

Using Genetic Lotteries within Families to Examine the Causal Impact of Poor Health on Academic Achievement^{*†}

Jason M. Fletcher
Yale University
jason.fletcher@yale.edu

Steven F. Lehrer
Queen's University and NBER
lehrers@queensu.ca

February 2008

Abstract

One of the most robust relationships in the social sciences is the large positive correlation between health and education, but establishing a causal link remains a substantial challenge. This paper exploits differences in genetic inheritance among children within the same family to estimate the impact of several poor health conditions on academic outcomes. We present evidence of large impacts of poor mental health on academic achievement. Our estimates suggest that accounting for family fixed effects is important but these strategies cannot fully account for the endogeneity of poor health. Finally, our results demonstrate that the presence of comorbid conditions presents immense challenges for empirical studies that aim to estimate the impact of specific health conditions.

^{*}We are grateful to Weili Ding, Robert McMillan, John Mullahy, Jody Sindelar and participants at the 2007 NBER Summer Institute, Yale Health Policy Colloquium, University of British Columbia, University of Connecticut, University of Tennessee, University of Toronto and Simon Fraser University for comments and suggestions that have improved this paper. We are both grateful to the CLSRN for research support. Lehrer also wishes to thank SSHRC for additional research support. We are responsible for all errors.

[†]This research uses data from Add Health, a program project designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris, and funded by a grant P01-HD31921 from the National Institute of Child Health and Human Development, with cooperative funding from 17 other agencies. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Persons interested in obtaining data files from Add Health should contact Add Health, Carolina Population Center, 123 W. Franklin Street, Chapel Hill, NC 27516-2524 (addhealth@unc.edu).

1 Introduction

One of the most controversial debates in academic circles concerns the relative importance of an individual's innate qualities ("nature") versus environmental factors ("nurture") in determining individual differences in physical and behavioral traits.¹ For many years, "nature" was a black box, forcing researchers to examine the relative importance of a multitude of environmental factors on various individual outcomes. Yet, with the decoding of the human genome, it is possible to enter this "black box," and recent years have been characterized by substantial amounts of research examining whether specific variants in genetic code (aka single nucleotide polymorphisms (SNPs)) between dizygotic twins (among other family-based samples) are associated with specific diseases and outcomes. Findings from these studies have not only led to new drug discoveries but also improved diagnostic tools, therapies, and preventive strategies for a number of complex medical conditions.² As clinical researchers identify unique genetic bases for many complex health behaviors, diseases and other outcomes, opportunities arise for social scientists to exploit this knowledge and use differences in specific sets of genetic information to gain new insights into a variety of questions.³

¹This debate has been traced back to 13th century France. The relative importance of nature and nurture is of particular relevance for public policy. For example, consider education policy: if nurture factors drives the success of children in school, inequality in educational opportunity may well come from sources such as failing capital markets suggesting that specific policies could reduce future inequalities in schooling. However, if inequality in educational opportunity reflects the distribution of innate ability among the population, there is fewer opportunities to design policies that can reduce future inequality. That being said the notion that nurture inputs are more easily susceptible to policy remediation relative to nature, is a non sequitur.

²For example, see Johnson (2003), Kelada et al. (2003), Goldstein et al. (2003), Zerhouni (2003) and Merikangas and Risch (2003).

³Ding et al. [2006] was the first empirical study within economics to explicitly use differences in genetic information across individuals as an instrumental variable in estimating the effects of poor health on high school grade point average (GPA). More recently, Norton and Han [2007] use genetic information to attempt to estimate the impact of obesity on employment. Neither study used variation within families (the "genetic lottery"), which we show to be

In this paper, we exploit differences in genetic inheritance among children within the same family to estimate the impact of several poor health conditions on academic outcomes via an instrumental variables strategy.⁴ Understanding the consequences of growing up in poor health for adolescent development has presented serious challenges to empirical researchers due to endogeneity that arises from both omitted variables and measurement error problems pertaining to health.⁵ Empirical research that has attempted to estimate a causal link have either used a within-family strategy (i.e. Currie and Stabile [2006], Fletcher and Wolfe [2007a,2007b], and Fletcher [2007]) or instrumental variables approach (i.e. Ding et al. [2006,2007], Behrman and Lavy [1998], Norton and Han [2007] as well as Glewwe and Jacoby [1995]) and in general researchers find large negative impacts of poor health on academic outcomes.⁶ Our empirical strategy combines both elements and identifies the causal impact of health on education by exploiting exogenous variation in genetic inheritance

important empirically and improve the plausibility of the exclusion restriction. Ding and Lehrer [2007] provide a summary and history of the interdisciplinary literature that describes how such genetic differences can be used as a source of identification for a multitude of traits and behaviors. Related, but distinct from this literature, a number of studies such as Cawley(2004) have used family background information on phenotypes to proxy for actual genetic endowments contained in an individual's DNA.

⁴Genes consist of two alleles, and a child randomly inherits one of the two alleles from each parent at the time of conception. Since alleles could differ by the particular building blocks (base pairs) that make up DNA, any difference in the coding of a specific marker between full siblings presents an experiment in "nature".

⁵Grossman and Kaestner [1997] and Strauss and Thomas [1998] present surveys of the literature of the impact of health on, respectively, education and income. The majority of empirical studies discussed in the surveys report correlational relationships.

⁶Two other studies that use alternative empirical approaches are worth noting. Kremer and Miquel [2004] randomly assign health treatments to primary schools in Kenya and find that health improvements from the clinical treatment significantly reduced school absenteeism but did not yield any gains in academic performance. Bleakley (2007) uses a quasi-experimental strategy that exploits different timing at which cohorts were exposed to a large scale public health intervention against hookworm in childhood. He finds that the treatment boosted health, was associated with larger gains in income and higher rates of return to schooling later in life.

among both siblings and dizygotic twins.

In particular, our empirical strategy permits us to elucidate the extent to which developmental differences between twins and siblings are due to differences in genetic inheritance. These differences occur at conception and remain fixed between family members at every point in time irrespective of all nurture investments. Since a great deal of variation in characteristics and outcomes is found within families, exploiting the genetic processes that affect development (but are not self-selected by the individuals' themselves) presents a potential strategy to identify differences within families. Since nearly every social, behavioral and health outcome has a unique genetic basis, this identification strategy can potentially shed light on a large number of questions.⁷

Within economics, the use of twin samples and twin fixed effects regressions to address questions related to differences in individual characteristics on outcomes has a long history.⁸ This method allows the researcher to simultaneously control (assuming constant impacts between twins) for both many common genetic factors and parental characteristics/behaviors, but does not provide any guidance as to why, within a twin pair, the subjects differed in explanatory characteristics and outcomes. Since even monozygotic twins are often discordant for many health conditions, it is important to state that the health outcomes that we consider in this paper are likely to be influenced by both multiple genetic factors as well as the environment an individual encounters throughout her life (as well as possible gene-environment interactions).⁹ However, only the genetic factors are

⁷These ideas are far from new as discussed in Harrison (1970) and Allen (1970).

⁸Similar researchers have undertaken a related empirical strategy with samples that only consist of siblings by estimating models with family fixed effects. The earliest attempt to look at siblings data in economics can be traced back to the dissertation of Gorseline [1932] who in the conclusion noted that twins may be a more desirable sample. Behrman and Taubman [1976], Taubman ([1976a], [1976b]), and Behrman et al. [1977] appear to be the first studies in economics to use data on twins.

⁹More recently evidence indicates that differences within families even among identical twins can exist because of epigenetic factors. Epigenetics refers to natural chemical modifications that occur in a person's genome shortly after conception and that act on a gene like a gas pedal or a brake, marking it for higher or lower activity. For instance,

acquired prior to any experience an individual faces (even in utero), and the specific set we use in our analysis have been demonstrated in the genetics literature to be predisposition genes with pleiotropic effects.¹⁰

Our empirical analysis reaches three major conclusions. First, we find that the impact of poor mental health outcomes on academic achievement is substantial. Inattention and depression both lead to decreases in academic performance. The significant negative impacts of inattention as well as ADHD on academic performance even appear significant if we examine the relationship only using a subsample of same-sex twins. In contrast, the estimated impact for our physical health measures of being overweight (or obese) rarely enters in a significant manner across samples and empirical strategies when we control for poor mental health in the specification. Second, we conduct a battery of tests that confirm that our set of genetic markers indeed have desirable properties to serve as an instrumental variable for our health outcomes. The individual markers have statistically significant correlations with each endogenous health variables and, consistent with Mendel's hypothesis that the hereditary factors for different genes are independent, statistical tests demonstrate that these markers are not related to each other and affect academic performance through health outcomes. Third, our results indicate that measuring health is a substantial challenge facing empirical researchers interested in identifying the impacts of specific disorders. The presence of comorbid conditions presents difficulties as one requires a source of exogenous variation that could identify a specific condition in the presence of unmeasured comorbid poor health conditions. While these findings are similar to Ding et al. [2006, 2007], our analysis also indicates that accounting for identical twins have different fingerprints. The general pattern of their fingerprints is determined by genetic factors and is initially identical, however the exact pattern changes in utero based on when and how each twin touched the amniotic sac (Jain et al. 2002).

