



Network for Studies on Pensions, Aging and Retirement

Jason M. Fletcher

Steven F. Lehrer

## **Genetic Lotteries within Families**

**Discussion Paper 08/2009 - 026**

August, 2009

# Genetic Lotteries within Families<sup>\*†</sup>

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

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

August 2009

## Abstract

Drawing on findings from the biomedical literature, this paper introduces the idea that specific exogenously inherited differences in the genetic code between full biological siblings can be used to test within-family estimators and potentially improve our understanding of economic relationships. These points are illustrated with an application to identify the causal impact of several poor health conditions on academic outcomes. We present evidence of large impacts of poor mental health on academic achievement and demonstrate that our results are robust to reasonable violations of the exclusion restriction assumption. Further, our estimates suggest that family fixed effects estimators by themselves cannot fully account for the endogeneity of poor health.

---

\*We are grateful to Ken Chay, Dalton Conley, Weili Ding, Ted Joyce, Robert McMillan, John Mullahy, Matthew Neidell, Jody Sindelar and participants at the 2007 NBER Summer Institute, Northwestern University, Brown University, CUNY, McGill University, University of Calgary, Tinbergen Institute, Institute for Fiscal Studies, Warwick University, University of Calgary, 2008 AHEC Conference at the University of Chicago, 2008 SOLE meetings, Yale Health Policy Colloquium, University of British Columbia, University of Connecticut, University of Saskatchewan, University of Tennessee, University of Toronto and Simon Fraser University for comments and suggestions that have improved this paper. Some of this research was conducted while Lehrer was a visitor at the department of economics of the Free University in Amsterdam and he thanks the department for providing a hospitable and productive environment as well as thanking NETSPAR for funding used to support this visit. 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.<sup>1</sup> For many years, researchers in the social sciences could only examine the relative importance of a multitude of environmental factors on various individual outcomes, as data on molecular genetic variation between individuals was unavailable. Yet, with the decoding of the human genome, this limitation no longer exists, and recent years have been characterized by substantial amounts of research in the biomedical literature examining whether specific point mutations 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.<sup>2</sup> As clinical researchers identify unique molecular genetic bases for many complex health behaviors, diseases and other outcomes,<sup>3</sup> 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 as well as allow explicit tests of economic models.

In this paper, we exploit differences in genetic inheritance among children within the same family to first test whether family fixed effects estimators by themselves can fully solve the endogeneity problem when estimating the impact of several poor health conditions on academic outcomes.<sup>4</sup> Family fixed effects estimators allow researchers to simultaneously control (assuming constant impacts between family members) 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. Differences in genetic inheritance occur at conception and remain fixed between family members at every point in the lifecycle, irrespective of all nurture investments an individual faces (even those that occur in utero).<sup>5</sup> 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.<sup>6</sup>

This study also contributes to a burgeoning literature that investigates what are the consequences of growing up in poor health for adolescent development. Prior empirical research in this area has attempted to estimate a causal link have either used a within-family strategy (i.e. Currie and Stabile [2006], Fletcher and Wolfe [2008, in press], and Fletcher [2008]) or instrumental variables approach (i.e. Ding et al. [2006,2009], Behrman and Lavy [1998], Norton and Han [2008] as well as Glewwe and Jacoby [1995]), and in general researchers find large negative impacts of poor health on academic outcomes.<sup>7</sup> The genetic lottery empirical strategy combines both elements and identifies the causal impact of several poor health conditions on academic outcomes by exploiting exogenous variation in genetic inheritance between full biological siblings including dizygotic twins via a family fixed effects instrumental variables estimator.

The genetic lottery identification strategy relies on knowledge of how specific genetic markers affect health and academic outcomes in adolescence. Despite a voluminous literature in several biomedical literatures, a consensus on how specific genetic markers operate still has not been reached. Thus, concerns could exist that, despite no detectable evidence in the biomedical literature to date,<sup>8</sup> the specific 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. In our analysis, we discuss the main threat arises from comorbid conditions and examine the sensitivity of our empirical results to the degree in which the exclusion restriction assumption is potentially violated, finding that our main results are not sensitive to the plausibility of the instruments at reasonable levels.<sup>9</sup>

Our empirical analysis reaches three major conclusions. First, we conduct a variety of specification tests which indicate that family fixed effects estimators by themselves cannot fully account for the endogeneity of poor health. This indicates that the commonly observed differences in health and education outcomes between full biological siblings should not be treated as random in empirical

analyses.

Second, we find that the impact of poor mental health outcomes on academic achievement is substantial. Our preferred estimates examine the relationship with a sample consisting only of same sex dizygotic twins, and they indicate that inattention leads on average to a one standard deviation decrease in academic performance.<sup>10</sup> The significant negative impacts of inattention on academic performance remain large and significant if we examine the relationship using other family-based samples. Sensitivity analyses indicate that our main results are robust to moderate violations of the exclusion restriction assumption.

Third, our results indicate that measurement error may be a serious concern for researchers examining the impact of health conditions.<sup>11</sup> The importance of the sensitivity analysis noted in the above paragraph should not be understated, since poor health conditions often occur simultaneously and it is hard to identify a unique source of genetic or environmental variation to identify the impact of specific disorders due to the potential presence of unmeasured comorbid conditions. We find that if one were to include only one poor health condition in the specification that the sensitivity analysis fails at very mild levels. The large difference in results from the sensitivity analyses as we move from our complete specification of the health vector to less rich subsets increases our confidence in both our main results and specification of the health vector.

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 dataset 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 dataset.<sup>12</sup> 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. To develop our identification strategy, we use a specific subsample of the respondents 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, 2,147, and 2,275 individuals who completed the survey at each interview point. Excluding those individuals for whom there is incomplete education, health and DNA measures for multiple family members reduces the sample to 1684 individuals.

The dataset 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 four-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 18-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,<sup>13</sup> we examine the effects of these different domains as well as the clinical measure of ADHD of any type.

Overweight and obesity are calculated from each individual’s self-reported height and weight applied to age and gender specific definitions obtained from the Center for Disease Control. Finally, in our analysis we control for birth weight.

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]) indicates that there is a strong correlation between measured and self-reported height (0.94), and between measured and self-reported weight (0.95). There is no evidence that reporting errors are correlated with observed variables such as race, parental education, and household income. Further, 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).

Regarding academic outcomes, the data contains information on GPA and an age standardized score on a common verbal test.<sup>14</sup> 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 dataset to serve as additional controls.

