I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14600

Health Endowments, Schooling Allocation in the Family, and Longevity: Evidence from US Twins

Peter A. Savelyev
Benjamin C. Ward
Robert F. Krueger
Matt McGue

JULY 2021

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14600

Health Endowments, Schooling Allocation in the Family, and Longevity: Evidence from US Twins

Peter A. Savelyev
College of William & Mary and IZA

Robert F. Krueger
University of Minnesota

Benjamin C. Ward
University of Georgia

Matt McGue
University of Minnesota

JULY 2021

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Health Endowments, Schooling Allocation in the Family, and Longevity: Evidence from US Twins*

We analyze data from the Minnesota Twin Registry (MTR), combined with the Socioeconomic Survey of Twins (SST), and new mortality data, and contribute to two bodies of literature. First, we demonstrate a beneficial causal effect of education on health and longevity in contrast to other twin-based studies of the US population, which show little or no effect of education on health. Second, we present evidence that parents compensate for differences in their children's health endowments through education, but find no evidence that parents reinforce differences in skill endowments. We argue that there is a bias towards detecting reinforcement both in this paper and in the literature. Despite this bias, we still find statistical evidence of compensating behavior. We account for observed and unobserved confounding factors, sample selection bias, and measurement error in education.

JEL Classification: I12, I140, I240, J130, J24

Keywords: health endowment, skill endowment, intrafamily resource allocation, education, health, longevity, twin study, Minnesota Twin Registry

Corresponding author:
Peter A. Savelyev
College of William & Mary
300 James Blair Dr.
Chancellor's Hall
Room 317
Williamsburg
VA 23185

E-mail: pasavelyev@wm.edu

* A version of this paper was presented to the 2019 NBER Health Economics Program Meeting, Boston, USA; The Third Meeting of the Society of Economics of the Household, Lisbon, Portugal; IZA World Labor Conference, Berlin, Germany; Federal Reserve Bank of Richmond Research Seminar; Southern Economic Association 87th Annual Meeting, Tampa, FL; and to the Economics Work-in-Progress Lunch Seminar at the College of William & Mary. We thank participants of these meetings for helpful suggestions and stimulating discussions. For their comments and suggestions, the authors wish to thank Kathy Anderson, Andrew Goodman-Bacon, Ana Balsa, Rong Hai, Anna Sanz-de-Galdeano, and anonymous NSF and journal referees. The authors are especially grateful to Jere Behrman for sharing the data from the Socioeconomic Survey of Twins and commenting on our drafts, and to Federico Gutierrez for his considerable help in conceiving and designing the analysis of schooling allocation in the family. The authors are grateful to Jack Buckman and Renee Garrow for their research assistance, and to William Anderson, Xiaoyu (Nancy) Chen, Isabel Haber, Eli Rothleder, Max Sacher, Katia Savelyeva, and Nathan Troutman for their proofreading of the manuscript. Ward and Savelyev gratefully acknowledge support from the NSF 1460003 grant, Vanderbilt University Department of Economics Faculty-Graduate Student Collaboration Grant, and the Kirk Dornbush Summer Research Grant. Savelyev also benefited from the financial support of the Grey Fund at Vanderbilt and research support from the College of William & Mary. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the funders.

1 Introduction

This paper contributes to understanding two empirical questions that are important for both economic theory and policy: (1) whether education causally affects health and longevity and (2) whether parents compensate for or reinforce differences in endowments among their children. It is natural to study these two separate research questions together, as a model that addresses question (1) is itself included within a more complex model that addresses question (2). Our analysis is based on three datasets that were gathered over multiple decades and that describe a a single sample of US twins. The initial dataset is the Minnesota Twin Registry (MTR). Its follow-up survey, conducted by economists, is the Socioeconomic Survey of Twins (SST). We add to these existing datasets our own, newly-collected individual mortality data.

The collection of the MTR data began in 1983 in order to identify and study twins born in Minnesota. The SST survey followed up with a subsample of the initial MTR participants. We analyze same-sex twin pairs, in which both twins participated in the SST and both provided information about their education levels. Our estimation sample contains 1,233 twin pairs born between 1936 and 1955. We match the data from SST to the data from MTR. We then match the resulting MTR-SST data to new data on mortality, which we gathered from the Social Security Death Master File, the National Death Index, and contact with surviving relatives.

We apply a linear probability model to within-pair differences between identical twins, which allows us to estimate the extent to which a twin with more years of education can be expected to outlive their less educated twin. From this estimate we are able to draw conclusions about the effect of education on longevity. This within-twin-pair approach leverages the common family and genetic background shared by identical twins. We use the same method to study effects of education on health and health behaviors.

To find whether parents use educational investments to compensate for or reinforce differences in *health* endowments between their children, we adopt the same method initially proposed by Behrman et al. (1994) for studying *wage-earning* endowments. For identification, we rely on the presence of identical genes in monozygotic (MZ, or "iden-

tical") twins, genetic variation in dizygotic (DZ, or "fraternal") twins, and shared family background for all observed twin pairs. In addition, we reanalyze results by Behrman et al. (1994) regarding wage earning endowments using direct measures of wages and earnings from SST (the original paper used earnings imputed from occupations reported in MTR).

We account for a number of econometric issues. By using twin fixed effects we not only control for major confounders, but also control for possible selection biases, including selective attrition (Behrman et al., 1994; Heckman and MaCurdy, 1980; Pitt and Rosenzweig, 1990). To account for measurement error in schooling, we use the well-established Ashenfelter and Krueger (1994) method, which takes advantage of each twins' reports about their own education and the education of their twin. Finally, we control for expected major confounders that could differ across twins: birth weight and disabling injury.

The causal effect of education on health and longevity, our first empirical research question, is still debated in the literature. As we discuss in Section 2.1, many authors support the causal effect, while many others argue against it. Because effects of education on health and longevity likely differ from country to country, the most relevant results for our paper to compare to are results based on US datasets. However, the popular compulsory schooling law instrumental variable is weak for the US (Galama et al., 2018). A sound alternative estimator is the twin fixed effect. However, a small number of papers based on existing US twin datasets reach conflicting conclusions. The results we present here, based on well-established methods and new quality data, support the claim that education has a causal effect on health and health behaviors. In addition, to our knowledge we are the first to demonstrate a causal effect of education on longevity by applying twin-first-difference methodology to US data.

Our second research question, whether parents of multiple children tend to reinforce or compensate for differences in their children's endowments, is also empirical, as both results are theoretically possible. The question is controversial, as for every possible result—whether it be compensation, reinforcement, both effects, or neither effect—there exists a group of papers that supports it, as summarized in Almond and Mazumder

(2013). This great variation in literature results is likely driven by differences in the type of endowment (e.g., skill vs. health endowment), type of investment (e.g., education vs. health investment), type of population (e.g., developed vs. developing country), and type of identification method, among other differences. We discuss the literature in Section 2.2 while taking these differences into account. We contribute to this literature by adding unique data and measures of latent health endowments that are new in this literature.

We argue that the literature is biased towards its most common finding—that parents typically reinforce the differences in their children's endowments. Hence, if the true prevailing behavior is compensation, it can be misclassified in a statistical analysis as either reinforcement or neutral behavior. We are aware of two sources of bias, both going in the same direction. One source of bias is the failure to account for a child's capacity to make their own decisions—beyond those decisions made by their parents on the child's behalf—about the total investments in them. For instance, a sicker child may end up with lower level of education not only because their parents may choose to reallocate educational investments to their healthier child, but also because the sicker child has reduced capabilities and economic incentives to study. Another source of bias comes from failing to account for the effect of endowment on investment costs: if a less favorable endowment leads to increased investment costs, the resulting parental choice may look like reinforcement despite parental preference for compensation other things kept equal (Sanz-de Galdeano and Terskaya, 2019; Terskaya, 2019). To the best of our knowledge, the rest of the literature does not address these types of bias (e.g., Almond and Mazumder, 2013). We return to these types of bias in Section 4.3.

Estimated parameters of our model nominally imply that parents compensate for health endowments by investing in additional education for the less-healthy child, and that parents are neutral with respect to differences in skill endowments. When taking the expected bias into account, we can still claim a compensation effect for health endowments, as this result is obtained *despite* the bias. Our estimates for skills do not support reinforcement even before taking the bias into account. After taking the bias into account, our results do not rule out a possible compensation for skills.

2 Related Literature

This section complements the introduction by describing results of the literature and our contribution to it in more detail.

2.1 The Effect of Education on Health and Lonegvity

There are two classical explanations of the possible effect of education on health and longevity: (1) productive and (2) allocative efficiencies. A productive efficiency mechanism suggests that education amplifies the health output from given amounts of endogenous inputs (Grossman, 2000). An allocative efficiency mechanism suggests that education leads to better allocation of resources devoted to health production as a result of superior information about the true effects of health investments (Kenkel, 2000). Cutler and Lleras-Muney (2008) suggest that the following mechanisms link education and health: (1) higher income; (2) safer jobs; (3) higher value of life; (4) better health knowledge and superior cognitive skills; (5) lower discount rate and increased risk aversion; (6) higher rank in the society; and (7) larger social networks, which provide financial, physical and emotional support. In line with the theory, empirical research has suggested that education affects health and longevity though productive health behaviors, healthy lifestyles, safer work conditions, and superior socioeconomic status (e.g., Balia and Jones, 2008; Brunello et al., 2015; Contoyannis and Jones, 2004; Cutler et al., 2011; Darden et al., 2018; Hong et al., 2000).

Despite a considerable body of literature, including papers about the mechanisms of health production mentioned in the previous paragraph, the question regarding the causal relationship between education and health or longevity remains unresolved. This question is often addressed in the literature by using changes in compulsory schooling laws or the birth of twins as natural experiments. Other methods include randomized controlled trials, which are usually feasible at low levels of education such as preschool (e.g., Conti et al., 2016), military draft used as an instrument for men's education (Buckles et al., 2016), and methods that explicitly model unobserved heterogeneity (e.g., Bijwaard et al., 2015; Conti and Heckman, 2010; Heckman et al., 2018; Hong et al., 2000; Savelyev,

2020; Savelyev and Tan, 2019). See Grossman (2015) and Galama et al. (2018) for recent surveys.

Compulsory schooling laws identify the effect only for those students who would not gain further education otherwise. The compulsory schooling instrument for the US is weak and therefore fails to reliably estimate of the effect of interest, while results for other countries are mixed (Galama et al., 2018). For instance, Lleras-Muney (2005) argues that schooling affects mortality in the US, but Mazumder (2008) shows that these results do not survive controlling for state-specific time-trends. Van Kippersluis et al. (2011) find beneficial effects of education on longevity for Dutch men, but Albouy and Lequien (2009) and Meghir et al. (2018) do not find an effect of schooling on longevity for France and Sweden. Similarly, Arendt (2005) and Clark and Royer (2013) find no effect of compulsory schooling on health-related outcomes in Denmark and the UK.

In contrast to compulsory schooling laws, papers that leverage twins data usually identify the average treatment effect of an additional year of schooling. The use of twin-based identification is limited to available twins registries, among which only a small number are large and old enough to reliably study longevity. Lundborg et al. (2016) use Swedish twins data and find strong effects of education on longevity for both men and women. Madsen et al. (2010) and Behrman et al. (2011) use Danish twins data and find no effects. However, van den Berg et al. (2015) use the same data but a different methodology and find an effect of education on mortality for men but not for women. Behrman et al. (2015) use the Chinese Adult Twins Survey to study effects of education on health and health-related behaviors for a pooled sample of men and women and find a number of effects on important determinants of mortality: improvements in general and mental health, reductions in smoking, and the number of chronic diseases.

