmer and Ferraro 2005; Holmes and Zajacova 2014; Shuey and Willson 2008). However, this is primarily the case among those from low-SES origins. Thus, the pattern of resource multiplication by race—whereby education exacerbates disparities—is also limited to a particular subset of the population. These results may be driven by the fact that the ability to use one's education in support of health is impacted by discrimination, exposure to which varies across the intersection of SES origin and race.

While the results of this empirical chapter demonstrate that a complex intersectional perspective can illuminate additional layers of heterogeneity in the relationship of education with health, it also underscores the need for research on heterogeneity to straddle two competing goals. Specifically, such studies must both highlight meaningful differences while also summarizing commonalities. It can be a challenge to strike the ideal balance.

A series of papers by Montez and colleagues nicely demonstrate how knowledge regarding heterogeneity in the relationship between education and health can effectively evolve. Montez and Berkman (2014) first document that educational disparities in mortality vary across regions of the U.S. Follow-up papers have used this as motivation to study heterogeneity in educational disparities across U.S. states, to study state-level mechanisms, to study differential impacts of state-level mechanisms by demographic characteristics, and then to study variation in disparities across states and over time (Montez et al. 2016; Montez et al. 2017a; Montez et al. 2017b; Montez et al. 2019). Each of these studies builds on those conducted previously using theory to inform new directions.

Future work could extend the same logic to the topic of this empirical chapter, using its results as motivation to study the interaction between educational attainment, SES origin, and race in greater depth. If the smaller effects of education on health for low-SES origin blacks that I observe is due to racial discrimination and related stressors, I would not expect to find a strong interaction in environments where discrimination is absent. Thus, studies could evaluate heterogeneity in the effects of education on health by SES origin and race across environments in which racial discrimination was likely to be more or less pervasive and damaging. For example, patterns in southern states during the Jim Crow era could be compared to those observed in later decades or to those observed in northern states.

In both of the previously mentioned empirical chapters of this dissertation, I assessed whether educational disparities varied across measures of health. Chapter 2 examined summary measures of physical and mental, cardiovascular, and metabolic health. Chapter 3 studied self-reported health and mortality. The final empirical chapter (*Chapter 4: Unconditional Quantile Regression and Educational Disparities in Biomarkers of Health Risk*), however, took this investigation a step further by assessing whether educational disparities in biomarkers of health risk vary across their distributions.

Drawing on fundamental cause theory (Link and Phelan 1995; Phelan et al. 2010), I anticipated that educational disparities in some biomarkers would be greatest at their least healthy levels, where unequal access to and efficacy of medical interventions might exacerbate the magnitude of educational disparities. Results for blood sugar and blood pressure are consistent with this hypothesis. The

implications of these results are discouraging, as the points at which educational disparities in blood sugar and blood pressure are largest are also the points at which the association with subsequent morbidity and mortality is greatest (Coutinho et al. 1999; Lewington et al. 2002).

Future research should assess the precise mechanisms underlying these patterns. Longitudinal data—ideally starting in adolescence, before diabetes, hypertension, and other chronic conditions emerge—could be used to assess whether differences in rates and timing of diagnosis and/or differences in treatment assignment and adherence across education levels lead to the large differences at unhealthy levels of blood sugar and blood pressure that I observe. Studies could also investigate educational differences in uptake of new health guidelines by leveraging the fact that clinical standards for blood sugar and blood pressure, as well as cholesterol, are occasionally changed. Such research is needed to identify which stage(s) of medical intervention should be targeted to reduce disparities in extreme health risk by education and SES more generally.

The results of this chapter are not only of theoretical and substantive interest; they also provide methodological guidance for future work using biomarkers. Common analytic strategies—including the analysis of conditional means—might mask disparities in biomarkers that exist only or primarily in particular segments of their distributions. That said, studies that dichotomize measures following clinical guidelines will overlook unequal outcomes at moderate levels, which also have consequences for subsequent health (Brunner et al. 2006; Coutinho et al. 1999; Lewington et al. 2002; Rapsomaniki et al. 2014; Selvin et al. 2010; Vasan et al. 2001).

These methodological implications are timely, given the recent proliferation of biomarkers in social science datasets (Harris and Schorpp 2018; McDade et al. 2007). The increasing availability of and attention to biomarkers reflects a broader shift away from the study of subjective measures of health. This is in part because the considerations involved in subjective perceptions of health have been found to vary across the same characteristics that are the subject of health disparities research. For example, among those rating their health categorically the same, the highly educated have healthier biomarker profiles, suggesting that high-SES individuals judge their health more harshly than those of low-SES (Dowd and

Zajacova 2010; see also, Dowd and Zajacova 2007; Zajacova and Dowd 2011).

Diagnoses present similar interpretational challenges, as they may be driven not only by underlying physiological functioning but also by access to and utilization of medical services. And, though mortality is a useful objective measure of health, it presents its own limitations. For example, all-cause mortality reflects many different dimensions of health, each of which evolves through unique biological mechanisms, and studying cause-specific mortality requires large samples for sufficient statistical power. This is particularly problematic in studies of younger populations, since mortality tends to occur in older age.

Biomarkers like those studied in Chapter 4 pose a unique solution to some of these issues (Harris and Schorpp 2018; McDade et al. 2007). They are objective and continuous, reflecting a range of risk to health, which, as shown in Chapter 4, enables modeling strategies that are not feasible using previously available binary and categorical measures of health. Biomarkers also demonstrate variation even in young, healthy populations, identifying risk to health before outward symptoms manifest. Because of this, in clinical settings they are often the first line of defense against chronic disease, morbidity, and mortality. Research on biomarkers thus has real potential for contributing to improved population health and reduced health disparities.

One of the largest issues confronting the study of biomarkers in social science research, in my view, is that the literature on these measures is largely found in medical journals. And, much like the language of sociology, the language of the medical establishment is largely foreign to those not explicitly instructed in it. Training opportunities for biomarker researchers may help to encourage a more widespread adoption of these measures in social science research. They would also likely reduce the number of flawed studies and conclusions that make their way into social science journals due to understandable misinterpretations of medical research.

SUMMARY

This dissertation moved the literature on education and health forward by estimating the utility of

a control for genetic selection, by highlighting heterogeneity in the relationship of education with health at the intersection of socioeconomic origin and race, and by demonstrating distributional variation in the magnitude of educational disparities in biomarkers of health risk. One overarching takeaway is that, consistent with prior observational research (Cutler and Lleras-Muney 2008; Hummer and Lariscy 2011), more education invariably predicts better health. In no chapter and in no analysis did I find that educational attainment predicted worsened health along any dimension.

In that sense, results suggest that investment in education as a tool for public health improvement—which has been suggested by many (Cohen and Syme 2013; Galea et al. 2011; Hahn and Truman 2015; Woolf et al. 2007)—may be worthwhile. That said, future research would do well to examine whether increasing educational attainment is the best way to achieve public health goals. Improving access to early schooling may also be vital for future life chances (Heckman 2006). The content and quality of schooling is also likely to matter both in addition to (Dudovitz et al. 2016; Johnson 2010) and in interaction with (Frisvold and Golberstein 2011; Sansani 2011) years of education.

My results do not offer a clear-cut response to the question, "Does education affect health?" Instead, my findings resonate with the idea, previously expressed by Montez and Friedman (2015), that this question itself is untenable. Effects of education on health vary across individuals and environments and across manifestations of health. Variation may be driven by cross-sectional differences and temporal changes in the social world, particularly those that render education-linked resources more or less relevant to particular health conditions.

Defining the concept of embodiment, epidemiologist Nancy Krieger (2001) states, "no aspect of our biology can be understood absent knowledge of history and individual and societal ways of living" (Pp.672). Like the social world that informs it, the relationship between education and health is diverse and dynamic. This both complicates and motivates attempts to further unravel the relationship between education and health.

Appendices

Appendix Table A2.1. Rotated[a] loadings of original measures of health on PCFA-constructed factors[b]

	WLS (n = 6,018)			HRS (n = 12,629)		
	Factor 1: Physical & mental health	**Factor 2:** Cardio-vascular health	**Factor 3:** Metabolic health	**Factor 1:** Physical & mental health	**Factor 2:** Cardio-vascular health	**Factor 3:** Metabolic health
Self-reported (good) health	**0.547**	0.213	**0.344**	**0.693**	0.139	0.216
Any basic (physical) limitations	**-0.632**	0.132	**-0.421**	**-0.748**	-0.041	-0.115
Any instrumental (cognitive) limitations	**-0.785**	-0.014	0.196	**-0.669**	-0.108	0.197
Depressive symptoms	**-0.598**	0.198	-0.161	**-0.794**	0.187	-0.086
Self-reported (good) memory	**0.716**	-0.002	-0.182	**0.573**	0.005	-0.119
Diagnosed high blood pressure	0.102	**-0.568**	**-0.380**	-0.054	**-0.374**	**-0.475**
Diagnosed or measured high cholesterol[c]	0.110	**-0.781**	0.051	-0.254	**0.724**	0.032
Diagnosed heart disease	-0.108	**-0.684**	0.028	-0.122	**-0.700**	0.035
Diagnosed stroke	**-0.459**	**-0.446**	0.146	-0.231	**-0.643**	0.220
Diagnosed diabetes or high blood sugar	-0.046	**-0.350**	**-0.541**	-0.052	**-0.311**	**-0.620**
Body mass index category	0.086	0.061	**-0.866**	-0.074	0.230	**-0.878**
% Variance explained after rotating[a]	24.41	20.71	18.36	26.40	20.53	15.85
Eigenvalue	3.31	1.72	1.05	3.36	1.52	1.17

Notes: Loadings of magnitude 0.3 or greater are bolded. WLS = Wisconsin Longitudinal Study; HRS = Health and Retirement Study.

[a] Factors are rotated obliquely, which allows them to be correlated.

[b] Both original measures of health and PCFA-constructed factors are standardized (mean = 0, SD = 1) across the observations eligible for the PCFA.

[c] In the WLS, respondents were asked whether they had ever been diagnosed with high cholesterol. In the HRS, respondents had their cholesterol measured directly; thus, low cholesterol readings in the HRS could be the result of a healthy lifestyle, genetics, or medications.

Appendix Table A2.2. Correlations between original and summary measures of health[a]

	WLS (n = 6,018)			HRS (n = 12,629)		
	Physical & mental health	Cardio-vascular health	Metabolic health	Physical & mental health	Cardio-vascular health	Metabolic health
Self-reported (good) health	**0.575**	**0.346**	**0.474**	**0.707**	**0.319**	**0.299**
Any basic (physical) limitations	**-0.547**	0.014	**-0.435**	**-0.653**	-0.162	-0.161
Any instrumental (cognitive) limitations	**-0.616**	-0.053	0.107	**-0.569**	-0.166	0.131
Depressive symptoms	**-0.592**	0.047	-0.217	**-0.700**	-0.002	-0.121
Self-reported (good) memory	**0.641**	0.070	-0.057	**0.537**	0.114	-0.055
Diagnosed high blood pressure	-0.028	**-0.561**	**-0.448**	-0.148	**-0.413**	**-0.515**
Diagnosed or measured high cholesterol[b]	-0.009	**-0.688**	-0.088	-0.064	**0.618**	0.140
Diagnosed heart disease	-0.185	**-0.621**	-0.109	-0.250	**-0.650**	-0.092
Diagnosed stroke	**-0.383**	**-0.390**	0.050	-0.271	**-0.558**	0.109
Diagnosed diabetes or high blood sugar	-0.139	**-0.391**	**-0.570**	-0.132	**-0.367**	**-0.631**
Body mass index category	-0.039	-0.115	**-0.799**	-0.096	0.032	**-0.796**

Notes: Correlations of magnitude 0.3 or greater are bolded. WLS = Wisconsin Longitudinal Study; HRS = Health and Retirement Study

[a] Both original and summary measures of health are standardized (mean = 0, SD = 1) across the observations included in the PCFA.

[b] In the WLS, respondents were asked whether they had ever been diagnosed with high cholesterol. In the HRS, respondents had their cholesterol measured directly; thus, low cholesterol readings in the HRS could be the result of a healthy lifestyle, genetics, or medications.

Appendix Text A2.1. PGS construction

PGSs for education based on Lee et al. (2018) were constructed by several of the GWAS's lead authors to be merged with the WLS (Okbay et al. 2018a) and HRS (Okbay et al. 2018b) samples.

In the WLS, genetic data was initially available from 9,109 individuals and for 713,014 SNPs. A total of 8,527 individuals and 604,710 SNPs remained after quality control procedures were completed.[26] Non-measured SNPs were imputed against the Haplotype Reference Consortium v1.1 European reference panel (McCarthy et al. 2016) using the Michigan Imputation Server. Genetic effects were then adjusted for linkage disequilibrium using LDpred (Vilhjálmsson et al. 2015) after restricting to HapMap3 SNPs, which are known to be well imputed and to provide good coverage of the genome in European-ancestry individuals (Altshuler et al. 2010). The PGSs incorporate information from 1,170,820 SNPs and are available for 8,527 individuals; 8,509 can successfully be merged with public WLS data.

HRS genetic data was originally available for 15,620 respondents across over 2 million SNPs. Imputation of non-measured genotypes and initial quality control procedures were undertaken at the University of Washington's Genetics Coordinating Center (GCC) (Health and Retirement Study 2012b, 2013b). Imputation was conducted for 12,454 individuals using the March 2012 release of the 1000 Genomes Phase 1 reference panel (Abecasis et al. 2012). Lee and colleagues' (2018) coauthors used this imputed data to construct PGSs for years of schooling (Okbay et al. 2018b), first adjusting genetic effects for linkage disequilibrium using Ldpred and restricting to HapMap3 SNPs. The HRS PGSs combine information from 1,104,681 SNPs across 8,652 European-ancestry respondents.

[26] Individuals were dropped from the data if they had non-European or outlying ancestry, a mismatch between reported and genetic sex, a mismatch between reported and genetic familial relationships, a high rate of missing genetic data (missingness exceeding 0.05 on any chromosome), or an outlying rate of heterozygosity/homozygosity (which suggests genotyping error). A SNP was removed from the data if its call rate was below 0.95 (i.e., if its rate of missingness across respondents was above 0.05), if its minor allele frequency was below 0.01 (i.e., if it demonstrated little variation across respondents), or if its Hardy-Weinberg exact test p-value was below 10^{-5} (which suggests genotyping error).

