The Effects of Adolescent Health on Educational Outcomes: Causal Evidence using 'Genetic Lotteries' between Siblings[•♦]

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

Steven F. Lehrer
Queen's University and NBER

Abstract

There has been growing interest in using specific genetic markers as instrumental variables in attempts to assess causal relationships between health status and socioeconomic outcomes, including human capital accumulation. In this paper we use a combination of family fixed effects and genetic marker instruments to show strong evidence that inattentive symptoms of ADHD in childhood and depressive symptoms as an adolescent are linked with years of completed schooling. Our estimates suggest that controlling for family fixed effects is important but these strategies cannot fully account for the endogeneity of poor mental heath. Finally, our results demonstrate that the presence of comorbid conditions present immense challenges for empirical studies that aim to estimate the impact of specific health conditions.

• The authors gratefully acknowledge research support from SSHRC (Lehrer) and CLSRN (Fletcher and Lehrer). We thank Erdal Tekin and participants at the 2008 Southern Economic Association Annual Meetings for helpful comments. 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).

Introduction

A large body of research documents that children who experience poor health have significantly worse adult outcomes, including lower educational attainment, adverse health conditions, and lower social status (Case et al., 2002; Case et al., 2005). A particularly potent conduit through which childhood health is linked to adult outcomes is education. Poor health impedes educational progress because a student with health problems is not prepared to fully engage in or take advantage of learning opportunities at school or at home (Hanson et al., 2004).

Schools have long recognized the relation between student health and educational progress, and have played a role in diagnosing and treating student health conditions such as vision, hearing, and speech impairments, as well as asthma, mental disorders, and more recently obesity (Council of Chief State School Officers, 1998).[1] While much of the research in the social sciences that links poor health in childhood to later outcomes is focused on low birth weight (e. g. see Almond et al. 2005 and the references therein), there has been less work focused on other conditions that arise in early childhood and adolescence.[2] As recent years have been characterized by schools scaling back health programs in order to devote more resources towards improving student test performance (Constante, 2002; Deutsch, 2000; Hanson et al., 2004), it remains an open question of whether improving health status due to disorders that arise while in school could have long run benefits.[3] Understanding whether poor health in childhood has impacts on future outcomes is important as a large body of research documents the strong association between economic growth and a variety of education measures that includes, the levels of

[1] Research from the medical community confirms that common health conditions can have negative consequences on children's ability to learn. Vision problems in children are associated with developmental delays and often require special education and additional services beyond childhood (Centers for Disease Control, 2004). Children with asthma miss more days of school than children without asthma, and experience restrictions in other daily activities, such as play and sports (Newacheck, 2000). Significant hearing loss among children can interfere with phonological and speech perception abilities required for language learning, which subsequently can lead to poor academic performance, especially in reading (National Institutes of Health, 1993). Children with speech impairments perform more poorly on reading tests than children in non-impaired comparison groups (Catts, 1993).

[2] A large body of literature (surveyed in the next section) examines the contemporaneous impact of health on academic outcomes.

[3] A large body of research has indeed demonstrated that improving schooling not only boosts future education outcomes but also the development of cognitive skills (Heckman et al. 2004) and non-cognitive skills (Cunha and Heckman, 2007)

specific cognitive skills, average years of education completed and academic performance on international tests or report cards.

While there is a well-established, large positive correlation between mental and physical health and education outcomes, establishing a causal link remains a substantial challenge. The estimated relationships between health and education found in the literature can arise through three mechanisms that are not mutually exclusive: (1) there could be a causal effect of health on education (Becker 1962) (2) there could be reverse causality where previous education outcomes led to health (Grossman 1972) and (3) there could be "third factors" such as time preference, unmeasured ability, or other omitted variables that leads to a spurious correlation (Fuchs 1982).

In this paper we employ the "genetic lottery" identification strategy introduced in Fletcher and Lehrer (2008) to examine the long run impacts of several poor childhood health measures on years of schooling. This strategy leverages the biological feature that siblings who share the same parents had an equal chance of inheriting specific genetic variation. Thus, our identification strategy exploits an "experiment in nature" by examining genetic differences between full biological siblings, including dizygotic twins. Intuitively, since nearly every behavior and health condition can be argued to have a specific genetic basis, exploiting differences in inheritance of a subset of these specific markers can potentially elucidate their causal impact on other factors. We term the manner in which a child inherits one copy from each parent's own genetic pair at conception, where this copy is selected from the parent randomly as the "genetic lottery."[4] While this paper uses this new research design to estimate the relationship between health and education outcomes, we believe that the strategy has much wider applicability and has the ability to generate estimates of causal effects and elucidate causal pathways in a number of critical areas in both social science and health services research.

[4] Essentially this strategy combines a family fixed effect with an instrumental variables estimator. However, unlike traditional econometric applications of instrumental variables it is insufficient just to examine the statistical properties of the instrument to argue their validity. Ding et al. (in press) discuss which scientific and statistical conditions must be verified to justify the use of specific genetic markers as instrumental variables. Ding et al. (in press) also discuss the importance of examining the sensitivity of one's estimates to the extent to which fundamentally untestable statistical assumptions are met. We conduct the full set of these tests in our analysis.