There are only a small number of papers on this topic that examine US twins, and the results of those papers are at odds with each other. In particular, Lundborg (2013) studies monozygotic twins based on the Midlife in the United States (MIDUS) survey and finds causal effects of education on health and health behaviors. The rest of this literature does not support such a causal effect, but this paper joins Lundborg (2013) in supporting the causal effects of education on health and health behaviors. Furthermore,

certain mediating effects that we estimate (alcohol problems and overweight) complement the types of effects estimated by Lundborg (smoking and exercise). Additional novel contributions of our paper relative to Lundborg's include the use of a different US dataset, a study of effect heterogeneity by subject's sex, and explicit controls for measurement error in education and thus for the attenuation bias that is associated with it. Finally, to our knowledge this paper is the first study in the US-twin-based literature to demonstrate the effect of education on longevity.

Turning to other related papers, Kohler et al. (2011) estimate models for MZ twins using SST data and conclude that there is no effect of education on health. However, the authors use only one outcome, self-rated health, and perform their estimation only for female twins. We use the same SST data, but instead we investigate both sexes and additional health-related outcomes, including newly-collected mortality data, so that we complement existing results from that paper with new estimates of our own. As a result, our overall conclusion stands in contrast to that of Kohler et al. (2011). While we confirm the statistically insignificant result reported by Kohler et al. (2011) for self-reported health of women, our estimates for a larger set of health-related outcomes and for both sexes provide evidence in favor of the effect of education on health-related outcomes.

Amin et al. (2015) also apply a twins-based approach to the same MTR data that we use here, but they combine the MTR data with another dataset called the Mid-Atlantic Twin Registry (MATR) to increase sample size. We compare their results with ours, which are derived from the MTR dataset and its follow-up surveys. The authors report estimates for the pooled sample of men and women only, and concentrate on three health outcomes: self-reported health, Body Mass Index (BMI), and overweight status. Our results for the overweight status based on the MTR data show a statistically significant sex difference in effects: we find a strong beneficial effect of education for men, no effect for women, and no statistically significant effect for the pooled sample. Thus, pooling data for overweight-related variables could mask important relationships. Our estimate for the pooled-sex sample of twins shows a beneficial effect of education on self-

reported health status; this matches the finding of Amin et al. (2015)[1] However, unlike our paper, their paper does not investigate mortality or occurrence of specific physical health problems.

2.2 Family Resource Allocation

Results in the literature on Family Resource Allocation differ greatly. Some papers show that parents reinforce differences in endowments by investing more in children who have more beneficial endowments (e.g., Adhvaryu and Nyshadham, 2014; Aizer and Cunha, 2012; Almond et al., 2009; Behrman, 1988; Behrman et al., 1982, 1994; Borga and Pidkuyko, 2018; Datar et al., 2010; Frijters et al., 2013; Parman, 2015; Rosenzweig and Zhang, 2009; Venkataramani, 2012). Other papers show that parents compensate for differences by investing more in children with less beneficial endowments (e.g., Bharadwaj et al., 2018; Black et al., 2010; Del Bono et al., 2012; Halla and Zweimüller, 2014; Pitt et al., 1990; Sanz-de Galdeano and Terskaya, 2019; Terskaya, 2019). There are also papers that show both compensating and reinforcing effects (e.g., Ayalew, 2005; Hsin, 2012; Restrepo, 2016; Yi, 2019; Yi et al., 2015) and papers that find little or no effect (e.g., Abufhele et al., 2017; Almond and Currie, 2011; Bharadwaj et al., 2013; Kelly, 2011; Nicoletti et al., 2017; Royer, 2009). Almond and Mazumder (2013) survey many of these papers.

However, many papers on family resource allocation are hardly comparable due to the different types of child endowments or parental investments they study. Therefore, we split this set of papers into more comparable groups and then discuss them respectively.

We distinguish *health* endowments and investments from *skill* endowments and investments, especially given the evidence that the behavior of parents could be either compensating or reinforcing depending on the children's endowment and investment type (e.g., Ayalew, 2005; Nicoletti and Tonei, 2017; Yi et al., 2015).

Since this paper deals with educational investments, we restrict our comparison to a group of papers that study investments in broadly defined skills. We further clas-

[1]Compared to the system that the authors use for these health ratings (bad = 1 and excellent = 5), we use the same survey responses but reverse the numerical order (bad = 5 and excellent = 1).

sify these papers into three groups by the type of measure that they use to capture a child's health endowment: (1) low birth weight or exposure to adverse environment (a pandemic, nuclear power plant accident, and the like) while in utero or in early childhood (Abufhele et al., 2017; Aizer and Cunha, 2012; Almond and Currie, 2011; Almond et al., 2009; Bharadwaj et al., 2018, 2013; Black et al., 2010; Datar et al., 2010; Halla and Zweimüller, 2014; Hsin, 2012; Kelly, 2011; Nicoletti et al., 2017; Parman, 2015; Restrepo, 2016; Venkataramani, 2012; Yi, 2019); (2) education polygenic score (PGS) (Sanz-de Galdeano and Terskaya, 2019); (3) endowments measured after birth (Ayalew, 2005; Behrman et al., 1982, 1994; Borga and Pidkuyko, 2018; Frijters et al., 2013; Nicoletti and Tonei, 2017; Rosales-Rueda, 2014; Terskaya, 2019; Yi et al., 2015).

These three groups of papers differ in their advantages and disadvantages. The advantage of group (1) is that many of them have a source of arguably exogenous variation, though possible confounders cannot always be ruled out. The advantage of group (2) is that genes are determined at conception. However, genes may be correlated with family background through the correlation with parental genes. Groups (1) and (2) arguably identify some unknown mixture of skill and health endowments. Health shocks in utero or early childhood may negatively affect not only health but also cognitive skills (e.g., O'Conner et al., 2000). Education PGS predicts not only education, but also health conditional on education (Bolyard and Savelyev, 2021). If health and skill endowments have different effects on investments, then the effect of a mixture depends on mixing weights.

In contrast, the measures of health after birth that are used in group (3) are at a bigger risk to be confounded, but usually it is clear whether each such measure is mostly related to health (e.g. physical disability) or to skill (e.g. IQ). This paper is most comparable to the third group of papers and contributes by adding measures of latent health endowments that are new in the literature. Model parameters that are related to our latent endowments are identified by leveraging within-twin-pair differences in education, health measures, and longevity. We account for potential confounders by controlling for the twin fixed effects, as well as for potentially confounding observables.

Our results show that families invest in each of their children's education in a way that compensates for health endowment differences between the children. In a robust-

ness check that is available from the authors upon request, we show that this result is not driven by differences in wage-earning endowments, which may correlate with our measures of health. In addition, we find that the combined MTR-SST data provide no evidence of reinforcement for differences in skill endowments. As we expect our results to be biased towards detecting reinforcement (see Section 4.3 for details), our compensation result for health is still valid, while our statistically insignificant result for skills does not rule out compensation. Since papers from groups (1) and (2) identify effects for some mixture of skill and health endowments, our findings are broadly consistent with those papers that find either evidence of compensation (Bharadwaj et al., 2018; Black et al., 2010; Del Bono et al., 2012; Halla and Zweimüller, 2014; Sanz-de Galdeano and Terskaya, 2019) or a small/negligible effect (Abufhele et al., 2017; Almond and Currie, 2011; Bharadwaj et al., 2013; Kelly, 2011; Nicoletti et al., 2017).

Two papers argue that low-educated mothers reinforce, while high-educated mothers compensate (Hsin, 2012; Restrepo, 2016). We lack statistical power to either confirm or reject this result using our data.

Below we compare our paper to the papers from group (3) in more detail, as this group is the most comparable to our paper. We begin with papers that are concerned with health endowment and then proceed to papers about skill endowment.

Health Investments and Postnatal Measures of Health Endowment Papers in this group show both compensation and reinforcement. However, we see no direct contradiction between our results and papers that find evidence of reinforcement. This is the case because, compared to our paper, each of these papers uses different measures of endowment or a different population, or both.

Rosales-Rueda (2014) uses US data and finds reinforcement for mental health, a separate endowment from the physical health endowment that we study. Ayalew (2005) and Borga and Pidkuyko (2018) study Ethiopia, a country in extreme poverty, where parents face sharp trade-offs. In Ethiopia, giving an extra health investment to a sick child, such as food, can make a difference between life and death, but skipping school has less dramatic consequences. So sick children tend to skip school but get extra food

(Ayalew, 2005).

Yi et al. (2015) analyze Chinese twins data and find that when a twin receives a negative health shock between ages 0–3, the other twin receives health investments worth 305 yuan less, but education investments worth 182 yuan more at around age 11. This result implies a compensation of the health shock with health investments, a reinforcement of the health shock with education investments, and a net compensation of the health shock.

There are major differences between developing countries and the US in terms of pension systems, credit constraints, wealth, culture, and financing of medicine and education. Yi et al. (2015) note that parents in developing countries may have an extra motivation to reinforce endowments, because they are more dependent on their children in retirement. Therefore, they may invest more in a child who is more likely to bring back high financial returns. In contrast, US parents are less dependent on their children in retirement and might be more inclined to care about the equality of their children's outcomes. This paper is consistent with this intuition, as it finds that US children with lower health endowments are allocated with more schooling.

Two papers from this group show compensation, the same result as in this paper. Terskaya (2019) uses Mexican data for identification of parental preference for equality and argues that even though parents act as if they wish to reinforce disability with less schooling, this result is confounded by different prices of educational investments for healthy and disabled children. Terskaya (2019) argues that parents actually have a preference for equality and that they compensate for disability conditional on the price effect.[2] Nicoletti and Tonei (2017) use Australian data and show that parents compensate for health shocks with parental time investments in developmental activities. We complement papers by Terskaya (2019) and Nicoletti and Tonei (2017) by using data from a different country (USA), a different identification method, and different measures of health endowment.

[2]See also Section 4.3 for a discussion of biases that may lead to false detection of reinforcement.

Skill Investments and Postnatal Measures of Skill Endowment In this group several papers find reinforcement (Ayalew, 2005; Behrman et al., 1994; Frijters et al., 2013; Nicoletti and Tonei, 2017), while results by Behrman et al. (1982) are nuanced. Our paper does not confirm reinforcement and, moreover, does not rule out compensation.

Nicoletti and Tonei (2017) use Australian data and argue that parents do not react to changes in cognitive skills but reinforce changes in socio-emotional skills. Ayalew (2005) uses data from Ethiopia, and shows that school attendance is reinforced by skill.

Papers from this group that are, like our paper, based on US data include Behrman et al. (1982, 1994) and Frijters et al. (2013). Frijters et al. (2013) use handedness (right handed vs left handed) as an instrument for cognitive ability differences. They use an index for various parental non-formal education investments.[3] Behrman et al. (1982) estimate a general preference model based on data describing US white male fraternal twin pairs, in which both twins served in the U.S. military forces, primarily during the Second World War. The result is nuanced: the authors estimate parameters of the model that are consistent with reinforcement. However, they find that parents do care about equality and that their behavior is significantly different from that in a pure investment model. We discuss results of another paper by Behrman et al. (1994) in more detail in Section 5.4, in which we argue that there is no contradiction between our results and results by Behrman et al. (1994) that are based on the MTR data.

Twins vs. Non-Twin Siblings The extent to which our results are applicable for non-twin siblings is an important question. Bharadwaj et al. (2018) conjecture that parents find it easier to compensate when siblings are not exactly the same age. Sanz-de Galdeano and Terskaya (2019) also find evidence consistent with this idea. We do not have data on siblings who are not twins, but we find evidence of compensation for twins. Therefore, we can expect compensation for non-twins as well.