Appendix Table A2.3. Percent reduction in the effect of a year of education on standardized summary measures of health when controlling for the top ten principal components (PCs) of the genetic data (measures of genetic ancestry) alone versus when controlling for the PCs as well as the education PGS

	% Effect of education reduced		
Dependent variable	**Model 1A vs. 1B**	**Model 2A vs. 2B**	**Model 3A vs. 3B**
WLS			
Physical & mental health			
Model B controls for PCs only	-0.5%	-0.3%	0.4%
Model B controls for PCs & PGS	4.8%	5.8%	4.0%
Cardiovascular health			
Model B controls for PCs only	0.9%	1.7%	1.3%
Model B controls for PCs & PGS	21.1%	22.8%	16.9%
Metabolic health			
Model B controls for PCs only	2.4%	1.5%	1.5%
Model B controls for PCs & PGS	15.8%	15.5%	12.1%
HRS			
Physical & mental health			
Model B controls for PCs only	0.2%	0.4%	-
Model B controls for PCs & PGS	10.5%	9.8%	-
Cardiovascular health			
Model B controls for PCs only	0.6%	2.8%	-
Model B controls for PCs & PGS	26.1%	34.6%	-
Metabolic health			
Model B controls for PCs only	0.8%	1.5%	-
Model B controls for PCs & PGS	18.3%	19.7%	-

Notes: Percent reductions are calculated by comparing estimates from Models 1A, 2A, and 3A to different versions of Models 1B, 2B, and 3B, respectively. Model 1 controls for demographic characteristics; Model 2 adds controls for family background and childhood health; and Model 3 adds a control for cognitive performance in adolescence. Rows labeled "Model B controls for PCs only" utilize a version of Model 1B, 2B, or 3B that adds controls for the top ten principal components of the genetic data (measures of genetic ancestry). Results found in rows labeled "Model B controls for PCs & PGS" utilize the version of Model 1B, 2B, or 3B that is used in the main analysis and that adds controls for the top ten principal components of the genetic data as well as a control for the education PGS. PCs = Principal components; PGS = Polygenic score; WLS = Wisconsin Longitudinal Study; HRS = Health and Retirement Study.

Appendix Table A2.4. Additional descriptive statistics

	WLS		HRS[a]	
	Mean (SD) or %	**Min, Max**	**Mean (SD) or %**	**Min, Max**
Family background controls				
Mother's education				
< 8 years	8.1%	-	14.4%	-
8 years	26.8%	-	18.5%	-
9-11 years	11.5%	-	14.0%	-
12 years	37.9%	-	36.9%	-
13-15 years	9.7%	-	9.0%	-
16+ years	6.1%	-	7.2%	-
Father's education				
< 8 years	19.67%	-	22.1%	-
8 years	29.0%	-	20.0%	-
9-11 years	10.9%	-	12.3%	-
12 years	25.1%	-	28.6%	-
13-15 years	7.2%	-	7.4%	-
16+ years	8.2%	-	9.7%	-
Parental income (100s of USD)	65.06 (60.4)	1, 998	-	-
Perceived SES				
Poor	-	-	27.7%	-
Average	-	-	65.8%	-
Well off	-	-	6.5%	-
Father's occupation				
Farming	22.6%	-	-	-
Unskilled	30.3%	-	-	-
Skilled	9.5%	-	-	-
White collar	21.6%	-	-	-
Professional	11.6%	-	-	-
Not in labor force	4.4%	-	-	-
Father experienced extended unemployment or was absent	-	-	26.4%	-
Lived with both parents	92.1%	-	-	-
Number of siblings	3.24 (2.4)	0, 15	-	-
Population of hometown				
Rural: < 1,000	29.4%	-	-	-
1,000 – 9,999	23.4%	-	-	-
10,000 – 49,999	24.9%	-	-	-
50,000 – 149,999	10.1%	-	-	-
Urban: 150,000 +	12.2%	-	-	-

Appendix Table A2.4 (Continued)

	WLS		HRS[a]	
	Mean (SD) or %	**Min, Max**	**Mean (SD) or %**	**Min, Max**
Region of birth				
Northeast	-	-	20.4%	-
Midwest	-	-	40.0%	-
South	-	-	29.3%	-
West	-	-	10.4%	-
Childhood health controls				
Childhood health				
Poor	0.5%	-	1.2%	-
Fair	3.1%		4.5%	
Good	12.4%	-	14.2%	-
Very good	34.4%	-	25.8%	-
Excellent	49.7%	-	54.4%	-
Number of health conditions	1.09 (1.0)	0, 5	-	-
Extended activity limitations	12.8%	-	-	-
Cognitive performance in adolescence, centile[b]	63.9 (25.3)	0, 100	-	-
N possible observations	6,018		12,629 (7,726 respondents)	
N complete observations [c]	4,322		10,076 (6,053 respondents)	

[a] Figures represent means across respondents.

[b] This is the centile rank, compared to national test-takers, of the respondent's score on the Henmon-Nelson test of mental ability (Henmon et al. 1957), taken in high school and adjusted for age.

[c] In the WLS, the most commonly-missing variables are number of health conditions experienced in childhood (n missing = 751), experienced extended activity limitations (n missing = 541), parental income in adolescence (n missing = 697), childhood health (n missing = 461), and adolescent cognitive performance (n missing = 284). No other variables listed above are missing for more than 200 respondents. In the HRS, parental education was often missing (n respondents missing maternal education = 1,096; paternal education = 1,291). Fewer than 200 respondents are missing information on any other independent variable.

CHAPTER 3 APPENDICES

Appendix Table A3.1. Effects of a one-year increase in education on odds of good self-reported health, by sociodemographic group

Sociodemographic group	OR (SE)	Pairwise comparisons of OR						
White males, high SES	1.221 (0.023) ***	Ref						
White males, low SES	1.301 (0.025) ***	*	Ref					
White females, high SES	1.239 (0.025) ***	NS	†	Ref				
White females, low SES	1.331 (0.031) ***	**	NS	*	Ref			
Black males, high SES	1.210 (0.042) ***	NS	†	NS	*	Ref		
Black males, low SES	1.159 (0.029) ***	†	***	*	***	NS	Ref	
Black females, high SES	1.224 (0.042) ***	NS	NS	NS	*	NS	NS	Ref
Black females, low SES	1.231 (0.027) ***	NS	†	NS	*	NS	†	NS

Notes: Group-specific odds ratios are calculated using Model 3 in Table 3.2. Model controls for age, age-squared, year of birth, the interaction between age and year of birth, the interaction between age-squared and year of birth, and childhood health. Standard errors are adjusted for clustering at the household level.
*** p < .001; ** p < .01; * p < .05; † p < .1; NS = Not significant, p > .1 (two-tailed test)
Ref = Reference group; OR = Odds ratio; SE = Standard error; SES = Socioeconomic origin

Appendix Table A3.2. Odds ratios (ORs) from a single logistic regression model of good self-reported health on highest degree attained, by sociodemographic group

	Model 3b: OR (SE), by sociodemographic group							
	White males, High SES	White males, Low SES	White females, High SES	White females, Low SES	Black males, High SES	Black males, Low SES	Black females, High SES	Black females, Low SES
Panel A. Reference: Less than high school								
< HS	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
HS/GED	2.143 ***	2.684 ***	2.398 ***	3.046 ***	1.773 **	1.536 **	2.150 ***	2.226 ***
	(0.349)	(0.298)	(0.379)	(0.290)	(0.378)	(0.215)	(0.448)	(0.247)
AA/Some college	2.861 ***	3.201 ***	3.412 ***	2.986 ***	2.197 *	3.039 **	1.933 *	3.356 ***
	(0.608)	(0.793)	(0.689)	(0.564)	(0.753)	(1.071)	(0.515)	(0.706)
BA/BS+	4.977 ***	8.760 ***	5.973 ***	10.140 ***	4.320 ***	3.829 ***	4.918 ***	5.452 ***
	(0.873)	(1.676)	(1.042)	(2.134)	(1.167)	(1.033)	(1.228)	(1.117)
Panel B. Reference: High school or GED								
< HS	0.467 ***	0.373 ***	0.417 ***	0.328 ***	0.564 **	0.651 **	0.465 ***	0.449 ***
	(0.076)	(0.041)	(0.066)	(0.031)	(0.120)	(0.091)	(0.097)	(0.050)
HS/GED	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
AA/Some college	1.335 †	1.193	1.423 *	0.980	1.239	1.979 †	0.899	1.507 *
	(0.207)	(0.282)	(0.206)	(0.173)	(0.368)	(0.692)	(0.187)	(0.304)
BA/BS+	2.322 ***	3.264 ***	2.491 ***	3.328 ***	2.436 ***	2.494 **	2.287 ***	2.449 ***
	(0.227)	(0.580)	(0.251)	(0.658)	(0.512)	(0.666)	(0.440)	(0.486)

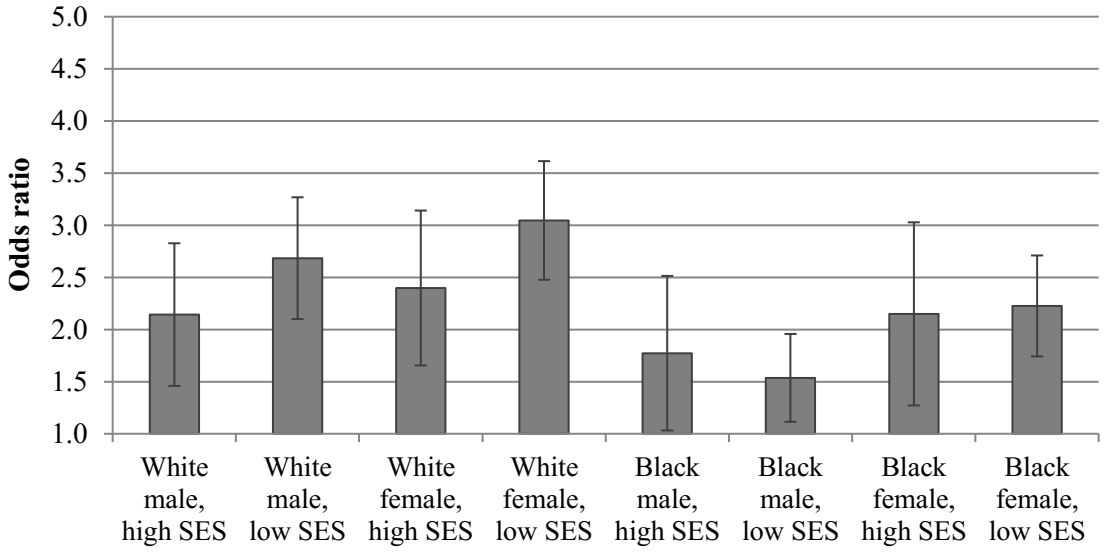
Notes: Models control for age, age-squared, year of birth, interactions between age and age-squared and year of birth, and childhood health. Standard errors are adjusted for clustering at the household level. N = 57,587 observations from 16,439 respondents; Pseudo $R^2 = 0.1079$.

*** p < .001; ** p < .01; * p < .05; † p < .1 (two-tailed test)

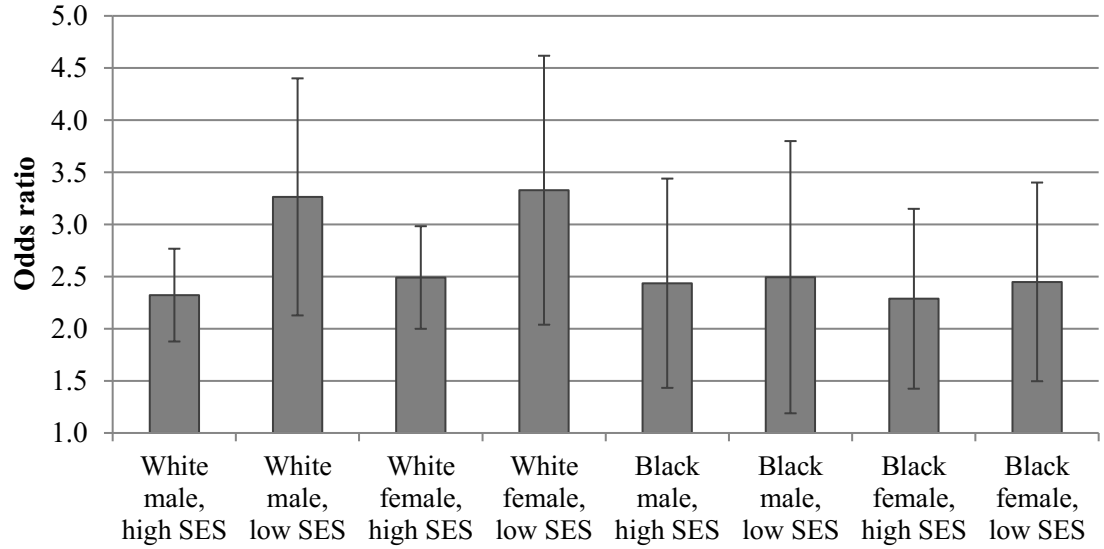
Ref = Reference group; OR = Odds ratio; SE = Standard error; SES = Socioeconomic origin; <HS = Less than high school; HS/GED = High school diploma or GED; AA/Some college = Associate's degree or some college; BA/BS+ = Bachelor's degree or higher

Appendix Figure A3.1. Effects of highest degree attained on odds of good self-reported health, by sociodemographic group

Panel A: Odds ratios for those with a high school diploma or GED compared to those with less than a high school education



Panel B: Odds ratios for those with a Bachelor's degree or higher compared to those with a high school diploma or GED



Notes: Group-specific odds ratios are calculated based on Model 3b in Appendix Table A3.2. Error bars represent 95% confidence intervals. SES = Socioeconomic origin.

Appendix Table A3.3. Effects of a one-year increase in education on odds of mortality, by sociodemographic group

Sociodemographic group	**OR (SE)**	**Pairwise comparisons of OR**						
White males, high SES	0.888 (0.017) ***	Ref						
White males, low SES	0.913 (0.016) ***	NS	Ref					
White females, high SES	0.849 (0.023) ***	NS	*	Ref				
White females, low SES	0.826 (0.018) ***	*	***	NS	Ref			
Black males, high SES	0.895 (0.038) **	NS	NS	NS	†	Ref		
Black males, low SES	0.979 (0.022)	**	*	***	***	†	Ref	
Black females, high SES	0.915 (0.052)	NS	NS	NS	†	NS	NS	Ref
Black females, low SES	0.906 (0.020) ***	NS	NS	†	**	NS	*	NS

Notes: Group-specific odds ratios are calculated using Model 3 in Table 3.3. Model controls for age, age-squared, year of birth, the interaction between age and year of birth, the interaction between age-squared and year of birth, and childhood health. Standard errors are adjusted for clustering at the household level.