Summary statistics on our sample are provided in Table 1. Household income for the full sample (column 1) is slightly higher than US averages and the majority of mothers have attended college. Both the sibling and twins subsamples respectively presented in columns 2 and 3 appear gender balanced. With the sole exception of race and birth weight variables, there are few differences in any of the summary statistics between the subsample of siblings and twins. While the mean verbal test score for each sample approximates the national average, the standard deviation of test scores is slightly smaller than those obtained with nationally representative samples.<sup>15</sup> Unlike the education and demographic variables that are similar to those obtained from nationally represented surveys, the incidence of poor mental health outcomes differ. On the one hand, roughly 8% of the sample

is coded with ADHD, which exceeds the 6% national average. On the other hand, 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. Similarly, both obesity rates and rates of being overweight rates fall slightly below the national average for this period. Only the separate diagnoses of AD and HD fall within standard ranges observed with adolescent samples. Lastly, individuals classified in poor health, either as depressed or as obese have significantly lower (one sided t-tests) verbal test scores.

## 2.1 Genetic Data

The DNA samples were drawn in the third collection and were genotyped for six candidate polymorphisms.<sup>16</sup> 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. It is important to state that these health outcomes are polygenic—they are affected by many mutations at many genetic loci (including many that are not collected in the study) as well as the environment an individual encounters throughout her life (as well as possible gene-environment interactions).<sup>17</sup> However, only an individual's genetic make-up is both assigned at conception prior to any interaction with the environment and remains invariant to all nurture investments over the life-cycle, eliminating concerns related to reverse causality.

The set of genetic markers we use in our analysis includes the dopamine transporter (DAT), dopamine D4 receptor (DRD4), serotonin transporter (5HTT), monoamine oxidase A (MAOA), dopamine D2 receptor (DRD2) and cytochrome P4502A6 (CYP2A6) gene. Mutations in the coding of these genes, not the genes themselves, are believed to impact multiple health outcomes and behaviors. Scientists hypothesize that these point mutations distort cell functions and/or processes, leading to the higher propensities for specific disorders. It is important to state explicitly that

individual point mutations can have phenotypic effects of any strength, including quite mild effects, and it is likely that each genetic marker has pleiotropic effects.<sup>18</sup>

The genetic markers collected in the Add Health study are primarily linked to the transmission of two specific neurotransmitters in the primitive limbic system of the brain: dopamine and serotonin.<sup>19</sup> The scientific hypothesis of how these genetic 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.<sup>20</sup> 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. An allelic variant of this gene is believed to have decreased productivity of this protein, thereby increasing the risk for 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. 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 bloodstream to the brain where they generally lead to neurotransmitters being released. Finally, in our analysis we will not only consider the allele variants by themselves but also allow for gene-gene interactions, which may also have potentially powerful effects.<sup>21</sup> We present and discuss the genetic characteristics of our sample and unconditional relationships with poor health outcomes in the results section of the paper.

### 3 Empirical Framework

The empirical framework that underlies our analysis involves the estimation of a system of equations generated from a simple extension to the model developed in Ding et al. [2009]. Ding et al. [2009] makes an important distinction between health behaviors such as smoking or drinking that are treated as control variables versus health conditions such as depression or ADHD which are state variables in the model.<sup>22</sup> We continue to assume that in each period, altruistic parents select inputs to maximize the household indirect utility function after receiving noisy signals of their children’s health status, health behaviors and ability endowment. Subsets of these inputs enter both an education production function and health production function, generating stocks of human capital for each child. The parents provide children who have different abilities and health outcomes with different inputs where we now require in equilibrium the marginal returns to investments in schooling of one child is equated to the marginal returns to investments in health in their sibling.

Formally, consider a linear representation of the child’s education production function, which translates a set of inputs into human capital as measured by a score on an achievement test as

$$A_{ifjT} = \beta_0 + \beta_1 X_{iT} + \beta_2 H_{iT} + \beta_3 Q_{jT} + \beta_4 N_{iT} + \beta_5 BW_{iT} + v_f + \varepsilon_{ifjT} \quad (1)$$

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 (child gender, race, parental education, birth order, family income and family structure),<sup>23</sup> the vector  $H$  consists of variables that capture health measures, the vector  $Q$  contains school quality variables, the vector  $N$  contains information on community and neighborhood inputs, the vector  $BW$  contains an individual’s birth weight up to a quartic,  $v_f$  is an unobserved family effect and  $\varepsilon_{ifjT}$  is an idiosyncratic error term. Notice that  $H_{iT}$  is directly included as an input to the education production function. We hypothesize that there are several possible channels under which health status potentially affects academic performance. First, it may affect the physical energy level of a child which determines the time (including classroom

attendance) that can be used for learning. Second, it affects the child's mental status that may have a direct impact on academic performance. Lastly, a child's health status may affect the way a child is treated by teachers, parents and peers; which can in part shape the learning environment that is encountered.

The major empirical challenge in estimating equation (1) is that the health vector ( $H_{iT}$ ) is likely to be endogenous.<sup>24</sup> 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. The inclusion of family fixed effects ( $v_f$ ) in equation (1) directly accounts for unobserved to the researcher family factors that are common across siblings and may be related to both individual health and education outcomes. This allows the researcher to simultaneously control (assuming constant impacts between family members) for many parental characteristics/behaviors and some genetic factors. However, it does not provide any guidance as to why, within a twin or sibling pair, the subjects differ in explanatory characteristics such as health status. Thus, estimating equation (1) using a family 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 remain between the health variables and the error term (i.e.  $Cov(H_{iT} - \bar{H}_f, \varepsilon_{ifjT} - \bar{\varepsilon}_f) \neq 0$ ).

Supplementing the family fixed effects strategy with instrumental variables can potentially overcome the endogeneity bias arising from  $Cov(H_{iT} - \bar{H}_f, \varepsilon_{ifjT} - \bar{\varepsilon}_f)$ . We propose to use exogenous variation from the "genetic lottery" between family members to identify the impact of poor health on measures of achievement. 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. Formally the first stage presents a linear representation

of the child’s health production function

$$H_{ifT} = \gamma_0 + \gamma_1 X_{iT} + \gamma_2 G_i^H + \gamma_3 Q_{jT} + \gamma_4 N_{iT} + \gamma_5 BW_{iT} + v_f + v_{ifjT}, \quad (2)$$

where  $G_i^H$  is a vector of genetic markers that may provide endowed predispositions to the current state of health status.

Our identification relies on the assumption that the vectors of genetic markers that impact health outcomes ( $G_i^H$ ) are unrelated to unobserved components ( $\varepsilon_{ifjT}$ ) 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).