¹⁰Pleiotropy refers to the heterogeneous impacts that a difference in specific genetic marker occurs. Intuitively the operation is similar to a "power grid", as a single-gene mutation also affects the expression of many other genes which together leads to changes in behaviors and outcomes.

family-specific unobserved heterogeneity is statistically important and that ignoring this factor on the one hand leads to estimates of the impacts of depression and hyperactivity that are over 50% smaller in magnitude but on the other hand leads to substantially large impacts of inattention.

The rest of the paper is organized as follows. In Section II, we provide an overview of the data we employ in the study. We also review the scientific literature linking the genes in our data set to health behaviors and health outcomes. The empirical framework that guides our investigation and our identification strategy is described in Section III. The empirical results are presented and discussed in Section IV. A concluding section summarizes our findings and discusses directions for future research.

2 Data

This project makes use of the National Longitudinal Study of Adolescent Health (Add Health), a nationally representative longitudinal data set.¹¹ The dataset was initially designed as a school-based study of the health-related behaviors of 12 to 18 year old adolescents who were in grades 7 to 12 in 1994/5. A large number of these adolescents have subsequently been followed and interviewed two additional times in both 1995/6, and 2001/2. Our project makes use of a specific subsample of the respondents that permit us to develop our identification strategy. Specifically, we analyze data for the sample for which DNA measures were collected during the 2001/2 interview and for which there were multiple family members in the survey. This specific subsample is composed of monozygotic twins, dizygotic twins, and full biological siblings and includes information on 2,101,

¹¹Add Health selected schools in 80 communities that were stratified by region, urbanicity, school type (public, private, or parochial), ethnic mix, and size. In each community, a high school was initially selected but since not all high schools span grades 7-12, a feeder school (typically a middle school) was subsequently identified and recruited. In total, there are 132 schools in the sample and additional details on the construction of the sample are provided in Harris et al. [2003].

2,147, and 2,275 individuals who completed the survey at each interview point. Excluding those individuals for whom there are incomplete education, health and DNA measures for multiple family members reduces the sample to 1684 individuals.¹²

The data set contains information on a number of health conditions, including depression, ADHD and obesity. Depression is assessed using 19 responses to the Center for Epidemiologic Studies-Depression Scale (CES-D), a 20-item self-report measure of depressive symptoms. Items on the CES-D are rated along a 4-point Likert scale to indicate how frequently in the past week each symptom occurred (0 = never or rarely; 3 = very often). The sum of these items is calculated to provide a total score where higher scores indicate a greater degree of depressive symptoms. To determine whether an individual may be depressed, we followed findings from earlier research with adolescent samples (Roberts, Lewinsohn, and Seeley [1991]) and use specific age and gender cutoffs. We also use adult-based cutoffs to capture a broader measure of depressive symptoms in our analyses. The primary indicator of childhood ADHD symptoms is taken from an eighteen-question retrospective rating collected during the third data wave. Since there is evidence that the effects of ADHD may vary by whether the symptoms are of the inattentive or hyperactive type,¹³ we examine the effects of these different domains as well as usual measures of ADHD of any type. Finally, overweight and obesity are calculated from each individual's self-reported height and weight applied to age and gender specific definitions obtained from the Centers for Disease Control¹⁴.

While concerns may exist regarding the use of self-reports to construct indicators for health measures such as ADHD or obesity, we believe this is a limited concern for our study. Not only are we using an instrumental variables approach but past research with this data (Goodman et al. [2000]) also indicate that there is a strong correlation between measured and self-reported height (0.94),

¹²We do not use sampling weights in our tables or analysis.

¹³For example, Babinski et al. [1999], Ding et al. [2006] and Fletcher and Wolfe [2007a] present empirical evidence of different impacts from these two diagnoses.

¹⁴See <http://www.cdc.gov/growthcharts/> for details.

and between measured and self-reported weight (0.95) and there is no evidence that reporting errors are correlated with observed variables such as race, parental education, and household income.¹⁵

Regarding academic outcomes, the data contains information on GPA and a score on a common verbal test.¹⁶ The data also provides a rich set of information on environmental and demographic variables (i.e. family income, gender, parental education, family structure, etc.) that are used as control variables in our analysis. Finally, the restricted Add Health data allows community-level variables from the Census Bureau and school input variables from the NCES common core of data to be matched to the individuals in the data set to serve as additional controls.

Summary statistics on our sample are provided in Table 1. The first column contains the full sample where the second and third columns only contain the subsets of siblings and twins, respectively. The verbal test score for the sample and each of the subsamples approximates the national mean but the standard deviation is slightly smaller than those obtained on national samples.¹⁷ Household income is slightly higher than US averages and the majority of mothers have attended college. The twins and sibling samples are both almost equally composed of males and females. African Americans and Hispanics account for approximately one third of the sample. With the exception of race, there are few differences in any of the summary statistics between the full sample and the subsample of siblings and twins. While many of the education and demographic variables fall within national averages, the rates of poor mental health outcomes are slightly higher. While the

¹⁵Retrospective ratings of previous ADHD are also likely measured with error. Fortunately, several reviews have concluded that childhood experiences are recalled with sufficient accuracy to provide useful information in retrospective studies (e.g. Kessler et al. 2005).

¹⁶The test is an abridged version of the Peabody Picture Vocabulary Test-Revised and consists of 78 items. The test was administered at the beginning of the in-home interview and first involves the interviewer reading a word aloud. The respondent then selects among four, simple black-and-white illustrations arranged in a multiple-choice format the illustration that is the closest match to the word. The test scores are standardized by age and some psychologists interpret the scores as a measure of verbal IQ.

¹⁷See <http://www.agsnet.com/assessments/technical/ppvt.asp> for details.

AD and HD subscale averages fell within standard ranges for adolescent samples, roughly 8% of the sample is coded with ADHD, which exceeds the 6% national average. Conversely those adolescents classified as being depressed in our sample is lower than the 1999 estimate of the fraction of the adolescent population being clinically depressed (12.5%) from the U.S. Department of Health and Human Services. Finally, overweight rates fall slightly below the national average for this period. The overclassification of ADHD could result from measurement error, an issue we will investigate in our empirical analysis.

The well-known positive association between good health and educational outcomes is also observed in the data. As indicated in Table 2, individuals classified as depressed and obese, respectively, have on average verbal test scores that are 7 and 3 points lower than their counterparts. These differences are statistically significant (one sided t-tests). Yet there does not exist a highly significant test score gap between those who smoke or are classified with ADHD, AD and HD and their classmates. In fact, individuals with HD score higher than those who are not coded with this disorder. There are very few changes in the significance of these differences in test performance if we restrict the sample to twins or siblings.

The DNA samples were drawn in the third collection and were genotyped for six candidate polymorphisms.¹⁸ The specific markers that have been collected in this study were selected based upon a large and growing body of research showing a strong correlation between their variation and health outcomes such as obesity, ADHD, and depression, controlling for other relevant factors. The genetic markers collected in the Add Health study are primarily linked to the transmission of two

¹⁸Complete details of the sampling and laboratory procedures for DNA extraction, genetic typing and analysis are provided in an online document prepared by Add Health Biomarker Team available at <http://www.cpc.unc.edu/addhealth/files/biomark.pdf>/ Note, that the the method to genotype varies across markers and different assays were conducted. In addition to reduce coding errors, genotypes were scored independently by two individuals. To control for potential genotyping errors, any analysis that is questionable for routine problems (i.e. poor amplification, gel quality, software problems, etc.) is repeated.

specific neurotransmitters in the primitive limbic system of the brain: dopamine and serotonin.¹⁹

The initially targeted candidates are the dopamine transporter (DAT), the dopamine D4 receptor (DRD4), the serotonin transporter (5HTT), monoamine oxidase A (MAOA), the dopamine D2 receptor (DRD2) and the cytochrome P4502A6 (CYP2A6) gene. Variants in the coding of these genes, not the genes themselves, are believed to impact multiple health outcomes and behaviors. Variants in the DNA base sequence are called single nucleotide polymorphisms (SNPs) and scientists hypothesize that the SNPs distort cell functions and/or processes, which leads to the higher propensities for specific disorders.

The scientific hypothesis of how these markers predispose individuals to poor health is that these genetic markers each impact the synaptic level of dopamine and serotonin, which provides larger signals of pleasure from the limbic system and leads individuals to forego other basic activities.²⁰ The specific markers are believed to achieve these impacts as follows: Individuals with the A1 allele variants of the DRD2 gene have fewer dopamine D2 receptors than those with the A2 allele, thereby requiring larger consumption of substances to achieve the same level of pleasure. The DAT and 5HTT genes code for proteins that lead to the reuptake of dopamine and serotonin respectively. For each of these genes, longer lengths are believed to affect the speed at which production of these proteins occur. The MAOA gene product is primarily responsible for the degradation of dopamine, serotonin, and norepinephrine in several regions of the brain. A SNP of this gene is

¹⁹The effect of a neurotransmitter comes about by its binding with receptor proteins on the membrane of the postsynaptic neuron. As long as the neurotransmitter remains in the synapse, they continue to bind its receptors and stimulate the postsynaptic neuron. In the brain, dopamine and serotonin function as a neurotransmitter as they are commonly believed to provide individuals feelings of enjoyment.