*** $p < .001$; ** $p < .01$; * $p < .05$; † $p < .1$; NS = Not significant, $p > .1$ (two-tailed test)

Ref = Reference group; OR = Odds ratio; SE = Standard error; SES = Socioeconomic origin

Appendix Table A3.4. Odds ratios (ORs) from a single logistic regression model of mortality on highest degree attained, by sociodemographic group

	Model 3b: OR (SE), by sociodemographic group							
	White males, High SES	White males, Low SES	White females, High SES	White females, Low SES	Black males, High SES	Black males, Low SES	Black females, High SES	Black females, Low SES
Panel A. Reference: Less than high school								
< HS	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
HS/GED	0.562 ***	0.625 ***	0.496 ***	0.489 ***	0.616	0.922	0.878	0.746 *
	(0.090)	(0.066)	(0.093)	(0.047)	(0.196)	(0.139)	(0.389)	(0.098)
AA/Some college	0.663 †	0.737	0.491 *	0.418 **	0.625	0.668	0.890	0.504
	(0.149)	(0.169)	(0.137)	(0.127)	(0.347)	(0.331)	(0.493)	(0.211)
BA/BS+	0.356 ***	0.407 ***	0.287 ***	0.280 ***	0.432 *	0.724	0.719	0.348 ***
	(0.061)	(0.068)	(0.063)	(0.063)	(0.175)	(0.218)	(0.341)	(0.105)
Panel B. Reference: High school or GED								
< HS	1.778 ***	1.600 ***	2.016 ***	2.043 ***	1.624	1.085	1.140	1.341 *
	(0.284)	(0.170)	(0.376)	(0.195)	(0.518)	(0.164)	(0.505)	(0.176)
HS/GED	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
AA/Some college	1.179	1.180	0.990	0.855	1.015	0.725	1.014	0.676
	(0.213)	(0.260)	(0.228)	(0.257)	(0.514)	(0.359)	(0.436)	(0.284)
BA/BS+	0.634 ***	0.651 **	0.578 ***	0.572 *	0.702	0.785	0.819	0.467 *
	(0.067)	(0.100)	(0.086)	(0.126)	(0.236)	(0.237)	(0.253)	(0.142)

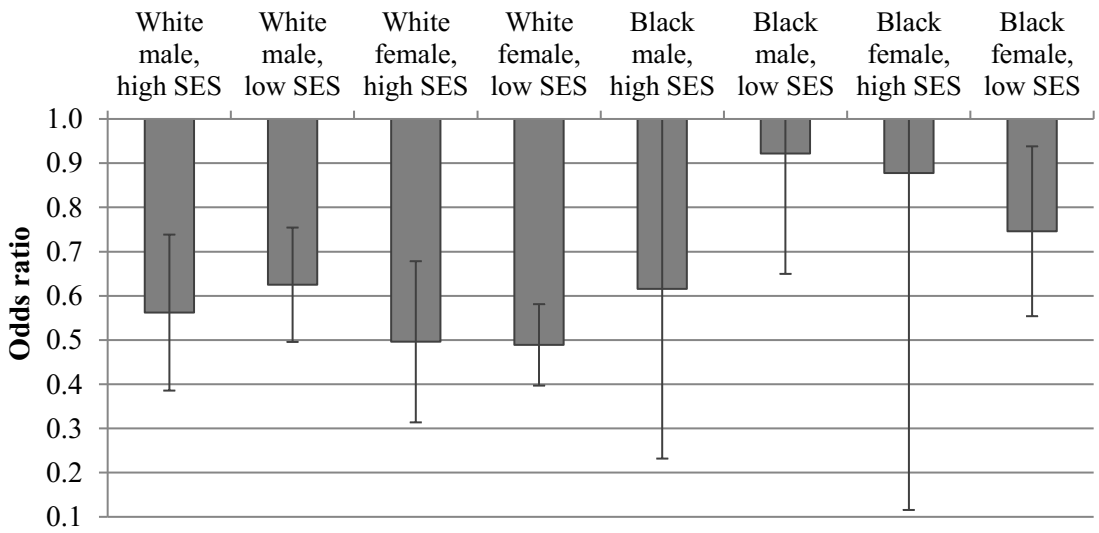
Notes: Models control for age, age-squared, year of birth, interactions between age and age-squared and year of birth, and childhood health. Standard errors are adjusted for clustering at the household level. N = 191,685 observations, 16,439 respondents; Pseudo R^2 = 0.0464.

*** p < .001; ** p < .01; * p < .05; † p < .1 (two-tailed test)

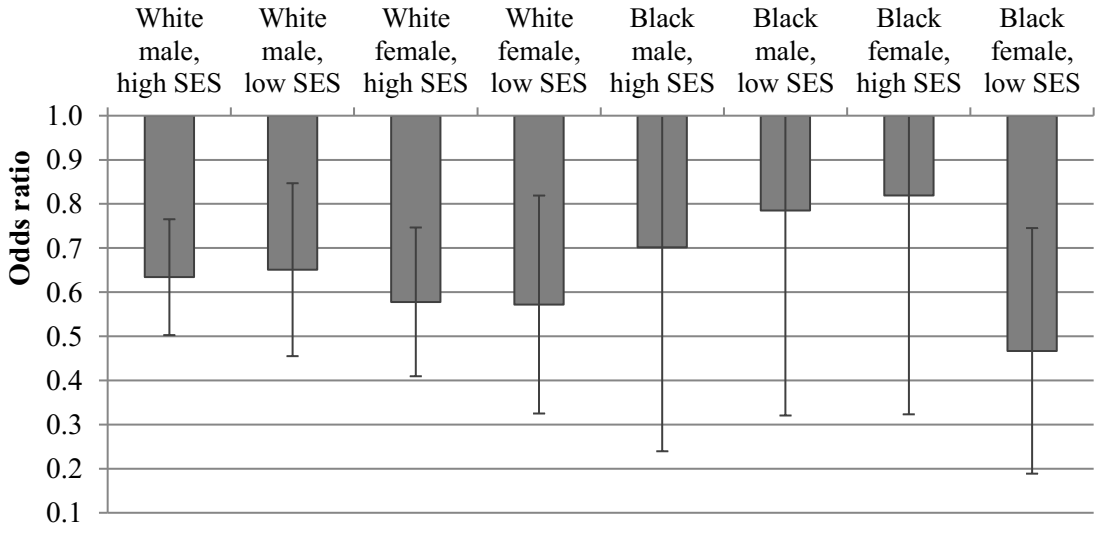
Ref = Reference group; OR = Odds ratio; SE = Standard error; SES = Socioeconomic origin; <HS = Less than high school; HS/GED = High school diploma or GED; AA/Some college = Associate's degree or some college; BA/BS+ = Bachelor's degree or higher

Appendix Figure A3.2. Effects of highest degree attained on odds of mortality, by sociodemographic group

Panel A: Odds ratios for those with a high school diploma or GED compared to those with less than a high school education

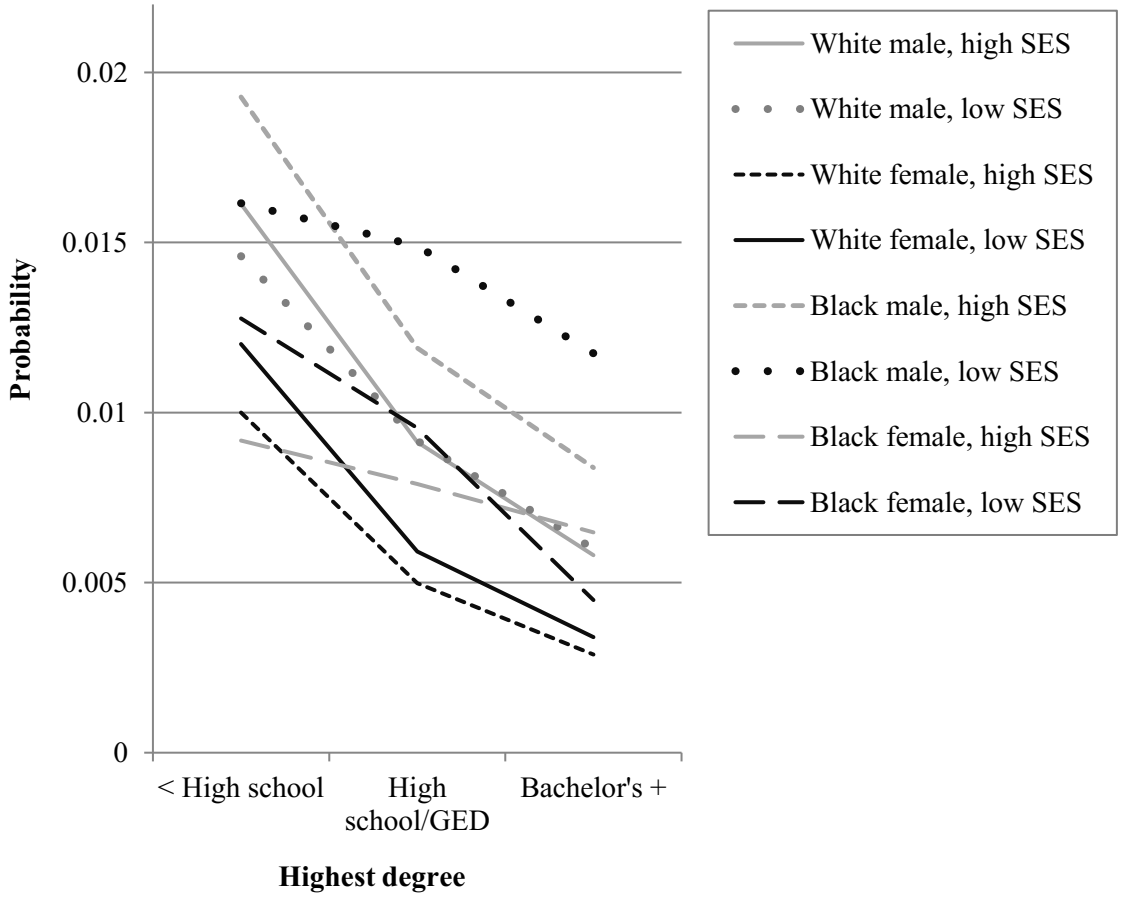


Panel B: Odds ratios for those with a Bachelor's degree or higher compared to those with a high school diploma or GED



Notes: Group-specific odds ratios are calculated based on Model 3b in Appendix Table A3.4. Error bars represent 95% confidence intervals. SES = Socioeconomic origin.

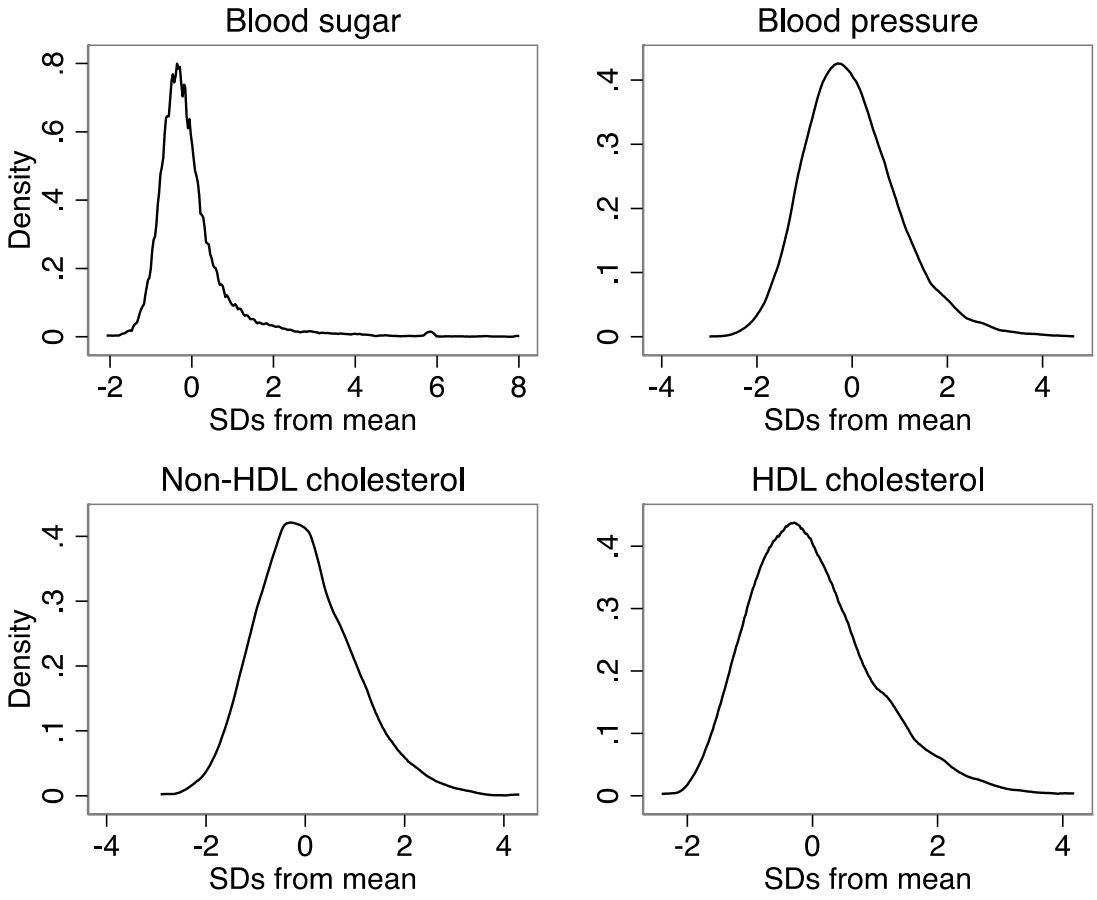
Appendix Figure A3.3. Predicted probability of mortality, by sociodemographic group and highest degree attained



Notes: Predicted probabilities are calculated using Model 3b in Appendix Table A3.4. Age is held at 58 years and covariates are held at their respondent-level means. SES = Socioeconomic origin.

CHAPTER 4 APPENDICES

Appendix Figure A4.1. Distributions of standardized biomarkers



Notes: Biomarkers are standardized across all n = 20,927 observations.

Appendix Table A4.1. Additional descriptive statistics

	Mean or %	(SD)	Min, Max
Observations (n = 20,927)			
Body mass index (BMI)	29.69	6.22	14.0, 61.7
Smoking status			
Never smoker	44.3%		
Former smoker	42.1%		
Current smoker	13.6%		
Respondents (n = 15,077)			
Mother's years of education			
Less than 8	22.2%		
8	14.0%		
Greater than 8, details unknown	5.0%		
9-11	12.5%		
12	32.4%		
13-15	7.4%		
16 or more	6.5%		
Perceived childhood SES			
Poor	30.0%		
About average	62.7%		
Pretty well-off	7.3%		
Father present and employed	72.5%		
Region of birth			
Northeast	17.9%		
Midwest	27.6%		
South	33.6%		
West	8.5%		
Outside U.S.	12.5%		
Childhood health			
Poor or fair	6.6%		
Good	15.8%		
Very good	24.5%		
Excellent	53.1%		

Appendix Table A4.2. Percentiles of biomarkers by years of education

	Years of education			
	< 12	12	13-15	≥16
Standardized biomarkers				
Blood sugar				
10th percentile	-0.75	-0.84	-0.84	-0.87
50th percentile	-0.07	-0.18	-0.26	-0.29
90th percentile	1.43	0.99	0.87	0.68
Blood pressure				
10th percentile	-1.07	-1.15	-1.21	-1.25
50th percentile	0.10	-0.05	-0.15	-0.25
90th percentile	1.59	1.36	1.23	1.03
Non-HDL cholesterol				
10th percentile	-1.22	-1.19	-1.19	-1.23
50th percentile	-0.12	-0.06	-0.06	-0.11
90th percentile	1.32	1.35	1.35	1.25
HDL cholesterol				
10th percentile	-1.25	-1.16	-1.16	-1.12
50th percentile	-0.26	-0.18	-0.09	-0.04
90th percentile	1.16	1.30	1.40	1.49
Unstandardized biomarkers				
Blood sugar (%)				
10th percentile	5.11	5.02	5.02	4.99
50th percentile	5.78	5.68	5.60	5.57
90th percentile	7.28	6.84	6.72	6.53
Blood pressure (mmHg)				
10th percentile	109.33	107.67	106.33	105.67
50th percentile	133.00	130.00	128.00	126.00
90th percentile	163.33	158.67	156.00	152.00
Non-HDL cholesterol (mg/dL)				
10th percentile	96.24	97.67	97.71	96.07
50th percentile	139.62	141.89	141.92	139.72
90th percentile	195.84	197.06	196.97	192.93
HDL cholesterol (mg/dL)				
10th percentile	34.45	35.88	35.88	36.48
50th percentile	50.28	51.59	53.01	53.73
90th percentile	73.00	75.14	76.80	78.29
Sample size	N = 3,884	N = 6,886	N = 4,976	N = 5,181