[3]The index for parental investment includes helping their child learn the alphabet, reading stories, having books, etc.

3 Data

We combine three datasets, which longitudinally describe the lives of twin pairs who were born in Minnesota. Each pair in the sample was raised together.

Minnesota Twin Registry As described in Krueger and Johnson (2002), the MTR was initiated in 1983 and includes data on twins born in Minnesota between 1936 and 1955. The MTR staff identified the twins retrospectively from their birth records and contacted twins to ask for their participation in surveys in person, by mail, and over the telephone. Approximately 80% of the identified twins were located. Among those located, approximately 80% agreed to participate. There were 4,307 twin pairs in which both twins participated. MTR participants answered survey questions about an array of topics, including their education and health backgrounds. The MTR gathered participants' birth weight data directly from their birth certificates.

Socioeconomic Survey of Twins In 1994, MTR respondents from same-sex twin pairs were resurveyed by economists in the SST, which gathered further information from each twin regarding their labor market participation, wages, health, and education. SST respondents were also asked to provide information about their parents, siblings, spouses, and children. Importantly for our analysis, the SST asked each twin to report education for both themself and the other twin, meaning that we have two separate observations on years of schooling for each twin. 1,325 intact twin pairs returned valid SST questionnaires (Behrman and Rosenzweig, 2002).[4] The SST has been used in a number of influential publications in economics (e.g., Antonovics and Town, 2004; Behrman and Rosenzweig, 2002, 2004).

Mortality Data To construct mortality data for these twins, we gathered data from both the Social Security Death Master File and the Centers for Disease Control and Prevention's National Death Index. Research found that over 90% of deaths are correctly identified by each database (Hauser and Ho, 2001; Wentworth et al., 1983). We improve

[4]See Behrman and Rosenzweig (1999) for a more thorough description of the SST.

the accuracy of the mortality data by comparing data from alternative sources, including data from contacting next-of-kin. The available mortality data up to year 2014 provides us with a 20-year risk period window between the initial date when living twins participated in the SST and the date when the mortality status of respondents was last observed.

Characteristics of the Twin Sample MTR twins are reasonably representative of their Minnesota birth cohort (Krueger and Johnson, 2002). Twins are known to have lower birth weight than singletons; however, this difference is a natural adjustment to limited space in utero. The MTR sample is almost entirely white, which is consistent with the historical demographics of Minnesota.[5] We exclude two twin pairs with at least one non-white parent from the estimation sample, since the data are insufficient for a reliable study of the minority population. Our twin sample thus consists of twin pairs who are white and of the same sex, who participated in the SST survey, and in which both twins provided education information in that survey. This gives us a sample of 834 male and 1,632 female twins, whose characteristics are defined and described in Table 1.

For coding the total effective years of schooling, we follow the same procedure that Antonovics and Goldberger (2005) designed for the same SST data, in which years of schooling are defined based on the highest degree achieved, as well as any additional reported schooling beyond the highest degree.[6] Although twins from the same pair are generally similar to one another, Panel B of Table 1 shows that on average they differ in education level by over one year.

Our main outcome of interest is mortality over a 20-year risk period from 1994 to 2014. In Panel B we can see a substantial variation of mortality within the twin pairs. For women and men respectively, both twins died in 2 and 3% of cases, which contrasts to a much larger percentage when only one twin died, 11 and 16% of cases. Paired

[5] In 1960, non-whites represented 1.2% of Minnesota's total population (US Census Bureau, 1960).

[6] For example, a high school degree is coded as 12 years of schooling or a college degree as 16. A twin who reports a high school degree plus one year of college will be coded 13 years of schooling. However, a twin who has not completed a particular degree is only at most coded with the years associated with that degree minus one, regardless of how many years they report. Thus a twin who reports a high school degree plus five years of college but no college degree will be coded with 15 years rather than 17.

Table 1: Summary Statistics

Variable	Year measured	Males Mean	Males Std. dev.	Females Mean	Females Std. dev.
(A) Individual Twins					
Year of birth	At birth	1947.7	5.4	1947.9	5.6
Birth weight in pounds	At birth	5.96	1.14	5.68	1.11
Monozygotic[a]	1983	0.61	0.49	0.58	0.49
Ever had a disabling injury?[b]	1983	0.32	0.47	0.33	0.47
Years of education[c]	1994	15.1	2.3	14.1	2.2
Physical health problems[d]	1994	0.51	0.50	0.51	0.50
Self-reported health[e]	1994	1.58	0.64	1.64	0.66
BMI[f]	1994	26.60	3.58	25.42	5.16
Overweight[g]	1994	0.66	0.47	0.42	0.49
Alcohol problems[h]	1994	0.045	0.206	0.016	0.124
Died before 2015	1994–2014	0.096	0.295	0.063	0.244
Age at death if died	1994–2014	62.8	7.6	62.9	8.8
(B) Twin Pairs					
Both died before 2015	1994–2014	0.03	0.17	0.02	0.14
At least one died before 2015	1994–2014	0.16	0.37	0.11	0.31
Absolute difference within pairs					
Birth weight, pounds	At Birth	0.68	0.77	0.69	0.61
Ever had a disabling injury?[b]	1983	0.13	0.33	0.12	0.32
Education, years[c]	1994	1.52	1.62	1.19	1.42
Physical health problems[d]	1994	0.42	0.49	0.38	0.49
Self-reported health[e]	1994	0.53	0.63	0.51	0.61
BMI[f]	1994	2.65	2.33	3.51	3.54
Overweight[g]	1994	0.29	0.46	0.30	0.46
Alcohol problems[h]	1994	0.07	0.26	0.03	0.17
Age at death if both died	1994–2014	4.90	5.16	8.00	6.30
Number of individuals[i]		834		1632	

Notes: "Std. dev." stands for "standard deviation." [a]Twin respondents are classified in the MTR data as monozygotic or dizygotic based on responses to a questionnaire for parents designed to determine the degree of similarity between twins. [b]Suffered from any disabling injury by 1983. [c]We use the code by Antonovics and Goldberger (2005) for calculating the effective years of schooling from the raw SST responses. [d]Respondents experienced any of the following: migraine headaches; hay fever; frequently occurring skin rash; hearing impairment; high blood pressure; heart condition; and loss of function in the neck, back, arms, or legs. [e] Respondents indicated their health status by selecting a number between 1 (bad) and 5 (excellent). However, in order to enable the signs on our regression coefficient estimates to be interpreted consistently across all outcomes, we reverse the numerical mapping of this scale, such that bad = 5 and excellent = 1. [f]Weight in kilograms divided by square of height in meters. [g]Overweight is a dummy=1 if BMI$>$ 25; [h]Any "family, job, or health problems due to alcohol use." [i]The count of all twins from same-sex twin pairs in which both twins reported their education level in the SST. The sizes of our various estimation samples vary, depending upon the number of twin pairs for which both twins have known values of each respective outcome variable. Here we report the count of the superset that comprises all of the various subsets that constitute our individual estimation samples.

with variation in education within twin pairs, the within-pair variation in mortality is essential for model identification and statistical power.

Other key health measures gathered in the SST include both a history of clinical health problems and a self-reported health rating (see the notes to Table 1 for details on variable definitions for these and other variables). Reporting a clinical physical health problem is a common outcome (seen in about half of the twins) and it shows substantial variation: for about 40% of twin pairs, one twin reported a clinical physical health problem, while the other reported none. The more subjective self-reported physical health measure is concentrated around good health (the average self-reported health is about 1.6 on a scale from 1 (excellent) to 5 (bad), which results in a rather small absolute difference within pairs, about 0.5 on average).

For our study of the mechanisms that link education and health we use a small set of available variables that are informative of lifestyles and health behaviors. We study the BMI, a dummy for overweight status that is based on BMI, and a dummy for problems at home or at work due to alcohol abuse. We consider both BMI and a dummy variable for overweight because the former has the benefit of being continuous and containing additional information, while the latter has the benefit of indicating a clinically important BMI threshold. Few people report alcohol abuse (1.6% of women and 4.5% of men); however, overweight is a frequent issue for both women and men (42% and 66%), and it shows substantial average absolute difference within twin pairs (about 0.3).

As one can see from Table 1, our sample is characterized by sex imbalance. The sex imbalance was already a feature of the original MTR survey, because females were more likely to cooperate in both recruitment and in answering additional questions or tests (Lykken et al., 1990).[7]

By the time of the SST follow-up, additional factors further contributed to the sex imbalance, as males had a larger dropout rate. Moreover, we lose a twin *pair* from our sample every time that a *single* twin fails to answer a given survey question, a feature

[7] Lykken et al. (1990) report that "... 92% of the individual females returned a Biographical Questionnaire (BQ) while 63% of the female pairs were concordant for returning all tests. For males, 77% of individuals returned a BQ but only 34% of the pairs were concordant for testing."

that widens the sex gap further. As a result, 66% of our estimation sample are females.

This selection on biological sex does not create a bias in our estimates given that all pairs in our estimates are same-sex and given that we we control for common traits of twins though differencing within twin pairs. At the same time we control for many other types of sample selection, as described in detail in Section 4.2.

4 Methodology

We apply a linear probability model (LPM) to within-twin-pair first differences among MZ twins in order to determine the effect of education on health outcomes. We then compare the results for MZ twins to those for DZ twins in order to characterize the manner in which families allocate educational resources among siblings with different health endowments.

We take advantage of properties of linear models that allow for implementation of well-established techniques: differencing between twins (to account for sample selection and unobserved heterogeneity) combined with instrumenting noisy differences (to account for measurement error). Not surprisingly, the linear probability model is a common choice in analyses of twin data, including studies of mortality outcomes (e.g., Almond et al., 2005; McGovern, 2019). We return to the discussion of the LPM in the end of Section 4.2, where we argue that the limitations of the LPM create neither identification nor estimation issues in our paper.

4.1 Model of Schooling Decision and Health

Behrman et al. (1994) lay out a model for determining schooling and wages for twins. We adapt this model to the case of schooling and health outcomes as follows.

Consider a family j with twin children i and k. The family allots S years of schooling to each twin according to the equations

$$S_{ij} = \alpha_1 a_{ij} + \alpha_2 a_{kj} + \delta h_j + f_j + \boldsymbol{\alpha}_3 \boldsymbol{x}_{ij} + u_{ij} \tag{1}$$

and

$$S_{kj} = \alpha_1 a_{kj} + \alpha_2 a_{ij} + \delta h_j + f_j + \boldsymbol{\alpha}_3 \boldsymbol{x}_{kj} + u_{kj}. \tag{2}$$

Here, a_{ij} and a_{kj} represent individual-specific genetic health endowments. However, because all outcomes in this paper are adverse (mortality, poor health identified by the doctor, and a self-reported health rating), variables a_{ij} and a_{kj} are interpreted as "negative health endowments" based on the structure of Equations (1–4): they positively affect adverse outcomes in equations (3) and (4). Variable h_j represents common endowments, and f_j, represents family environment. Variables $\boldsymbol{x}_{ij}$ and $\boldsymbol{x}_{kj}$ are vectors of possible confounders—in our case, birth weight and history of disabling injury. Finally, u_{ij} and u_{kj} are random shocks to educational attainment.

Mortality outcomes M for each twin are determined by equations:

$$M_{ij} = \beta_1 S_{ij} + a_{ij} + h_j + \gamma f_j + \boldsymbol{\beta}_2 \boldsymbol{x}_{ij} + v_{ij} \tag{3}$$

and

$$M_{kj} = \beta_1 S_{kj} + a_{kj} + h_j + \gamma f_j + \boldsymbol{\beta}_2 \boldsymbol{x}_{kj} + v_{kj}, \tag{4}$$

where v_{ij} and v_{kj} are random shocks to health.