²⁰The limbic system is highly interconnected with the region of the brain associated with reward and pleasure. This region was initially discovered in Olds and Milner [1954] who reported that if given the choice of food versus stimulation by electrodes of the neurons within this region of the brain, rodents ended up dying from starvation and exhaustion, rather than lessening the stimulation of their pleasure center. Recent studies using mice whose genes have been mutated to affect dopamine and serotonin production have confirmed that these markers affect basic activities.

believed to have decreased productivity of this protein, thereby increasing the risk of a number of poor outcomes. Individuals with a longer version of the DRD4 gene are more inclined to partake in additional novelty or sensation-seeking activities to achieve similar levels of reward as those with shorter variants. Finally, the CYP2A6 gene is primarily located in the liver and affects the rate of metabolism for tobacco, drugs and other toxins. Once these compounds are broken down, they travel in the blood stream to the brain where they generally lead to neurotransmitters being released.

SNPs of these genes may independently affect the propensity to develop a poor health outcome, but gene-gene interactions can also have potentially powerful effects. For example, Dremencov et al. [2004] present evidence that the SNPs of the 5HTT gene interacts with genes that release dopamine and suggest this channel could impact the speed at which certain pharmaceutical treatments become effective. Similarly since many addictions stimulate dopamine release in the nucleus accumbens, it is likely that the rate of metabolism of these drugs (which is in part determined by the CYP2A6 gene) interacts with the DRD2 genes. Thus, in our analysis we will not only consider the SNPs by themselves but also their interactions. The genetic characteristics of our sample and unconditional relationships with poor health outcomes are discussed in the results section of the paper.

3 Empirical Framework

Both the well-known association between health and education outcomes and the reasons for heterogeneity in health behaviors across individuals have been discussed extensively in the economics, psychology, and health sciences literatures. The empirical framework that underlies our analysis involves the estimation of a system of equations generated from a model developed in Ding et al. [2006].²¹ This system of equations includes both a health production function and education

²¹This model departs from earlier research that seeks to explain the association between health and education as it assumes that neither the adolescent nor her altruistic parent chooses by themselves all of the inputs that enter both

production function.

Consider a linear representation of the education production function which translates a set of inputs into human capital as measured by a score on an achievement test or report card as

$$A_{ifjT} = \beta_0 + \beta_1 X_{iT} + \beta_2 H_{iT} + \beta_3 Q_{jT} + \beta_4 N_{iT} + v_f + \varepsilon_{ifjT} \quad (1)$$

where A_{ifjT} is a measure of achievement for child i in family f , in school j in year t , the vector X contains individual and family characteristics (gender, race, parental characteristics), the vector H consists of variables that captures health measures, the vector Q contains school quality variables, the vector N contains information on community and neighborhood inputs, v_f is an unobserved family effect and ε_{ifjT} is an idiosyncratic error term.

The major challenge with estimating the causal effect of poor health from the above equation is that the health variables are likely to be endogenous.²² That is, individuals with a higher health "endowment" could obtain improved academic performance because of genetic characteristics or parental investments that are also unobserved to the analyst. By including family fixed effects (v_f) in the estimation we can directly account for unobserved to the researcher family factors that are common across siblings and may be related to both individual health and education outcomes. While using a family fixed effects strategy allows the researcher to simultaneously control (assuming constant impacts between family members) for many parental characteristics/behaviors and some

the health production function and an education production function. Rather, both the adolescent and her parent each can make a subset of decisions regarding inputs to maximize household indirect utility subject to a series of standard constraints. For example, the authors postulate that a teenager would make decisions such as whether or not to smoke or use narcotic substances, while their parents makes decisions related to which neighborhood to reside in, which school their child should be sent to, the type of health insurance to purchase and number of visits to health care providers. This distinction between health behaviors and health states is important empirically.

²²An equally important challenge occurs in measuring the health vector from omitted variables. If the researcher omits comorbid conditions one will recover biased estimates of the impacts of poor health on academic outcomes. This empirical challenge is discussed in detail in Section 4.4 of the text.

genetic factors, it does not provide any guidance as to why, within a twin or sibling pair, the subjects differ in explanatory characteristics and outcomes. That is, a fixed effects approach may overcome biases from correlations between the health vector and the family effect v_f , but it may not completely solve the endogeneity problem, as correlations may exist between the health variables and the error term (i.e. $Cov(H_{iT} - \bar{H}_f, \varepsilon_{ifjT} - \bar{\varepsilon}_f) \neq 0$).

To remove these additional correlations between the health variables and the idiosyncratic error term we use the method of instrumental variables to supplement the fixed effects strategy. Specifically, we use exogenous variation from the "genetic lottery" between family members to identify the impact of poor health on measures of achievement. That is, we estimate the linear representation of the health production function

$$H_{ifT} = \gamma_0 + \gamma_1 X_{iT} + \gamma_2 G_i^H + \gamma_3 Q_{jT} + \gamma_4 N_{iT} + v_f + v_{ifjT} \quad (2)$$

as the first stage regression in our fixed effects instrumental variables strategy, where G_i^H is a vector of genetic markers that may provide endowed predispositions to the current state of health status. Thus, in the first stage equation we explain differences in health outcomes between family members using differences in the coding of specific genetic markers between family members as an instrumental variable, while controlling for other individual and family characteristics that affect health and education outcomes.

In our analysis, we consider two different health vectors that consist of multiple health problems. The first health vector includes depression, overweight, and ADHD. The second health vector includes depression and overweight but decomposes ADHD into being inattentive (AD) or hyperactive / impulsive (HD). We make this distinction as ADHD is often denoted by AD/HD since, as defined in the American Psychiatric Association's Diagnostic and Statistical Manual, it encompasses the "Inattentive Type" marked by distractibility, difficulty following through on tasks as well as the "Hyperactive Type," which includes excessive talking, impulsivity and restlessness. It is not uncommon for people to be diagnosed with the "Combined Type," showing a history of both features,

but ex-ante we would imagine that inattention and hyperactivity could have different impacts on academic performance as well as other human capital outcomes.

In our empirical specifications, we control for child gender, race, parental education, birth order, family income and family structure. Ex ante, one could hypothesize that parental education and family income are positively associated with measures of academic performance. In genetic studies, controlling for ethnicity and race are important as it has been hypothesized that there are differences in allele frequencies across race and ethnic groups (e.g. Cooper et al. [2003]). Within families, birth order effects could exist as higher rank children are more likely to have older parents at birth, which could affect the amount of time invested by parents. Similarly, across families, higher rank children are more likely to be born into larger families, which can also capture family size effects. For robustness checks, we can include summary measures on a number of school quality variables and neighborhood controls that are matched to an individual based on zip code information (but these measures will be subsumed in the family fixed effect in many of our results).

Our identification relies on the assumption that the vectors of genetic markers that impact health outcomes (G_i^H) are unrelated to unobserved components (ε_{ifjT}) of the achievement equation. While there might not be any existing evidence that the markers considered in this study have any impact on the education production process, it remains possible. Additionally, our strategy is valid as long as this set of genetic markers only affects A_{ifjT} via the health outcomes we consider, and not through some other channel. Using multiple genetic instruments also allows the use of over-identification tests of the validity of our choice of instruments. Finally, an additional advantage of our identification strategy is that there are no concerns regarding reverse causality, as these genetic markers are assigned at conception, prior to any health outcome or selection of any parental choice input to the health production function (even in utero).

In our analysis, we consider estimation of the system of equations (1) and (2) via fixed effects instrumental variables methods but also consider OLS and family fixed effects estimation of equation

(1) as well as instrumental variables estimation of the system of equations described above. We use estimates from these alternative approaches to conduct specification tests that can shed light on the source of the endogeneity of health outcomes in education production functions. Finally, using these alternative methods will allow us to relate our findings with the existing literature that uses one (or two) of these three alternative estimation approaches.

4 Results

4.1 Genetic Associations

Our empirical identification relies on the validity of the ‘genetic lottery’ to serve as a source to identify the impact of adolescent health on education outcomes. Statistically, for the genetic markers to serve as instruments they must possess two properties. First, they must be correlated with the potentially endogenous health variables. Second, they must be unrelated to unobserved determinants of the achievement equation.

Prior to describing our instrument set and conducting formal tests, we present some summary information in our data that motivates the notion that these markers and their two-by-two polygenic interactions are good candidates to serve as instruments for adolescent health outcomes. Table 3 contains the conditional mean, standard deviation and odds ratio of alternative poor health outcomes for individuals that possess a particular marker. For each genetic marker, we use at most three discrete indicators that are defined by specific allelic combinations.²³

²³The DAT genotypes are classified with indicator variables for the number of 10-repeat alleles (zero, one, or two). The MAOA genotypes is classified with indicator variables for the number of 4-repeat alleles (zero, one, or two). Similarly, the DRD4 genotype is classified with indicator variables for the number of 7-repeat alleles (zero, one, or two). The DRD2 gene is classified as A1/A1, A1/A2 or A2/A2 where the A1 allele is believed to code for reduced density of D2 receptors. The SLC6A4 gene is classified as SS, SL or LL where S denotes short and L denotes long. A2/A2. Finally, we include indicator variables for the two possible variants of the CYP gene. We organize the genetic

For each poor health outcome and behavior, there is at least one gene in which a specific SNP exhibits a higher propensity. Statistically different odds ratios in Table 3 are denoted with an asterisk. For depression, individuals with the A2A2 allele of the DRD2 gene and two 7 repeats of the DRD4 gene have significantly lower odds. For ADHD, individuals with 2 four repeats of the MAOA gene have greater odds and individuals with 1 four repeat of the MAOA gene have lower odds. These relationships also show up for inattention (AD) and hyperactivity (HD). For obesity, those with no repeats of the DAT1 gene have substantially lower odds.