Appendix Table A4.3. Association of highest degree with standardized biomarkers across their distributions (Model 1b)

	Mean	**Quantile .10**	**Quantile .50**	**Quantile .90**
Reference category: < High school	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)
Blood sugar				
High school or GED	-0.196 (0.026) ***	-0.039 (0.015) *	-0.090 (0.016) ***	-0.490 (0.077) ***
Associate's or some college	-0.261 (0.027) ***	-0.062 (0.017) ***	-0.136 (0.017) ***	-0.639 (0.081) ***
Bachelor's or higher	-0.361 (0.027) ***	-0.077 (0.016) ***	-0.197 (0.015) ***	-0.855 (0.081) ***
Blood pressure				
High school or GED	-0.169 (0.024) ***	-0.081 (0.027) **	-0.160 (0.027) ***	-0.279 (0.057) ***
Associate's or some college	-0.234 (0.025) ***	-0.117 (0.029) ***	-0.223 (0.028) ***	-0.373 (0.057) ***
Bachelor's or higher	-0.323 (0.025) ***	-0.166 (0.029) ***	-0.301 (0.026) ***	-0.543 (0.058) ***
Non-HDL cholesterol				
High school or GED	0.029 (0.022)	0.006 (0.032)	0.022 (0.028)	0.008 (0.046)
Associate's or some college	-0.004 (0.024)	-0.013 (0.033)	0.004 (0.028)	-0.019 (0.046)
Bachelor's or higher	-0.064 (0.024) **	-0.048 (0.033)	-0.054 (0.028)	-0.123 (0.049) *
HDL cholesterol				
High school or GED	0.109 (0.022) ***	0.084 (0.027) **	0.077 (0.027) **	0.116 (0.043) **
Associate's or some college	0.165 (0.024) ***	0.075 (0.025) **	0.147 (0.029) ***	0.220 (0.044) ***
Bachelor's or higher	0.226 (0.024) ***	0.122 (0.025) ***	0.207 (0.028) ***	0.305 (0.050) ***

Notes: Standard errors are clustered at the household level and standard errors for quantile regression estimates are bootstrapped across 200 iterations. Control variables include age, age-squared, year of birth, the interaction between age and year of birth, and the interaction between age-squared and year of birth.

*** p < .001; ** p < .01; * p < .05 (two-tailed test); N = 20,927 observations from 15,077 respondents

Appendix Table A4.4. Association of highest degree with standardized biomarkers across their distributions (Model 2b)

	Mean	**Quantile .10**	**Quantile .50**	**Quantile .90**
Reference category: < High school	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)
Blood sugar				
High school or GED	-0.065 (0.028) *	-0.017 (0.016)	-0.024 (0.016)	-0.168 (0.085) *
Associate's or some college	-0.123 (0.029) ***	-0.040 (0.020) *	-0.068 (0.016) ***	-0.307 (0.084) ***
Bachelor's or higher	-0.189 (0.030) ***	-0.043 (0.019) *	-0.108 (0.018) ***	-0.461 (0.081) ***
Blood pressure				
High school or GED	-0.115 (0.025) ***	-0.035 (0.026)	-0.110 (0.029) ***	-0.211 (0.055) ***
Associate's or some college	-0.180 (0.027) ***	-0.062 (0.030) *	-0.172 (0.030) ***	-0.313 (0.060) ***
Bachelor's or higher	-0.276 (0.028) ***	-0.126 (0.030) ***	-0.260 (0.032) ***	-0.471 (0.059) ***
Non-HDL cholesterol				
High school or GED	-0.003 (0.023)	-0.031 (0.034)	0.005 (0.030)	-0.020 (0.050)
Associate's or some college	-0.044 (0.026)	-0.054 (0.036)	-0.024 (0.032)	-0.043 (0.055)
Bachelor's or higher	-0.095 (0.027) ***	-0.081 (0.040) *	-0.076 (0.031) *	-0.115 (0.057) *
HDL cholesterol				
High school or GED	0.073 (0.023) **	0.042 (0.028)	0.037 (0.026)	0.067 (0.044)
Associate's or some college	0.122 (0.025) ***	0.021 (0.029)	0.103 (0.030) **	0.166 (0.054) **
Bachelor's or higher	0.242 (0.027) ***	0.091 (0.031) **	0.224 (0.029) ***	0.335 (0.052) ***

Notes: Standard errors are clustered at the household level and standard errors for quantile regression estimates are bootstrapped across 200 iterations.

Control variables include age, age-squared, year of birth, the interaction between age and year of birth, the interaction between age-squared and year of birth, sex and race/ethnicity, maternal education, perceived socioeconomic status in childhood, an indicator of paternal presence/employment, region of birth, and health in childhood.

*** p < .001; ** p < .01; * p < .05 (two-tailed test); N = 20,927 observations from 15,077 respondents

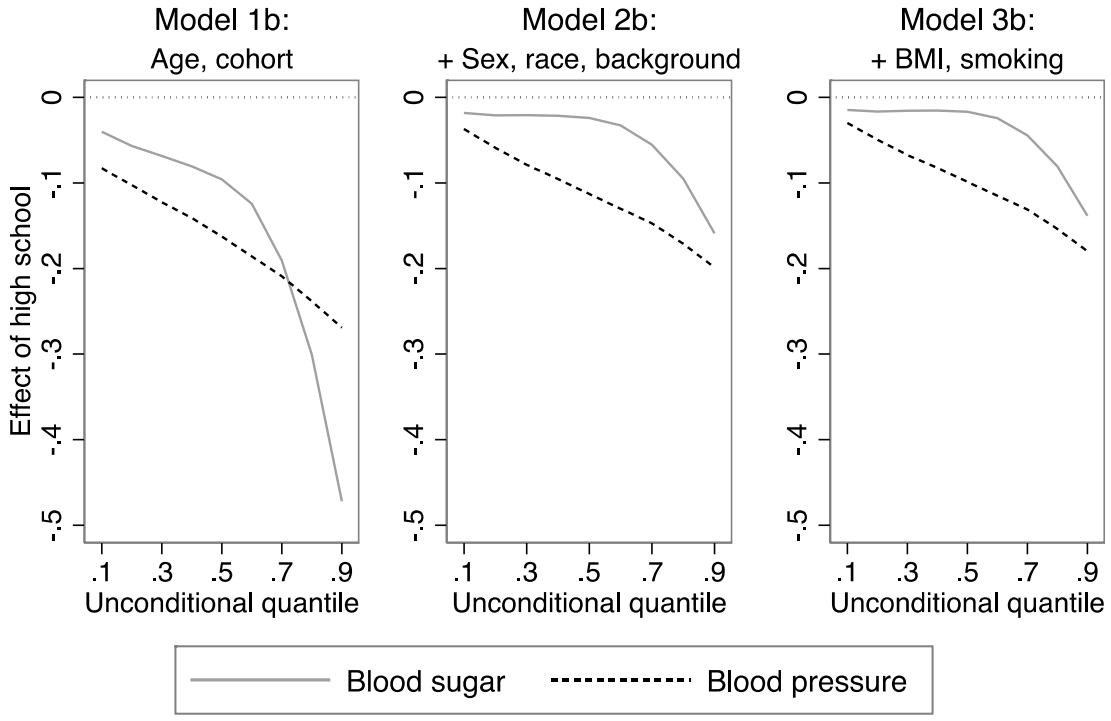
Appendix Table A4.5. Association of highest degree with standardized biomarkers across their distributions (Model 3b)

	Mean	**Quantile .10**	**Quantile .50**	**Quantile .90**
Reference category: < High school	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)
Blood sugar				
High school or GED	-0.055 (0.027) *	-0.014 (0.015)	-0.017 (0.015)	-0.146 (0.084)
Associate's or some college	-0.107 (0.029) ***	-0.034 (0.018)	-0.056 (0.016) ***	-0.271 (0.081) **
Bachelor's or higher	-0.146 (0.030) ***	-0.026 (0.019)	-0.077 (0.018) ***	-0.364 (0.083) ***
Blood pressure				
High school or GED	-0.100 (0.025) ***	-0.028 (0.026)	-0.096 (0.026) ***	-0.192 (0.058) **
Associate's or some college	-0.159 (0.027) ***	-0.051 (0.033)	-0.152 (0.030) ***	-0.287 (0.063) ***
Bachelor's or higher	-0.231 (0.028) ***	-0.097 (0.034) **	-0.216 (0.033) ***	-0.415 (0.064) ***
Non-HDL cholesterol				
High school or GED	0.002 (0.023)	-0.028 (0.032)	0.009 (0.028)	-0.009 (0.050)
Associate's or some college	-0.037 (0.026)	-0.050 (0.034)	-0.019 (0.028)	-0.029 (0.055)
Bachelor's or higher	-0.081 (0.027) **	-0.075 (0.036) *	-0.066 (0.031) *	-0.084 (0.056)
HDL cholesterol				
High school or GED	0.053 (0.022) *	0.029 (0.027)	0.018 (0.026)	0.040 (0.045)
Associate's or some college	0.093 (0.024) ***	0.005 (0.026)	0.075 (0.026) **	0.124 (0.054) *
Bachelor's or higher	0.172 (0.026) ***	0.050 (0.030)	0.156 (0.027) ***	0.236 (0.055) ***

Notes: Standard errors are clustered at the household level and standard errors for quantile regression estimates are bootstrapped across 200 iterations. Control variables include age, age-squared, year of birth, the interaction between age and year of birth, the interaction between age-squared and year of birth, sex and race/ethnicity, maternal education, perceived socioeconomic status in childhood, an indicator of paternal presence/employment, region of birth, health in childhood, body mass index (BMI), BMI-squared, and smoking status.

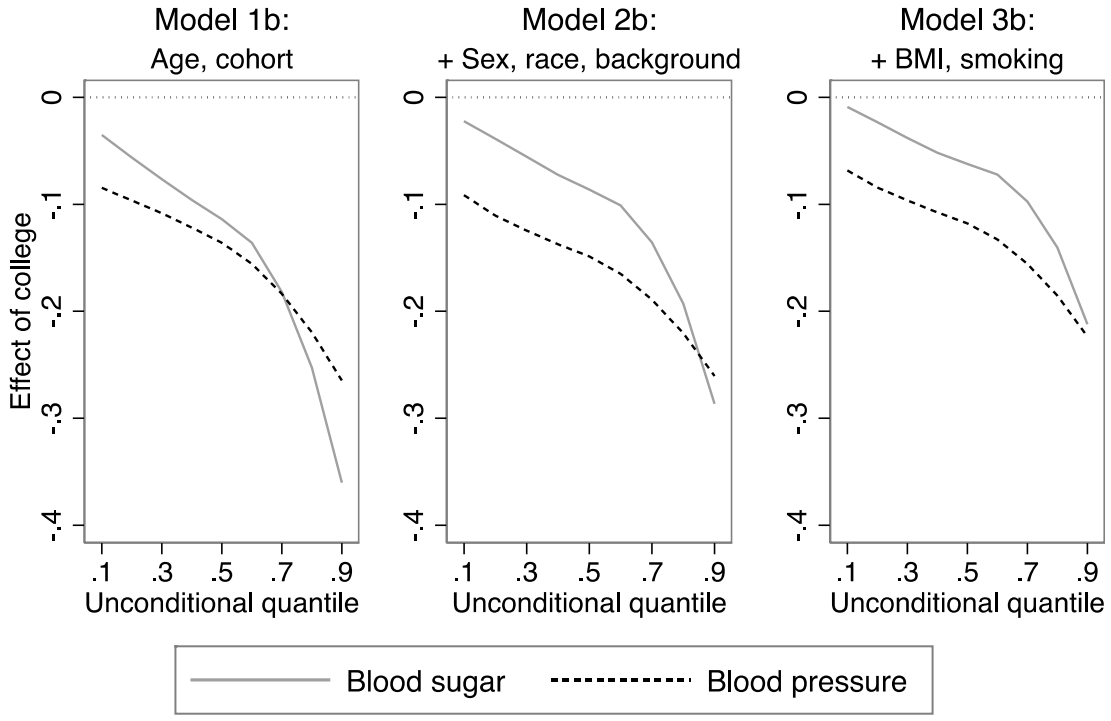
*** p < .001; ** p < .01; * p < .05 (two-tailed test); N = 20,972 observations from 15,077 respondents

Appendix Figure A4.2. Estimated disparities in standardized biomarkers across their distributions between those who did not complete high school (reference category) and those with a high school diploma or GED, from Models 1b, 2b, and 3b



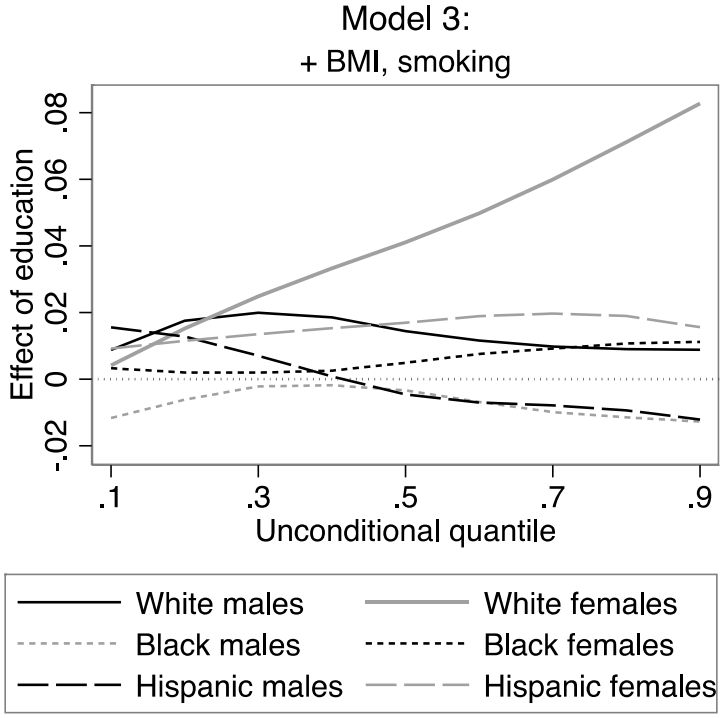
Notes: Estimates are based on unconditional quantile regression models, as shown in Appendix Tables A4.3-A4.5. Model 1b controls for age, age-squared, year of birth, the interaction between age and year of birth, and the interaction between age-squared and year of birth. Model 2b adds controls for sex and race/ethnicity, maternal education, perceived socioeconomic status in childhood, an indicator of paternal presence/employment, region of birth, and health in childhood. Model 3b adds controls for body mass index (BMI), BMI-squared, and smoking status. Lowess smoothing functions are applied. Biomarkers are standardized across the n = 20,972 observations.