In this paper, we focus on the link between health and education because unraveling the mechanisms linking health and educational outcomes is of primary importance in our ability to suggest policies to increase population health and the stock of human capital. We present compelling evidence of large impacts of poor mental health on years of schooling completed. Our estimates suggest that accounting for family fixed effects is important but these strategies cannot fully account for the endogeneity of poor mental health. Finally, our results demonstrate that the presence of comorbid conditions present immense challenges for empirical studies that aim to estimate the impact of specific health conditions.

This paper is organized as follows. In the next section we summarize the related literature that examines how researchers estimate the causal impact of health on socioeconomic outcomes. The economic model that underlies our analysis is introduced in Section 3. We draw a distinction between early measures of human capital that were produced on the basis of parental decisions and subsequent human capital investment decisions, which are made by individuals themselves. The model suggests that early human capital measures that include the stock of child and adolescent health may affect the productivity of education investment decisions and subsequent earnings in adulthood and as the stock of these variables are a result of behavioral choices they should be treated as endogenous in the estimation of earnings equations and when considering schooling decisions.[5] We introduce the data and discuss the econometric methods we use to estimate the empirical model in Section 4. A brief overview of the scientific literature that suggests there exists valid genetic instruments to identify the causal impact of health conditions on a series of years of schooling is also provided. We present and discuss the empirical results in Section 5. A concluding section summarizes our findings and suggests the broad applicability of our research design.

2 Background

This study builds on an incredibility rich and voluminous literature linking health and education (and education and health). As ethical considerations would prevent

[5] In his investigation of the effect of health on schooling, Perri (1984) constructs a simple model in the spirit of Weiss (1971) that treats the stock of health as exogenous.

researchers from conducting randomized experiments that permit causal relationships to be estimated between developing poor health conditions and other human capital outcomes, there are few empirical studies in the literature that would meet the "gold standard" of a randomized control trial.[6] Thus, researchers have generally pursued the further examination of the links between health and education using data with rich control variables, quasi-natural experiments, instrumental variables, and/or sibling/twin differences. Grossman and Kaestner (1997) provide the first comprehensive survey of this literature and conclude that the preponderance of research examines the effect of education on health, but among those papers that investigated the effects of health on education, they suggest large causal effects linking health and education.[7] Cutler and Lleras-Muney (2006) update this survey and provide an important overview of the short and long-term effects of education on many measures of health status.

Recent years have been characterized by an increasing number of studies that have attempted to investigate the impact of adolescent health on academic outcomes in adolescence. Several studies have focused on the effect of obesity and relied on including rich controls. Datar and Sturm (2006) and Datar et al. (2004) find that being overweight is associated with lower test scores in elementary school. Similarly, Sabia (2007) finds a negative correlation between being overweight and grade point average. In contrast, Kaestner and Grossman (2008) find that, in general, children who are overweight or obese have test scores that are about the same as children with average weight.

As noted above, experimental and quasi-experimental approaches have been examined, providing some limited information on the effect of several specific health outcomes on education outcomes in certain narrow populations. 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

[6] It is feasible to conduct randomized experiments that provide access to treatment for disorders and we survey papers that use this approach below.
[7] Strauss and Thomas (1998) provide a similar survey linking health and income.

return to schooling later in life. While this set of papers has contributed to our knowledge of whether a health-education link exists, there are far fewer experiments that can inform us on the impacts of common mental and physical conditions that currently affect children and adolescents in North America.

Researchers have used within-family comparisons to examine the effects of each of t he health status variables used in this study (ADHD, depression, and overweight) on education, to some extent. This approach eliminates any shared environmental factors of siblings, including much or all genetic endowments and exposure to other family or neighborhood factors that are common to the siblings or twins. Currie and Stabile (2006) use between-sibling comparisons in the US and Canadian versions of the NLSY to link childhood ADHD symptoms with lower childhood education outcomes such as grade repetition and achievement on test scores. Currie and Stabile (2006) find that ADHD has large negative effects on test scores and schooling attainment. Fletcher and Wolfe (2008a) use a similar approach with the Add Health data and show that the longer term educational effects of childhood ADHD are more modest.[8] Fletcher (2008a,b) uses the Add Health data to link adolescent depression with several educational outcomes, such as high school dropout, years of schooling, and college enrollment. Kaestner and Grossman (2008) use individual fixed effects rather than sibling fixed effect to show that overweight status likely has little impact of educational outcomes. Fletcher and Lehrer (2008) present evidence that using sibling fixed effects by themselves would lead one to conclude that there is no link between overweight and education outcomes.[9]

The main conceptual limitation of the family fixed effects approach is that after removing the common family effect, the remaining source of variation that leads one sibling to have poor health and the other sibling to have good health is largely unknown and is implicitly assumed to be conditionally random to variations in education outcomes between family members. Consistent with this conjecture, Fletcher and Lehrer (2008)

[8] Fletcher and Wolfe (2008b) use the same approach and data to show a link between ADHD and criminal activities as a young adult.