The significant correlations between the SNPs and the health outcomes are also consistent with the scientific hypotheses outlined in Section 2. Each of the health disorders we consider in this paper is believed to have a large genetic component and be polygenic.²⁴ To date, the scientific literature has not identified a unique depression, ADHD, or obesity gene. Concerns could exist that the genetic markers we use in our analysis are not only related to poor health in adolescence but also to genetic factors that directly impact education outcomes.²⁵ To examine this concern, we first present evidence that there are no direct links between the inheritance of the specific genetic markers in our study with other portions of the genetic codes, and second in our empirical analysis we use a procedure developed in Conley, Hansen and Rossi (2007) to examine the sensitivity of our estimates to the degree in which the exclusion restriction assumption is violated.

data reported in empirical table in order of the raw number of individuals who possess each particular marker within that gene from lowest frequency to most common.

²⁴Polygenic refers to a phenotype that is determined by multiple genes. For example, the ninth annual Human Obesity Gene Map released in 2006 identified more than 300 genes and regions of human chromosomes linked to obesity in humans. Several of the genetic markers contained in Add Health are listed but one should reasonably expect that they only account for a limited amount of variation in the health outcomes.

²⁵Plomin et al. [2006] and de Quervain and Papassotiropoulos [2006] present recent surveys on which genes are believed to be directly associated with intelligence and memory ability respectively. Researchers have found no direct links between several of the genes in this study and either intelligence (i.e. Moises et al. [2001]) or cognitive ability (e.g. Petrill et al. [1997]) and we hypothesize that the link operates through specific health measures.

Regarding whether the inheritance of different portions of the genetic code are correlated, we examine the extent to which genetic linkages occurs in our sample.²⁶ Table 4 presents cross-tabulations of different genetic combinations for both the full sample as well as by the first and second family member in the data. We constructed the sample of single family members based on relative age of the family members since one could expect linkages within families, but whether Mendel’s law of independent assortment is violated can only be tested across families. Each cell in Table 4 provides the raw count of people and conditional probability (based on possessing the gene given by the row variable) of possessing that specific genetic combination. Statistical analyses of odds ratios does not find evidence of statistical links at the 5% level or lower between the markers in this study for either sample that contains only one family member, lessening concerns regarding linkage.²⁷ All the cells in Table 4 are populated with multiple individuals, which indicates that the polygenic interactions can be identified across families.

To construct the instrument set, we only included genetic markers or their interactions that had statistically significant (at the 2% level) differences in the odds ratio of suffering from one of the four conditions.²⁸ It is unlikely that the majority of these unconditional relationships are due to chance

²⁶Examining whether genetic linkages occur is an active area of study as it presents a test of whether Mendel’s law of independent assortment is supported. This law suggests that different genes are inherited independently of each other, and scientists have essentially concluded that there is an independent assortment of chromosomes during meiosis but alleles that are in close proximity on the same chromosome may be inherited as a group. Studies finding small links in genetic assortment have been obtained from samples consisting only of family members. However, there appears to be evidence that different groups of alleles are transmitted together across families, when many of these studies and samples are examined jointly. Thus, violations are not systematic.

²⁷As discussed in the preceding footnote, this result is consistent with a large scientific literature.

²⁸Recall, Table 3 demonstrated that significant correlations indeed exist between health outcomes and the genetic markers in our data. To construct the instrument set, we considered two alternative strategies to construct the instrument set that are available from the authors. First, we followed Klepinger, Lundberg and Plotnick [1999], who used forward stepwise estimation to select a subset of these markers and their interactions. This implementation is identical to Ding et al. [2006] and this approach has the advantage in making it easier to replicate the study. The

and we also considered whether the direction of the odds ratio was biologically plausible. We do not vary our instrument set across samples so that any observed difference in terms of health effects is not the result of the selection of different instrument sets that vary based on genetic similarity between family members. It is worth repeating that these genes are pleiotropic and cannot credibly account for the majority of the variation in these health disorders. Thus, even if two siblings had the same markers for many of these six genes, this would neither guarantee that they suffer from the same disorders nor that these particular genes would affect the siblings in a similar fashion.

4.2 Estimates of the Empirical Model

We now examine whether poor health is related to academic outcomes in adolescence. Table 5 presents estimates of equation (1) for the full sample. In the odd columns results are presented for the first health vector which includes depression, overweight and ADHD. The even columns decompose the classification of ADHD into being inattentive (AD) or hyperactive / impulsive (HD) in the health vector. The first four columns of Table 5 presents OLS and family fixed effects which either assume that health is exogenous or that health is only correlated with the family-specific component of the residual. We find that depression is strongly negatively correlated with academic performance, however, the estimated magnitude diminishes by over 50% when family fixed effects

scientific literature provides some (arguably weak) guidance for selecting particular markers, as the evidence tends to be inconsistent across studies, which tend to use very small unrepresentative clinical samples. We examined the robustness of our results by using the complete set of the markers in our study. The general pattern of IV and fixed effects IV results are robust to the instrument set for the full sample. The first stage properties are particularly weak for the full set of markers and their two by two interactions, yet the partial R-squared for that instrument set is substantially larger than studies using dates of birth in the labor economics literature. Finally, at the request of a seminar participant, we considered 5 other strategies based on either stepwise regression using different criteria or retaining those markers with significant relationships at the 5% level. Again the pattern of results was fairly consistent.

are included in the specification. While the impacts of depression in the OLS specifications are fairly large relative to the other health variables, they remain approximately half of the estimated magnitude of the race variables. In addition to depression, the two other mental health conditions enter the equation in a significant manner. AD is strongly negatively correlated and HD is positively correlated with academic performance when family fixed effects are not included. Despite the evidence in Table 2 that overweight and obese students score significantly lower than non-overweight and non-obese students, this health state does not significantly affect verbal test scores in any of the specifications in Table 5, which is consistent with Kaestner and Grossman (2008). The OLS results also indicate that both African Americans and Hispanics score substantially lower on the verbal test than Caucasian and Asian students, the oldest child performs slightly better than her siblings and that parental education and family income are positively correlated with test scores. There does not appear to be any evidence indicating that gender differences exist once family fixed effects are controlled.

Instrumental variable and family fixed effects IV estimates of the impacts of poor health on education are presented in the last four columns of Table 5. The IV estimated impacts of depression, AD and HD are very large relative to the OLS results and the latter two are marginally significant. As to the size of the impact, the results indicate that both depression and inattention lead to substantial decreases in test scores whereas HD leads to a marked increase. The inclusion of family fixed effects leads the IV point estimate of HD and depression to become statistically insignificant in both health vectors. Notice in the last column, that the magnitude of the coefficient on depression and HD diminishes substantially as we add the family fixed effects into the IV analysis. Only the IV fixed effects estimate of AD remains statistically significant once we account for family fixed effects. It also increases by over 40% in magnitude. Focusing on the fixed effects IV-2SLS specification in column 8 as a benchmark, the point estimate indicates that suffering from inattention would lead to roughly a 26 point decline in academic performance. We note that the parameters in Table 5 are

reduced-form estimates. Since we have instrumented for poor health outcomes, we make the causal assertion that AD significantly decreases verbal tests scores, while a range of other demographic variables excluding race, birth order and maternal education have at best a tenuous impact on test score performance.²⁹

Attenuation bias due to measurement error in the AD and HD variables could account for some of the difference between the OLS and instrumental variable estimates in Table 5. Recall that these classifications are based on answers to retrospective questions, which are thought to be recorded with error. By including statistical controls for common family influences, the fixed effects strategy only uses information within families, attenuating the variance in the regressors. Thus, measurement error imposes a degradation in the signal to noise ratio and a variable measured with error will be severely biased toward zero. Yet, interestingly, only the estimates on HD and depression becomes substantially smaller when including family fixed effects to the estimation of equation (1) whereas AD and obesity increase in magnitude.

The estimates from Table 5 can also be used to examine the source of the endogeneity in the health variables. Tests of joint significance of the family effects are statistically significant for all specifications. This indicates that one should account for family-specific heterogeneity. Random effect estimates (not reported) were used to conduct Hausman tests of the endogeneity of the health variables and the results suggest fixed effects indeed removes some of the endogeneity. We next examined whether accounting for family fixed effects eliminates the need to treat the health vector as endogenous by testing the null hypothesis that the IV estimates and the fixed effects IV estimates are similar using a Hausman-Wu test. If the Null is accepted, this would suggest there are efficiency

²⁹While the estimated effect for AD is quite large (approximately two standard deviations in the test score) in comparison to the estimated effects of depression and obesity, the effect size differences are consistent with differences in the typical age of onset of the health outcomes. For AD and HD, symptoms occur at a young age, typically during elementary school or earlier. In contrast, the age of onset for symptoms of depression is typically during middle adolescence, peaking at approximately age 16.

gains from conducting family fixed effects estimates. For both health vectors, we can reject the Null, suggesting that the family fixed effects do not fully remove the sources of endogeneity that bias estimates of the impacts of poor health.