Appendix Figure A4.3. Estimated disparities in standardized biomarkers across their distributions between those with a high school diploma or GED (reference category) and those with a Bachelor's degree or higher, from Models 1b, 2b, and 3b



Notes: Estimates are based on unconditional quantile regression models, as shown in Appendix Tables A4.3-A4.5. Model 1b controls for age, age-squared, year of birth, the interaction between age and year of birth, and the interaction between age-squared and year of birth. Model 2b adds controls for sex and race/ethnicity, maternal education, perceived socioeconomic status in childhood, an indicator of paternal presence/employment, region of birth, and health in childhood. Model 3b adds controls for body mass index (BMI), BMI-squared, and smoking status. Lowess smoothing functions are applied. Biomarkers are standardized across the n = 20,972 observations.

Appendix Figure A4.4. Association of a year of education with standardized HDL cholesterol across its distribution, by sex and race



Notes: Estimates are based on unconditional quantile regression (UQR) models (not shown) that incorporate an interaction between sex/race and years of schooling. Lines for those of "other" race are not shown as they were imprecisely estimated. Model 3 controls for age, age-squared, year of birth, the interaction between age and year of birth, the interaction between age-squared and year of birth, maternal education, perceived socioeconomic status in childhood, an indicator of paternal presence/employment, region of birth, health in childhood, body mass index (BMI), BMI-squared, and smoking status. Lowess smoothing functions are applied. Biomarkers are standardized across the n = 20,972 observations.

References

Abecasis, Gonçalo R., *et al*. 2012. "An Integrated Map of Genetic Variation from 1,092 Human Genomes." *Nature* 491:56-65.

Acevedo-Garcias, Dolores, Mah-J. Soobader, and Lisa F. Berkman. 2007. "Low Birthweight Among US Hispanic/Latino Subgroups: The Maternal Foreign-Born Status and Education." *Social Science & Medicine* 65(12):2503-2516.

Albouy, Valerie and Laurent Lequien. 2009. "Does Compulsory Education Lower Mortality?" *Journal of Health Economics* 28:155-168.

Allard-Ratick, Marc P., Pratik Sandesara, Arshed A. Quyyumi, and Laurence S. Sperling. 2019. "Everything in Moderation: Investigating the U-Shaped Link Between HDL Cholesterol and Adverse Outcomes." *US Cardiology Review* 13(1):49-53.

Altshuler, David M., Richard A. Gibbs, and Leena Peltonen. 2010. "Integrating Common and Rare Genetic Variation in Diverse Human Populations." *Nature* 467:52-58.

American Diabetes Association (ADA). 2018. "2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes—2018." *Diabetes Care* 41(Supplement 1):S13-S27.

Amin, Vikesh, Jere R. Behrman, and Hans-Peter 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.

Andersson, Matthew A. and Vida Maralani. 2015. "Early-Life Characteristics and Educational Disparities in Smoking." *Social Science & Medicine* 144:138-147.

Andersson, Matthew A. and Kenneth Vaughan. 2017. "Adult Health Returns to Education by Key Childhood Social and Economic Indicators: Results from Representative European Data." *SSM – Population Health* 3:411-418.

Aune, Dagfinn, Abhijit Sen, Manya Prasad, Teresa Norat, Imre Janszky, Serena Tonstad, Pål Romundstad, and Lars J. Vatten. 2016. "BMI and All Cause Mortality: Systematic Review and Non-Linear Dose-Response Meta-Analysis of 230 Cohort Studies with 3.74 Million Deaths among 30.3 Million Participants." *BMJ* 353:i2156.

Ayanian, John Z., Joel S. Weissman, Eric C. Schneider, Jack A. Ginsburg, and Alan M. Zaslavsky. 2000. "Unmet Health Needs of Uninsured Adults in the United States." *JAMA* 284(16):2061-2069.

Ayanian, John Z., Alan M. Zaslavsky, Joel S. Weissman, Eric C. Schneider, and Jack A. Ginsburg. 2003. "Undiagnosed Hypertension and Hypercholesterolemia Among Uninsured and Insured Adults in the Third National Health and Nutrition Examination Survey." *American Journal of Public Health* 93(12):2051-2054.

Barcellos, Silvia H., Leandro S. Carvalho, and Patrick Turley. 2019. "Distributional Effects of Education on Health." *NBER Working Paper No. 25898*.

Batty, G. David, Geoff Der, Sally Macintyre, and Ian J. Deary. 2006. "Does IQ Explain Socioeconomic Inequalities in Health? Evidence from a Population Based Cohort Study in the West of Scotland." *British Medical Journal* 332:1-5.

Bauldry, Shawn. 2014. "Conditional Health-Related Benefits of Higher Education: An Assessment of Compensatory versus Accumulative Mechanisms." *Social Science & Medicine* 111:94-100.

Bauldry, Shawn. 2015. "Variation in the Protective Effect of Higher Education Against Depression." *Society and Mental Health* 5(2):145-161.

Behrman, Jere R., Hans-Peter Kohler, Vibeke Myrup Jensen, Dorthe Pedersen, Inge Petersen, Paul Bingley, and Kaare Christensen. 2011. "Does More Schooling Reduce Hospitalization and Delay Mortality? New Evidence Based on Danish Twins." *Demography* 48:1347-1375.

Belsky, Daniel W., Benjamin W. Domingue, Robbee Wedow, Louise Arsenault, Jason D. Boardman, Avshalom Caspi, Dalton Conley, Jason M. Fletcher, Jeremy Freese, Pamela Herd, Terrie E. Moffitt, Richie Poulton, Kamil Sicinski,Jasmin Wertz, and Kathleen Mullan Harris. 2018. "Genetic Analysis of Social-Class Mobility in Five Longitudinal Studies." *PNAS* 115(31):E7275-E7284.

Belsky, Daniel W., Terrie E. Moffitt, David L. Corcoran, Benjamin Domingue, Hona Lee Harrington, Sean Hogan, Renate Houts, Sandhya Ramrakha, Karen Sugden, Benjamin S. Williams, Richie Poulton, and Avshalom Caspi. 2016. "The Genetics of Success: How Single-Nucleotide Polymorphisms Associated with Educational Attainment Relate to Life-Course Development." *Psychological Science* 27(7):957-972.

Ben-Shlomo, Yoav and Diana Kuh 2002. "A Life Course Approach to Chronic Disease Epidemiology: Conceptual Models, Empirical Challenges and Interdisciplinary Perspectives." *International Journal of Epidemiology* 31:285-93.

Bertrand, Marianne and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review* 94(4):991-1013.

Blau, Peter and Otis Dudley Duncan. 1967. *The American Occupational Structure*. New York: Free Press.

Boardman, Jason D. and Jason M. Fletcher. 2015. "To Cause of Not to Cause? That is the Question, but Identical Twins Might Not Have All of the Answers." *Social Science & Medicine* 127:198-200.

Borrell, Luisa N., Catarina I. Kiefe, David R. Williams, Ana V. Diez-Roux, Penny Gordon-Larsen. 2006. "Self-Reported Health, Perceived Racial Discrimination, and Skin Color in African Americans in the CARDIA Study." *Social Science & Medicine* 63:1415-1427.

Bottai, Matteo, Edward A. Frongillo, Xuemei Sui, Jennifer R. O'Neill, Robert E. McKeown, Trudy L. Burns, Angela D. Liese, Steven N. Blair, and Russell R. Page. 2014. "Use of Quantile Regression to Invesetigate the Longitudinal Association Between Physical Activity and Body Mass Index." *Obesity* 22(5):E149-E156.

Bourdieu, Pierre. 1984. *Distinction: A Social Critique of the Judgment of Taste*. Cambridge, MA: Harvard University Press.

Bowles, Samuel and Herbert Gintis. 1976. *Schooling in Capitalist America*. New York: Basic Books.

Braakman, Nils. 2011. "The Causal Relationship Between Education, Health, and Health-Related Behavior: Evidence From a Natural Experiment in England." *Journal of Health Economics* 30:753-763.

Branigan, Amelia R., Kenneth J. McCallum, and Jeremy Freese. 2013. "Variation in the Heritability of Educational Attainment: An International Meta-Analysis." *Social Forces* 92(1):109-140.

Brody, Gene H., Tianyi Yu, Edith Chen, Gregory E. Miller, Steven M. Kogan, and Steven R. H. Beach. 2013. "Is Resilience Only Skin Deep? Rural African Americans' Socioeconomic Status-Related Risk and Competence in Preadolescence and Psychological Adjustment and Allostatic Load at Age 19." *Psychological Science* 24(7):1285-1293.

Brummett, Beverly H., Michael A. Babyak, Ilene C. Siegler, Michael Shanahan, Kathleen Mullan Harris, Glen H. Elder, and Redford B. Williams. 2011. "Systolic Blood Pressure, Socioeconomic Status, and Biobehavioral Risk Factors in a Nationally Representative U.S. Young Adult Sample." *Hypertension* 58(2):161-166.

Brunner, Eric J., Martin J. Shipley, Daniel R. Witte, John H. Fuller, and Michael G. Marmot. 2006. "Relation Between Blood Glucose and Coronary Mortality Over 33 Years in the Whitehall Study." *Diabetes Care* 29(1):26-31.

Bulik-Sullivan, Brendon, Hilary K. Finucane, Verneri Anttila, Alexander Gusev, Felix R. Day, Po-Ru Loh, ReproGen Consortium, Psychiatric Genomics Consortium, Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3, Laramie Duncan, John R. B. Perry, Nick Patterson, Elise B. Robinson, Mark J. Daly, Alkes L. Price, and Benjamin M. Neale. 2015. "An Atlas of Genetic Correlations Across Human Diseases and Traits." *Nature Genetics* 47:1236-1241.

Buttorff, Christine, Teague Ruder, and Melissa Bauman. 2017. *Multiple Chronic Conditions in The United States*. Santa Monica, CA: RAND Corporation.

Carroll, Margaret D., Cheryl Fryar, and Duong T. Nguyen. 2017. "Total and High-Density Lipoprotein Cholesterol in Adults in the United States, 2015-2016." *NCHS Data Brief No. 290*.

Carson, April P., Caroline S. Fox, Darren K. McGuire, Emily B. Levitan, Martin Laclaustra, Devin M. Mann, and Paul Muntner. 2010. "Low Hemoglobin A1c and Risk of All-Cause Mortality Among US Adults Without Diabetes." *Circulation: Cardiovascular Quality and Outcomes* 3(6):661-667.

Case, Anne, Angela Fertig, and Christina Paxson. 2005. "The Lasting Impact of Childhood Health and Circumstance." *Journal of Health Economics* 24:365-389.

Case, Anne and Christina Paxson. 2005. "Sex Differences in Morbidity and Mortality." *Demography* 42(2):189-214.

Centers for Disease Control (CDC). 2016. *Defining Adult Overweight and Obesity*. Atlanta, GA: Centers for Disease Control, U.S. Department of Health & Human Services.

Centers for Disease Control (CDC). 2017. *National Diabetes Statistics Report, 2017: Estimates of Diabetes and Its Burden in the United States*. Atlanta, GA: Centers for Disease Control and Prevention, U.S. Department of Health and Human Services.

Cesarini, David and Peter M. Visscher. 2017. “Genetics and Educational Attainment.” *NPJ Science of Learning* 2.

Chabris. Christopher F., James J. Lee, David Cesarini, Daniel J. Benjamin, and David I. Laibson. 2015. “The Fourth Law of Behavior Genetics.” *Current Directions in Psychological Science* 24(4):304-312.

Chang, Virginia W. and Diane S. Lauderdale. 2009. “Fundamental Cause Theory, Technological Innovation, and Health Disparities: The Case of Cholesterol in the Era of Statins.” *Journal of Health and Social Behavior* 50:245-260.

Chaparro, M.Pia, Amanda Hughes, Meena Kumari, and Michaela Benzeval. 2019. “The Association Between Self-Rated Health and Underlying Biomarker Levels is Modified by Age, Gender, and Household Income: Evidence from Understanding Society—The UK Household Longitudinal Study.” *SSM – Population Health* 8:100406.

Chen, Edith, Gregory E. Miller, Gene H. Brody, and ManKit Lei. 2015. “Neighborhood Poverty, College Attendance, and Diverging Profiles of Substance Use and Allostatic Load in Rural African American Youth.” *Clinical Psychological Science* 3(5):675-685.

Choo, Hae Yeon and Myra Marx Ferree. 2010. “Practicing Intersectionality in Sociological Research: A Critical Analysis of Inclusions, Interactions, and Institutions in the Study of Inequalities.” *Sociological Theory* 28:129-149.

Clark, Damon and Heather Royer. 2013. “The Effect of Education on Adult Health and Mortality: Evidence from Britain.” *American Economic Review* 103(6):2087-2120.

Clouston, Sean A. P., Marcie S. Rubin, Jo C. Phelan, and Bruce G. Link. 2016. “A Social History of Disease: Contextualizing the Rise and Fall of Social Inequalities in Cause-Specific Mortality.” *Demography* 53(5):1631-1656.

Cockerham, William C. 2005. “Health Lifestyle Theory and the Convergence of Agency and Structure.” *Journal of Health and Social Behavior* 46(1):51-67.

Cohen, Alison Klebanoff and S. Leonard Syme. 2013. “Education: A Missed Opportunity for Public Health Intervention.” *American Journal of Public Health* 103(6):997-1001.

Colen, Cynthia G., David M. Ramey, Elizabeth C. Cooksey, and David R. Williams. 2018. “Racial Disparities in Health Among Nonpoor African Americans and Hispanics: The Role of Acute and Chronic Discrimination.” *Social Science & Medicine*:167-180.

Collins, Francis S., Eric D. Green, Alan E. Guttmacher, and Mark S. Guyer. 2003. “A Vision for the Future of Genomics Research.” *Nature* 422:835-847.

Collins, Patricia Hall. 1990. *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*. Boston, MA: Unwin Hyman.

Conley, Dalton. 2016. “Socio-Genomic Research Using Genome-Wide Molecular Data.” *Annual Review of Sociology* 42(1):275-299.

Conley, Dalton, Benjamin W. Domingue, David Cesarini, Christopher Dawes, Cornelius Rietveld, and Jason D. Boardman. 2015. “Is the Effect of Parental Education on Offspring Biased or Moderated by Genotype?” *Sociological Science* 2:82-105.

Conley, Dalton and Jason Fletcher. 2017. *The Genome Factor: What the Social Genomics Revolution Reveals about Ourselves, Our History, and the Future*. Princeton, NJ: Princeton University Press.

Conti, Gabriella, James J. Heckman, and Sergio Urzua. 2010. "The Education-Health Gradient." *American Economic Review* 100(2):234-238.

Coutinho, Mario, Yong Wang, Hertzel C. Gerstein, and Salim Yusuf. 1999. "The Relationship Between Glucose and Incident Cardiovascular Events." *Diabetes Care* 22(2):233-240.

Crenshaw, Kimberle. 1989. "Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics." *The University of Chicago Legal Forum*:139-167.

Crimmins, Eileen M., Jessica Faul, Jung Ki Kim, Heidi Guyer, Kenneth Langa, Mary Beth Ofstedal, Amanda Sonnega, Robert Wallace, and David Weir. 2013. *Documentation of Biomarkers in the 2006 and 2008 Health and Retirement Study*. Ann Arbor, MI: Survey Research Center, University of Michigan.

Crimmins, Eileen M., Jessica Faul, Jung Ki Kim, and David Weir. 2015. *Documentation of Biomarkers in the 2010 and 2012 Health and Retirement Study*. Ann Arbor, MI: Survey Research Center, University of Michigan.