We not only estimate the system of equations (1) and (2) via fixed effects instrumental variables methods, but also consider family fixed effects estimation of equation (1) as well as both OLS and instrumental variables estimation of the system of equations described above where  $v_f = 0$ . Estimates from these alternative approaches are used to conduct specification tests that can shed light on the source of the endogeneity in estimating the impact of poor health on academic outcomes. For example, Hausman test exploits the fact that the two estimators have different probability limits. Rejecting the Null hypothesis relies in part on the genetic markers meeting the statistical properties of an instrumental variable as if they had weak links to health conditions the denominator term in the test statistic would be exceptionally large. Similarly if the instruments were invalid the fixed effects instrumental variables estimates would be inconsistent and the bias in the estimated coefficients is the same as that of the fixed effects estimator and hence we would be unable to reject

the Null. Yet as we will demonstrate in the next section, we will clearly reject the consistency of both the OLS and family fixed effects estimates for all samples.<sup>25</sup>

In both equations (1) and (2) we directly control for an individual’s birth weight up to a quartic to proxy for an individual’s initial stock of health capital. It is well documented by many authors that better health early in life is associated with higher educational attainment (e.g. (Grossman [1975], Perri [1984]) and that more educated individuals in turn have better health later in life (e.g. Grossman and Kaestner [1997], and Cutler and Lleras-Muney [2007]). While birth weight is known to have a large genetic component (e.g. Lunde et al. [2007]) it is well established to differ even among monozygotic twins. Royer [2009] presents evidence that these birth weight differences between twins have impacts on educational attainment and Christensen et al. [2001] demonstrates differences in birth weight also affects health later in life between twins. We believe that accounting for non-linear differences in birth weight can capture additional differences in both genetic factors and pre-natal environments between full biological siblings.<sup>26</sup>

In the 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 and 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.

## 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.<sup>27</sup>

For each poor health outcome and behavior, there is at least one allele variant of a the markers we consider that 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 two 4-repeats of the MAOA gene have greater odds and individuals with one 4-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 allele variants 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.<sup>28</sup> 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.<sup>29</sup> To examine this concern, we first examined maps of the location between the specific genetic markers in our study and those which have presently been hypothesized to be linked to education outcomes (Plomin et al. [2007], see footnote 8 for more details), we find no evidence that they are located closely on the genome, suggesting that linkage in inheritance is unlikely. Second, we present over-identification tests of our instrument sets for each health outcome. Last, 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.

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.<sup>30</sup> It is unlikely that the majority of these unconditional relationships are due to chance 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 3 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 3 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 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. Consistent with Kaestner and Grossman [2008] we do not find a significant relationship between obesity and academic performance. The OLS results also indicate that both African Americans and Hispanics score substantially lower on the verbal test than Caucasian and Asian students, the children who are older in their families perform slightly better than their 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 3. 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 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 3 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.<sup>31</sup>

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 3. 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. Interestingly, only the estimates on two health conditions, HD and depression, become smaller when family fixed effects are accounted for when estimating equation (1), suggesting this is not the explanation for the large difference in the impact of AD.

The estimates from Table 3 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 gains from conducting family fixed effects estimates. For both health vectors, we can reject the Null of IV and IV/FE coefficient equality, 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 IV estimates. In the event of weak instruments (as well as overfitting), the fixed effects IV 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.

We replicated the analysis on various subsets of the data based on family relationships, zygosity and gender as well as additional controls for health endowments. We considered these family relationship breakdowns as the inclusion of family fixed effects ensures that only the dizygotic twins and siblings identify the fixed effect IV estimates of  $\beta_2$ . The measure of genetic relatedness does not differ in theory between dizygotic twins and full siblings since 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 for 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.<sup>32</sup>

While one could imagine that data on the subsample of twins would provide the strongest 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.<sup>33</sup> 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 4.

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

in Table 4. 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. The change in the sign of the estimated impact of HD on test scores between Tables 3 and 4 is suggestive of other inputs in the production process being increased in response to the disorder within families.<sup>34</sup> While neither depression or obesity enter the equation in a statistically significant manner, the coefficient estimates for these are practically identical in magnitude and sign to those presented in Table 3 and it is important to stress that we have a very small sample size in which we are able to identify effects, leading to larger standard errors. Lastly, Hausman tests between columns 2 and 6 of Table 4 reject both the exogeneity of the health vector and that family fixed effects estimators can fully solve the endogeneity problem.

We believe that the estimates in Table 4 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, race and gender. With the exception of health and education outcomes, the only other measures contained in our data for which there are different values within kids in these families are genetic markers. As noted above, these results are also robust to including birth weight controls. The fixed effect-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.<sup>35</sup>

### 4.3 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 IV 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. For the full sample and subsample of twins, the smallest of the p-values for these tests is respectively 0.29 and 0.35, providing little evidence against the overidentifying restrictions.<sup>36</sup>

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 calculated the test statistic proposed by Stock and Yogo [2005].<sup>37</sup> 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 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.<sup>38</sup> 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. For the full sample and 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.<sup>39</sup> 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.<sup>40</sup> 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. Finally, tests of the validity of the instrument set for both the subsample of same-sex twins and siblings continue to suggest that this set of genetic markers has good statistical properties and

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 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.<sup>41</sup> 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. We successively increased by 5% increments the amount of exogeneity error from 0% to 90% of the reduced form impacts. At levels below 45% 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 the extent of deviations from the exact exclusion restrictions are assumed to be above 65% 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 given the structure of an education production function derived from the underlying model as well as the inclusion of birth weight variables in the estimating equations. The sensitivity analysis suggests that our quantitative results are robust to potentially mild and moderate violations of the exogeneity assumption, further increasing our confidence in both Tables 3 and 4. As we discuss in the next subsection, this finding is important as there is a potential additional concern from omitting comorbid conditions.

#### **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. Using only a single health outcome to proxy for health could lead to different results, since health disorders and risky health behaviors are known in the medical literature to be more common among individuals with one particular disorder than among the remaining population. Table 5 demonstrates the substantial presence of comorbidities in our sample. Column 1 of Table 5 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 any health disorder are more likely to smoke cigarettes.