β_1 is a key parameter of interest. Establishing that $\beta_1 < 0$ would imply that additional years of schooling reduce mortality. Other key parameters are α_1 and α_2, which describe the own- and cross-effects of individual health endowments on the family's distribution of educational resources between twins. If $\alpha_1 < 0$ and $\alpha_2 > 0$ (case 1), this implies that families reinforce differences in health endowments by increasing years of schooling for the better-endowed twin (the twin with the lower value of a) at the expense of the worse-endowed twin.[8] If we have $\alpha_1 > 0$ and $\alpha_2 < 0$ (case 2), this would imply that families instead compensate for differences in health endowments. Finally, $\alpha_1 = \delta$ and $\alpha_2 = 0$ (case 3) would imply that the educational investment for each twin is set individually and is unrelated to the other twin's endowment. Other health-related outcomes are

[8]The signs of these relationships are reversed from those described in Behrman et al. (1994), since we normalize the latent health endowment associated with the adverse outcome of mortality, while they normalize the latent wage-earning endowment associated with the beneficial outcome of wage.

modeled using the same type of equations that we use for mortality.

In this classification of cases 1–3 we follow the standard implicit assumption of the literature that schooling outcomes represent parental resource allocation decisions and that costs of investments are independent of endowments. In Section 4.3 we discuss the bias towards detecting reinforcement that is induced by these assumptions.

One limitation of this model is that regressions (3) and (4) do not allow for possible interactions between schooling and endowments (otherwise, we would lose our identification.) However, the literature on this subject presents mixed evidence on the interaction of schooling with genetic endowments: it could be positive, negative, nonexistent overall, or negligible, because interactions with specific components of the overall endowment can cancel each other.[9] The investigation of heterogeneity is an important task that we leave to future research based on data that have observable measures of genetic endowments.

4.2 Within-Twin-Pair First Differences

MZ twins have identical genetic endowments, so that for MZ twin pairs in our model, $a_{ij} = a_{kj}$. Taking the difference between Equations (1) and (2) yields the following equations for the difference in years of schooling and for the difference in mortality among MZ twin pairs:

$$\Delta S_j^M = \boldsymbol{\alpha}_3 \Delta \boldsymbol{x}_j^M + \Delta u_j^M \tag{5}$$

[9] Conti and Heckman (2010) suggest that the effect of education on self-reported health for men does not differ by health endowment, while its interaction with cognitive skills have the opposite sign to its interaction with noncognitive skills. For women, the corresponding interaction effects appear considerably weaker than for men. Barcellos et al. (2018) study an interaction between education and genetic endowment in predicting body size, lung function, and blood pressure. They find that for those with higher genetic risk of obesity, the beneficial effects of education on health outcomes are stronger. However, they do not find such interaction with genetic predisposition for educational attainment, which is known to be strongly predictive of both early skills and early health (Bolyard and Savelyev, 2021). In the Web Appendix B to their paper studying the effect of education on longevity, Hong et al. (2000) test for interactions of education with two measures of endowments, IQ and a latent propensity for educational achievement, and they cannot reject the hypothesis that these interactions equal to zero. Auld and Sidhu (2005) use parental schooling as an instrumental variable for own schooling affecting health and argue that schooling is only beneficial for health of low-ability individuals. In contradiction to this result, Savelyev and Tan (2019) and Savelyev (2020) find strong positive effects of education on health and longevity for individuals with extraordinarily high cognitive ability.

and

$$\Delta M_j^M = \beta_1 \Delta S_j^M + \boldsymbol{\beta}_2 \Delta \boldsymbol{x}_j^M + \Delta v_j^M, \tag{6}$$

from which we are able to identify β_1.

DZ twins have different genetic endowments, so the analogous equations for DZ twin pairs are:

$$\Delta S_j^D = (\alpha_1 - \alpha_2)\Delta a_j + \boldsymbol{\alpha}_3 \Delta \boldsymbol{x}_j^D + \Delta u_j^D \tag{7}$$

and

$$\Delta M_j^D = \beta_1 \Delta S_j^D + \Delta a_j + \boldsymbol{\beta}_2 \Delta \boldsymbol{x}_j^D + \Delta v_j^D. \tag{8}$$

In the system of equations represented by Equations (5)–(8), α_1 and α_2 are not individually identified. However, as shown by Behrman et al. (1994), if we assume that the individual-specific stochastic components v_{ij} and u_{ij} are drawn from the same distribution for both MZ and DZ twins, then the difference $(\alpha_1 - \alpha_2)$ is identified and can be calculated in the following way:

$$\alpha_1 - \alpha_2 = \frac{1 - R}{\beta_1^D - \beta_1^M}\ , \tag{9}$$

where $0 < R \equiv \frac{var(\Delta S^M)}{var(\Delta S^D)} < 1$, and β_1^D and β_1^M represent the estimates from Equation (6) for the DZ and MZ twin subsamples, respectively. (Here we ignore the important issue of measurement error that we address below in Section 4.4.)

As $R < 1$, the sign of $(\alpha_1 - \alpha_2)$ matches the sign of $(\beta_1^D - \beta_1^M)$. Identification of this difference is sufficient to determine the type of allocation behavior of families, as discussed in the end of Section 4.1. However, we recommend taking into account not only the estimated sign of $(\alpha_1 - \alpha_2)$, but also the bias towards detecting reinforcement, as we discuss in Section 4.3.

Another benefit of using within-twin pair differences is the avoidance of selection biases (Behrman et al., 1994; Heckman and MaCurdy, 1980; Pitt and Rosenzweig, 1990). A selection bias can be viewed as an omitted variable bias (Heckman, 1979). As long as the omitted variable that affects sample selection is family-specific, the twin first differences model controls for it. For instance, twin pairs might be selectively attritted

due to low common health or skill endowments, or biological sex (we use a sample of same-sex twins). Selection into birth in Minnesota may depend on parental background, such as Scandinavian origin. Selection into being a twin could be affected by maternal genes, health, or prenatal environment.[10]

A possible attrition caused by education largely falls into the same category. Twins typically have either the same or very similar levels of education, as confirmed by this paper. Therefore, schooling is mostly affected by family-specific variables, which are common genetic endowment and common home environment. We control for these common causes of possible selection using the twin first differences technique.

However, it is still possible, though theoretically unlikely, that attrition of pairs depends on the *difference* in education between the twins in each pair. This potential issue would constitute an endogenous selection for our education model in differences. In Table A-1 we test, and fail to reject, the hypothesis that the absolute value of the within-pair difference in education is unrelated to the attrition of twin pairs.

Implications of Relying on the Linear Probability Model The linear probability model (LPM) that we rely on for modeling outcomes, such as mortality, is known to be only an approximation to a true nonlinear probability model. However, an LPM has important advantages, which is why it is so widely used in the literature. Advantages of the LPM are crucial in our specific case, as we must use linear equations in order to leverage the within-twin-pair differencing that underpins our identification strategy.

Two questions arise regarding our LPM: (1) whether we can expect our LPM estimates to be accurate and (2) whether heteroscedasticity issues associated with the LPM create a problem either with model identification or with standard error estimation. (In their paper introducing this technique, Behrman et al. (1994) considered continuous outcomes, such as income, for which homoscedastic errors are possible.)

We address the first question based on both theoretical and empirical arguments. Moffitt (1999) provides a detailed theoretical discussion of the LPM. According to this discussion, LPM is often a reasonable approximation. However, its accuracy varies across

[10] E.g., through genetic predisposition for multiple ovulation or through selective miscarriage (Bhalotra and Clarke, 2019).

the probability range due to the limited ability of a linear function to approximate a non-linear relationship. Importantly, Moffitt (1999) reports that the accuracy of the LPM is the highest not only around probability 1/2, but also in the tails.[11] Our data are typically in the tails, and so we can expect a reasonably high accuracy of our approximation.

Our empirical argument is based on verifying these theoretical expectations using binary outcomes available in our data including low-probability outcomes, such as mortality, and an outcome that occurs with sizable probabilities for both men and women, which is the overweight status. We estimate associations between years of education for each *individual* twin from our sample and binary outcomes using LPM, logit, and probit models, and find that the marginal associations are almost identical among these three models (see Table A-2 in the Web Appendix).

To address the potential heteroscedasticity issue, we note that the identifying formulas for our model, which we adapted from Behrman et al. (1994) while keeping the same model structure, are based on unconditional variances and covariances, whereas homoscedasticity is an assumption about a conditional variance.[12] Just as OLS estimates of regression coefficients are robust to heteroscedasticity because they are based on unconditional variances and covariances, ours have the same property.

However, standard errors of an LPM do require adjusting for heteroscedasticity. Therefore, we calculate Huber-White robust standard errors rather than classical errors in our main model. In addition, we present robustness checks in Tables A-3–A-5 of the Web Appendix, where we compare our Huber-White asymptotic standard errors with bootstrap standard errors and with classical (asymptotic) standard errors. The results of this check is that standard errors of these three types are almost identical. Both Huber-White and bootstrap errors are heteroscedasticity-robust, but use very different estimation procedures. Therefore, these two methods support each other's validity. The closeness of classical errors to heteroscedasticity-robust errors suggests that heteroscedasticity issues are negligible in the case of our model.

[11] See the bottom of p. 1377 of Moffitt (1999).
[12] See formulas 13–15 in Behrman et al. (1994).

4.3 Biases towards Detecting Reinforcement that are Inherent in the Literature

We argue that there are two types of bias towards detecting reinforcement, which is the outcome most commonly found in related literature.[13] Our study is subject to both of these biases, as we describe below. However, our findings of compensating behavior are only strengthened by the presence of biases *against* detecting compensating behavior.

The first type of bias is due to *a child's own influence on decisions about investment in their skills*. Implicit in the methodology of many papers in the related literature is the assumption that skill investments are purely determined by parental choice. While this assumption might be true for certain specific parental investments, such as reading to a baby, we argue that, in general, the child's own actions affect the total skill investment.

For instance, the total number of effective years of education, the metric we consider in this paper, is arguably affected by both parental and children's decisions. On their part, parents can encourage children to study, help them with schoolwork, hire a tutor, support their college education, and redistribute resources to either compensate or reinforce endowments. However, a less favorable endowment for a child may make studying a more costly activity for that child. A student with lower health or intelligence may find it more difficult to concentrate on learning. Therefore, this student may have higher psychological cost of studying: studying might be less enjoyable (or more disliked). Second, such a student may need additional study time to achieve the same educational goal as others with more favorable endowments. Longer time spent studying takes away time from paid work or leisure. In accordance with the law of demand, we would therefore expect a reduction in that child's personal investment in education, other things being held equal.

Regarding the expected benefits of studying, a poor endowment may imply shorter life expectancy, leading to smaller expected lifetime return from education and therefore reduced economic incentives to study (Becker, 2007).

The complementarity between a person's own endowment and educational invest-

[13]The literature on intrafamily resource allocation is surveyed in Section 2.2. The two types of biases are not necessarily present in every single study, but are typical for existing studies.

ments is consistent with theory (Becker, 2007) and with empirical evidence (e.g., Case et al., 2005; Ding et al., 2009; Lehrer and Fletcher, 2009, 2011).

Therefore, while parental propensity to allocate resources for skill investments could be either compensating, reinforcing, or neutral, a child's own contributions to the final own-skill investment can be expected to be complementary with the child's own endowment. When observational data are used to analyze household decision-making, the result of this pattern is a bias towards detecting parental reinforcement.