[9] The authors show that the results are sensitive to controlling for comorbid conditions, including a change in the sign of the estimated coefficient across specifications. The difference in these results suggest that the sensitivity in the estimated associations between obesity and education outcomes that have been found in the literature could be due to differences in which other health conditions are being controlled for..

present evidence from Hausman tests that rejects the exogeneity of the family fixed effect estimates.

The third empirical approach often used to estimate the effects of health on education is the instrumental variables method. The great challenge for researchers who adopt this approach is locating an instrument that is both (1) strongly associated with the health status measure (e.g. overweight) and (2) validly excluded from the education production function (except through its relationship with health). One instrument that has been proposed for the case of overweight is sibling/mother's overweight status, although Kaestner and Grossman (2008) among others, question the validity of the exclusion restriction. Instruments for ADHD and depression seem to be even more difficult to locate.[10]

Lately, progress on the use of instruments has been made in the use of genetic markers, which have been shown in the medical literature to be strongly associated with many measures of health status. Ding et al. (2006) was the first paper within economics to use this strategy. These authors argue that these associations in conjunction with demonstrating that the specific markers are not subject to linkage disequilibrium can justify their use to identify the causal impacts of poor health on high school grade point average using a data set of adolescents in five schools in a single county in Virginia. The authors further discuss that resulting estimates may include dynastic effects but remains policy relevant since children are not randomly assigned to families. Norton and Han (2008) have subsequently used genetic markers as instruments to estimate the effects of obesity on early labor market outcomes. Fletcher and Lehrer's (2008) use of the genetic lottery strategy can be viewed as an extension of this approach as it combines the instrumental variables with the family fixed effects methodology. This approach (implicitly assuming equal impacts across siblings) eliminates all shared family factors (including those which may be unobserved such as the dynastic genetic endowment) that could cause concerns with the exclusion restrictions in previous work. The authors find strong negative effects of inattentive symptoms on achievement tests and suggestive evidence of negative effects from depression systems.

[10] Although see Tefft (2008) for an innovative idea of using exposure to daylight as an instrument for depressive symptoms (as measured by poor mental health status).

To summarize, the findings from these studies have nearly universally suggested a strong link between health and education outcomes. However, Grossman and Kaestner (1997) and Cutler and Lleras-Muney (2006) both conclude their surveys by stating there remains a need for additional research that attempts to use more persuasive empirical methodologies to address limitations with prior studies. The credibility of past empirical designs are often debated, as critics have concerns regarding either the plausibility of the unconfoundedness assumption or availability of valid instruments and question the source and exogeneity of the identifying variation in studies that use sibling/twin fixed effects to estimate the health-education link. As discussed in the introduction, we follow the combined IV/FE strategy of Fletcher and Lehrer (2008) and extend the results to examine the effects of depression, ADHD, and overweight on years of completed schooling.

3. Economic Model

This paper tests the hypothesis that health status in childhood affects human capital investment decisions. A simple two stage model of schooling choice motivates our empirical hypotheses. In the first stage, we adopt a paternalistic approach and assume that until a child reaches the age of eighteen, her altruistic parents select inputs to maximize the household indirect utility function. Subsets of these inputs enter both education production functions and health production functions generating stocks of human capital that are captured in a vector H_o. We assume that the value that parents place on the human capital of each child is an increasing and concave function of the number of children.

At the age of eighteen, each child endows a level of 'basic' human capital (H_o) — determined from previous parental decisions—which determines the productivity of later investments as an adult. In the second stage of the model, each individual faces constraints in that they (1) have to allocate their total lifetime (y) between working and investing in their own education and (2) they have to allocate their lifetime income between their own per-period consumption. Borrowing from future generations and bequests are not allowed.

Formally an adult decides on how to allocate their time across educational investments (*h*) and working (*l)* in order to maximize lifetime utility. The lifetime utility function can be expressed as

$$V = \int_{18}^{T} e^{-rt} F(c_t) dt \qquad (1)$$

where r is the real rate of interest and c_t is per period consumption. The time and goods constraint is given by

$$T \geq l + h \qquad (2)$$

$$y \geq \int_{18}^{T} e^{-rt} c_t dt \qquad (3)$$

Adult education (H_a) determines productivity in the labor market, and the stock of adult human capital one accumulates is determined by the time invested in education together with the human capital endowment. We assume that human capital and time are complements in all production functions, such that adult human capital increases the productivity of time in the labor market and basic human capital including health increases the productivity of education in generating adult human capital. Production functions take on simple multiplicative forms, with constant returns to human capital.

$$H_a = ahH_0 + d \qquad (4)$$

$$y = \beta H_a \qquad (5)$$

where α, d and β are non-negative constants. The level of adult human capital (H_a) is selected by individuals themselves and is directly used to produce goods. The constant d can be viewed as innate skills since this is the stock of adult human capital that an individual would have if they did not make any explicit investments in education. Thus, as in the early Becker (1964) and Mincer (1974) models, individuals are making schooling decisions to maximize lifetime earnings.