Similarly, we conducted Hausman tests between the simple OLS and 2SLS estimates. In the event of weak instruments (as well as overfitting), the fixed effects 2SLS estimates would be biased towards the OLS estimates. We can reject the Null of exogeneity of health outcomes for each health vector with each sample at the 5% level.

Testing the Validity of the Instruments

We considered several specification tests that examine the statistical performance of the instruments for each health equation and sample. Since our 2SLS estimates are over-identified, we use a J-test to formally test the overidentifying restrictions. This test is the principal method to test whether a subset of instruments satisfy the orthogonality conditions. The smallest of the p-values for these tests is 0.29, providing little evidence against the overidentifying restrictions.³⁰

In order to further examine whether these genetic markers are valid instruments, we considered several specification tests to be used with multiple endogenous regressors. First, we used the Cragg–Donald (1993) statistic to examine whether the set of instruments is parsimonious (i.e. the matrix is of full rank) and has explanatory power. Second, in order to examine whether weak instruments are a concern, we calculate the test statistic proposed by Stock and Yogo (2005).³¹ To demonstrate the strength of the instruments we considered the most difficult test with our data is using the full set of genetic instruments. That is, since using a large number of instruments or moment conditions

³⁰Many of the p-values are large and exceed 0.5. P-values are computed from Sargan tests of the joint null hypothesis that the excluded instruments are valid instruments for the health variables in the achievement equation. Similarly with other instrument sets that we explored, we found evidence of large p-values above 0.2.

³¹This is an F-statistic form of the Cragg and Donald (1993) statistic and requires an assumption of i.i.d. errors, which is more likely to be met in the specifications with family fixed effects. We are not aware of any studies on testing for weak instruments in the presence of non-i.i.d. errors.

can cause the estimator to have poor finite sample performance we will demonstrate results using the full set of genetic instruments and their polygenic interactions. Our preferred instrument sets are a subset, and one could argue that we achieved strong results in those contexts since we dropped redundant instruments, thereby leading to more reliable estimates.³² The critical value for the Stock and Yogo (2005) test is determined by the number of instruments, endogenous regressors and the amount of bias (or size distortion) one is willing to tolerate with their IV estimator. With the full set of instruments, the critical value increases substantially and we find that the Cragg-Donald statistic is 45.73 and 46.11 in health vector 2 with and without family fixed effects respectively, which exceeds the critical value.³³ This suggests that even with this large set of instruments the estimator will not perform poorly in finite samples and that with or without family fixed effects, we can reject the null hypothesis, suggesting an absence of a weak instruments problem. We also considered more traditional F-statistics with our preferred set, to test for the joint significance of the full set of instruments in each first stage equation. The first stage F-statistics indicate that in each equation the full set of instruments is jointly significant in both the specifications that include and exclude family fixed effects.³⁴ We also examined the partial R-squared for each outcome and they ranged between 2.3% - 5.1%, which fit our prior, that since these disorders are polygenic, it would be unlikely that these genes would account for more than 5% of the variation in the disorders.

To examine the sensitivity of both our IV and family fixed effect IV estimates to the degree in which the exclusion restriction assumption is potentially violated, we considered the local to zero approximation sensitivity analysis proposed in Conley, Hansen and Rossi (2007). This analysis

³²We did conduct Kleibergen and Paap (2006) tests for the preferred instrument set reported in table 5 and the Kleibergen-Paap we can reject the Null hypothesis at the 10% level. This suggests the matrix is of full rank and while overidentified the set does provide identification of the health variables.

³³For health vector 1 the results are 48.03 and 51.62.

³⁴The F-statistics also suggest that our empirical results in Table 5 are not driven by the instruments performing well in certain health equations and not in others.

involves making an adjustment to the asymptotic variance matrix, thereby directly affecting the standard errors. While the variance matrix continues to account for the usual sampling behavior, Conley, Hansen and Rossi (2007) suggest including a term that measures the extent to which the exogeneity assumption is erroneous.³⁵ The amount of uncertainty about the exogeneity assumption is constructed from prior information regarding plausible values of the impact of genetic factors on academic performance that are obtained from the reduced form. In our analysis, we consider increasing the exogeneity error from 0% to 90% of the reduced form impacts. At levels, below 40% of the reduced form impacts, our results are robust as inattention continues to have a statistically significant negative impact on verbal test scores. Our full set of results become statistically insignificant only if we assume the extent of deviations from the exact exclusion restrictions are assumed to be above 60% of the reduced form impacts. Since there does not exist any scientific evidence that these specific markers directly affect academic achievement, the sensitivity analysis indicates the levels at which our results are sensitive to the exclusion restriction assumption appear highly implausible. Further, increasing our confidence in Table 5, the sensitivity analysis suggests that our quantitative results are robust to potentially mild and moderate violations to the exogeneity assumption.

4.3 Robustness

In order to demonstrate the robustness of our empirical findings, we replicated the analysis on various subsets of the data based on family relationships, zygosity and gender. We considered

³⁵Essentially, the procedure involves estimates of the second stage equation with the instrumented health vector where the instruments are additionally included in the specification. If the exclusion restriction assumption is satisfied, the coefficients on the instrument are not identified. To conduct the analysis, we assume a prior distribution for the estimated impact of these coefficients. In our analysis, the impacts are distributed $N(0, \mathbf{q}^2 \mathbf{I})$, where \mathbf{q} is the $q\%$ percentage of the reduced form impact obtained from an OLS regression of academic achievement on the instruments and exogenous factors. We vary q to conduct our sensitivity analysis.

these breakdowns as the inclusion of family fixed effects ensures that only the dizygotic twins and siblings identify the fixed effect IV estimates of β_2 . The measure of genetic relatedness does not differ in theory between dizygotic twins and full siblings because dizygotic twins come from different eggs they are as genetically similar as any other non-twin sibling and have a genetic correlation of approximately half that of monozygotic twins. However, the inclusion of family fixed effects also imposes an equal environment assumption on the family members. That is 1) family inputs that are unobserved to the analyst do not differ between family members and 2) these factors have the same impact on achievement between relations. This assumption of equal impacts from family factors is more likely to be satisfied with data on twins than siblings as one could imagine that 1) parents make differential time-varying investments across siblings, and 2) the impacts of particular family factors may differ on children of different ages. In addition, sibling models do not effectively deal with endogeneity bias that could result from parents adjusting their fertility patterns in response to the (genetic) quality of their earlier children.³⁶

While one could imagine that data on the subsample of twins would provide the most accurate robustness check, we imposed an additional sample restriction that the pairs (or trios) of children are of the same gender. It is more likely that parents will make the same investments in the children who are most similar.³⁷ We replicate the above analysis only on the subsample of twins of the same gender and the results from all four estimation approaches are presented in Table 6.

Notice, the OLS estimates (column 2) suggest a substantially larger role for ADHD (column 1) and AD (column 2), whose magnitude is nearly twice as large as that for the full sample presented

³⁶A large empirical literature has documented that subsequent fertility decisions are influenced by prior birth outcomes. For example, Angrist and Evans [1998] and Preston [1985], among others, have established that fertility decisions are influenced by sex composition of existing children as well as past neo-natal or infant mortality.

³⁷For example birth order, birth spacing and sex composition have been shown to affect differential levels of investment by parents into children (e. g. Hanushek [1992], Black, Devereux, and Salvanes [2005] and Conley and Glauber [2005]).

in Table 5. On average, inattention leads to a six point decline in verbal test scores. Depression no longer enters the equation in a significant manner, though the magnitude is similar, and the impact of being overweight on academic performance leads to a small decrease in academic performance that is statistically significant at the 10% level. None of the health variables enter the equation in a significant manner once we either include family fixed effects or use traditional IV analysis. However, once we account for family fixed effects and also instrument the health conditions AD continues to enter the equation in a significant manner, on average, a child with AD scores almost 14 points lower. ADHD also now enters significantly in these specification and HD now enters in a marginally significant manner but the sign of the coefficient has changed. The large impact of both AD and HD are identified from dizygotic twin pairs which differ in these classifications but this is the only specification in which the impacts of AD and HD enter in a significant manner and are not significantly different. While neither depression or obesity enter the equation in a statistically significant manner, it is important to stress that we have a very small sample size in which we are able to identify effects and approximately 60% of the twin pairs are monozygotic, leading to larger standard errors.³⁸ However, the coefficient estimates for depression and overweight are practically identical in magnitude and sign to those presented in Table 5. Additionally, tests of the validity of the instrument continue to suggest that this set of genetic markers has good statistical properties and Hausman tests between columns 2 and 6 of Table 6 reject the exogeneity of the health vector.