Crimmins, Eileen M., Jung Ki Kim, and Aïda Solé-Auró. 2011. "Sex Differences in Health: Results from SHARE, ELSA, and HRS." *The European Journal of Public Health* 21(1):81-91.

Cross, Christina J. 2018. "Extended Family Households among Children in the United States: Differences by Race/Ethnicity and Socioeconomic Status." *Population Studies* 72(2):235-251.

Currie, Craig J., John R. Peters, Aodán Tynan, Marc Evans, Robert J. Heine, Oswaldo L. Bracco, Tony Zagar, and Chris D. Poole. 2010. "Survival as a Function of HbA1C in People with Type 2 Diabetes: A Retrospective Cohort Study." *The Lancet* 375(9713):481-489.

Currie, Janet and Enrico Moretti. 2003. "Mother's Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings." *The Quarterly Journal of Economics* 118(4):1495-1532.

Cutler, David and Adriana Lleras-Muney. 2008. "Education and Health: Evaluating Theories and Evidence." In Robert F. Schoeni, James S. House, George A. Kaplan, and Harold Pollack (Eds.) *Making Americans Healthier: Social and Economic Policy as Health Policy*. New York: Russell Sage Foundation.

Cutler, David and Adriana Lleras-Muney. 2010. "Understanding Differences in Health Behaviors by Education." *Journal of Health Economics* 29(1):1-28.

Denney, Justin T., Richard G. Rogers, Robert A. Hummer, and Fred C. Pampel. 2010. "Education Inequality in Mortality: The Age and Gender Specific Mediating Effects of Cigarette Smoking." *Social Science Research* 39(4):662-673.

Domingue, Benjamin W., Daniel W. Belsky, Dalton Conley, Kathleen Mullan Harris, and Jason D. Boardman. 2015. "Polygenic Influence on Educational Attainment: New Evidence from the National Longitudinal Study of Adolescent to Adult Health." *AERA Open* 1(3).

Domingue, Benjamin W., Dalton Conley, Jason Fletcher, and Jason D. Boardman. 2016. "Cohort Effects in the Genetic Influence on Smoking." *Behavior Genetics* 46(1):31-42.

Dowd, Jennifer Beam and Anna Zajacova. 2007. "Does the Predictive Power of Self-Rated Health for Subsequent Mortality Risk Vary by Socioeconomic Status in the US?" *International Journal of Epidemiology* 36(6):1214-1221.

Dowd, Jennifer Beam and Anna Zajacova. 2010. "Does Self-Rated Health Mean the Same Thing Across Socioeconomic Groups? Evidence from Biomarker Data." *Annals of Epidemiology* 20(10):743-749.

Downey, Douglas B. and Shana Pribesh. 2004. "When Race Matters: Teachers' Evaluations of Students' Classroom Behavior." *Sociology of Education* 77(4):267-282.

Dudbridge, Frank. 2013. "Power and Predictive Accuracy of Polygenic Risk Scores." *PLoS Genetics* 9(3).

Dudovitz, Rebecca N, Bergen B. Nelson, Tumaini R. Coker, Christopher Biely, Ning Li, Lynne C. Wu, and Paul J. Chung. 2016. "Long-Term Health Implications of School Quality." *Social Science & Medicine* 158:1-7.

Durkheim, Emile. 1956[1922]. *Education and Society*. Glencoe, IL: Free press.

Duster, Troy. 2003. *Backdoor to Eugenics*. New York: Routledge.

Evans, Clare R., David R. Williams, Jukka-Pekka Onnela, and S. V. Subramanian. 2018. "A Multilevel Approach to Modeling Health Inequalities at the Intersection of Multiple Social Identities." *Social Science & Medicine* 203:64-73.

Farkas, George. 2003. "Cognitive Skills and Noncognitive Traits and Behaviors in Stratification Processes." *Annual Review of Sociology* 29:541–562.

Farmer, Melissa M. and Kenneth F. Ferraro. 2005. "Are Racial Disparities in Health Conditional on Socioeconomic Status?" *Social Science & Medicine* 60(1):191-204.

Firpo, Sergio, Nicole M. Fortin, and Thomas Lemieux. 2009. "Unconditional Quantile Regressions." *Econometrica* 77(3):953-973.

Fischer, Mary J. and Douglas S. Massey. 2004. "The Ecology of Racial Discrimination." *City & Community* 3(3):221-241.

Flaster, Allyson. 2018. "Kids, College, and Capital: Parental Financial Support and College Choice." *Research in Higher Education*:1-42.

Fletcher, Jason. 2015. "New Evidence of the Effects of Education on Health in the US: Compulsory Schooling Laws Revisited." *Social Science & Medicine* 127:101-107.

Fletcher, Jason M. and Steven F. Lehrer. 2011. "Genetic Lotteries within Families." *Journal of Health Economics* 30(4):647-659.

Fletcher, Jenna. 2017. "What Should My Cholesterol Level Be at My Age?" *Medical News Today*.

Freese, Jeremy. 2018. "The Arrival of Social Science Genetics." *Contemporary Sociology: A Journal of Reviews* 47(5):524-536.

Freese, Jeremy and Karen Lutfey. 2011. "Fundamental Causality: Challenges of an Animating Concept for Medical Sociology." In Bernice A. Pescosolido, Jack K. Martin, Jane McLeod, and Anne Rogers (Eds.) *Handbook of the Sociology of Health, Illness, and Healing*. New York: Springer.

Frisvold, David and Ezra Golberstein. 2011. "School Quality and the Education-Health Relationship: Evidence from Blacks in Segregated Schools." *Journal of Health Econonomics* 30(6):1232-1245.

Fujimura, Joan H., Deborah A. Bolnick, Ramya Rajagopalan, Jay S. Kaufman, Richard C. Lewontin, Troy Duster, Pilar Ossorio, and Jonathan Marks. 2014. "Clines Without Classes: How to Make Sense of Human Variation." *Sociological Theory* 32(3):208-227.

Fujiwara, Takeo and Ichiro Kawachi. 2009. "Is Education Causally Related to Better Health? A Twin Fixed-Effect Study in the USA." *International Journal of Epidemiology* 38:1310–1322.

Gaddis, S. Michael. 2017. "How Black are Lakisha and Jamal? Racial Perceptions of Names Used in Correspondence Audit Studies." *Sociological Science* 4:469-489.

Gage, Timothy B., Fu Fang, Erin O'Neill, and Greg DiRienzo. 2013. "Maternal Education, Birth Weight, and Infant Mortality in the United States." *Demography* 50(2):615-635.

Galama, Titus J., Adriana Lleras-Muney, and Hans van Kippersluis. 2018. "The Effect of Education on Health and Mortality: A Review of Experimental and Quasi-Experimental Evidence." *NBER Working Paper No. 24225*.

Galea, Sandro, Melissa Tracy, Katherine J. Hoggatt, Charles DiMaggio, and Adam Karpati. 2011. "Estimated Deaths Attributable to Social Factors in the United States." *American Journal of Public Health* 101(8):1456-1465.

Galobardes, Bruna, John Lynch, and George Davey Smith. 2004. "Childhood Socioeconomic Circumstances and Cause-Specific Mortality in Adulthood: Systematic Review and Interpretation." *Epidemiologic Reviews* 26:7-21.

Gaydosh, Lauren, Kristen M. Schorpp, Edith Chen, Gregory E. Miller, and Kathleen Mullan Harris. 2018. "College Completion Predicts Lower Depression but Higher Metabolic Syndrome Among Disadvantaged Minorities in Young Adulthood." *PNAS* 115(1):109-114.

Gerstel, Naomi. 2011. "Rethinking Families and Community: The Color, Class, and Centrality of Extended Kin Ties." *Sociological Forum* 26(1):1-20.

Gillespie, Kathleen D. and Kimberly A. Hurvitz. 2013. "Prevalence of Hypertension and Uncontrolled Hypertension—United States, 2007-2010." *Morbidity and Mortality Weekly Report* 62(03):144-148.

Glied, Sherry and Adriana Lleras-Muney. 2008. "Technological Innovation and Inequality in Health." *Demography* 45:741-761.

Goesling, Brian. 2007. "The Rising Significance of Education for Health?" *Social Forces* 85(4):1621-1644.

Goosby, Bridget J., Jacob E. Cheadle, and Colter Mitchell. 2018. "Stress-Related Biosocial Mechanisms of Discrimination and African American Health Inequities." *Annual Review of Sociology* 44:319-340.

Gottfredson, Linda S. 2004. “Intelligence: Is It the Epidemiologists’ Elusive ‘Fundamental Cause’ of Social Class Inequalities in Health?” *Journal of Personality and Social Psychology* 86(1):174-199.

Gottfredson, Linda S. and Ian J. Deary. 2004. “Intelligence Predicts Health and Longevity, but Why?” *Current Directions in Psychological Science* 13(1):1-4.

Grossman, Michael. 2015. “The Relationship Between Health and Schooling: What’s New?” *Nordic Journal of Health Economics* 3(1):7-17.

Gruenewald, Tara L., Sheldon Cohen, Karen A. Matthews, Russell Tracy, and Teresa E. Seeman. 2009. “Association of Socioeconomic Status with Inflammation Markers in Black and White Men and Women in the Coronary Artery Risk Development in Young Adults (CARDIA) Study.” *Social Science & Medicine* 69(3):451-459.

Grundy, Scott M., *et al*. 2018. “2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA Guideline on the Management of Blood Cholesterol: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines.” *Circulation* 000:e000-e000.

Haas, Steven A. 2007. “The Long-Term Effects of Poor Childhood Health: An Assessment and Application of Retrospective Reports.” *Demography* 44(1):113-135.

Hahn, Robert H. and Benedict I. Truman. 2015. “Education Improves Public Health and Promotes Health Equity.” *International Journal of Health Services* 45(4):657-678.

Hanas, Ragnar and Garry John. 2010. “2010 Consensus Statement on the Worldwide Standardization of the Hemoglobin A1C Measurement.” *Diabetes Care* 33(8):1903-1904.

Hardy, Melissa A., Francesco Acciai, and Adriana M. Reyes. 2014. “How Health Conditions Translate into Self-Ratings: A Comparative Study of Older Adults Across Europe.” *Journal of Health and Social Behavior* 55(3):320-341.

Harris, Kathleen Mullan and Kristen M. Schorpp. 2018. “Integrating Biomarkers in Social Stratification and Health Research.” *Annual Review of Sociology* 44:361-386.

Hauser, Robert M. and Alberto Palloni. 2011. “Adolescent IQ and Survival in the Wisconsin Longitudinal Study.*” The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences* 66B(Suppl 1):i91-i101.

Hauser, Robert M. and R. J. Willis. 2005. “Survey Design and Methodology in the Health and Retirement Study and the Wisconsin Longitudinal Study.” In Linda J. Waite (Ed.) *Aging, Health, and Public Policy: Demographic and Economic Perspectives*. New York: Population Council.

Hayward, Mark D. and Bridget K. Gorman. 2004. “The Long Arm of Childhood: The Influence of Early-Life Social Conditions on Men’s Mortality.” *Demography* 41(1):87-107.

Health and Retirement Study. 2012a. *Quality Control Report for Genotypic Data, 2012*. Seattle, WA: University of Washington.

Health and Retirement Study. 2012b. “CIDR Health Retirement Study: Imputation Report – 1000 Genomes Project Reference Panel.” Available at:

http://hrsonline.isr.umich.edu/sitedocs/genetics/HRS_1000G_IMPUTE2_REPORT_AUG2012.pdf

Health and Retirement Study. 2013a. *Quality Control Report for Genotypic Data, 2013*. Seattle, WA: University of Washington.

Health and Retirement Study. 2013b. "CIDR Health and Retirement Study, Combined Phases 1-3: Imputation Report – 1000 Genomes Project Reference Panel." Available at: http://hrsonline.isr.umich.edu/sitedocs/genetics/1000G_IMPUTE2report_HRS2_2006_2008_2010.pdf?_ga=2.211200278.2080554925.1540313968-768581060.1537968482

Health and Retirement Study. 2016. *RAND HRS Data, Version P: Public Use Dataset. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740).* Ann Arbor, MI: University of Michigan.

Heath, A.C., K. Berg, L.J. Eaves, M.H. Solaas, L.A. Corey, J. Sundet, P. Magnus, and W.E. Nance. 1985. "Education Policy and the Heritability of Educational Attainment." *Nature* 314(25):734-736.

Heckman, James J. 2006. "Skill Formation and the Economics of Investing in Disadvantaged Children." *Science* 312(5782):1900-1902.

Henmon, V. A. C., M. J. Nelson, and Tom A. Lamke. 1957. *The Henmon-Nelson Tests of Mental Ability.* Boston: Houghton-Mifflin Company.

Henretta, John C., Douglas A. Wolf, Matthew F. Van Voorhis, and Beth J. Soldo. 2012. "Family Structure and the Reproduction of Inequality: Parents' Contribution to Children's College Costs." *Social Science Research* 41(4):876-887.

Herd, Pamela, Deborah Carr, and Carol Roan. 2014. "Cohort Profile: Wisconsin Longitudinal Study (WLS)." *International Journal of Epidemiology* 43:34-41.

Hernández-Alava, Mónica and Gurleen Popli. 2017. "Children's Development and Parental Input: Evidence from the UK Millenium Cohort Study." *Demography* 54(2):485-511.

Hochschild, Jennifer L. 1995. *Facing Up to the American Dream: Race, Class, and the Soul of the Nation*. Princeton, NJ: Princeton University Press.

Holmes, Christopher J. and Anna Zajacova. 2014. "Education as 'the Great Equalizer': Health Benefits for Black and White Adults." *Social Science Quarterly* 95(4):1064-1085.

Hout, Michael. 1984. "Status, Autonomy, and Training in Occupational Mobility." *American Journal of Sociology* 89:1379-1409.

Hout, Michael. 1988. "More Universalism, Less Structural Mobility: The American Occupational Structure in the 1980s." *American Journal of Sociology* 93:1358-1400.

Hout, Michael. 2012. "Social and Economic Returns to College Education in the United States." *Annual Review of Sociology* 38(1):379-400.

Hummer, Robert A. and Joseph T. Lariscy. 2011. "Educational Attainment and Adult Mortality." In Richard G. Rogers and Eileen M. Crimmins (Eds.) *International Handbook of Adult Mortality*. New York, NY: Springer.

Idler, Ellen L. and Yael Benyamini. 1997. "Self-Rated Health and Mortality: A Review of Twenty-Seven Community Studies." *Journal of Health and Social Behavior* 38(1):21-37.

Jackson, Margot I. 2009. "Understanding Links Between Adolescent Health and Educational Attainment." *Demography* 46(4):671-694.

Jacobs, David, Henry Blackburn, Millicent Higgins, Dwayne Reed, Hiroyasu Iso, Gardner McMillan, James Neaton, James Nelson, John Potter, Basil Rifkind, Jacques Rossouw, Richard Shekelle, and Salim Yusuf. 1992. "Report on the Conference on Low Blood Cholesterol: Mortality Associations." *Circulation* 86(3):1046-1060.

James, Sherman A. 1994. "John Henryism and the Health of African-Americans." *Culture, Medicine and Psychiatry* 18:163-182.