As in the standard human capital framework (Becker 1964, Card 1999) an adult will attend school until the expected future benefits of attendance is equal to the current (opportunity) costs of spending time in the classroom rather than working. In this model, childhood health affects the amount of human capital an adult develops in a

multiplicative manner. In other words, early measures of health raise lifetime utility by increasing the productivity from educational investments.[11]

4. Data

This project makes use of the National Longitudinal Study of Adolescent Health (Add Health), a nationally representative longitudinal data set.[12] The dataset was initially designed as a school-based study of the health-related behaviors of adolescents who were in grades 7 to 12 in 1994/5. A large number of these adolescents have subsequently been followed and interviewed two additional times in both 1995/6, and 2001/2. Our project makes use of a specific subsample of the respondents that permit us to develop our identification strategy. Specifically, we analyze data for the sample for which DNA measures were collected during the 2001/2 interview—by design, all of these individuals also had siblings 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 are incomplete education, health and DNA measures for multiple family members reduces the sample to approximately 1,620 individuals

The data set contains information on a number of health conditions, including depression, ADHD and obesity. Depression is assessed using 19 responses to the Center for Epidemiologic Studies-Depression Scale (CES-D), a 20-item self-report measure of depressive symptoms. Items on the CES-D are rated along a 4-point Likert scale to indicate how frequently in the past week each symptom occurred (0 = never or rarely; 3 = very often). The sum of these items is calculated to provide a total score where higher

[11] An alternative mechanism by which one can link early measures of health to adult human capital investment decisions is provided in the models of Kalemli-Ozcan et al. (2000) and Tamura (2001). In these models, the mechanism explored is increased longevity which could appear from early health investments. As adults live longer, their productive horizon is enhanced, the period over which the gains from investments in human capital can be reaped is extended. Therefore, the rate of return to investments in human capital increases and educational attainment tends to rise.

[12] 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 and additional details on the construction of the sample are provided in Harris et al. (2003).

scores indicate a greater degree of depressive symptoms. To determine whether an individual may be depressed, we followed findings from earlier research with adolescent samples (Roberts, Lewinsohn, and Seeley (1991)) and use specific age and gender cutoffs. We also use adult-based cutoffs to capture a broader measure of depressive symptoms in our analyses. The primary indicator of childhood ADHD symptoms is taken from an eighteen-question retrospective rating collected during the third data wave. Since there is evidence that the effects of ADHD may vary by whether the symptoms are of the inattentive or hyperactive type[13], we examine the effects of these different domains as well as usual definition of ADHD of any type. Finally, overweight and obesity are calculated from each individual's self-reported height and weight applied to age and gender specific definitions obtained from the Centers for Disease Control.

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. Past research with this data (Goodman et al. (2000)) indicate that there is a strong correlation between measured and self-reported height (0.94), and between measured and self-reported weight (0.95) and there is no evidence that reporting errors are correlated with observed variables such as race, parental education, and household income.[14] Importantly, since we are using instrumental variables methods, our approach will allow us to reduce any measurement error in the health outcomes (Bound et al. 2001), which will allow us to attain consistent estimates.

Regarding educational outcomes, the data contains information on whether the respondent completed high school, years of completed schooling at Wave 3, as well as many other measures. The data also provides a rich set of information on environmental and demographic variables (i.e. family income, gender, parental education, family structure, etc.) that are used as control variables in our analysis. Finally, the restricted Add Health data allows community-level variables from the Census Bureau and school input variables from the NCES common core of data to be matched to the individuals in the data set to serve as additional controls.

---

[13] For example, Babinski et al. (1999), Ding et al. (in press) and Fletcher and Wolfe (2008a) present empirical evidence of different impacts from these two diagnoses on academic performance.

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

Summary statistics on our sample are provided in Table 1. The first column contains the full sample where the second and third columns only contain the subsets of siblings and twins, respectively. The respondents have completed over 13 years of education on average by the Wave 3 data collection, when they were on average 22 years old. Eleven percent of the sample did not finish high school at Wave 3. Household income is slightly higher than US averages and the majority of mothers have attended college. The twins and sibling samples are both almost equally composed of males and females. African Americans and Hispanics account for approximately one third of the sample. With the exception of race, there are few differences in any of the summary statistics between the full sample and the subsample of siblings and twins. While many of the education and demographic variables fall within national averages, the rates of poor mental health outcomes are slightly higher. While the AD and HD subscale averages fell within standard ranges for adolescent samples, roughly 8% of the sample is coded with ADHD, which exceeds the 6% national average. Conversely those adolescents classified as being depressed in our sample is lower than the 1999 estimate of the fraction of the adolescent population being clinically depressed (12.5%) from the U.S. Department of Health and Human Services. Finally, overweight rates fall slightly below the national average for this period. The over-classification of ADHD could result from measurement error, an issue we will investigate in our empirical analysis.