We believe that the estimates in Table 6 present the strongest possible robustness check for our empirical evidence of causal impacts of poor mental health on academic achievement as the family members are of the same age, same race and same gender and with the exception of health and education outcomes the only other measures contained in our data for which they have different

³⁸For example birth order, birth spacing and sex composition have been shown to affect differential levels of investment by parents into children (e. g. Hanushek [1992], Black, Devereux, and Salvanes [2005] and Conley and Glauber [2005]).

values are genetic markers. The fixed effects-IV estimates presented in the last column continue to suggest that poor mental health impacts academic performance, whereas our physical health measure has no significant impact. \

Since one must always be cautious in attributing external validity to an analysis with twins data, we replicate the analysis that correspond to Table 5 where we only utilize the subsamples of siblings in Appendix Table 1. As discussed above, the equal family environment assumption is inconsistent with many models of family behavior³⁹ and the likelihood that the assumption is valid is higher with the subsample of twins (of the same gender) versus siblings.⁴⁰ However, results with the sibling sample are likely of increased external validity (presented in Appendix table 1), so there is a clear trade-off. In the sibling sample, it is interesting to note that the AD condition continues to lead to a significant decrease in test scores (column 8). The large penalty on academic performance to a sibling with AD is striking particularly if the assumption that parents are making equal investments to their children holds. None of the other health variables enter the equation in a significant manner in the family fixed effects and IV analyses. Ignoring family fixed effects, the IV estimates indicate that both hyperactivity (HD) has a positive impact on test score performance and depression has a negative impact that is marginally significant when we exclude family fixed effects from the IV analysis. Finally, in this subsample, the instrument set continues to have good first stage properties, the p-values of the overidentification tests are above 0.35 and Hausman tests suggest that the health vector should be treated as endogenous where family fixed effects by themselves do not remove all of the potential biases.

³⁹See Rosenzweig and Wolpin (2000) for a discussion.

⁴⁰Results for the full subsample of twins (n=617) are available upon request. There are few differences in the significance and magnitude of the impacts from health variables.

4.4 Comorbidity and Measurement Error

In our study, we used a rich vector of health outcomes in part to ensure that the exclusion restriction property of the instrument holds. Ding et al. [2006] argue that the major challenge for studies that estimate the impacts of health is how to what measures should be included in this vector. Using only a single health outcome to proxy for health could lead to different results. Table 7 demonstrates the substantial presence of comorbordities in our sample. Column 1 of Table 7 displays the number of individuals (and marginal distribution) in each wave who smoke or have been classified with either AD, HD, ADHD, obesity or depression. Across each row, we present the number of individuals (and conditional frequency) who also engage in smoking or suffer other poor health outcomes. Not only are adolescents with ADHD more likely to smoke but they also have a higher rate of being classified as either depressed or obese than their cohorts (one sided t-tests). This result is not unique to ADHD as we find that individuals with any of these health disorders are significantly more likely to have a second disorder. In addition, those with a health disorder are more likely to smoke cigarettes.

Since health disorders and risky health behaviors are more common among individuals with one particular disorder than among the remaining population, we investigate whether estimates of the impacts of a specific disorder vary in sign, significance and magnitude if we do not control for comorbidities. The majority of the literature on the impacts of health generally include only a single outcome measure such as obesity, smoking or birthweight in their analysis. We consider what would happened if we followed the usual practice of ignoring comorbid conditions and only included one health outcome in the achievement equation. One could imagine that in OLS and family fixed effects strategies omitted variable bias could occur since many of the neglected health conditions would be correlated with both the included health condition as well as verbal test scores. The IV estimates may not overcome bias in this setting unless the genetic instruments are unique to specific disorders.⁴¹ This is unlikely as they are associated with the same region of the brain

⁴¹In our context the genes may also be associated with schizophrenia and Tourette’s syndrome. These two disorders

and gene-gene interactions are likely to be substantial. Excluding significant comorbid conditions potentially leads to problems with sets of genetic instruments, as it is hard to imagine that any nurture or environmental factor could break the statistical association between these disorders.⁴²

Table 8 presents OLS, family fixed effect, IV and fixed effects-IV estimation of equation (1) where the health vector includes only a single specific disorder at a time. Thus, each entry in Table 8 refers to the point estimate of that specific health outcome on verbal achievement, controlling for the same set of observed controls as in Table 5. The empirical estimates of several disorders differ from that obtained using the full health vector reported in Table 5. In the OLS regressions reported in Table 8, HD no longer enters significantly and the magnitude of the impact of AD is substantially smaller. The fixed effects results in Table 8 are very similar to those obtained in Table 5 which could suggest that there are limited sets of twins/siblings that are discordant for multiple health problems. Interestingly the impact of depression does not vary substantially between Table 8 and Table 5 in the OLS and fixed effects analysis. The instrumental variables estimates in Table 8 differ greatly when only one health variable is to be included. One would conclude that each health variable with the exception of AD has a significant impact on academic performance. Depression is negatively and significantly related to verbal test scores, but the estimated impact of hyperactivity changes signs from that reported in Table 5. ADHD is highly negatively related to test scores and enters in a significant manner at the 15% level. Finally, the estimated impact of being overweight now becomes significant at the 15% level and leads to a 7 point increase in test scores on average when estimating equation (1) using IV analysis. Finally, regarding the preferred fixed effects IV specifications, we would conclude that AD and ADHD each has a negative and significant impact on

have both low prevalence rates and also low discordance rates within families. We do not believe that this is a major issue with the IV fixed effects specification reported earlier but this remains an empirical question.

⁴²For example, Chou et al. [2004] and Gruber and Frakes [2006] examine whether higher cigarette prices affected relative prices, thereby reducing smoking but increasing obesity. The former study finds evidence and the latter examines the robustness and suggests that much of the results are implausible.

academic performance. The sign of the estimated impact on HD changes from negative to positive. Interestingly the addition of family fixed effects leads the estimated signs of the impacts of ADHD, HD and obesity to change signs when instruments are also employed. Similar to Table 5, the estimated impact of depression decreases substantially when family fixed effects and instrumental variables are used to estimate equation (1).

Overall, there are a number of differences in the estimated impacts of mental health disorders between when estimating equation (1) by OLS, IV and family fixed effects with IV. Constructing an appropriate health vector presents a substantial empirical challenge and the omission of comorbid conditions could lead to biases in coefficient estimates.

5 Conclusions

Numerous studies have reported that within families, siblings and twins are often radically different in personality traits, health, education and labor market outcomes. Researchers have traditionally examined whether different environmental factors account for the development of these differences within families but have concluded that these factors can only account for a limited amount of the variation in outcomes within families. Each time a new sibling is conceived, a "genetic lottery" occurs and roughly half of the genes from each parent are passed on to the child in a random process. With recent scientific discoveries (most notably the decoding of the human genome), it is now possible to collect data that provides a precise measure of specific genetic markers, permitting researchers to enter what traditionally has been a blackbox in empirical research. In this paper, we exploit variation within siblings and within twins from the "genetic lottery" to identify the causal effect of several poor health conditions on academic outcomes via a family fixed effect / instrumental variables strategy.

We find evidence of large impacts from poor mental health to lower academic achievement.

Inattention and depression both lead to large decreases in individual performance on verbal IQ tests within families. Our results indicate that when estimating education production functions, researchers should treat health as an endogenous input. Further, bias from endogeneity cannot be fully removed using family fixed effects estimators. Our results suggest that while these family factors should be accounted for, an instrumental variable that varies between family members is required to overcome endogeneity biases. We present evidence that using differences in genetic inheritance are valid instrumental variables with good statistical properties, allowing us to identify the causal impact of poor health on education. Finally, we find that measurement error could potentially present a serious challenge to empirical researchers due to the prevalence of comorbid conditions.

The quantitative and qualitative patterns of our empirical results are robust to multiple sample definitions including the restriction to using only dizygotic twins of the same gender. One potential limitation of this study deals with external validity. It is important to consider whether our analysis of family members can be generalized to larger populations of interest.

We believe that there is substantial potential from explicitly using data on genetic markers in social science research. As the scientific literature is developing an ever-increasing understanding of how genetic inheritance relates to individual (health) outcomes, this knowledge can be used to refine searches for potential genetic markers to serve as instrumental variables. Genetic markers have a great deal of conceptual validity as instruments for many (health) outcomes since i) the markers are inherited at conception prior to any interaction with the environment, eliminating concerns related to reverse causality, ii) large literatures exist that report robust correlations between specific markers and individual (health) outcomes, iii) studies of genetic inheritance and measures of genetic distance from maps of the human genome are available to investigate whether genetic linkage is a valid concern, and iv) most genes are pleiotropic so that a predisposition can be viewed as a form of inherited encouragement. In addition, researchers could investigate the sources of pleiotropy by

examining how different environmental disturbances affect gene expression and how that relates to a variety of economic outcomes. In summary, we believe that integrating biological findings into the social sciences has potential to not only address open research questions but also help develop policies that can promote human capital development. However, unlike biological measures such as height, weight, blood pressure, blood alcohol content, cholesterol levels or hormones whose measures are influenced by behavioral inputs, genetic markers are time-invariant and cannot be modified by environmental influences, but within families, any differences in the inheritance of specific markers presents the opportunity for additional experiments in "nature".