James, Sherman A., David S. Strogatz, Steven B. Wing, and Diane L. Ramsey. 1987. "Socioeconomic Status, John Henryism, and Hypertension in Blacks and Whites." *American Journal of Epidemiology* 126(4):664-673.

Jencks, Christopher. 1980. "Heredity, Environment, and Public Policy Reconsidered." *American Sociological Rev*iew 45(5):723-736.

Johnson, Rucker C. 2010. "The Health Returns of Education Policies from Preschool to High School and Beyond." *American Economics Review* 100(2).

Johnston, David W., Grace Lordan, Michael A. Shields, and Agne Suziedelyte. 2015. "Education and Health Knowledge: Evidence from UK Compulsory Schooling Reform." *Social Science & Medicine* 127:92-100.

Jylhä, Marja. 2011. "Self-Rated Health and Subjective Survival Probabilities as Predictors of Mortality." In Richard G. Rogers and Eileen M. Crimmins (Eds.) *International Handbook of Adult Mortality*. New York: Springer.

Kang, Yuan-Yuan and Ji-Guang Wang. 2016. "The J-Curve Phenomenon in Hypertension." *Pulse* 4(1):49-60.

Killewald, Alexandra and Jonathan Bearak. 2014. "Is the Motherhood Penalty Larger for Low-Wage Women? A Comment on Quantile Regression." *American Sociological Review* 79(2):350-357.

Kimbro, Rachel Tolbert, Sharon Bzostek, Noreen Goldman, and Germán Rodríguez. 2008. "Race, Ethnicity, and the Education Gradient in Health." *Health Affairs* 27(2).

Kohler, Hans-Peter, Jere R. Behrman, and Jason Schnittker. 2011. "Social Science Methods for Twins Data: Integrating Causality, Endowments, and Heritability." *Biodemography and Social Biology* 57(1):88-141.

Krapohl, Eva, Kaili Rimfeld, Nicholas G. Shakeshaft, Maciej Trzaskowski, Andrew McMillan, Jean-Baptiste Pingault, Kathryn Asbury, Nicole Harlaar, Yulia Kovas, Philip S. Dale, and Robert Plomin. 2014. "The High Heritability of Educational Achievement Reflects Many Genetically Influenced Traits, Not Just Intelligence." *PNAS* 111(42):15273-15278.

Krieger, Nancy. 2001. "Theories for Social Epidemiology in the 21st Century: An Ecosocial Perspective." *International Journal of Epidemiology* 30:668-677.

Krieger, Nancy. 2011. *Epidemiology and the People's Health: Theory and Context*. New York: Oxford University Press.

Lam, Jack, Martin O'Flaherty, and Janeen Baxter. 2019. "The Scars of the Past? Childhood Health and Health Differentials in Later Life*." SSM – Population Health* 7.

Lareau, Annette. 2003. *Unequal Childhoods: Class, Race, and Family Life*. Berkeley, CA: University of California Press.

Lee, James J., Robbee Wedow, Aysu Okbay, Edward Kong, Omeed Maghzian, Meghan Zacher, Tuan Anh Nguyen-Viet, Peter Bowers, Julia Sidorenko, Richard Karlsson Linner, *et al*., Philipp D. Koellinger, Peter M. Visscher, Patrick Turley, Daniel J. Benjamin, and David Cesarini. 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.

Lewington, Sarah, *et al*. 2002. "Age-Specific Relevance of Usual Blood Pressure to Vascular Mortality: A Meta-Analysis of Individual Data for One Million Adults in 61 Prospective Studies." *The Lancet* 360:1903-1913.

Lewis, Tené T., Courtney D. Cogburn, and David R. Williams. 2015. "Self-Reported Experiences of Discrimination and Health: Scientific Advances, Ongoing Controversies, and Emerging Issues." *Annual Review of Clinical Psychology* 11(1):407-440.

Link, Bruce G. and Jo C. Phelan. 1995. "Social Conditions as Fundamental Causes of Disease." *Journal of Health and Social Behavior*:80-94.

Link, Bruce G., Jo C. Phelan, Richard Miech, and Emily L. Westin. 2008. "The Resources that Matter: Fundamental Social Causes of Health Disparities and the Challenge of Intelligence." *Journal of Health and Social Behavior* 49(1):72-91.

Liu, Hui and Robert A. Hummer. 2008. "Are Educational Differences in U.S. Self-Rated Health Increasing? An Examination by Gender and Race." *Social Science & Medicine* 67(11):1898-1906.

Lleras-Muney, Adriana. 2005. "The Relationship Between Education and Adult Mortality in the United States." *Review of Economic Studies* 72:189-221.

Lleras, Christy. 2008. "Do Skills and Behaviors in High School Matter? The Contribution of Noncognitive Factors in Explaining Differences in Educational Attainment and Earnings." *Social Science Research* 37(3):888-902.

Lloyd, G. E. R. (Ed.) 1983. *Hippocratic Writings*. London: Penguin Books.

Long, J. Scott and Jeremy Freese. 2014. *Regression Models for Categorical Dependent Variables Using Stata, Third Edition*. College Station, TX: Stata Press.

Long, J. Scott and Sarah A. Mustillo. 2018. "Using Predictions to Compare Groups in Regression Models for Binary Outcomes." Available at: http://www.indiana.edu/~jslsoc/files_research/groupdif/long-mustillo-comparing-groups-brm-2018-03-05.pdf

Lundborg, Petter. 2013. "The Health Returns to Schooling—What Can We Learn from Twins?" *Journal of Population Economics* 26(2):673-701.

Lundborg, Petter, Carl Hampus Lyttkens, and Paul Nystedt. 2016. "The Effect of Schooling on Mortality: New Evidence from 50,000 Swedish Twins." *Demography* 53(4):1135-1168.

Luo, Ye and Linda J. Waite. 2005. "The Impact of Childhood and Adult SES on Physical, Mental, and Cognitive Well-Being in Later Life." *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences* 60(2):S93-S101.

Lutfey, Karen and Jeremy Freese. 2005. "Toward Some Fundamentals of Fundamental Causality: Socioeconomic Status and Health in the Routine Clinic Visit for Diabetes." *American Journal of Sociology* 110(5):1326-1372.

Lynch, Scott M. 2006. "Cohort and Life-Course Patterns in the Relationship between Education and Health: A Hierarchical Approach." *Demography* 40(2):309-331.

Mackenbach, JP, J van den Bos, H van de Mheen, and K Stronks. 1994. "The Determinants of Excellent Health: Different from the Determinants of Ill-Health?" *International Journal of Epidemiology* 23(6):1273-1281.

Madsen, Mia, Anne-Marie Nybo Anderson, Kaare Christensen, Per Kragh Andersen, and Merete Osler. 2010. "Does Educational Status Impact Adult Mortality in Denmark? A Twin Approach." *American Journal of Epidemiology* 172(2):225-234.

Martin, Alicia R., Christopher R. Gignoux, Raymond K. Walters, Genevieve L. Wojcik, Benjamin M. Neale, Simon Gravel, Mark J. Daly, Carlos D. Bustamante, and Eimear E. Kenny. 2017. "Demographic History Impacts Genetic Risk Prediction Across Populations." *The American Journal of Human Genetics* 100:635-649.

Martinez-Sanchez, Eva and Enrique Regidor. 2002. "Self-rated Health by Education Level in Persons with and without Health Problems." *Journal of Health Psychology* 7:459–468.

Masters, Ryan K., Robert A. Hummer, and Daniel A. Powers. 2012. "Educational Differences in U.S. Adult Mortality: A Cohort Perspective." *American Sociological Review* 77(4):548-572.

Masters, Ryan K., Bruce G. Link, and Jo C. Phelan. 2015. "Trends in Education Gradients of 'Preventable' Mortality: A Test of Fundamental Cause Theory." *Social Science & Medicine* 127:19-28.

McCall, Leslie. 2005. "The Complexity of Intersectionality." *Signs: Journal of Women in Culture and Society* 30:1771-1800.

McCarthy, Mark I., Gonçalo R. Abecasis, Lon R. Cardon, David B. Goldstein, Julian Little, John P.A. Ioannidis, and Joel N. Hirschhorn. 2008. "Genome-Wide Association Studies for Complex Traits: Consensus, Uncertainty and Challenges." *Nature Reviews Genetics* 9:356-369.

McCarthy, Shane, *et al*. 2016. "A Reference Panel of 64,976 Haplotypes for Genotype Imputation." *Nature Genetics* 48:1279-1283.

McDade, Thomas W., Sharon Williams, and J. Josh Snodgrass. 2007. "What a Drop Can Do: Dried Blood Spots as a Minimally Invasive Method for Integrating Biomarkers into Population-Based Research." *Demography* 44(4):899-925.

Meara, Ellen, Seth Richards, and David Cutler. 2008. "The Gap Gets Bigger: Changes in Mortality and Life Education, 1981-2000." *Health Affairs* 27(2):350-360.

Miech, Richard, Fred Pampel, Jinyoung Kim, and Richard G. Rogers. 2011. "The Enduring Association Between Education and Mortality: The Role of Widening and Narrowing Disparities." *American Sociological Review* 76(6):913-934.

Miller, Gregory E., Tianyi Yu, Edith Chen, and Gene H. Brody. 2015. "Self-Control Forecasts Better Psychosocial Outcomes but Faster Epigenetic Aging in Low-SES Youth." *PNAS* 112(33):10325-10330.

Mirowsky, John and Catherine E. Ross. 2003. *Education, Social Status, and Health*. New York, NY: de Gruyter.

Mollborn, Stephanie and Elizabeth Lawrence. 2018. "Family, Peer, and School Influences on Children's Developing Health Lifestyles." *Journal of Health and Social Behavior* 59(1):133-150.

Montez, Jennifer Karas and Lisa F. Berkman 2014. "Trends in the Educational Gradient of Mortality Among US Adults Aged 45 to 84 Years: Bringing Regional Context Into the Explanation." *American Journal of Public Health* 104(1):e82-e90.

Montez, Jennifer Karas and Esther M. Friedman. 2015. "Educational Attainment and Adult Health: Under What Conditions is the Association Causal?" *Social Science & Medicine* 127:1-7.

Montez, Jennifer Karas and Mark D. Hayward. 2011. "Early Life Conditions and Later Life Mortality." In Richard G. Rogers and Eileen M. Crimmins (Eds.) *International Handbook of Adult Mortality*. New York: Springer.

Montez, Jennifer Karas and Mark D. Hayward. 2014. "Cumulative Childhood Adversity and Active Life Expectancy Among U.S. Adults." *Demography* 51(2):413-435.

Montez, Jennifer Karas, Mark D. Hayward, and Douglas A. Wolf. 2017a. "Do U.S. States' Socioeconomic and Policy Contexts Shape Disability?" *Social Science & Medicine* 178:115-126.

Montez, Jennifer Karas, Robert A. Hummer, and Mark D. Hayward. 2012. "Educational Attainment and Adult Mortality in the United States: A Systematic Analysis of Functional Form." *Demography* 49:315-336.

Montez, Jennifer Karas, Robert A. Hummer, Mark D. Hayward, Hyeyoung Woo, and Richard G. Rogers. 2011. "Trends in the Educational Gradient of U.S. Adult Mortality From 1986 to 2006 by Race, Gender, and Age Group." *Research on Aging* 33(2):145-171.

Montez, Jennifer Karas, Anna Zajacova, and Mark D. Hayward. 2016. "Explaining Inequalities in Women's Mortality Between U.S. States." *SSM – Population Health* 2:561-571.

Montez, Jennifer Karas, Anna Zajacova, and Mark D. Hayward. 2017b. "Disparities in Disability by Educational Attainment Across US States." *American Journal of Public Health* 107(7):1101-1108.

Montez, Jennifer Karas, Anna Zajacova, Mark D. Hayward, Steven H. Woolf, Derek Chapman, and Jason Beckfield. 2019. "Educational Disparities in Adult Mortality Across U.S. States: How Do They Differ, and Have They Changed Since the Mid-1980s?" *Demography* 56(2):621-644.

Mood, Carina. 2010. "Logistic Regression: Why We Cannot Do What We Think We Can Do, And What We Can Do About It." *European Sociological Review* 26(1):67-82.

Moon, Kilson, Carolin Krems, Thorsten Heuer, Alexander Roth, and Ingrid Hoffmann. 2017. "Predictors of BMI Vary Along the BMI Range of German Adults: Results of the German National Nutrition Survey II." *Obesity Facts* 10(1):38-49.

Mostafavi, Hakhamanesh, Arbel Harpak, Dalton Conley, Jonathan K. Pritchard, and Molly Przeworski. 2019. "Variable Prediction Accuracy of Polygenic Scores within an Ancestry Group." *bioRxiv*.

Nandi, Arijit, M. Maria Glymour, Ichiro Kawachi, and Tyler VanderWeele. 2012. "Using Marginal Structural Models to Estimate the Direct Effect of Adverse Childhood Social Conditions on Onset of Heart Disease, Diabetes, and Stroke." *Epidemiology* 23:223-232.

National Center for Health Statistics. 2014. *Health, United States, 2013: With Special Feature on Prescription Drugs*. Hyattsville, MD: National Center for Health Statistics.

Navarese, Eliano P., Jennifer G. Robinson, Mariusz Kowalewski, Michalina Kolodziejczak, Felicita Andreotti, Kevin Bliden, Udaya Tantry, Jacek Kubica, Paolo Raggi, and Paul A. Gurbel. 2018. "Association Between Baseline LDL-C Level and Total and Cardiovascular Mortality After LDL-C Lowering: A Systematic Review and Meta-Analysis." *JAMA* 319(15):1566-1579.

O'Brien, Rourke L. 2012. "Depleting Capital? Race, Wealth, and Informal Financial Assistance." *Social Forces* 91(2):375-396.

Okbay, Aysu, Jonathan P. Beauchamp, Mark Alan Fontana, James J. Lee, Tune H. Pers, Cornelius A. Rietveld, Patrick Turley, *et al*., Peter M. Visscher, Tõnu Esko, Philipp D. Koellinger, David Cesarini, and Daniel J. Benjamin. 2016. "Genome-Wide Association Study Identifies 74 Loci Associated with Educational Attainment." *Nature* 533:539-42.

Okbay Aysu, Daniel Benjamin, and Peter Visscher. 2018a. *Documentation*. Available at: https://www.ssc.wisc.edu/wlsresearch/documentation/GWAS/

Okbay, Aysu, Daniel Benjamin, and Peter Visscher. 2018b. *SSGAC Educational Attainment: GWAS and MTAG Polygenic Scores (V.1.0)*. Available at: https://ssl.isr.umich.edu/hrs/start.php

Olson, Julie Skalamera, Robert A. Hummer, and Kathleen Mullan Harris. 2017. "Gender and Health Behavior Clustering among U.S. Young Adults." *Biodemography and Social Biology* 63(1):3-20.

Osterberg, Lars and Terrence Blaschke. 2005. "Adherence to Medication." *The New England Journal of Medicine* 353:487-497.

Palloni, Alberto. 2006. "Reproducing Inequalities: Luck, Wallets, and the Enduring Effects of Childhood Health." *Demography* 43:587-615.

Pampel, Fred C., Patrick M. Krueger, and Justin T. Denney. 2010. "Socioeconomic Disparities in Health Behaviors." *Annual Review of Sociology* 36:349-370.