In Table 2, we present simple t-tests of equality of means that show the well-known significant associations between poor health and poor education outcomes. Individuals with inattentive symptoms (AD) complete nearly 1 fewer years of schooling. Individuals with hyperactive symptoms (HD) complete fewer years of schooling. Individuals we classify as depressed complete nearly 1 fewer years of schooling. Finally, overweight individuals complete approximately ½ fewer years of schooling.

4.1 Genetic Data

The DNA samples were drawn in the third collection and were genotyped for six candidate polymorphisms that are expressed generally in the primitive limbic system of

the brain.[15] The initially targeted candidates are the dopamine transporter (DAT), the dopamine D4 receptor (DRD4), the serotonin transporter (5HTT), monoamine oxidase A (MAOA), the dopamine D2 receptor (DRD2) and the cytochrome P4502A6 (CYP2A6) gene. Variants in the DNA base sequence (single nucleotide polymorphisms) are hypothesized to directly affect the synaptic level two neurotransmitters, dopamine and serotonin, who each provide signals of pleasure from the limbic system and leads individuals to forego other basic activities. Specific polymorphisms are believed to independently affect the propensity to develop a poor health outcome over the lifecycle and interactions between the genetic markers may also have potentially powerful effects.[16] Essentially, the biomedical literature postulates that poor health outcomes and behaviors are associated with markers that indication fewer dopamine receptors, diminished synthesis of serotonin and diminished production of the reuptake protein for dopamine.[17]

The identification strategy proposed in this paper relies on there being a significant association between differences in the occurrence of poor health outcomes between individuals with different genetic markers. We conducted tests for homogeneity of odds ratios to see whether possessing a given polymorphism increased the odds of a particular health outcomes or behavior occurring. For each genetic marker there exists a statistically different odds ratio in the occurrence rate of at least one of the health outcomes and behaviors. These findings are all consistent with the biomedical literature

[15] The limbic system is highly interconnected with the region of the brain associated with reward and pleasure. 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.

[16] In our analysis we code for the three most common polymorphisms of each genetic marker. 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.. Finally, we include indicator variables for the two possible variants of the CYP gene.

[17] Full details on how these markers affect the health conditions examined in this paper are provided in Fletcher and Lehrer (2008).

which indicate that poor health outcomes and behaviors are associated with markers that indication fewer dopamine receptors, diminished synthesis of serotonin and diminished production of the reuptake protein for dopamine.

Not only did we investigate whether the genetic markers in our study are linked to measures of health but also if their inheritance was linked. Using one sibling (selected by having the lowest record number in the dataset) we conducted tests for homogeneity of odds ratios to see whether possessing a polymorphism in one genetic marker increases the odds of possessing a specific polymorphism in a different genetic marker. We did not find any evidence indicating a systematic relationship between markers of any two of the genes. This was not a surprise as linkage was highly unlikely due to the location of these markers on the genome. Additionally, using maps of the location between the specific genetic markers in our study and those which have been hypothesized to be linked to education outcomes (Plomin et al. 2007), we find no evidence that they are located closely on the genome, suggesting that linkage in inheritance is unlikely.

4.2 Estimation

To investigate how health status in childhood affects human capital investment decisions we essentially estimate standard equations derived from our model of returns to education. Recall, as in Becker (1964) or Card (1999), individuals choose the optimal number of years of schooling to maximize the utility function except now the benefits from increases in adult education are also affected by the stock of health the parents provide to their child upon entering adulthood.

Formally, in our analysis we estimate a linear equation that relates individual years of schooling completed of child i in family *f* (Y_{if}) to individual covariates (X_{if}), a vector of individual child and adolescent health measures (H_{if}), unobserved family factors (v_f) as follows:

$$Y_{if} = \beta_0+\beta_1 X_{if} +\beta_2 H_{if} + v_f + \varepsilon_{if} \qquad (1)$$

where ε_{if} is a independently distributed residual with mean zero. Controlling for unobserved family factors provides a rich measure of contextual variables and parental genetic and background factors and our interest is whether differences in the early health measures of the child (β_2) within families can explain differences in a variety of

outcomes in adulthood. Since the stock of health conditions upon entering adulthood reflects parental behavioral choices, they should be treated as endogenous in the empirical analysis. As discussed, we will use exogenous variation from a set of genetic markers and their interactions to identify β_2. Thus, the empirical model contains multiple equations and consistent estimates are obtained by using GMM to estimate this system of equations. If we define G_i^H to be the vector of genetic markers that provide endowed predispositions to the state of health conditions, identification relies on the assumption that G_i^H is unrelated to the structural errors in the system. While there might not be any existing evidence in the scientific literature that the markers considered in this study have any impact on schooling choices, it remains possible.[18] Statistically, to convince the reader that G_i^H is unrelated to ε_{if} we will conduct a variety of instrument validity tests including, 1) examine the first stage regression for weak instruments, 2) conduct overidentification tests and 3) perform sensitivity analysis with respect to violations of the exogeneity assumption of the instruments using approaches developed in Conley et al., (2007).