The majority of the empirical literature that estimates the impact or association of health with socioeconomic outcomes generally include only a single explanatory measure such as obesity, smoking or birth weight in their analysis. We considered what would happen to the sign, significance

and magnitude of the estimated impact of each specific disorder if we followed the usual practice and did not control for comorbidities in the achievement equation. It is reasonable to hypothesize that in OLS and family fixed effects strategies, omitted variable bias would arise, since many of the neglected health conditions would be correlated with both the included health condition as well as verbal test scores. Further, in these specifications, IV or family fixed effects IV estimates may not overcome these biases, unless a subset of the genetic instruments are known to be scientifically unique to that included health condition to ensure the plausibility of the exclusion restriction assumption. Excluding significant comorbid conditions potentially leads to problems not only with sets of genetic markers as instruments, but makes it equally difficult to imagine that any nurture or environmental factor could break the statistical association between those included and excluded to the estimating equation measures of poor health.<sup>42</sup>

Table 6 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.<sup>43</sup> Thus, each entry in Table 6 refers to the point estimate of that specific health outcome on verbal achievement, controlling for the same set of observed controls as in Table 3. The empirical estimates of several disorders differ from that obtained using the full health vector reported in Table 3. In the OLS regressions reported in Table 6, HD no longer enters significantly and the magnitude of the impact of AD is substantially smaller. The fixed effects results in Table 6 are very similar to those obtained in Table 3, 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 6 and Table 3 in the OLS and fixed effects analysis.

The IV estimates in Table 6 differ greatly and it could be concluded 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 3. ADHD is highly negatively related to test scores and enters in a significant manner at the 15% level. The estimated impact of being overweight now be-

comes significant at the 15% level and leads to a seven point increase in test scores on average when estimating equation (1) using IV analysis. Regarding the preferred fixed effects IV specifications from Table 6, we would conclude that AD and ADHD each has a negative and significant impact on 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 3, the estimated impact of depression decreases substantially when family fixed effects and instrumental variables are used to estimate equation (1). Finally, sensitivity analysis for all IV and family fixed effects IV estimates in Table 6 indicate that they are extremely sensitive to the degree in which the exclusion restriction assumption is potentially violated. None of the results remain significant at very low levels of exogeneity error (5-10% of the reduced form impacts), confirming that ignoring comorbid conditions leads to the exclusion restriction assumption becoming implausible.

In our application, there may be a concern that the genetic markers used in the above analysis may also be associated with health measures not available in the data. Recall in the prior subsection we reported that our full set of results in Tables 3 and 4 become statistically insignificant only if the extent of deviations from the exact exclusion restrictions are assumed to be above 65% of the reduced form impacts is suggestive that if there are other channels they only can account for roughly one-third of these markers impacts. This is nearly a seven-fold increase then when we examine one of the health conditions in isolation. While there clearly may be other channels through which these markers affect academic outcomes other than the poor health outcomes the extent to which they appear to be a threat appears quite limited. An initial exhaustive survey of PubMed indicates two potential disorders: schizophrenia and Tourette's syndrome. We believe that this concern is unlikely to be a serious threat to our main results as schizophrenia does not manifest itself among adolescents and Tourette's syndrome is extremely uncommon, with current estimates indicating that it affects approximately 0.5 to 3 people in 1000.<sup>44</sup> A subsequent search further identified other poor health outcomes that either have low prevalence rates among youth such as panic disorder and psychoses,

or low discordance rates within families such as disorganized attachment, cholesterol (including HDL and triglyceride levels) and personality disorders where heritability has been reported to be .78 and .79 for narcissistic and obsessive-compulsive disorders. Thus, we do not believe that this issue is a major concern for either the IV or fixed effects IV estimates of the specifications with the full health vectors reported earlier, but it remains an empirical question.

Overall, this investigation clearly demonstrates that controlling for comorbid conditions is an important issue to credibly estimate the impact of specific health conditions on educational outcomes. We find that there are numerous differences in the estimated impacts of mental health disorders when estimating equation (1) by OLS, IV and family fixed effects with IV, depending on whether one comorbid conditions are accounted for in the specifications. To summarize, constructing an appropriate health vector presents an additional challenge for empirical researchers, as the omission of comorbid conditions could lead to either biases in coefficient estimates or invalidate exclusion restriction assumptions.

## 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 directly explore a variable that empirical researchers traditionally viewed as unobserved heterogeneity. In this paper, we exploit variation within siblings and twins from the "genetic

lottery" to test whether family fixed effects estimators by themselves can fully solve the endogeneity problem when estimating the impact of several poor health conditions on academic outcomes via a family fixed effects instrumental variables strategy.

Our results indicate that, while researchers should treat health as an endogenous input when estimating education production functions, family fixed effects estimators by themselves cannot fully remove the endogeneity bias. We present evidence that differences in genetic inheritance have desirable properties to identify the impact of poor health on education within families as there are, consistent with the biomedical literature, statistically significant correlations with each endogenous health variables and sensitivity analyses indicate that our results are robust to reasonable violations of the exclusion restriction assumption. Lastly, our results underscore the challenge facing empirical researchers interested in identifying the impact of specific health conditions that arises due to comorbidities.

In summary, we believe that integrating biological findings into the social sciences has the potential to not only address open research questions, test economic models, 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. However, within families, any differences in the inheritance of specific markers present the opportunity for additional experiments in "nature".

## Notes

<sup>1</sup>This debate has been traced back to 13th-century France and the field of quantitative behavioural genetics basically compares trait similarities across individuals that systemically differ in the genetic or environmental influences they have in common (e.g. identical vs. fraternal twins, adoptive vs. biological children), to decompose the variation of quantitative traits, and their covariances with other traits, into genetic and environmental (co)variance components. Within economics, Cesarini et al. [2008, 2009] utilize these methodologies to demonstrate that preferences for cooperative behavior, risk and giving have a significant genetic component.

<sup>2</sup>For example, see Johnson [2003], Kelada et al. [2003], Goldstein et al. [2003], Zerhouni [2003] and Merikangas and Risch [2003].

<sup>3</sup>Using similar methodologies, economists have begun to explore whether specific genetic loci are associated with financial risk preferences (e.g. Dreber et al [2009], Benjamin et al. [2009]).

<sup>4</sup>Within economics, the use of family fixed effects regressions to address questions related to differences in individual characteristics on outcomes has a long history that 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.

<sup>5</sup>Genes consist of two alleles, and a child randomly inherits one of the two alleles from each parent at the time of conception. The child's genome consists of approximately 3.2 billion base pairs, along which there are 9.2 million candidate SNPs (International HapMap Consortium, 2005), which are specific locations where a mutation in the genetic code is known to occur in the population. This variability in the genetic code may influence an individual's susceptibility to various developmental outcomes such as developing an illness. In other words, our empirical strategy exploits these differences in the coding of a specific marker between full siblings and can intuitively be viewed as an experiment in "nature".

<sup>6</sup>Ding et al. [2006, 2009] 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 [2008] use genetic information to attempt to estimate the impact of obesity on employment. Neither study exploited variation in genetic inheritance within families (the “genetic lottery”), which we show to be important empirically and improves the plausibility of the exclusion restriction.