References

- [1] Allen, Gordon (1970). "Within group and between group variation expected in human behavioral characters." *Behavior Genetics*, 1(3-4), 175-194.
- [2] Angrist, J. D. and W. Evans. (1998). "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size." *American Economic Review*, 88, 450-477.
- [3] Babinski, L. M., C. S. Hartsough and N. M. Lasbert (1999). "Childhood conduct problems, hyperactivity-impulsivity, and inattention as predictors of adult criminal activity," *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 40(3), 347-355.
- [4] Behrman, J. R. and P. Taubman (1976). "Intergenerational Transmission of Income and Wealth." *American Economic Review*, 66(2), 436-440.
- [5] Behrman, Jere R., P. Taubman, T. Wales, and Z. Hrubec. (1977). "Inter- and Intragenerational Determination of Socioeconomic Success with Special Reference to Genetic Endowment and Family and Other Environment." *mimeo*, University of Pennsylvania.
- [6] Behrman, J. R. and V. Lavy. (1998). "Child Health and Schooling Achievement: Association, Causality and Household Allocations." *CARESS Working Papers 97-23*, University of Pennsylvania.
- [7] 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, 1131-1174.
- [8] Black, S., P. Devereux, and K. Salvanes. (2005). "The More the Merrier? The Effect of Family Size and Birth Order on Children's Education." *Quarterly Journal of Economics*, 120, 669-700.
- [9] Bleakley, H. C. (2007). "Disease and Development: Evidence from Hookworm Eradication in the American South." *Quarterly Journal of Economics*, 122(1), 73-117.
- [10] Cawley, J. (2004). "The Impact of Obesity on Wages." *Journal of Human Resources*, 39(2), 451-474.
- [11] Conley, T., C. Hansen and P. E. Rossi. (2007). "Plausibly Exogenous." *mimeo*, University of Chicago.
- [12] Cooper R. S., J. S. Kaufman and R. Ward. (2003). "Race and Genomics." *The New England Journal of Medicine*, 348(12), 1166-1170.

- [13] Cragg, J. G., and S. G. Donald. (1993). "Testing Identifiability and Specification in Instrumental Variables Models." *Econometric Theory* 9, 222–240.
- [14] Chou, S.-Y., M. Grossman and H. Saffer. (2004). "An Economic Analysis of Adult Obesity: Results from the Behavioral Risk Factor Surveillance System." *Journal of Health Economics*, 23, 565–587.
- [15] Conley, D. and R. Glauber. (2005). "Parental Education Investment and Children's Academic Risk: Estimates of the Impact of Sibship Size and Birth Order from Exogenous Variation in Fertility." *NBER Working Paper w11302*.
- [16] Currie, J. and M. Stabile. (2006). "Child Mental Health and Human Capital Accumulation: The Case of ADHD." *Journal of Health Economics*, 25(6), 1094–1118.
- [17] de Quervain, D. J.-F. and A. Papassotiropoulos. (2006). "Identification of a Genetic Cluster Influencing Memory Performance and Hippocampal Activity in Humans." *Proceedings of the National Academy of Sciences USA*, 103, 4270–4274.
- [18] Ding, W., and S. F. Lehrer. (2007). "The Promise of Using Neural Correlates to Establish Causal Relationships in Social Science Research." *mimeo*, Queen's University.
- [19] Ding, W., S. F. Lehrer, J. N. Rosenquist and J. Audrain-McGovern. (2007). "The Impact of Poor Health on Education: New Evidence Using Genetic Markers." *mimeo*, Queen's University.
- [20] Ding, W., S. F. Lehrer, J. N. Rosenquist and J. Audrain-McGovern. (2006). "The Impact of Poor Health on Education: New Evidence Using Genetic Markers." *NBER Working Paper w12304*.
- [21] Dremencov, E., I. Gispan-Herman, M. Rosenstein, A. Mendelman, D.H. Overstreet, J. Zohar and G. Yadid. (2004). "The Serotonin–Dopamine Interaction is Critical for Fast-Onset Action of Antidepressant Treatment: In Vivo Studies in an Animal Model of Depression." *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 28, 141–147.
- [22] Fletcher, J.M. (2007). "Adolescent Depression and Educational Attainment: Evidence from Sibling Fixed Effects." *mimeo*, Yale University.
- [23] Fletcher, J.M. and B.L. Wolfe. (2007a). "Long-term Consequences of Childhood ADHD on Criminal Activities." *mimeo*, Yale University.
- [24] Fletcher, J.M. and B.L. Wolfe. (2007b). "Child Mental Health and Human Capital Accumulation: The Case of ADHD Revisited." forthcoming in the *Journal of Health Economics*.

- [25] Glewwe, P. and H. Jacoby. (1995). "An Economic Analysis of Delayed Primary School Enrollment in a Low-Income Country-the Role of Early Childhood Nutrition." *Review of Economics and Statistics*, 77, 156-169.
- [26] Goldstein D. B., S. K. Tate and S. M. Sisodiya. (2003). "Pharmacogenetics Goes Genomic." *Nature Reviews Genetics*, 4, 937-947.
- [27] Goodman E., B. R. Hinden and S. Khandelwal. (2000). "Accuracy of Teen and Parental Reports of Obesity and Body Mass Index." *Pediatrics*, 106(1), 52-58.
- [28] Gorseline, D.W. The Effect of Schooling Upon Income. (1932). (Bloomington: Indiana University Press).
- [29] Grossman, M. and R. Kaestner. (1997). "Effects of Education on Health," in J. R. Behrman and N. Stacey eds. *The Social Benefits of Education*, University of Michigan Press, Ann Arbor.
- [30] Grossman, M. (1975). "The Correlation between Health and Schooling," in *Household Production and Consumption*, Ed N. E. Terleckyj, Studies in Income and Wealth, Vol. 40, Conference on Research in Income and Wealth. New York: Columbia University Press for the National Bureau of Economic Research.
- [31] Gruber, J. and M. Frakes. (2006). "Does Falling Smoking Lead to Rising Obesity?" *Journal of Health Economics*, 25, 183-197.
- [32] Hanushek, E. (1992). "The Trade-off between Child Quantity and Quality." *Journal of Political Economy*, 100 84-117.
- [33] Harris, K. M., F. Florey, J. Tabor, P. S. Bearman, J. Jones and J. R. Udry. (2003). "The National Longitudinal Study of Adolescent Health: Research Design," www document available at <http://www.cpc.unc.edu/projects/addhealth/design>, Carolina Population Center, University of North Carolina, Chapel Hill, NC.
- [34] Harrison, Ainsworth Gordon. (1970). "Human Variation and Its Social Causes and Consequences." *Proceedings of the Royal Anthropological Institute of Great Britain and Ireland*, 1970, 5-13.
- [35] Jain, A.K., S. Prabhakar, and S. Pankanti. (2002). "On the similarity of identical twin fingerprints." *Pattern Recognition*, 35:2653-2663.
- [36] Johnson JA. (2003). "Pharmacogenetics: potential for individualized drug therapy through genetics." *Trends Genet*, 19:660-66.

- [37] Kelada S. N., D. L. Eaton, S. S. Wang, N. R. Rothman and M. J. Khoury. (2003). "The Role of Genetic Polymorphisms in Environmental Health." *Environmental Health Perspectives*, 111, 1055–1064.
- [38] Kaester, R., M. Grossman. (2008). "Effects of Weight on Children’s Educational Achievement." NBER Working Paper 13764.
- [39] Kessler, R. at al. (2005). “Patterns and Predictors of Attention-Deficit/Hyperactivity Disorder Persistence into Adulthood: Results from the National Co-morbidity Survey Replication.” *Biological Psychiatry*, 57, 1442-1451.
- [40] Kleibergen, F., and R. Paap. (2006). "Generalized reduced rank tests using the singular value decomposition." *Journal of Econometrics* 127(1), 97–126.
- [41] Klepinger, D. S. Lundberg and R. Plotnick. (1999). "How Does Adolescent Fertility Affect the Human Capital and Wages of Young Women?" *The Journal of Human Resources*, 34(3), 421-448.
- [42] Kremer M. and E. Miguel. (2004). “Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities.” *Econometrica*, 72, 159-217.
- [43] Merikangas K. R. and N. Risch. (2003). "Genomic Priorities and Public Health." *Science* 302, 599–601.
- [44] Moises H. W., R. M. Frieboes, P. Spelzhaus, L. Yang, M. Kohnke, O. Herden-Kirchhoff, P.Vetter, J. Neppert, and I. Gottesman. (2001). “No Association between Dopamine D2 Receptor Gene (DRD2) and Human Intelligence.” *Journal of Neural Transmission*, 108, 115-121.
- [45] Neumark, D. (1999). “Biases in Twin Estimates of the Return to Schooling.” *Economics of Education Review*, 18, 143-148.
- [46] Norton, E.C. and E. Han. (2007). "Genetic Information, Obesity, and Labor Market Outcomes." *mimeo*, University of North Carolina at Chapel Hill.
- [47] Olds, J. (1967). “The Limbic System and Behavioral Reinforcement.” *Progress in Brain Research*, 27, 144-64.
- [48] Olds, J., Milner, P. (1954). :Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain." *J.Comp. Physiol. Psycholo*, 47, 419–427.
- [49] Perri, T. J. (1984). “Health Status and Schooling Decisions of Young Men.” *Economics of Education Review*, 3, 207-213.