In our analysis, we will not only estimate equation (1) using a family fixed effects/instrumental variable strategy but also by OLS, and family fixed effects and IV by themselves. Specification tests between the estimates from the different empirical strategies will be conducted to determine whether one should account for family unobserved heterogeneity and determine if family fixed effects by themselves remove all the endogeneity.[19]

For robustness, we will follow Fletcher and Lehrer (2008) and conduct the family fixed effects IV analysis described above with sub-samples of our data defined by different family relationships. Implicitly both twin and sibling models assume that family fixed effects have the same impact on both individuals and ex-ante this assumption is

[18] Plomin et al. (2006) and de Quervain et al. (2006) present surveys on which genes are believed to be associated with intelligence and ability. None of the markers we consider are listed in these reviews or several earlier reviews in the scientific literature and as noted earlier maps of the genome indicate they do are not located in close proximity on the chromosome indicating that linkage disequilibrium is not a serious concern.

[19] While a family fixed effects strategy allows the researcher to simultaneously control (assuming constant impacts between family members) for many parental characteristics/behaviors, it does not provide any guidance as to why within a twin or sibling pair the subjects differed in explanatory characteristics. It is well known that even monozygotic twins are often discordant for health conditions including those being investigated in this study.

more plausible with twins as they are of the same age and in general of the same gender. We will re-estimate equation (1) using only the subset of twins of the same gender to examine the sensitivity of our results.

As our interest in the empirical analysis is focused on changes in several poor health conditions determined from genetic factors, it is worth to state explicitly that we do not deny that there are many other dimensions of health that are endogenous to individuals' schooling decisions. The available evidence in the scientific and medical literature supports the idea that the genetic markers we consider have links to two additional disorders; Tourette's syndrome and schizophrenia. These disorders have low prevalence rates. so ex-ante it appears doubtful that their exclusion would dominate the estimated relationships between measures of childhood and adolescent health and later outcomes. However, previous work by both Ding et al. (in press) and Fletcher and Lehrer (2008) have clearly illustrated that due to the high comorbidity in health conditions and the lack of exogenous variations that can explain one particular condition only, it is important to control for a rich health vector in the analyses to present the most accurate set of estimates. As such, in our analysis we will consider two vectors for H_{if} (vector 1 consists of ADHD, obesity and depression, vector 2 consists of AD, HD, obesity and depression).

5. Results

In this section we present and discuss estimates of equation 1. We subsequently discuss the statistical performance of the genetic lottery identification strategy. We conclude the section by discussing how one should interpret our estimates.

Estimates of equation 1 are presented in Table 3 where results that separate AD and HD symptoms are presented in the odd columns and combined ADHD status in the even columns. The first two columns show results for our baseline OLS specification. Results show that individuals with inattentive symptoms (AD) complete over ½ fewer years of schooling, depressed individuals complete nearly 2/3 fewer years of schooling[20], and overweight individuals complete nearly 1/3 fewer years of schooling. We do not find

[20] Fletcher (2008) also finds similar results for depression.

a statistically significant effect for hyperactive symptoms. Column 2 shows that combined ADHD status is associated with nearly ½ fewer years of schooling.

We next presents results using family fixed effects in columns 3 and 4 in order to control for potential omitted variables at the family level that could link health status with educational outcomes. As expected, we find that all of the estimated relationships shrink considerably with these controls. The effect of AD is reduced by 50%, the effect of depression is reduced by over 50%, and the effect of overweight is reduced by nearly 2/3rds. None of the relationships are statistically significant.

In order to correct for endogeneity of health status as well as measurement error in our health status variables, we present results from instrumental variables specification in columns 5 and 6. In comparison to OLS, the results become much larger. We now find the AD symptoms reduce years of schooling by nearly 3 years. HD symptoms are positively related to years of schooling (increasing schooling by 4 years). Depression is negatively associated with year of schooling, decreasing schooling by nearly 0.80 years, although the result is imprecisely estimated. Overweight status is shown to decrease years of schooling by nearly 1.5 years and is not statistically significant.

For our preferred set of results, we use a combined instrumental variables/fixed effects specification in order to take advantage of the "genetic lottery" that occurs between siblings. We find a large statistically significant effect of inattentive symptoms (AD), decreasing years of schooling by over 3.5 years. While this effect is very large in magnitude, it is important to note that AD symptoms reflect health status in elementary school, which could lead to accumulated negative effects on education outcomes over time. Fletcher and Lehrer (2008) also find very large negative effects of AD on achievement test scores, suggesting a potential mechanism of these large effects on schooling. The positive education effect for hyperactive symptoms (HD) is reduced but still large, increasing years of schooling by 2.7 years, though the result is not statistically significant. We also find large negative effects of depressions, which reduces schooling by 1 year, though it is also imprecisely measured. Our estimates for overweight status now change to a positive effect, which suggests caution in interpreting this relationship (as also found in Fletcher and Lehrer 2008).