<sup>7</sup>Two 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.

<sup>8</sup>Plomin 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. Using maps of the location between these genes and the specific genetic markers in our study, we find no evidence that they are located closely on the genome, suggesting that linkage in inheritance is unlikely. Researchers have found no direct links between several of the genes in this study and intelligence (i.e. Moises et al. [2001]) or cognitive ability (e.g. Petrill et al. [1997]), and we hypothesize that if a link exists, that it operates through specific health measures.

<sup>9</sup>In the results section, we present evidence from a battery of statistical tests that the genetic lottery has good properties to identify the impact of poor health conditions.

<sup>10</sup>Similarly large negative impacts of poor health on measures of later cognitive achievement have been found in studies that exploit shocks to an individual’s prenatal conditions such as in utero exposure to the flu (Almond, 2006) and low levels of radiation (Almond, Edlund and Palme, 2008).

<sup>11</sup>This result may also have implications for genetic association studies that typically examine

specific health conditions or endophenotypes in isolation.

<sup>12</sup>Add 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. Additional details on the construction of the sample are provided in Harris et al. [2003].

<sup>13</sup>For example, Babinski et al. [1999], Ding et al. [2009], and Fletcher and Wolfe [in press] present empirical evidence of different impacts from these two diagnoses.

<sup>14</sup>The 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 the illustration that is the closest match to the word from four simple black-and-white illustrations. The test is arranged in a multiple-choice format.

<sup>15</sup>See <http://www.agsnet.com/assessments/technical/ppvt.asp> for details.

<sup>16</sup>Complete 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 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.

<sup>17</sup>More recently, evidence indicates that differences within families, even among identical twins, can exist because of epigenetic factors. Epigenetics refer 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, 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).

<sup>18</sup>Pleiotropy 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 may also affect the expression of other genes, which together leads to changes in behaviors and outcomes.

<sup>19</sup>The 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, it continues 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 with feelings of enjoyment.

<sup>20</sup>The 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.

<sup>21</sup>For example, Dremencov et al. [2004] present evidence that allele variants 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 addicts 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 allele variants of the DRD2 genes.

<sup>22</sup>Ding et al. [2009] present evidence that this distinction is important empirically and that one should account for the endogeneity of health behaviors when estimating health production functions. As our focus in this manuscript is on contrasting the family fixed effects estimator to the family fixed effects IV estimator of the education production function, we will not discuss estimates of the health production function.

<sup>23</sup>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.

<sup>24</sup>An equally important challenge occurs in measuring the health vector from omitted variables. If the researcher omits comorbid conditions, biased estimates of the impacts of poor health on academic outcomes will be recovered. This empirical challenge is discussed in detail in Section 4.4 of the text.

<sup>25</sup>It is worth stating explicitly that since Hahn and Hausman [2003] show that for a given correlation between invalid instruments and the error term, the maximal asymptotic bias is inversely related to the partial R-squared from the instruments, and in our setting these partial R-squared values are not small, resulting in these concerns to be not relevant in our application.

<sup>26</sup>Our results are robust to both including an individual's birth weight linearly as well as excluding birthweight from both estimating equations. Fletcher and Lehrer [2009] present the full analysis where birth weight is excluded from the estimating equation.

<sup>27</sup>The 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 data reported in the empirical table in order of the raw number of individuals who possess each

particular marker within that gene from lowest frequency to most common.

<sup>28</sup>Polygenic 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.

<sup>29</sup>To examine this concern we additionally examined the extent to which genetic linkages occurs in our sample. Cross-tabulations of different genetic combinations for both the full sample as well as by the first and second family member in the data are presented in Appendix Table 1 of Fletcher and Lehrer (2009) who report from tests for homogeneity of odds ratios to see whether possessing a specific allelic variant in one marker increases the odds of possessing a specific variant in a different genetic marker that there is no evidence indicating a systematic relationship between markers of any two of the genes for either sample that contains only one family member, lessening concerns regarding linkage.

<sup>30</sup>Recall that Table 3 demonstrated that significant correlations do indeed exist between health outcomes and the genetic markers in our data. To construct the instrument set, we considered two alternative strategies. 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, 2009] and this approach has the advantage of making it easier to replicate the study. The 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 five 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. These results are available from the authors upon request.

<sup>31</sup>While 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. There is also emerging evidence that children seem to outgrow HD symptoms to some extent but not AD symptoms.

<sup>32</sup>A 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.

<sup>33</sup>For 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]).

<sup>34</sup>We are grateful to Richard Blundell for identifying this difference.

<sup>35</sup>Results for the subsample of siblings are available upon request (and appear as Appendix Table 2 in Fletcher and Lehrer [2009]). continue to find that the AD condition leads to a significant decrease in test scores, the instrument set continues to have good first stage properties and Hausman tests suggest that the health vector should be treated as endogenous, and that family fixed effects by themselves do not remove all of the potential biases.

<sup>36</sup>Many 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.

<sup>37</sup>This 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.

<sup>38</sup>We did conduct Kleinbergen and Paap (2006) tests for the preferred instrument set reported in table 5 and 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.

<sup>39</sup>For health vector 1, the results are 48.03 and 51.62.

<sup>40</sup>The 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.

<sup>41</sup>Essentially, 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  $N(0, \delta^2)$ , where  $\delta$  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.

<sup>42</sup>For 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.

<sup>43</sup>The results in this subsection are robust to examining only the same-sex twin subsample.

<sup>44</sup>Note the Add Health data contains information on other health conditions including diabetes and asthma that do not have strong links to the genetic markers we used in Tables 3 and 4. Our full set of results were invariant to their inclusion in the estimating equations but they were weakly identified.

## References

- [1] Almond, D., Edlund, L. and M. Palme (2008). "Chernobyl's Subclinical Legacy: Prenatal Exposure to Radioactive Fallout and School Outcomes in Sweden." forthcoming in the *Quarterly Journal of Economics*.
- [2] Almond D. (2006). "Is the 1918 Influenza Pandemic Over? Long-term Effects of In Utero Influenza Exposure in the Post-1940 U.S. Population." *Journal of Political Economy*, 114(4), 672-712.
- [3] 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.
- [4] 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.
- [5] Behrman, J. R. and P. Taubman. (1976). "Intergenerational Transmission of Income and Wealth." *American Economic Review*, 66(2), 436-440.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] Benjamin, D., C. Chabris, E. I. Glaeser and D. Laibson. (2009). "Genetic Influences on Economic Outcomes." paper presented at 2009 AEA Annual meeting, San Francisco.
- [10] 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.
- [11] Bleakley, H. C. (2007). "Disease and Development: Evidence from Hookworm Eradication in the American South." *Quarterly Journal of Economics*, 122(1), 73-117.