The second type of bias is due to the *unaccounted-for price effect*. Terskaya (2019) provides both theoretical and empirical arguments that inequality-averse parents may nevertheless reinforce the differences in their children's endowments due to the higher relative price of investing in children with lower endowments. The dependence of parental investment costs on the child's endowment was first modeled by Becker and Tomes (1976). Terskaya (2019) and Sanz-de Galdeano and Terskaya (2019) offer empirical approaches that control for the price effect and thus disentangle parental inequality aversion from the price effect. In contrast, traditional methods based on use of family fixed effects alone are only able to identify an aggregate measure which fails to separate the effects of inequality aversion from the price effect. However, the decision to control for the price effect comes with an inherent econometric cost: additional assumptions must be made, and such an approach precludes the use of family fixed effects or twin fixed effects to account for the endogeneity problem.[14]

4.4 Addressing Measurement Error through Instrumental Variables

As Griliches (1979) points out, the attenuation bias due to measurement error is particularly troublesome in estimates derived from twin data. For instance, when we take the difference of education levels of twins, we can expect a small level of the difference (as twins tend to have similar education levels) combined with an amplified measurement error. Ashenfelter and Krueger's (1994) elegant instrumental variables approach is de-

[14] Terskaya (2019) makes a parallel paths assumption, which is similar to the one used in the difference-in-difference model, by assuming that schooling level of disabled individuals changes with family size by the same amount as the schooling level of non-disabled individuals. Sanz-de Galdeano and Terskaya (2019) rely on the OLS model combined with estimates of the omitted variable bias due to likely confounders.

signed to correct the large attenuation bias that we can expect in this situation. This approach uses one twin's report of the intra-twin-pair difference in education as an instrument for the other twin's report of the same difference. We apply this IV approach in our analysis, using the 1994 SST survey data in which each twin reported both their own and their twin's education backgrounds.

Consider twins 1 and 2 from a same-sex pair j. Let S_k^i represent twin i's report of twin k's years of schooling, and let $\Delta_{Si} = (S_1^i - S_2^i)$, $i = 1, 2$, which is how many more years of schooling twin 1 had than twin 2 based on twin i's reports. Then the first stage of the two-stage least squares (2SLS) method for each sex in this two-stage least squares framework can be written as

$$\Delta_{S1} = a_0 + b\Delta_{S2} + \boldsymbol{c}\Delta\boldsymbol{x}_j + \epsilon_j. \tag{10}$$

In the second stage of this approach, the observed difference in mortality outcomes is regressed on the predicted value of the difference in education $\widehat{\Delta_{S1}}$, as calculated in the first stage regression:

$$\Delta M_j = \beta_1\widehat{\Delta_{S1}} + \boldsymbol{\beta}_2\Delta\boldsymbol{x}_j + \Delta v_j. \tag{11}$$

Ashenfelter and Krueger demonstrate that this approach generates consistent estimates of β_1, the coefficient of interest, even when a twin's reports of her own education and of her twin's education have measurement errors that are correlated with one another. We estimate β_1 using a 2SLS estimator, which we estimate using a single formula rather than actually running two steps. We report Huber-White standard errors and demonstrate that our statistical inference is robust to using an alternative bootstrap approach.[15]

We estimate versions of (10) and (11) under each of two alternative specifications: (A) no controlling for $\Delta\boldsymbol{x}_j$, and (B) controlling for $\Delta\boldsymbol{x}_j$, with the missing values for $\Delta\boldsymbol{x}_j$ imputed using Markov Chain Monte Carlo (MCMC) multiple imputation as described in Rubin (1987) and Schafer (1997), a method that preserves the variance-covariance matrix of variables in the data. The results from estimating approaches (A) and (B) are

[15]See Tables A-3–A-5 of the Web Appendix for almost identical standard errors calculated using alternative methods.

very close to one another and both support the conclusions of this paper. We report results for conditional specification (B) in the main text and show a comparison with the unconditional specification (A) in Table A-6 in the Web Appendix. The nonessential role of confounding factors that include birth weight in twin fixed effects estimates of the effect of education on mortality is consistent with results by Lundborg et al. (2016).

5 Empirical Results

5.1 Health Outcomes

The results from our estimation of the first stage regression (10), run separately for MZ and DZ sub-samples, are shown in Table 2. Unsurprisingly, one twin's report of the intra-pair difference in years of schooling is a strong predictor for the other twin's report of the same difference: F-statistics range from 176 to 595, which is evidence of an exceptionally strong instrument given the standard threshold for a strong instrument of approximately nine.[16]

Then we use a (one-step) 2SLS estimator to identify the effect of education on health-related outcomes based on the sample of MZ twins. Specifically, we consider the outcomes of mortality (death within 20 years after the SST survey in 1994), any report of physical health problems, and overall self-reported health rating.

In light of abundant evidence in the literature that the effects of education on health, longevity, and a number of health behaviors are non-harmful, we primarily use one-tailed tests for the estimated coefficients on years of schooling in our regressions.[17]

The second stage results are shown in Table 3. For males, each additional year of schooling yields a 3.1 percentage point (PP) drop in the probability of death during the 20 years following the 1994 SST survey. For women, each additional year of schooling is associated with a 3.4 PP decline in the probability of reported physical health problems. Other estimates by sex are not precisely determined, but the signs of these estimated

[16]For instance, Stock et al. (2002) suggest a threshold of 8.96 for the case of one instrumental variable.

[17]See, for example, Grossman (2015) and Galama et al. (2018), in addition to the discussion of the literature in this paper. Despite a common misconception, it is sensible to use all available information to maximize statistical power.

Table 2: First Stage of the 2SLS Approach: Intra-Pair Difference in Years of Schooling as Reported by Twin One Regressed on Intra-Pair Difference as Reported by Twin Two

		Pooled sexes	Males	Females
Monozygotic	coefficient	0.776 ***	0.827 ***	0.745 ***
	standard error	*(0.036)*	*(0.061)*	*(0.041)*
	F-statistic	*459*	*187*	*327*
	# of twin pairs	694	244	450
Dizygotic	coefficient	0.840 ***	0.875 ***	0.819 ***
	standard error	*(0.041)*	*(0.036)*	*(0.062)*
	F-statistic	*425*	*595*	*176*
	# of twin pairs	539	173	366

Notes: We use specification (10) for each sex. For the pooled sample we additionally control for a sex dummy. Huber-White standard errors are shown in parentheses. Asterisks represent statistical significance levels: *** 1%, ** 5%, * 10%.

Table 3: Effects of Education on Mortality and Health

		Pooled sexes	Males	Females	Sex difference
Mortality	coefficient	-0.023 **	-0.031 **	-0.016	-0.015
	standard error	*(0.012)*	*(0.018)*	*(0.015)*	*(0.023)*
	# of twin pairs	558	204	354	
Physical health problems	coefficient	-0.028 *	-0.021	-0.033 *	0.011
	standard error	*(0.019)*	*(0.027)*	*(0.025)*	*(0.037)*
	# of twin pairs	694	244	450	
Self-reported health[(a)]	coefficient	-0.044 *	-0.061	-0.022	-0.039
	standard error	*(0.034)*	*(0.049)*	*(0.047)*	*(0.068)*
	# of twin pairs	680	241	439	

Notes: Results of the 2SLS model reported. Huber-White standard errors are shown in parentheses. One-tailed p-values reported for estimated coefficients on years of education, two-tailed for difference in coefficient estimates across sexes. Asterisks represent statistical significance levels: *** 1%, ** 5%, * 10%. [(a)]In order that the signs on our regression coefficient estimates can be interpreted consistently across all outcomes, we reversed the numerical mapping of this scale, such that bad = 5 and excellent = 1.

coefficients are all in the direction of health improvement.

However, when we increase statistical power by pooling sexes, the effects on all three health outcomes become statistically significant at least at the ten percent level. On average, one year of schooling decreases mortality by 2.3 PP, decreases the reported physical health problems by 2.8 PP, and improves self-reported health by 0.044 standard deviations.

Since standard errors are rather large, these estimates imply a large range of possible population coefficients. Therefore, we do not focus our analysis on the numerical estimate of the effect, but rather on testing whether the effect is different from zero, a question that is a subject of a major debate in the literature, as we describe in Section 2.1 above. In the same section we analyse how our contribution is related to contributions of other papers.

Prior research established that individuals with higher income and education levels consume more healthcare, all else equal (Strauss and Thomas, 1998), a result we need to take into account when interpreting our estimates. Indeed, our physical problems dummy is based on subjects' knowledge of their medical conditions. Some medical conditions, such as high blood pressure or heart disease, are likely unknown to the respondent without a diagnosis from a medical professional. Accordingly, our coefficient estimates for the physical health problem outcomes, which describe the protective effect of education on the probability of reporting *awareness* of having experienced a physical health problem, likely understate education's effect on the probability of truly *experiencing* a physical health problem. If more-educated respondents had the same amount of health conditions as did less-educated respondents but were more likely to be aware of them, we would then find a positive effect of education on the probability of reporting health problems. Despite this expected bias toward finding a positive effect, we still find a negative effect, suggesting a substantial true beneficial effect of education on the probability of experiencing health problems. In contrast to health measures reported by respondents, mortality is a fully objective measure and is not susceptible to the same type of bias.

5.2 Suggestive Mechanisms of the Effect of Education on Health and Longevity

We have data on BMI and on alcohol problems, which are among possible mechanisms that drive the effects of education on health and longevity. Twins indicated their height and weight at the time of the SST, from which we generate both the continuous BMI variable and a dummy variable for being overweight that is defined as BMI > 25. The twins also indicated whether they had ever experienced "family, job, or health problems due to alcohol use," which can be viewed as a proxy for alcohol addiction or abuse.

2SLS estimates for the effects of education on BMI, overweight status, and alcohol problems are presented in Table 4. Among men, each additional year of education decreases the BMI index by 0.20 and decreases the likelihood of being overweight by 4.4 PP.[18] We find no statistically significant relationship between education and BMI or being overweight for women. We also find a small and borderline statistically significant ($p = 0.123$) 1% reduction in alcohol-related problems for the pooled sample of men and women. Just as estimates in Table 3, estimates in Table 4 have large standard errors and so suffer from low accuracy. Therefore, just as above, we concentrate not on the effect estimate, but on testing the sign of the relationship, whether the sign supports the qualitative results from Table 3 or not.

Overall, based on our very limited data on the potential mechanisms, we find solid evidence of the excessive body weight channel for men but not for women and weak evidence of alcohol abuse channel for the pooled sample.

Results regarding differences in the effect of education on body weight between the sexes differ in the literature, as some papers find them either only for men or primarily for men (e.g., Bockerman and Maczulskij, 2016; James, 2015; Kemptner et al., 2011; Webbink et al., 2010), while some others find them either only for women or primarily for women, (e.g., Atella and Kopinska, 2014; Brunello et al., 2013; Grabner, 2009; Sassi et al., 2009).

Differences in results might be related to differences in countries studied in these

[18]The effect of years of education on BMI for men was estimated by Behrman et al. (1994) using the same data but the authors found no statistically significant results.

Table 4: Suggestive Mechanisms of the Effect of Education on Health and Longevity

		Pooled sexes	Males	Females	Sex difference
Overweight[(a)]	coefficient	-0.010	-0.044 **	0.024	-0.069 **
	standard error	*(0.014)*	*(0.022)*	*(0.017)*	*(0.027)*
	# of twin pairs	670	240	430	
BMI	coefficient	-0.030	-0.199 *	0.127	-0.326
	standard error	*(0.113)*	*(0.145)*	*(0.166)*	*(0.220)*
	# of twin pairs	670	240	430	
Alcohol[(b)]	coefficient	-0.012	-0.016	-0.009	-0.007
	standard error	*(0.008)*	*(0.013)*	*(0.008)*	*(0.015)*
	# of twin pairs	694	244	450	

Notes: Effects of education on suggestive mechanisms reported based on the 2SLS model. Huber-White standard errors are shown in parentheses. One-tailed tests used for overweight and BMI, and two-tailed for sex differences and for alcohol (there is no consensus in the health economics literature regarding the sign of the relationship between educational attainment and alcohol use). Asterisks represent statistical significance levels: *** 1%, ** 5%, * 10%. [(a)]Refers to $BMI > 25$. [(b)]Refers to "family, job, or health problems due to alcohol use."

papers, in methodologies used, and in cohorts. Different cohorts are associated with different societal attitudes towards gender, which have greatly evolved over the course of the 20th century.