In order to assess the robustness of our findings, we present results that only using the sub-sample of same-sex twins in Table 4. This subsample is used because the equal environment assumption that the family fixed effects implicitly impose is most likely to hold. This sample also eliminates the possibility of any gender-related biases within families that may affect how parents respond to adolescent mental and physical health status. Lastly, this sample also increases our confidence that parents are not compensating (or reinforcing) health status differences between siblings as it would be most difficult for them to differentiate between children of the same age and gender.

The empirical results are presented in Table 4. Even though the sample size is significantly reduced and there are arguably a limited number of discordant health status cases within families we continue to find significant relationships. Most importantly, the main results are quite robust to this subsample and we continue to find that childhood inattentive symptoms lead to reduced schooling. We find small effects from depression, and inconsistent results for overweight status.

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 GMM estimates of equation (1) are over-identified, we use a J-test to formally test the overidentifying restrictions. The smallest of the p-values for these tests is 0.32, providing little evidence against the overidentifying restrictions. To examine whether these genetic markers are valid instruments, we first calculated the Cragg—Donald (1993) statistic to examine whether the set of instruments is parsimonious and has explanatory power and then used it in computing the test statistic proposed by Stock and Yogo (2005) that demonstrates the strength of the instruments. With both health vectors, the Cragg-Donald statistic is large and we are able to reject the null hypothesis, suggesting an absence of a weak instruments problem thereby indicating that the estimator will not perform poorly in finite samples and that with or without family fixed effects.

Finally, 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 by including a term that measures the extent to which the exogeneity assumption is potentially erroneous, thereby directly affecting the standard errors. The amount of uncertainty about the exogeneity assumption is constructed from prior information regarding plausible values of the impact of genetic factors on academic performance that are obtained from the reduced form. In our analysis, we consider increasing the exogeneity error from 0% to 70% of the reduced form impacts. At levels, below 35% of the reduced form impacts, our results are robust as inattention continues to have a statistically significant negative impact on years of schooling. Since there does not exist any scientific evidence that these specific markers directly affect any of these outcomes, the sensitivity analysis suggests the levels at which our results are sensitive to the exclusion restriction assumption appear highly implausible.

Discussion

As in Ding et al. (in press) and Fletcher and Lehrer (2008), we interpret the estimates presented in both Tables 3 and 4 as reduced form coefficients. We cannot disentangle the impact of the health condition as explained by genes from that of the response from the environment to the health conditions as explained by genes. The notion that parents, peers and teachers respond differently to children with health disorders has substantial support in the developmental psychology literature. Fortunately, the Add-health database provides some information allowing us to conduct a slightly more rigorous evaluation of the link between health conditions and inputs from other actors. In Table 5, we present simple estimates from family fixed effects models of the extent to which the subject reports a variety of peer and parental inputs on the same set of explanatory variables as included in equation (1). Notice that children who suffer from both inattention and depression receive fewer inputs, though many of the results are not statistically significant.

The large significant impact from inattention on years of schooling may also arise since our measure of inattention is taken at a early age. This could either reflect 1) the notion that since early learning begets later learning (Heckman et al. (2007)) the children experience setbacks that affect their permanent developmental trajectory, 2) as inattention affects the ability to concentrate, the large impact on school achievement is carried over to schooling choices, 3) the reliability of that coefficient is dependent on the rich controls

we have on most of the comorbid health conditions and there is an important omitted comorbid condition, 4) there is a strong link between early health and later health (Persico et al. (2005)) and the estimate of inattention is either proxying for a current disorder or lower test scores. To sort between these completing explanations, future research that can estimates richer system of equations could separate out the relative impacts of health on each independent outcome and control for larger pre-determined time-varying health and education vectors is needed.

Conclusions

In this paper, we use the genetic lottery identification strategy that combines within-family comparisons with instrumental variables estimator to estimate the causal effects of poor adolescent mental and physical health status on years of completed schooling. We present evidence that inattentive symptoms in early childhood have large lasting effects in reducing completed schooling. We also find little consistent evidence that adolescent overweight status influences years of schooling completed. Our estimates suggest that accounting for family fixed effects is important but these strategies cannot fully account for the endogeneity of poor mental heath. Finally, this strategy further confirms earlier work the effects of health on education by demonstrating that the presence of comorbid conditions present immense challenges for empirical studies that aim to estimate the impact of specific health conditions.

Health and productivity are often argued to have a complex interdependence in the modern workplace. Much of this research has focused on how negative factors in the work environment such as stress affect both measures of worker health and worker productivity. Yet, surprisingly there is very little research on how health measures that precede the work place affect choice in further education. We find very large impacts of the effects of inattention and argue that more careful and insightful evaluation in this area is needed since there are potentially large benefits from childhood and adolescent health interventions that are so far unidentified through evidence based research. As a large number of school-based programs have recently been introduced to prevent obesity through lifestyle changes it is important for policymakers to target health conditions that are not the easiest to identify but rather may have the largest impacts on one's future.