- [12] Cesarini D, Dawes C. T., Johannesson M, Lichtenstein P, Wallace B. 2009. “Genetic Variation in Preferences for Giving and Risk-Taking.” *Quarterly Journal of Economics*, 124(2), 809–842.
- [13] Cesarini D, C. T. Dawes, J. H. Fowler, M. Johannesson, P. Lichtenstein, B. Wallace. (2008). “Heritability of Cooperative Behavior in the Trust Game.” *Proceedings of the National Academy of Sciences*, 105, 3721-3726.
- [14] Christensen, K., A. Wienke, A. Skytthe, N. V. Holm, J. W. Vaupel, and A. I. Yashin (2001), “Cardiovascular mortality in twins and the fetal origins hypothesis.” *Twin Research* 4, 344–349.
- [15] Conley, T., C. Hansen and P. E. Rossi. (2007). “Plausibly Exogenous.” *mimeo*, University of Chicago.
- [16] Cooper R. S., J. S. Kaufman and R. Ward. (2003). “Race and Genomics.” *The New England Journal of Medicine*, 348(12), 1166–1170.
- [17] Cragg, J. G., and S. G. Donald. (1993). “Testing Identifiability and Specification in Instrumental Variables Models.” *Econometric Theory* 9, 222–240.
- [18] 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.
- [19] 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*.
- [20] 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.
- [21] Cutler, D. and A. Lleras-Muney. (2007). “Education and Health: Evaluating Theories and Evidence.” *NBER Working Paper w12352*.
- [22] 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.
- [23] Ding, W., S. F. Lehrer, J. N. Rosenquist and J. Audrain-McGovern. (2009). “The Impact of Poor Health on Academic Performance: New Evidence Using Genetic Markers.” *Journal of Health Economics*, 28(3), 578–597.

- [24] 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*.
- [25] Dreber, A., C. L. Apicella, D. T. A. Eisenberg, J. R. Garcia, R. Zamore, J. K. Lum and B. C. Campbell. (2009). “The 7R Polymorphism in the Dopamine Receptor D4 Gene (DRD4) is Associated with Financial Risk-Taking in Men.” *Evolution and Human Behavior*, 30(2), 85–92.
- [26] 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.
- [27] Fletcher, J. M. and S. F. Lehrer. (2009). “Using Genetic Lotteries within Families to Examine the Causal Impact of Poor Health on Academic Achievement.” *NBER Working Paper w15148*.
- [28] Fletcher, J. M. (2008). “Adolescent Depression and Educational Attainment: Evidence from Sibling Fixed Effects.” *Health Economics*, 17: 1215-1235
- [29] Fletcher, J. M. and B. L. Wolfe. (in press). “Long-term Consequences of Childhood ADHD on Criminal Activities.” *Journal of Mental Health Policy and Economics*
- [30] Fletcher, J. M. and B. L. Wolfe. (2008). “Child Mental Health and Human Capital Accumulation: The Case of ADHD Revisited.” *Journal of Health Economics*, 27(3): 794-800
- [31] 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.
- [32] Goldstein D. B., S. K. Tate and S. M. Sisodiya. (2003). “Pharmacogenetics Goes Genomic.” *Nature Reviews Genetics*, 4, 937-947.
- [33] 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.
- [34] Gorseline, D.W. (1932). *The Effect of Schooling Upon Income*. (Bloomington: Indiana University Press).
- [35] 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.

- [36] 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.
- [37] Gruber, J. and M. Frakes. (2006). "Does Falling Smoking Lead to Rising Obesity?" *Journal of Health Economics*, 25, 183–197.
- [38] Hahn, J., Hausman, J. 2003. "Instrumental Variable Estimation with Valid and Invalid Instruments." *MIT Department of Economics Working Paper No. 03-26*.
- [39] Hanushek, E. (1992). "The Trade-off between Child Quantity and Quality." *Journal of Political Economy*, 100 84-117.
- [40] 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.
- [41] The International HapMap Consortium. (2005). "A Haplotype Map of the Human Genome." *Nature*, 437 1299-1320.
- [42] Jain, A.K., S. Prabhakar, and S. Pankanti. (2002). "On the Similarity of Identical Twin Fingerprints." *Pattern Recognition*, 35:2 653-2663.
- [43] Johnson J. A. (2003). "Pharmacogenetics: Potential for Individualized Drug Therapy Through Genetics." *Trends Genetics*, 19:6 60–66.
- [44] 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.
- [45] Kaester, R., M. Grossman. (2008). "Effects of Weight on Children's Educational Achievement." NBER Working Paper 13764.
- [46] 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.
- [47] Kleibergen, F., and R. Paap. (2006). "Generalized Reduced Rank Tests Using the Singular Value Decomposition." *Journal of Econometrics* 127(1), 97–126.

- [48] 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.
- [49] Kremer M. and E. Miguel. (2004). “Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities.” *Econometrica*, 72, 159-217.
- [50] Lunde, A., K. K. Melve, H. K. Gjessing, R. Skjaerven, and L. M. Irgens (2007). “Genetic and Environmental Influences on Birth Weight, Birth Length, Head Circumference, and Gestational Age by Use of Population-based Parent-Offspring Data.” *American Journal of Epidemiology* 165(7): 734–741.
- [51] Merikangas K. R. and N. Risch. (2003). “Genomic Priorities and Public Health.” *Science* 302, 599–601.
- [52] 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.
- [53] Neumark, D. (1999). “Biases in Twin Estimates of the Return to Schooling.” *Economics of Education Review*, 18, 143-148.
- [54] Norton, E.C. and E. Han. (2008). “Genetic Information, Obesity, and Labor Market Outcomes.” *Health Economics*, 17(9), 1089–1104.
- [55] Olds, J., Milner, P. (1954). “Positive Reinforcement Produced by Electrical Stimulation of Septal Area and Other Regions of Rat Brain.” *Journal of Comparative and Physiological Psychology*, 47, 419–427.
- [56] Perri, T. J. (1984). “Health Status and Schooling Decisions of Young Men.” *Economics of Education Review*, 3, 207-213.
- [57] 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.
- [58] 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.