Many people consider healthy BMI as a factor of physical attractiveness, especially for women. Wilson (2012) reports that healthy body weight is highly observable, highly valued, and has a profound effect in the marriage market. The author also finds an effect that is specific for women and not for men: women have lower success (measured by marital status) on the marriage market as a result of being either overweight or obese. The high societal expectations for female beauty are consistent with the sex differences in the prevalence of overweight BMI reported in Table 1 for our sample: 42% for women and 66% for men. For the cohorts that we study (the average birth year is 1948), it is possible that women cared about their body weight no matter their education, while educated men cared more about their weight for health reasons.

The situation could have changed for later cohorts with increasing emancipation of women and with progressing changes on the marriage market. For instance, Brunello

et al. (2013), suggest that education leads to a greater increase in income for women than men, and so women benefit more from the protective effects of income against BMI. They also suggest that more educated women are less likely to get pregnant, thus avoiding pregnancy-related weight gain. All these factors are typical for career-focused women. Meanwhile, among the earlier cohort that we study, even highly-educated women are often homemakers.

In addition, as described by Chiappori et al. (2009), the role that education plays in the marriage market has changed significantly. In the old equilibrium, men were more educated than women, and so many educated men married down with respect to education level. Therefore, both more-educated and less-educated women could compete on the same marriage market. In the new equilibrium, more women get a college education than men, which leaves less-educated women little chance to marry up. Therefore, by obtaining additional education, a woman gets access to a more competitive marriage market, which could create additional incentives for educated women to care about their body weight. In contrast, for educated men, who face a favorable sex ratio in this equilibrium, the incentives for weight control might be less pronounced.

In line with our hypothesis, the results of our paper, based on Americans born around the 1940s, are consistent with results by Bockerman and Maczulskij (2016), obtained for Finns born before 1958. Our hypothesis is also consistent with results by Baum (2017), who finds based on the 1997 wave of the National Longitudinal Survey of Youth that a decrease in weight caused by college completion is larger in absolute value for females between ages 25 and 31 but larger in absolute value for males between ages 45 and 52. Therefore, it is possible that we find a historic sex difference in the effect of education on body weight that may have reversed for newer cohorts.

5.3 Intra-Household Allocation of Resources

As explained in Section 4.2, we can determine whether parents compensate for or reinforce endowment differences in siblings by establishing the sign of $(\beta_1^D - \beta_1^M)$. Our estimates for this difference are shown in Table 5. The statistically significant positive

numbers that we find are indicative of compensating behavior, in which the unhealthier twin receives more education. Our estimation results provide some evidence of compensating behavior when health is measured in terms of mortality, and strong evidence of compensating behavior when health is measured in terms of physical health problems. The outcome of self-reported health provides no precisely determined estimate of the difference, however, this outcome shows little sampling variation and is, arguably, the most subjective.

Table 5: Difference in Education Coefficients between DZ and MZ Twins, $(\beta_1^D - \beta_1^M)$

		Pooled sexes	Males	Females
Mortality	coefficient	0.026 *	0.037 *	0.019
	standard error	*(0.014)*	*(0.022)*	*(0.018)*
Physical health problems	coefficient	0.065 ***	0.059 *	0.070 **
	standard error	*(0.024)*	*(0.035)*	*(0.032)*
Self-reported health[(a)]	coefficient	-0.028	-0.027	-0.044
	standard error	*(0.045)*	*(0.065)*	*(0.062)*

Notes: Difference in results of the 2SLS coefficients reported. Huber-White standard errors are shown in parentheses. Asterisks represent statistical significance levels: *** 1%, ** 5%, * 10%. [(a)]In order that the signs on our regression coefficient estimates can be interpreted consistently across all outcomes, we reversed the numerical mapping of this scale, such that bad = 5 and excellent = 1.

It is possible that parents respond to differences in children's health endowments by reinforcing differences on one margin, while simultaneously compensating on another. This is the finding of Yi et al. (2015), who use Chinese twins data to show that when one twin experiences a negative health shock in childhood, the parents divert educational investments toward the healthier twin (reinforcement) but divert health investments toward the less healthy twin (compensation). Parents compensate more than they reinforce, so the authors find that families' overall behavior is compensatory. In our sample we do not observe information about childhood health investments, e.g. physician visits, so we are unable to identify this kind of multidimensional response to differences in health endowments. However, our findings for US twins indicate that parents divert

educational resources toward the less healthy twin, which is the opposite of the finding by Yi et al. (2015) for Chinese twins. Taken together, these results suggest differences in the objective function for the representative Chinese household versus the representative American one, as we discuss in Section 2.2.

Apart from the differences in countries studied in these papers, there are also differences in measures of educational investments and health differences. Yi et al. (2015) analyze payments for schooling at age 11, while we analyze the total effective years of schooling. They consider responses to early life health shocks, while we consider latent health endowments measured using a structure that includes education, mid-life health, and longevity as outcomes.

A Data Limitation Since cognitive and noncognitive skills may affect health and longevity, our estimates of $(\alpha_1 - \alpha_2)$ could capture the influence of not only health endowments, but also skill endowments. The data available for our sample do not include early-life measures of ability, which would allow us to model a multi-dimensional vector of endowments. Therefore, a conservative interpretation of our findings would be that we analyze a fully-unknown mixture of cognitive, noncognitive, and health traits. However, there are three considerations that point to interpretation of our latent endowment as primarily health-related: (1) Our endowment is directly loaded on an assortment of health outcomes, which makes health endowments a direct influence, while every non-health influence is indirect and uncertain. (2) In a robustness check (available from the authors upon request), we find that results of the model are robust to controlling for wages on the right-hand side, as if wages were a background control variable. Even though wage is endogenous and these results should therefore be interpreted with caution, we argue that if the endowment that we identify were in fact highly loaded on cognitive and noncognitive skills, our results would not be robust to the inclusion of wages as control, as wages are highly loaded on both cognitive and noncognitive skills (Heckman et al., 2006). (3) When we reanalyze the original Behrman et al. (1994) model in Section 5.4 below, by using wage and income as outcomes, our results are very differ-

ent from the results that we get using health outcomes.[19] If cognitive and noncognitive skills were the driving force behind our results for health outcomes, then we would expect the results for wage outcomes to be similar to the analogous results for health outcomes.

5.4 Reanalysis of Intrafamily Resource Allocation Induced by Skills

In this section we apply the same model as above but use earnings and wages as outcomes instead of health and longevity. By doing so we reanalyze a portion of results by Behrman et al. (1994) based on the same method and the same data as in the original paper, but using superior data on the earnings of men. Behrman et al. (1994) did not have access to earnings data and had to impute earnings from occupational data. Since using earnings imputed from occupations as regression outcomes may lead to biased estimates (Saavedra and Twinam, 2020) in addition to reduced statistical power, we use directly measured data on both earnings and wages from the SST survey.

Table 6 presents estimates of $(\beta_1^D - \beta_1^M)$ for earnings and wage models. The sign of $(\beta_1^D - \beta_1^M)$ is informative of the sign of $(\alpha_1 - \alpha_2)$, as discussed in Section 4.2, but a positive sign should now be interpreted as reinforcement rather than compensation.[20] The table reports parameters for two model specifications that differ by outcome: (1) log(earnings), and (2) log(wage). Specification (2) allows us to better capture skills than specification (1) by separating productivity per hour from the time spent working. We are unable to reject the null hypothesis for any of these two specifications.

Results of Table 6 confirm the original results by Behrman et al. (1994) that are based on the MTR data, as the estimated parameters $(\beta_1^D - \beta_1^M)$ are statistically insignificant.[21]

[19] Our estimates derived from wage and income, while not precisely determined, consistently show the opposite direction of effect when compared to the corresponding estimates derived from health endowments (taking into account the need to reverse the sign when the outcome proxies a positive rather than a negative endowment).

[20] As discussed in Section 4.1, this interpretation depends on whether the outcome is adverse (like mortality) or beneficial (like wage).

[21] Behrman et al. (1994) cannot reject neutral behavior based on the MTR data, but find reinforcement based on the National Academy of Science-National Research Council (NAS-NRC) Twin Data, as well as on NAS-NRC data merged with MTR data. Reanalyzing the full set of results from Behrman et al. (1994) is beyond the scope of this paper, as we focus on complementary results based on newly collected mortality data for the MTR sample and on understanding health endowments.

Table 6: Estimates of $(\beta_1^D - \beta_1^M)$ for Models using Earnings and Wages as Outcomes, Males

	Log of earnings	Log of wage
coefficient	0.033	0.044
standard error	*(0.049)*	*(0.066)*

Notes: Differences in 2SLS coefficients reported. Huber-White standard errors are shown in parentheses. Two-tailed tests are used. Lack of asterisks implies that no results are statistically significant even at the 10% level.

However, given the bias towards detecting reinforcement that we discuss in Section 4.3, the statistically insignificant results for $(\beta_1^D - \beta_1^M)$ that we obtain do not necessarily imply neutral behavior of parents. In fact, while this result does not show evidence of a reinforcing behavior, it does not rule out a compensating one.

6 Conclusions

Using newly collected mortality data for the largest survey of US twins, we provide new evidence that education affects health-related outcomes for both men and women. We also study educational investments as a parental response to their children's endowments of two kinds: health and skills. We find that parents compensate for differences in their children's health endowments. We do not find evidence that parents reinforce differences in their children's skill endowments, and, given the expected bias, we cannot rule out that they in fact compensate for such differences.

Our results are relevant for both theory and economic policy. The existence of a causal effect of education on health makes education a useful health policy variable in the cases of sub-optimal educational investments due to market failure. Parental responses should be taken into account when the government wishes to design programs that either compensate for or reinforce endowment differences in an optimal way. Our results are consistent with the notion that some inequality reduction is already occurring

naturally at the family level.

References

Abufhele, A., J. Behrman, and D. Bravo (2017). Parental preferences and allocations of investments in children's learning and health within families. *Social Science & Medicine 194*, 76–86.

Adhvaryu, A. and A. Nyshadham (2014). Endowments at birth and parents' investments in children. *The Economic Journal 126*, 781–820.

Aizer, A. and F. Cunha (2012, September). The production of human capital: Endowments, investments and fertility. Working Paper 18429, National Bureau of Economic Research.

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

Almond, D., K. Y. Chay, and D. S. Lee (2005, August). The costs of low birth weight. *Quarterly Journal of Economics 120*(3), 1031–1083.

Almond, D. and J. Currie (2011). Human capital development before age five. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 4B, Chapter 15, pp. 1315–1486. North Holland: Elsevier.

Almond, D., L. Edlund, and M. Palme (2009). Chernobyl's subclinical legacy: Prenatal exposure to radioactive fallout and school outcomes in Sweden. *The Quarterly Journal of Economics 124*(4), 1729–1772.

Almond, D. and B. Mazumder (2013). Fetal origins and parental responses. *Annual Review of Economics 5*, 37–56.