- [50] Petrill, S. A., R. Plomin, G. E. McClearn, D. L. Smith, S. Vignetti, M. J. Chorney, K. Chorney, L. A. Thompson, D. K. Detterman, C. Benbow, D. Lubinski, J. Daniels, M. Owen and P. McGuffin. (1997). "No Association between General Cognitive Ability and the A1 Allele of the D2 Dopamine Receptor Gene." *Behavior Genetics*, 27(1), 29-31.
- [51] Plomin, R., J. K. J. Kennedy and I. W. Craig. (2006). "The Quest for Quantitative Trait Loci Associated with Intelligence." *Intelligence*, 34(6), 513-526.
- [52] Preston, S. H. (1985). "Mortality in Childhood: Lessons from WFS," in J. G. Cleland and J. Hobcraft (eds.), *Reproductive Change in Developing Countries*, Oxford: Oxford University Press, pp. 46-59.
- [53] Rosenzweig, M. R. and K. I. Wolpin. (2000). "Natural "Natural Experiments" in Economics." *Journal of Economic Literature*, 38, 827-874.
- [54] Stock, J. H., and M. Yogo. (2005). "Testing for Weak Instruments in Linear IV Regression," in D.W. Andrews and J. H. Stock (eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, Cambridge University Press.
- [55] Strauss, J. and D. Thomas. (1998). "Health, Nutrition, and Economic Development." *Journal of Economic Literature*, 36(2), 766-817.
- [56] Taubman, P. (1976a). "The Determinates of Earnings: Genetics, Family and Other Environments, a Study of White Male Twins." *American Economic Review*, 66(5), 858-870.
- [57] Taubman, P. (1976b). "Earnings, Education, Genetics, and Environment." *Journal of Human Resources*, 11(4), 447-461.
- [58] Zerhouni E. (2003). "Medicine. The NIH Roadmap." *Science*, 302, 63-72.

Table 1: Summary Statistics

Variable	Full Sample	Sibling Sample	Twin Sample
Test Score	100.552 (13.564)	100.794 (13.324)	100.107 (13.984)
AD	0.050 (0.218)	0.049 (0.215)	0.056 (0.229)
HD	0.049 (0.215)	0.052 (0.223)	0.043 (0.203)
ADHD	0.077 (0.266)	0.077 (0.266)	0.078 (0.268)
Depression	0.062 (0.241)	0.067 (0.251)	0.052 (0.223)
Obesity	0.072 (0.258)	0.081 (0.272)	0.060 (0.238)
Age in Initial Data Collection	17.03 (1.687)	17.054 (1.700)	16.990 (1.667)
Male	0.489 (0.500)	0.479 (0.500)	0.504 (0.500)
African American	0.169 (0.375)	0.131 (0.338)	0.234 (0.424)
Hispanic	0.141 (0.348)	0.140 (0.348)	0.145 (0.352)
Family Income (*\$1,000)	46.807 (40.158)	45.206 (30.734)	49.828 (53.873)
Mother's Education	13.200 (2.203)	13.166 (2.105)	13.232 (2.356)
Parental Age	41.850 (5.337)	41.382 (5.017)	42.527 (5.750)
Observations	1684	1068	629

Note: Standard deviations in parentheses.

Table 2: Summary Statistics on Peabody Verbal Test Score Performance Conditional on Health Disorder and Health Behavior

	Full sampling	Sibling	Twin
Depression	92.00 (14.19)	94.03 (13.53)	91.63 (15.87)
No depression	101.03 (13.38)	101.23 (13.16)	100.70 (13.73)
T-statistic	5.705	4.44	3.66
ADHD	100.19 (12.336)	101.5 (12.167)	98.06 (12.44)
No ADHD	100.58 (13.664)	100.68 (13.40)	100.40 (14.09)
T-statistic	0.312	-0.527	1.13
HD	102.18 (11.550)	103.11 (11.77)	100.39 (11.09)
No HD	100.49 (13.657)	100.62 (13.38)	100.22 (14.10)
T-statistic	-1.112	-1.34	-0.06
AD	98.45 (12.41)	99.56 (11.92)	96.84 (13.11)
No AD	100.66 (13.62)	100.81 (13.38)	100.42 (14.01)
T-statistic	1.456	0.646	1.46
Obese	98.00 (12.755)	98.84 (13.22)	96.02 (11.50)
Not obese	100.74 (13.68)	100.91 (13.31)	100.48 (14.08)
T-statistic	2.14	1.37	1.86
Overweight	100.798 (13.44)	99.70 (14.42)	97.32 (13.76)
Not overweight	98.92 (14.22)	100.92 (13.12)	100.61 (13.97)
T-statistic	1.92	1.02	1.89
Smoke Cigarettes	100.12 (12.22)	100.65 (11.93)	99.27 (12.69)
Does not smoke cigarettes	100.71 (13.97)	100.79 (13.73)	100.57 (14.38)
T-statistic	0.757	0.14	1.01

Note: Most cells present the mean verbal test score and standard deviations in parentheses for individuals by health category.

Table 6: Estimates of the Achievement Equation for the Sample of Twins of the Same Gender

Estimation Approach	OLS		Family Fixed Effects		Instrumental Variables		Family Fixed Effects Instrumental Variables	
AD	N/A	-5.957 (2.297)**	N/A	-3.049 (2.552)	N/A	-4.292 (6.218)	N/A	-14.991 (7.475)*
HD	N/A	2.061 (2.592)	N/A	-0.172 (2.749)	N/A	-4.213 (8.633)	N/A	-15.994 (10.828)
ADHD	-4.538 (1.812)*	N/A	-2.155 (2.153)	N/A	-6.643 (14.245)	N/A	-18.075 (6.473)**	N/A
Depression	-3.184 (2.969)	-3.306 (2.928)	0.738 (2.493)	0.734 (2.498)	-7.181 (17.247)	-4.161 (15.283)	-12.229 (21.557)	-11.27 (17.456)
Obesity	-2.853 (1.427)*	-2.93 (1.421)*	0.007 (1.81)	0.059 (1.81)	-3.379 (9.682)	-3.25 (8.718)	-3.884 (6.880)	-1.61 (6.261)
Male	3.597 (1.127)**	3.483 (1.125)**			3.641 (1.670)*	3.619 (1.515)*		
African American	-8.318 (1.463)**	-8.311 (1.463)**			-8.464 (2.009)**	-8.345 (1.970)**		
Hispanic	-6.894 (1.757)**	-6.93 (1.735)**			-6.895 (2.733)*	-6.974 (2.643)**		
Family Income	0.012 (0.004)**	0.013 (0.004)**			0.012 (0.007)	0.012 (0.007)+		
Maternal Years of Education	1.275 (0.240)**	1.249 (0.240)**			1.233 (0.363)**	1.26 (0.346)**		
Parents Age	0.184 (0.099)+	0.184 (0.099)+			0.197 (0.134)	0.187 (0.134)		
Parents Married	-1.659 (1.263)	-1.657 (1.268)			-1.795 (1.652)	-1.776 (1.680)		
Observations	469	469	469	469	469	469	469	469

Note: Corrected standard errors in parentheses. ***, **, * denote statistical significance at 1%, 5%, 10% level respectively.

Table 7: Relationship Between Health Behaviors and Health Outcomes During Adolescence

Behavior	Total Number	Nothing Else ¹	Also ADHD	Also AD	Also HD	Also Obese	Also Depressed	Also Smokes
Full Sample								
Nothing	975 [58.24]	***	***	***	***	***	***	***
ADHD	129 [7.66]	67 (51.94)	-----	-----	-----	16 (13.22)	11 (8.53)	46 (35.66)
AD	84 [4.99]	40 (47.62)	-----	-----	37 (44.05)	11 (13.10)	8 (9.52)	33 (39.29)
HD	82 [4.87]	41 (50.00)	-----	37 (45.12)	-----	11 (13.41)	5 (6.10)	30 (36.59)
Obese	121 [7.19]	69 (57.50)	16 (12.40)	11 (9.09)	11 (9.09)	-----	14 (11.57)	32 (26.67)
Depression	104 [6.18]	48 (46.15)	11 (11.93)	8 (7.69)	5 (4.81)	14 (13.46)	-----	44 (42.31)
Smokes Cigarettes	404 [24.08]	297 (73.51)	46 (11.39)	33 (8.17)	30 (7.43)	32 (7.92)	44 (10.89)	-----

Note: Each cell contains the number of individuals diagnosed with the respective row and column combination. The conditional frequency of dual diagnoses is presented in round parentheses. The marginal probability of being diagnosed with each outcome is presented in square [] parentheses.

¹ For ADHD nothing else excludes AD and HD.

Table 8: Estimates of the Achievement Equation Where We Include Only a Single Health Condition by Itself

Estimation Approach	OLS	Family Fixed Effects	Instrumental Variables	Family Fixed Effects and Instrumental Variables
AD	-2.275 (1.176)+	-0.737 (1.352)	-0.904 (6.040)	-15.050 (9.790)
HD	1.106 (1.142)	1.356 (1.408)	13.510 (9.600)	-7.353 (8.846)
ADHD	-1.208 (0.981)	0.317 (1.142)	3.304 (7.077)	-12.303 (8.532)
Depression	-4.473 (1.285)**	-2.193 (1.209)+	-23.265 (11.010)*	-5.742 (8.625)
Obesity	-0.846 (0.741)	-0.06 (0.877)	7.879 (5.308)	-6.887 (4.328)

Estimates from Specifications which only include AD and HD separate diagnoses.

AD	-3.289 (1.289)*	-1.424 (1.457)	-19.900 (12.456)	-17.164 (11.401)
HD	2.495 (1.302)+	1.912 (1.519)	31.573 (14.986)*	7.415 (12.557)

Note: Corrected standard errors in parentheses. Each cell of the table corresponds to a separate regression. The dependent variable of the regression differs by row. Columns reflect different estimation approaches as denoted in the first row. Regressions control for the same set of non-health inputs as in Table 5, including student demographics, parental characteristics and home environment variables. ***, **, * denote statistical significance at 1%, 5%, 10% level respectively.