- [59] 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.
- [60] Roberts, R.E., Lewinsohn, P.M., and J.R. Seeley. (1991). "Screening for Adolescent Depression: A Comparison of Depression Scales." *Journal of the American Academy of Child & Adolescent Psychiatry*. 30(1): 58-66
- [61] Rosenzweig, M. R. and K. I. Wolpin. (2000). "Natural" Natural Experiments in Economics." *Journal of Economic Literature*, 38, 827-874.
- [62] Royer, H. (2009). "Separated at Girth: US Twin Estimates of the Effects of Birth Weight." *American Economic Journal: Applied Economics*, 1(1), 49-85.
- [63] 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.
- [64] Strauss, J. and D. Thomas. (1998). "Health, Nutrition, and Economic Development." *Journal of Economic Literature*, 36(2), 766-817.
- [65] 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.
- [66] Taubman, P. (1976b). "Earnings, Education, Genetics, and Environment." *Journal of Human Resources*, 11(4), 447-461.
- [67] 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)
Birth Weight (lbs)	6.68 (1.494)	7.30 (1.229)	5.60 (1.269)
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: Relationship between Genetic Markers and Health Outcomes

Gene	Variant	ADHD	AD	HD	Obese	Depression	Smoking
DRD2	A1A1	0.076 (0.266) [0.987]	0.038 (0.192) [0.734]	0.053 (0.224) [1.103]	0.061 (0.240) [0.822]	0.053 (0.225) [0.840]	0.220 (0.416) [0.879]
	A1A2	0.071 (0.257) [0.876]	0.054 (0.225) [1.130]	0.038 (0.191) [0.671]+	0.072 (0.259) [1.014]	0.071 (0.257) [1.280]	0.237 (0.426) [0.967]
	A2A2	0.081 (0.273) [1.136]	0.049 (0.216) [0.963]	0.056 (0.229) [1.398]+	0.073 (0.260) [1.041]	0.057 (0.231) [0.827]+	0.246 (0.431) [1.071]
SLC6A4	Two short alleles	0.058 (0.234) [0.700]	0.032 (0.176) [0.576]*	0.038 (0.191) [0.726]	0.067 (0.250) [0.912]	0.076 (0.265) [1.328]	0.223 (0.417) [0.882]
	One short/one long allele	0.084 (0.278) [1.218]	0.058 (0.234) [1.362]	0.051 (0.221) [1.111]	0.072 (0.259) [1.017]	0.054 (0.226) [0.781]	0.230 (0.421) [0.900]
	Two long alleles	0.077 (0.267) [1.016]	0.050 (0.218) [0.998]	0.052 (0.221) [1.097]	0.074 (0.262) [1.047]	0.064 (0.244) [1.049]	0.265 (0.442) [1.222]*
DAT1	No 10 repeats	0.065 (0.247) [0.823]	0.032 (0.178) [0.621]	0.043 (0.204) [0.872]	0.032 (0.178) [0.416]+	0.054 (0.227) [0.856]	0.194 (0.397) [0.745]
	One ten repeat	0.088 (0.284) [1.279]	0.059 (0.236) [1.324]	0.059 (0.236) [1.381]	0.078 (0.268) [1.147]	0.062 (0.242) [1.017]	0.241 (0.428) [1.005]
	Two ten repeats	0.071 (0.257) [0.822]	0.046 (0.210) [0.832]	0.043 (0.204) [0.754]	0.072 (0.259) [1.005]	0.062 (0.241) [1.016]	0.244 (0.430) [1.057]
DRD4	No seven repeats	0.082 (0.274) [1.125]	0.052 (0.223) [1.172]	0.051 (0.219) [1.128]	0.073 (0.260) [1.039]	0.066 (0.249) [1.256]	0.242 (0.429) [1.025]
	One seven repeat	0.070 (0.255) [0.866]	0.047 (0.212) [0.919]	0.045 (0.208) [0.896]	0.068 (0.252) [0.917]	0.058 (0.235) [0.920]	0.242 (0.428) [1.006]
	Two seven repeats	0.044 (0.207) [0.546]	0.029 (0.170) [0.567]	0.044 (0.207) [0.898]	0.088 (0.286) [1.263]	0.015 (0.121) [0.219]*	0.209 (0.410) [0.827]
CYP	Main SNP	0.076 (0.265) [0.822]	0.049 (0.215) [0.604]	0.049 (0.216) [1.275]	0.073 (0.260) [1.433]	0.061 (0.239) [0.769]	0.237 (0.426) [0.687]+
MAOA	No four repeats	0.075 (0.264) [0.973]	0.046 (0.209) [0.875]	0.050 (0.217) [1.025]	0.075 (0.264) [1.074]	0.069 (0.254) [1.198]	0.235 (0.424) [0.953]
	One four repeat	0.046 (0.209) [0.507]***	0.028 (0.165) [0.477]**	0.030 (0.172) [0.546]*	0.061 (0.239) [0.795]	0.081 (0.273) [1.491]*	0.218 (0.414) [0.848]
	Two four repeats	0.093 (0.291) [1.547]**	0.064 (0.245) [1.735]**	0.057 (0.233) [1.420]+	0.075 (0.264) [1.100]	0.047 (0.212) [0.616]**	0.256 (0.437) [1.169]

Note: Each cell presents the conditional mean, the standard deviation in round parentheses and the odds ratio for outcomes (excluding BMI) in square parentheses. \*\*\*, \*\*, \*, +, denote the Null of homogeneity of odds across markers by genotype from a chi-squared test is rejected at the 1%, 5%, 10%, and 15% level respectively. The tests were conducted with the same sample used to construct Table 1.