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

Antonovics, K. L. and A. S. Goldberger (2005). Does increasing women's schooling raise the schooling of the next generation? Comment. *The American Economic Review 95*, 1738–1744.

Antonovics, K. and R. Town (2004). Are all the good men married? Uncovering the sources of the marital wage premium. *The American Economic Review 94*, 317–321.

Arendt, J. N. (2005). Does education cause better health? A panel data analysis using school reforms for identification. *Economics of Education Review 24*(2), 149–160.

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

Atella, V. and J. Kopinska (2014). Body weight, eating patterns, and physical activity: The role of education. *Demography*.

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

Ayalew, T. (2005). Parental preference, heterogeneity, and human capital inequality. *Economic Development and Cultural Change 53*(2), 381–407.

Balia, S. and A. Jones (2008). Mortality, lifestyle, and socio-economic status. *Journal of Health Economics 27*, 1–26.

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

Baum, C. L. I. (2017). The effects of college on weight: Examining the "freshman 15" myth and other effects of college over the life cycle. *Demography*.

Becker, G. S. (2007). Health as human capital: Synthesis and extensions. *Oxford Economic Papers 59*(3), 379–410.

Becker, G. S. and N. Tomes (1976). Child endowments and the quantity and quality of children. *Journal of Political Economy 84*(4), S143–S162.

Behrman, J. R. (1988). Intrahousehold allocation of nutrients in rural India: Are boys favored? Do parents exhibit inequality aversion? *Oxford Economic Papers 40*(1), 32–54.

Behrman, J. R., H.-P. Kohler, V. M. Jensen, D. Pedersen, I. Petersen, P. Bingley, and K. Christensen (2011). Does more schooling reduce hospitalization and delay mortality? New evidence based on Danish twins. *Demography 48*(4), 1347–1375.

Behrman, J. R., R. A. Pollak, and P. Taubman (1982). Parental preferences and provision for progeny. *Journal of Political Economy 90*(1), 52–73.

Behrman, J. R. and M. R. Rosenzweig (1999). "Ability" biases in schooling returns and twins: A test and new estimates. *Economics of Education Review 18*, 159–167.

Behrman, J. R. and M. R. Rosenzweig (2002). Does increasing women's schooling raise the schooling of the next generation? *The American Economic Review 92*, 323–334.

Behrman, J. R. and M. R. Rosenzweig (2004). Returns to birthweight. *The Review of Economics and Statistics 86*, 586–601.

Behrman, J. R., M. R. Rosenzweig, and P. Taubman (1994). Endowments and the allocation of schooling in the family and in the marriage market: The twins experiment. *Journal of Political Economy 102*(6), 1131–1174.

Behrman, J. R., Y. Xiong, and J. Zhang (2015). Cross-sectional schooling-health associations misrepresented causal schooling effects on adult health and health-related behaviors: Evidence from the Chinese Adults Twins Survey. *Social Science & Medicine 127*, 190–197.

Bhalotra, S. and D. Clarke (2019). Twin birth and maternal condition. *The Review of Economics and Statistics*. Accepted for publication.

Bharadwaj, P., J. P. Eberhard, and C. A. Neilson (2018). Health at birth, parental investments, and academic outcomes. *Journal of Labor Economics 36*(2), 349–394.

Bharadwaj, P., K. V. Loken, and C. Neilson (2013). Early life health interventions and academic achievement. *The American Economic Review 103*(5), 1862–1891.

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

Black, S. E., P. J. Devereux, and K. G. Salvanes (2010). Small family, smart family? Family size and the IQ scores of young men. *Journal of Human Resources 45*(1), 33–58.

Bockerman, P. and T. Maczulskij (2016). The education-health nexus: Fact and fiction. *Social Science & Medicine*.

Bolyard, A. and P. A. Savelyev (2021). Understanding the educational attainment polygenic score and its interactions with SES in determining health in young adulthood. HCEO Working Paper 2021-026.

Borga, L. G. and M. Pidkuyko (2018). Whoever has will be given more: Child endowment and human capital investment. *CERGE-EI Working Paper Series No. 616*.

Brunello, G., D. Fabbri, and M. Fort (2013). The causal effect of education on body mass: Evidence from Europe. *Journal of Labor Economics*.

Brunello, G., M. Fort, N. Schneeweis, and R. Winter-Ebmer (2015). The causal effect of education on health: What is the role of health behaviors. *Health Economics*, 314–336.

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

Case, A., A. Fertig, and C. Paxson (2005, March). The lasting impact of childhood health and circumstance. *Journal of Health Economics 24*(2), 365–389.

Chiappori, P.-A., M. Iyigun, and Y. Weiss (2009, December). Investment in schooling and the marriage market. *The American Economic Review 99*(5), 1689–1713.

Clark, D. and H. Royer (2013). The effect of education on adult mortality and health: Evidence from Britain. *American Economic Review 103*(6), 2087–2120.

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

Conti, G., J. J. Heckman, and R. Pinto (2016). The effects of two influential early childhood interventions on health and healthy behaviour. *The Economic Journal 126*(596), F28–F65.

Contoyannis, P. and A. M. Jones (2004). Socio-economic status, health, and lifestyle. *Journal of Health Economics 23*, 965–995.

Cutler, D. M. and A. Lleras-Muney (2008). Education and health: Evaluating theories and evidence. In R. F. Schoeni, J. S. House, G. A. Kaplan, and H. Pollack (Eds.), *Making Americans Healthier: Social and Economic Policy as Health Policy*, pp. 29–60. New York: Russell Sage Foundation.

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

Darden, M., D. B. Gilleskie, and K. Strumpf (2018). Smoking and mortality: New evidence from a long panel. *International Economic Review 59*(3), 1571–1619.

Datar, A., M. R. Kilburn, and D. S. Loughran (2010). Endowments and parental investments in infancy and early childhood. *Demography 47*(1), 145–162.

Del Bono, E., J. Ermisch, and M. Francesconi (2012). Intrafamily resource allocations: a dynamic structural model of birth weight. *Journal of Labor Economics 30*(3), 657–706.

Ding, W., S. F. Lehrer, J. N. Rosenquist, and J. Audrain-McGovern (2009, May). The impact of poor health on education: New evidence using genetic markers. *Journal of Health Economics 28*(3), 578–597.

Frijters, P., D. W. Johnston, M. Shah, and M. A. Shields (2013). Intrahousehold resource allocation: Do parents reduce or reinforce child ability gaps? *Demography 50*(6), 2187–2208.

Galama, T. J., A. Lleras-Muney, and H. van Kippersluis (2018, September). The effect of education on health and mortality: A review of experimental and quasi-experimental evidence. *The Oxford Research Encyclopedia, Economics and Finance (oxfordre.com/economics)*, 1–96.

Grabner, M. (2009). The causal effect of education on obesity: Evidence from compulsory schooling laws. Working paper.

Griliches, Z. (1979). Sibling models and data in economics: Beginnings of a survey. *The Journal of Political Economy 87*, S37–S64.

Grossman, M. (2000). The human capital model. In A. J. Culyer and J. P. Newhouse (Eds.), *Handbook of Health Economics*, Volume 1, pp. 347–408. Amsterdam: Elsevier.

Grossman, M. (2015). The relationship between health and schooling: What's new? *Nordic Journal of Health Economics 3*(1), 7–17.

Halla, M. and M. Zweimüller (2014). Parental response to early human capital shocks: Evidence from the Chernobyl accident. *IZA Discussion Paper No. 7968*.

Hauser, T. H. and K. K. L. Ho (2001). Accuracy of on-line databases in determining vital status. *Journal of Clinical Epidemiology 54*, 1267–1270.

Heckman, J. J. (1979, January). Sample selection bias as a specification error. *Econometrica 47*(1), 153–162.

Heckman, J. J., J. E. Humphries, and G. Veramendi (2018). The causal effects of education on earnings, health, and smoking. *Journal of Political Economy 126*(S1), S197–246.

Heckman, J. J. and T. E. MaCurdy (1980, January). A life cycle model of female labour supply. *Review of Economic Studies 47*(1), 47–74.

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

Hong, K., P. A. Savelyev, and K. Tan (2000). Understanding the mechanisms linking education with longevity. *Journal of Human Capital 14*(3), 371–400.

Hsin, A. (2012). Is biology destiny? Birth weight and differential parental treatment. *Demography 49*(4), 1385–1405.

James, J. (2015). Health and education expansion. *Economics of Education Review*.

Kelly, E. (2011). The scourge of asian flu: In utero exposure to pandemic influenza and the development of a cohort of British children. *The Journal of Human Resources 46*(4), 669–694.

Kemptner, D., H. Jurges, and S. Reinhold (2011). Changes in compulsory schooling and the causal effect of education on health: Evidence from Germany. *Journal of Health Economics*.

Kenkel, D. S. (2000). Prevention. In A. J. Culyer and J. P. Newhouse (Eds.), *Handbook of Health Economics* (1 ed.), Volume 1, Chapter 31, pp. 1675–1720. Elsevier.

Kohler, H.-P., J. R. Behrman, and J. Schnittker (2011). Social science methods for twins data: Intergrating causality, endowments, and heritability. *Biodemography and Social Biology* (57(1)), 88–141.

Krueger, R. F. and W. Johnson (2002). The Minnesota Twin Registry: Current status and future directions. *Twin Research 5*, 488–492.

Lehrer, S. F. and J. M. Fletcher (2009). The effects of adolescent health on educational outcomes: Causal evidence using 'genetic lotteries' between siblings. *Forum for Health Economics & Policy 12*(2).

Lehrer, S. F. and J. M. Fletcher (2011). Genetic lotteries within families. *Journal of Health Economics 30*(4), 647–659.

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

Lundborg, P. (2013). The health returns to schooling—what can we learn from twins? *Journal of Population Economics* (26), 673–701.

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

Lykken, D., T. J. Bouchard, M. McGue, and A. Tellegen (1990). The Minnesota Twin Family Registry: Some initial findings. *Acta Genet Med Gemellol 39*, 35–70.

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

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

McGovern, M. E. (2019). How much does birth weight matter for child health in developing countries? Estimates from siblings and twins. *Health Economics 28*, 3–22.

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

Moffitt, R. (1999). New developments in econometric methods for labor market analysis. In O. C. Ashenfelter and R. Layard (Eds.), *Handbook of of Labor Economics*, Volume 3, pp. 1367–1397. Amsterdam: Elsevier.

Nicoletti, C., K. G. Salvanes, and E. Tominey (2017). Response of parental investments to child's health endowment at birth. Working Paper.

Nicoletti, C. and V. Tonei (2017). The response of parental time investments to the child's skills and health. *University of York: Discussion Papers in Economics*.

O'Conner, T., M. Rutter, C. Beckett, L. Keaveney, and J. Kreppner (2000). The effects of global severe privation on cognitive competence: Extension and longitudinal follow-up. *Child Development 71*(2), 376–390.

Parman, J. (2015). Childhood health and sibling outcomes: Nurture reinforcing nature during the 1918 influenza pandemic. *Explorations in Economic History 58*, 22–43.

Pitt, M. M. and M. R. Rosenzweig (1990). Estimating the intrahousehold incidence of illness: Child health and gender-inequality in the allocation of time. *International Economic Review 31*(4), 969–989.

Pitt, M. M., M. R. Rosenzweig, and M. N. Hassan (1990). Productivity, health, and inequality in the intrahousehold distribution of food in low-income countries. *The American Economic Review*, 1139–1156.

Restrepo, B. J. (2016). Parental investment responses to a low birth weight outcome: Who compensates and who reinforces? *Journal of Population Economics 29*, 969–989.

Rosales-Rueda, M. F. (2014). Family investment responses to childhood health conditions: Intrafamily allocation of resources. *Journal of Health Economics 37*, 41–57.