Parsons, Talcott. 1961. "The School Class as a Social System." In J. F. A.H. Halsey, and C.A. Anderson (Eds.), *Education, Economy and Society*. New York: Free Press.

Paulose-Ram, Ryne, Qiuping Gu, and Brian K. Kit. 2017. "Characteristics of U.S. Adults with Hypertension who are Unaware of their Hypertension, 2011-2014." *NCHS Data Brief, No. 278.* Hyattsville, MD: National Center for Health Statistics.

Perkins, Kristin L. 2019. “Changes in Household Composition and Children’s Educational Attainment.” *Demography* 56:525-548.

Pew Research Center. 2016. “On Views of Race and Inequality, Blacks and Whites are Worlds Apart.” Available at: http://www.pewsocialtrends.org/2016/06/27/on-views-of-race-and-inequality-blacks-and-whites-are-worlds-apart/

Pfeffer, Fabian T. and Florian R. Hertel. 2015. “How Has Educational Expansion Shaped Social Mobility Trends in the United States?” *Social Forces* 94(1):143-180.

Pfeffer, Fabian T. and Alexandra Killewald. 2015. “How Rigid is the Wealth Structure and Why? Intergenerational and Multigenerational Associations in Family Wealth.” Rep. No. 15-845, Pop. Stud. Cent., University of Michigan, Ann Arbor, MI.

Phelan, Jo C. and Bruce G. Link. 2005. “Controlling Disease and Creating Disparities: A Fundamental Cause Perspective.” *The Journals of Gerontology* 60B(special issue II):27-33.

Phelan, Jo C., Bruce G. Link, Ana Diez-Roux, Ichiro Kawachi, and Bruce Levin. 2004. “Fundamental Causes of Social Inequalities in Mortality: A Test of the Theory.” *Journal of Health and Social Behavior* 45:265-285.

Phelan, Jo C., Bruce G. Link, and Parisa Tehranifar. 2010. “Social Conditions as Fundamental Causes of Health Inequalities: Theory, Evidence, and Policy Implications.” *Journal of Health and Social Behavior* 51(S):S28-S40.

Rader, Daniel J. and G. Kees Hovingh. 2014. “HDL and Cardiovascular Disease.” *The Lancet* 384(9943):16-22.

RAND. 2016. *RAND HRS Data, Version P. Produced by the RAND Center for the Study of Aging, with funding from the National Institute on Aging and the Social Security Administration*. Santa Monica, CA: RAND.

Rapsomaniki, Eleni, Adam Timmis, Julie George, Mar Pujades-Rodriguez, Anoop D. Shah, Spiros Denaxas, Ian R. White, Mark J. Caulfield, John E. Deanfield, Liam Smeeth, Bryan Williams, Aroon Hingorani, and Harry Hemingway. 2014. “Blood Pressure and Incidence of Twelve Cardiovascular Diseases: Lifetime Risks, Healthy Life-Years Lost, and Age-Specific Associations in 1.25 Million People.” *The Lancet* 383(9932):1899-1911.

Reigle-Crumb, Catherine and Melissa Humphries. 2012. “Exploring Bias in Math Teachers’ Perceptions of Students’ Ability by Gender and Race/Ethnicity.” *Gender & Society* 26(2):290-322.

Reile, Rainer and Mall Leinsalu. 2013. “Differentiating Positive and Negative Self-Rated Health: Results from a Cross-Sectional Study in Estonia.” *International Journal of Public Health* 58(4):555-564.

Rietveld, Cornelius A., Dalton Conley, Nicholas Eriksson, Tõnu Esko, Sarah E. Medland, Anna A.E. Vinkhuyzen, Jian Yang, Jason D. Boardman, Christopher F. Chabris, Christopher T. Dawes, Benjamin W. Domingue, David A. Hinds, Magnus Johannesson, Amy K. Kiefer, David Laibson, Patrick K.E., Magnusson, Joanna L. Mountain, Sven Oskarsson, Olga Rostapshova, Alexander Teumer, Joyce Y. Tung, Peter M. Visscher, Daniel J. Benjamin, David Cesarini, Philipp D. Koellinger, and the Social Science Genetic Association Consortium. 2014. “Replicability and Robustness of GWAS for Behavioral Traits.” *Psychological Science* 25(1):1975-1986.

Rietveld, Cornelius A., *et al*. 2013. "GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment." *Science* 340(6139):1467-1471.

Ross, Catherine E. and John Mirowsky. 2006. "Sex Differences in the Effect of Education on Depression: Resource Multiplication or Resource Substitution?" *Social Science & Medicine* 63(5):1400-1413.

Ross, Catherine E. and John Mirowsky. 2010. "Gender and the Health Benefits of Education." *The Sociological Quarterly* 51(1):1-19.

Ross, Catherine E. and John Mirowsky. 2011. "The Interaction of Personal and Parental Education on Health." *Social Science & Medicine* 72(4):591-599.

Ross, Catherine E., Ryan K. Masters, and Robert A. Hummer. 2012. "Education and the Gender Gaps in Health and Mortality." *Demography* 49(4):1157-1183.

Rostron, Brian L., John L. Boies, and Elizabeth Arias. 2010. "Education Reporting and Classification on Death Certificates in the United States." *Vital and Health Statistics Series 2, No. 151*. Hyattsville, MD: National Center for Health Statistics.

Salami, Joseph A., Haider Warraich, Javier Valero-Elizondo, Erica S. Spatz, Nihar R. Desai, Jamal S. Rana, Salim S. Virani, Ron Blankstein, Amit Khera, Michael J. Blaha, Roger S. Blumenthal, Donald Lloyd-Jones, and Khurram Nasir. 2017. "National Trends in Statin Use and Expenditures in the US Adult Population from 2002 to 2013." *JAMA Cardiology* 2(1):56-65.

Sampson, Robert J. and Alix Winter. 2016. "The Racial Ecology of Lead Poisoning: Toxic Inequality in Chicago Neighborhoods, 1995-2013." *DuBois Review: Social Science Research on Race* 13(2):261-283.

Sansani, Shahar. 2011. "The Effects of School Quality on Long-Term Health." *Economics of Education Review* 30(6):1320-1333.

Schaan, Barbara. 2014. "The Interaction of Family Background and Personal Education on Depressive Symptoms in Later Life." *Social Science & Medicine* 102:94-102.

Schafer, Markus H., Lindsay R. Wilkinson, and Kenneth F. Ferraro. 2013. "Childhood (Mis)fortune, Educational Attainment, and Adult Health: Contingent Benefits of a College Degree?" *Social Forces* 91(3):1007-1034.

Schnittker, Jason. 2005. "Cognitive Abilities and Self-Rated Health: Is There a Relationship? Is it Growing? Does it Explain Disparities?" *Social Science Research* 34(4):821-842.

Selvin, Elizabeth, Michael W. Steffes, Hong Zhu, Kunihiro Matsushita, Lynn Wagenknecht, James Pankow, Josef Coresh, and Frederick L. Brancati. 2010. "Glycated Hemoglobin, Diabetes, and Cardiovascular Risk in Nondiabetic Adults." *The New England Journal of Medicin*e 362:800-811.

Sewell, William H., Archibald O. Haller, and Alejandro Portes. 1969. "The Educational and Early Occupational Attainment Process." *American Sociological Review* 34:82-92.

Shields, Margot and Shahin Shooshtari. 2001. "Determinants of Self-Perceived Health." *Health Reports* 13(1):35-52.

Shonkoff, Jack, W. Thomas Boyce, and Bruce S. McEwen. 2009. "Neuroscience, Molecular Biology, and the Childhood Roots of Health Disparities: Building a New Framework for Health Promotion and Disease Prevention." *JAMA* 301:2252-9.

Shooshtari, Shahin, Verena Menec, and Robert Tate. 2007. "Comparing Predictors of Positive and Negative Self-Rated Health Between Younger (25-54) and Older (55+) Canadian Adults: A Longitudinal Study of Well-Being." *Research on Aging* 29(6):512-554.

Shuey, Kim M. and Andrea E. Willson. 2008. "Cumulative Disadvantage and Black-White Disparities in Life-Course Health Trajectories." *Research on Aging* 30(2):200-225.

Spencer, Karen Lutfey and Matthew Grace. 2016. "Social Foundations of Health Care Inequality and Treatment Bias." *Annual Review of Sociology* 42:101-120.

Stamler, Jeremiah, Martha L. Daviglus, Daniel B. Garside, Alan R. Dyer, Philip Greenland, and James K. Neaton. 2000. "Relationship of Baseline Serum Cholesterol Levels in 3 Large Cohorts of Younger Men to Long-Term Coronary, Cardiovascular, and All-Cause Mortality and to Longevity." *JAMA* 284(3):311-318.

StataCorp. 2017. *Stata Statistical Software: Release 15*. College Station, TX: StataCorp LLC.

Stephens, Nicole M., Stephanie A. Fryberg, Hazel Rose Markus, Camille S. Johnson, and Rebecca Covarrubias. 2012a. "Unseen Disadvantage: How American Universities' Focus on Independence Undermines the Academic Performance of First-Generation College Students." *Journal of Personality and Social Psychology* 102(6):1178-1197.

Stephens, Nicole M., Sarah S.M. Townsend, Hazel Rose Markus, and L. Taylor Phyillips. 2012b. "A Cultural Mismatch: Independent Cultural Norms Promote Greater Increases in Cortisol and More Negative Emotions Among First-Generation College Students." *Journal of Experimental Social Psychology* 48(6):1389-1393.

Tehranifar, Parisa, Alfred I. Neugut, Jo C. Phelan, Bruce G. Link, Yuyan Liao, Manisha Desai, and Mary Beth Terry. 2009. "Medical Advances and Racial/Ethnic Disparities in Cancer Survival." *Cancer Epidemiology, Biomarkers and Prevention* 18:2701-2708.

Tomaskovic-Devey, Donald, Melvin Thomas, and Kecia Johnson. 2005. "Race and the Accumulation of Human Capital across the Career: A Theoretical Model and Fixed Effects Approach." *American Journal of Sociology* 111(1):58-89.

Torche, Florencia. 2011. "Is a College Degree Still the Great Equalizer? Intergenerational Mobility across Levels of Schooling in the United States." *American Journal of Sociology* 117:763-807.

Toth, Peter P., Philip J. Barter, Robert S. Rosenson, William E. Boden, M. John Chapman, Marina Cuchel, Ralph B. D'Agostino Sr., Michael H. Davidson, W. Sean Davidson, Jay W. Heinecke, Richard H. Karas, Anatol Kontush, Ronald M. Krauss, Michael Miller, and Daniel J. Rader. 2013. "High-Density Lipoproteins: A Consensus Statement from the National Lipid Association." *Journal of Clinical Lipidology* 7(5):484-525.

Turkheimer, Eric. 2000. "Three Laws of Behavior Genetics and What They Mean." *Current Directions in Psychological Science* 9(5):160-164.

van Ryn, Michelle, Diana Burgess, Jennifer Malat, and Joan Griffin. 2006. "Physicians' Perceptions of Patients' Social and Behavioral Characteristics and Race Disparities in Treatment Recommendations for Men With Coronary Artery Disease." *American Journal of Public Health* 969(2):351-357.

van Ryn, Michelle and Jane Burke. 2000. "The Effect of Patient Race and Socio-economic Status on Physicians' Perceptions of Patients." *Social Science & Medicine* 50(6):813-828.

Vasan, Ramachandran S., Martin S. Larson, Eric P. Leip, Jane C. Evans, Christopher J. O'Donnell, William B. Kannel, and Daniel Levy. 2001. "Impact of High-Normal Blood Pressure on the Risk of Cardiovascular Disease." *The New England Journal of Medicine* 345:1291-1297.

Vilhjálmsson, Bjarni J., *et al*. 2015. "Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores." *The American Journal of Human Genetics* 97(4):576–592.

Virani, Salim S., LeChauncy D. Woodard, Cassie R. Landrum, Kenneth Pietz, Degang Wang, Christie M. Ballantyne, and Laura A. Petersen. 2011. "Institutional, Provider, and Patient Correlates of Low-Density Lipoprotein and Non-High-Density Lipoprotein Cholesterol Goal Attainment According to the Adult Treatment Panel III Guidelines." *American Heart Journal* 161(6):1140-1146.

Visscher, Peter M., Naomi R. Wray, Qian Zhang, Pamela Sklar, Mark I. McCarthy, Matthew A. Brown, and Jian Yang. 2017. "10 Years of GWAS Discovery: Biology, Function, and Translation." *The American Journal of Human Genetics* 101(1):5-22.

Walsemann, Katrina M., Bridget J. Goosby, and Deeonna Farr. 2016. "Life Course SES and Cardiovascular Risk: Heterogeneity Across Race/Ethnicity and Gender." *Social Science & Medicine* 152:147-155.

Weber, Max. 1978[1922]. *Economy and Society*. Berkeley, CA: University of California Press.

Wedow, Robbee, Meghan Zacher, Brooke M. Huibregtse, Kathleen Mullan Harris, Benjamin W. Domingue, and Jason D. Boardman. 2018. "Education, Smoking, and Cohort Change: Forwarding a Multidimensional Theory of the Environmental Moderation of Genetic Effects." *American Sociological Review* 83(4):803-832.

Whelton, Paul K., *et al*. 2018. "2017. ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/ NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines." *Hypertension* 71:e13-e115.

Williams, David R. and Chiquita Collins. 2001. "Racial Residential Segregation: A Fundamental Cause of Racial Disparities in Health." *Public Heath Reports* 116(5):404-416.

Williams, David R. and Selina A. Mohammed. 2009. "Discrimination and Racial Disparities in Health: Evidence and Needed Research*." Journal of Behavioral Medicine* 32(1):20-47.

Williams, David R., Selina A. Mohammed, Jacinta Leavell, and Chiquita Collins. 2010. "Race, Socioeconomic Status, and Health: Complexities, Ongoing Challenges, and Research Opportunities." *Annals of the New York Academy of Sciences* 1186:69-101.

Willson, Andrea E., Kim M. Shuey, and Glen H. Elder, Jr. 2007. "Cumulative Advantage Processes as Mechanisms of Inequality in Life Course Health." *American Journal of Sociology* 112(6):1886-1924.

Wisconsin Longitudinal Study. 2016. *Quality Control Report for Genotypic Data*. Seattle, WA: University of Washington. Available at: https://www.ssc.wisc.edu/wlsresearch/documentation/GWAS/Herd_QC_report.pdf

Woolf, Steven H., Robert E. Johnson, Robert L. Phillips, Jr., and Maike Philipsen. 2007. "Giving Everyone the Health of the Educated: An Examination of Whether Social Change Would Save More Lives Than Medical Advances." *American Journal of Public Health* 97(4):679-683.

Zajacova, Anna and Jennifer Beam Dowd. 2011. "Reliability of Self-Rated Health in US Adults." *American Journal of Epidemiology* 174(8):977-983.

Zajacova, Anna and Robert A. Hummer. 2009. "Gender Differences in Education Effects on All-Cause Mortality for White and Black Adults in the United States." *Social Science & Medicine* 69(4):529-537.

Zheng, Hui. 2017. "Why Does College Education Matter? Unveiling the Contributions of Selection Factors." *Social Science Research* 68:59-73.