Table 3: Estimates of the Achievement Equation for the Full Sample

Estimation Approach	OLS		Family Fixed Effects		Instrumental Variables		Family Fixed Effects Instrumental Variables	
AD	N/A	-3.627*** (1.356)**	N/A	-2.179 (1.451)	N/A	-22.61* (12.49)	N/A	-23.28* (12.87)
HD	N/A	2.277* (1.363)	N/A	1.885 (1.730)	N/A	20.73 (17.60)	N/A	8.982 (15.59)
ADHD	-1.227 (1.021)	N/A	-0.176 (1.247)	N/A	-4.312 (12.67)	N/A	-9.659 (14.53)	N/A
Depression	-4.341*** (1.365)**	-4.291*** (1.364)*	-2.288 (1.437)	-2.319 (1.444)	-10.046 (17.953)	-2.288 (1.437)	9.057 (9.609)	-8.508 (15.57)
Obesity	-0.599 (0.737)	-0.581 (0.735)	-0.823 (0.911)	-0.799 (0.907)	3.335 (7.661)	-0.823 (0.911)	3.712 (4.019)	7.844 (7.478)
Age	4.838 (3.141)	4.815 (3.137)	1.582 (3.963)	1.282 (3.956)	4.659 (3.829)	1.582 (3.963)	-0.113 (0.128)	1.141 (5.462)
Age squared	-0.153 (0.0985)	-0.151 (0.0984)	-0.0393 (0.124)	-0.0286 (0.124)	-0.141 (0.115)	-0.0393 (0.124)	1.028 (1.182)	-0.00533 (0.172)
Male	1.144* (0.603)	1.117* (0.602)	-0.512 (0.759)	-0.528 (0.759)	1.668 (1.076)	-0.512 (0.759)	-9.087*** (1.214)	-0.623 (1.140)
African American	-8.854*** (0.872)	-8.883*** (0.869)			-9.461 (1.130)**	-8.125*** (2.087)		
Hispanic	-7.224*** (0.972)	-7.232*** (0.970)			-7.755 (1.668)**	0.178 (1.063)		
Sibling	0.390 (0.771)	0.363 (0.768)			0.237 (0.934)	-1.180*** (0.429)		
Birth order	-1.145*** (0.317)	-1.153*** (0.317)	-1.413 (0.871)	-1.381 (0.872)	-1.240 (0.398)**	-1.413 (0.871)	-1.344 (1.159)	-0.869 (1.146)
Birth weight	27.54 (20.40)	27.31 (20.36)	40.36 (26.53)	41.34 (26.67)	34.231 (31.27)	23.201 (35.53)	61.93 (37.76)	52.33 (37.42)
Birth weight squared	-7.066 (4.997)	-7.013 (4.988)	-10.10 (6.603)	-10.32 (6.631)	-8.874 (7.578)	-6.156 (8.632)	-15.41* (9.303)	-13.22 (9.171)
Birth weight cubed	0.776 (0.519)	0.771 (0.518)	1.094 (0.693)	1.116 (0.695)	0.977 (0.775)	0.699 (0.884)	1.633* (0.965)	1.419 (0.948)
Birth weight <sup>4</sup>	-0.0307 (0.0194)	-0.0305 (0.0194)	-0.0431* (0.0261)	-0.0438* (0.0262)	-0.029 (0.032)	-0.0386 (0.0284)	0.0626* (0.036)	-0.0550 (0.0352)
Family Income	0.0209*** (0.00620)	0.0205*** (0.00612)			0.021 (0.008)**	1.295*** (0.409)		
Maternal Years of Education	1.167*** (0.155)	1.162*** (0.155)			1.301 (0.371)**	0.223** (0.0901)		
Parents Age	0.261*** (0.0643)	0.258*** (0.0643)			0.249 (0.080)**	-0.482 (1.127)		
Parents Married	-0.308 (0.762)	-0.297 (0.762)			-0.007 (0.953)	-6.686 (20.82)		
Observations	1684	1684	1684	1684	1684	1684	1684	1684

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

Table 4: 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	-6.277** (2.464)	N/A	-3.662 (2.401)	N/A	-6.001 (6.438)	N/A	-18.44** (7.523)
HD	N/A	2.597 (2.811)	N/A	-0.576 (4.051)	N/A	-7.150 (9.408)	N/A	-14.38 (8.944)
ADHD	-3.961 (1.949)	N/A	-2.881 (2.671)	N/A	-15.66 (21.10)	N/A	-25.82*** (8.477)	N/A
Depression	-1.539 (2.565)	-1.606 (2.537)	2.014 (2.148)	1.876 (2.078)	-9.268 (18.84)	-4.095 (13.89)	-3.639 (19.93)	-6.567 (19.56)
Obesity	-3.532*** (1.362)	-3.571*** (1.360)	-0.945 (1.860)	-0.877 (1.852)	-6.357 (9.405)	-1.653 (8.169)	-0.133 (4.845)	3.596 (5.129)
Male	4.579*** (1.148)	4.515*** (1.143)			4.804*** (1.629)	4.718*** (1.442)		
African American	-6.843*** (1.501)	-6.809*** (1.497)			-7.207*** (2.360)	-7.192*** (2.269)		
Hispanic	-7.134*** (1.915)	-7.238*** (1.908)			-7.164** (3.115)	-7.706** (3.052)		
Birth weight	121.9 (77.09)	123.1 (77.38)	140.3 (88.05)	147.4* (88.00)	145.1 (95.35)	136.7 (83.16)	162.9 (119.4)	174.2 (120.1)
Birth weight squared	-32.79 (21.66)	-33.18 (21.76)	-38.43 (25.10)	-40.48 (25.11)	-39.24 (26.93)	-36.95 (23.34)	-43.12 (33.67)	-46.90 (33.91)
Birth weight cubed	3.770 (2.595)	3.818 (2.609)	4.455 (3.027)	4.707 (3.029)	4.535 (3.238)	4.263 (2.791)	4.830 (4.025)	5.360 (4.057)
Birth weight <sup>4</sup>	-0.157 (0.112)	-0.159 (0.113)	-0.185 (0.131)	-0.196 (0.131)	-0.190 (0.141)	-0.178 (0.121)	-0.194 (0.173)	-0.220 (0.175)
Family Income	0.0108** (0.00476)	0.0110** (0.00480)			0.00736 (0.00806)	0.0116* (0.00675)		
Maternal Years of Education	1.104*** (0.256)	1.083*** (0.256)			1.078*** (0.401)	1.094*** (0.368)		
Parents Age	0.266** (0.115)	0.273** (0.115)			0.282* (0.147)	0.264* (0.148)		
Parents Married	-0.666 (1.412)	-0.735 (1.410)			-1.016 (1.899)	-1.005 (1.863)		
Observations	469	469	469	469	469	469	469	469

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

Table 5: Relationship Between Health Behaviors and Health Outcomes During Adolescence

Behavior	Total Number	Nothing Else <sup>1</sup>	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.

<sup>1</sup> For ADHD nothing else excludes AD and HD.

Table 6: 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	-3.879*** (1.368)	-2.254 (1.462)	-25.58** (12.31)	-19.17* (10.56)
HD	2.309* (1.378)	1.787 (1.725)	25.69 (16.71)	3.960 (13.38)
ADHD	-1.440 (1.028)	-0.287 (1.251)	-9.881 (6.450)	-15.39* (8.873)
Depression	-0.727 (0.737)	-0.787 (0.915)	11.15 (8.220)	-1.478 (6.224)
Obesity	-4.437*** (1.361)	-2.306 (1.438)	-22.24 (15.01)	-8.236 (12.29)

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

AD	-2.900** (1.205)	-1.614 (1.338)	-12.89* (6.944)	-17.63* (9.366)
HD	0.673 (1.186)	0.906 (1.579)	1.663 (10.40)	-11.78 (11.68)

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.