Rosenzweig, M. R. and J. Zhang (2009). Do population control policies induce more human capital investment? Twins, birth weight and China's "one-child" policy. *The Review of Economic Studies 76*(3), 1149–1174.

Royer, H. (2009). Separated at birth: US twin estimates of the effects of birth weight. *American Economic Journal: Applied Economics 1*, 49–85.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley.

Saavedra, M. and T. Twinam (2020). A machine learning approach to improving occupational income scores. *Explorations in Economic History 75*. Article 101304.

Sanz-de Galdeano, A. and A. Terskaya (2019). Parental human capital investment responses to childen's disability. Universidad de Alicante Working Paper.

Sassi, F., M. Devaux, M. Cecchini, and E. Rusticelli (2009). The obesity epidemic: Analysis of past and projected future trends in selected OECD countries. OECD working paper.

Savelyev, P. A. (2020). Conscientiousness, Extraversion, college education, and longevity of high-ability individuals. *Journal of Human Resources*. doi: 0918-9720R2.

Savelyev, P. A. and K. Tan (2019). Socioemotional skills, education and health-related outcomes of high-ability individuals. *American Journal of Health Economics 5*(1), 250–280.

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. Chapman & Hall/CRC.

Stock, J. H., J. H. Wright, and M. Yogo (2002, October). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics 20*(4).

Strauss, J. and D. Thomas (1998, June). Health, nutrition, and economic development. *Journal of Economic Literature 36*(2), 766–817.

Terskaya, A. (2019). Parental human capital investment responses to childen's disability. Universidad de Alicante Working Paper.

US Census Bureau (1960). Decennial census. https://www.census.gov/prod/www/decennial.html.

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

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

Venkataramani, A. (2012). Early life exposure to malaria and cognition in adulthood: Evidence from Mexico. *Journal of Health Economics 31*, 767–780.

Webbink, D., N. Martin, and P. Visscher (2010). Does education reduce the probability of being overweight? *Journal of Health Economics*.

Wentworth, D. N., J. D. Neaton, and W. L. Rasmussen (1983). An evaluation of the Social Security Administration Master Beneficiary Record File and the National Death Index in the ascertainment of vital status. *American Journal of Public Health 73*, 1270–1274.

Wilson, S. (2012). Marriage, gender, and obesity later in life. *Economics and Human Biology*.

Yi, J. (2019). Endogenous altruism: Theory and evidence from Chinese twins. *Journal of Labor Economics 37*(1), 247–295.

Yi, J., J. J. Heckman, J. Zhang, and G. Conti (2015, November). Early health shocks, intra-household resource allocation and child outcomes. *The Economic Journal 125*, F347–F371.

[FOR ONLINE PUBLICATION]

Web Appendix to "Health Endowments, Schooling Allocation in the Family, and Longevity: Evidence from US Twins"

Peter A. Savelyev* Benjamin C. Ward†

Robert F. Krueger‡ Matt McGue§

July 20, 2021

*The corresponding author and an Assistant Professor of Economics at The College of William & Mary. 300 James Blair Dr., Chancellor's Hall, Room 317, Williamsburg, VA 23185. E-mail: pasavelyev@wm.edu.

†Lecturer in the Department of Economics at The University of Georgia.

‡Hathaway Distinguished Professor of Clinical Psychology in the Department of Psychology at The University of Minnesota, Twin Cities.

§Regents Professor of Psychology in the Department of Psychology at The University of Minnesota, Twin Cities.

A Appendix

Table A-1: Marginal Coefficients for the Association Between Absolute Differences in Schooling within Twin Pairs and the Attrition of Pairs, Logit and Probit Models

	Logit	Probit
Marginal association	0.00147	0.00143
Standard error	*(0.0050)*	*(0.0050)*
p-value	0.777	0.776
Sample size	3327	3327

Notes: Huber-White standard errors are shown in parentheses. Absolute differences in the total years of schooling are based on MTR data.

Table A-2: Conditional Associations between Education and Selected Health-Related Outcomes, A Comparison of Marginal Coefficients Between LPM, Logit, and Probit Modles

		Males			Females		
		LPM	Logit	Probit	LPM	Logit	Probit
Mortality	coefficient	-0.0078 *	-0.0076 *	-0.0075 *	-0.0044 *	-0.0042 *	-0.0044 *
	std. error	*(0.0052)*	*(0.0050)*	*(0.0050)*	*(0.0031)*	*(0.0028)*	*(0.0030)*
Alcohol Problems	coefficient	-0.0069	-0.0067 *	-0.0067 *	0.0005	0.0006	0.0006
	std. error	*(0.0043)*	*(0.0039)*	*(0.0040)*	*(0.0017)*	*(0.0017)*	*(0.0017)*
Overweight	coefficient	-0.0270 ***	-0.0270 ***	-0.0272 ***	-0.0330 ***	-0.0340 ***	-0.0336 ***
	std. error	*(0.0092)*	*(0.0093)*	*(0.0094)*	*(0.0087)*	*(0.0093)*	*(0.0091)*

Notes: Asterisks represent statistical significance levels: *** 1%, ** 5%, * 10%. Estimation based on pooled zygosities (both MZ and DZ twins). Observation is the individual twin. Robust Huber-White standard errors clustered by family/twin-pair. One-tailed tests are used for mortality and overweight, two-tailed for alcohol problems. All regressions control for birthweight and injuries.

Table A-3: Effects of Education on Mortality and Health

		Pooled sexes	Males	Females	Sex difference
Mortality	coefficient	-0.023	-0.031	-0.016	-0.015
	bootstrap std. error	*(0.012)* **	*(0.018)* **	*(0.015)*	*(0.024)*
	robust asymptotic std. error	*(0.012)* **	*(0.018)* **	*(0.015)*	*(0.023)*
	classical std. error	*(0.011)* **	*(0.017)* **	*(0.014)*	*(0.022)*
	# of twin pairs	558	204	354	
Physical health problems	coefficient	-0.028	-0.021	-0.033	0.011
	bootstrap std. error	*(0.019)* *	*(0.028)*	*(0.025)* *	*(0.037)*
	robust asymptotic std. error	*(0.019)* *	*(0.027)*	*(0.025)* *	*(0.037)*
	classical std. error	*(0.018)* *	*(0.026)*	*(0.026)*	*(0.036)*
	# of twin pairs	694	244	450	
Self-reported health[(a)]	coefficient	-0.044	-0.061	-0.022	-0.039
	bootstrap std. error	*(0.033)* *	*(0.049)*	*(0.046)*	*(0.068)*
	robust asymptotic std. error	*(0.034)* *	*(0.049)*	*(0.047)*	*(0.068)*
	classical std. error	*(0.035)*	*(0.050)*	*(0.049)*	*(0.070)*
	# of twin pairs	680	241	439	

Notes: Results of the 2SLS model reported. Standard errors are shown in parentheses. One-tailed *p*-values reported for estimated coefficients on years of education, two-tailed for difference in coefficient estimates across sexes. Asterisks represent statistical significance levels: *** 1%, ** 5%, * 10%. [(a)]In order that the signs on our regression coefficient estimates can be interpreted consistently across all outcomes, we reversed the numerical mapping of this scale, such that bad = 5 and excellent = 1.

Table A-4: Suggestive Mechanisms of the Effect of Education on Health and Longevity

		Pooled sexes	Males	Females	Sex difference
Overweight[(a)]	coefficient	-0.010	-0.044	0.024	-0.069
	bootstrap std. error	*(0.014)*	*(0.024)* **	*(0.017)*	*(0.029)* **
	robust asymptotic std. error	*(0.014)*	*(0.022)* **	*(0.017)*	*(0.027)* **
	classical std. error	*(0.014)*	*(0.021)* **	*(0.019)*	*(0.029)* **
	# of twin pairs	670	240	430	
BMI	coefficient	-0.030	-0.199	0.127	-0.326
	bootstrap std. error	*(0.113)*	*(0.150)* *	*(0.165)*	*(0.223)*
	robust asymptotic std. error	*(0.113)*	*(0.145)* *	*(0.166)*	*(0.220)*
	classical std. error	*(0.107)*	*(0.123)* *	*(0.165)*	*(0.205)*
	# of twin pairs	670	240	430	
Alcohol[(b)]	coefficient	-0.012	-0.016	-0.009	-0.007
	bootstrap std. error	*(0.008)*	*(0.013)*	*(0.009)*	*(0.015)*
	robust asymptotic std. error	*(0.008)*	*(0.013)*	*(0.008)*	*(0.015)*
	classical std. error	*(0.006)* *	*(0.012)*	*(0.007)*	*(0.014)*
	# of twin pairs	694	244	450	

Notes: Effects of education on suggestive mechanisms reported based on the 2SLS model. Standard errors are shown in parentheses. One-tailed tests used for overweight and BMI, and two-tailed for sex differences and for alcohol (there is no consensus in the health economics literature regarding the sign of the relationship between educational attainment and alcohol use). Asterisks represent statistical significance levels: *** 1%, ** 5%, * 10%. [(a)]Refers to $BMI > 25$. [(b)]Refers to "family, job, or health problems due to alcohol use."

Table A-5: Difference in Education Coefficients between DZ and MZ Twins, $(\beta_1^D - \beta_1^M)$

		Pooled sexes	Males	Females
Mortality	coefficient	0.026	0.037	0.019
	bootstrap std. error	*(0.015)* *	*(0.022)* *	*(0.018)*
	robust asymptotic std. error	*(0.014)* *	*(0.022)* *	*(0.018)*
	classical std. error	*(0.014)* *	*(0.022)* *	*(0.018)*
Physical health problems	coefficient	0.065	0.059	0.070
	bootstrap std. error	*(0.024)* ***	*(0.035)* *	*(0.032)* **
	robust asymptotic std. error	*(0.024)* ***	*(0.035)* *	*(0.032)* **
	classical std. error	*(0.024)* ***	*(0.036)* *	*(0.032)* **
Self-reported health[(a)]	coefficient	-0.028	-0.027	-0.044
	bootstrap std. error	*(0.044)*	*(0.064)*	*(0.062)*
	robust asymptotic std. error	*(0.045)*	*(0.065)*	*(0.062)*
	classical std. error	*(0.046)*	*(0.071)*	*(0.063)*

Notes: Difference in results of the 2SLS coefficiants reported. Standard errors are shown in parentheses. Asterisks represent statistical significance levels: *** 1%, ** 5%, * 10%. [(a)]In order that the signs on our regression coefficient estimates can be interpreted consistently across all outcomes, we reversed the numerical mapping of this scale, such that bad = 5 and excellent = 1.

Table A-6: Effects of Education on Health Outcomes, Omitting Background Controls

		With controls	With no controls
Mortality	coefficient	-0.023 **	-0.021 **
	standard error	*(0.012)*	*(0.012)*
	# of twin pairs	558	558
Physical Health Problems	coefficient	-0.028 *	-0.028 *
	standard error	*(0.019)*	*(0.019)*
	# of twin pairs	694	694
Self-Reported Health[(a)]	coefficient	-0.044 *	-0.043 *
	standard error	*(0.034)*	*(0.033)*
	# of twin pairs	680	680

Notes: 2SLS estimates reported. Results are of the second stage of the 2SLS model reported. Results from pooled sample of male and female MZ twins. Huber-White standard errors are shown in parentheses. One-tailed tests are used for coefficients on years of education. Asterisks represent statistical significance levels: *** 1%, ** 5%, * 10%. [(a)]In order that the signs on our regression coefficient estimates can be interpreted consistently across all outcomes, we reversed the numerical mapping of this scale, such that bad = 5 and excellent = 1.