In this paper, we outline a new research design to identify causal relationships that potentially has wide applicability to a host of important questions in both the social science and health services research literatures. As the existing scientific evidence suggests that there is a strong genetic component in determining nearly all important economic and social outcomes studied by social scientists, our approach could be applied quite broadly. This voluminous literature from behavioral genetics and medicine coupled with advances in methods of collecting and sequencing genetic data as well as the growing availability of combined socioeconomic/genetic datasets suggest that efforts to leverage genetic information in social science investigations has the ability to considerably increase our knowledge of critical social and health policy issues. By exploiting the genetic variation in inheritance within families, our method presents a unique opportunity to isolate the variation in genetic factors from other dynastic and family characteristics. This identification strategy can be viewed as analyzing data from an experiment in "nature". This paper represents a step in this direction by illustrating the empirical approach, and we argue that a broad set of questions could be addressed using "genetic lotteries".

Table 1
Descriptive Statistics
Add Health Siblings with DNA Information

	Full Sample			Twin Sample			Sibling Sample		
Variable	Observations	Mean	Std. Dev	Observations	Mean	Std. Dev	Observations	Mean	Std. Dev
Years of Schooling	1618	13.31	1.93	604	13.37	1.86	1014	13.27	1.98
Drop Out	1596	0.11	0.31	600	0.09	0.28	996	0.12	0.32
AD	1618	0.05	0.21	604	0.05	0.21	1014	0.05	0.21
HD	1618	0.05	0.21	604	0.04	0.20	1014	0.05	0.22
ADHD	1618	0.07	0.26	604	0.07	0.26	1014	0.08	0.27
Depressed	1618	0.06	0.24	604	0.05	0.23	1014	0.07	0.25
Obese	1618	0.20	0.40	604	0.17	0.38	1014	0.22	0.41
Age	1618	17.03	1.68	604	17.01	1.65	1014	17.05	1.71
Male	1618	0.49	0.50	604	0.50	0.50	1014	0.48	0.50
Black	1618	0.17	0.37	604	0.23	0.42	1014	0.13	0.34
Hispanic	1618	0.14	0.35	604	0.14	0.34	1014	0.14	0.35
Birth Order	1618	2.04	1.32	604	1.89	1.42	1014	2.13	1.25
Family Income ($10,000s)	1618	46.90	40.75	604	50.10	54.98	1014	45.00	29.01
Maternal Education	1618	13.21	2.21	604	13.26	2.36	1014	13.17	2.12
Parent Age	1618	41.84	5.26	604	42.45	5.52	1014	41.47	5.07
Married Parents	1618	0.68	0.46	604	0.65	0.47	1014	0.70	0.45
Missing Parent Information	1618	0.32	0.47	604	0.33	0.47	1014	0.32	0.47

Table 2
Differences in Education Outcomes by Health Status

	Years of Schooling Full Sample	Years of Schooling Sibling Sample	Years of Schooling Twin Sample
Inattentive (AD = 1)	12.56	12.63	12.45
	(2.26)	(2.06)	(2.58)
Not Inattentive (AD = 0)	13.34	13.3	13.42
	(1.91)	(1.97)	(1.81)
T-Statistic	3.43	2.24	2.76
Hyperactive (HD =1)	12.79	12.81	12.77
	(2.00)	(1.87)	(2.29)
Not Hyperactive (HD = 0)	13.33	13.29	13.4
	(1.93)	(1.98)	(1.83)
T-Statistic	2.4	1.73	1.69
AD/HD = 1	12.83	12.88	12.72
	(2.12)	(1.98)	(2.37)
AD/HD = 0	13.35	13.3	13.42
	(1.91)	(1.98)	(1.80)
T-Statistic	2.85	1.78	2.39
Depressed = 1	12.45	12.38	12.58
	(1.88)	(1.88)	(1.89)
Depressed = 0	13.36	13.33	13.42
	(1.92)	(1.97)	(1.85)
T-Statistic	4.65	3.85	2.54
Obese = 1	12.83	12.74	13.01
	(1.81)	(1.85)	(1.73)
Obese = 0	13.43	13.41	13.45
	(1.94	(1.99)	(1.87)
T-Statistic	5.03	4.51	2.2

Note: Each health condition = 1 if the individuals has symptoms above the threshold described in the text.
T-statistics are for a t-test of differences between groups who have or do not have the health condition

Table 3
The Effects of Health Status on Years of Schooling Completed

Estimation Method	OLS	OLS	Family Fixed Effects	Family Fixed Effects	Instrumental Variables	Instrumental Variables	Genetic Lottery IV Fixed Effects	Genetic Lottery IV Fixed Effects
AD	-0.554**		-0.206		-2.773*		-3.792*	
	(0.24)		(0.36)		(1.66)		(2.00)	
HD	-0.271		-0.0113		4.019*		2.736	
	(0.23)		(0.37)		(2.16)		(2.36)	
ADHD		-0.439**		-0.154		1.157		-2.739
		(0.18)		(0.29)		(1.83)		(3.04)
Depressed	-0.656***	-0.657***	-0.246	-0.245	-0.798	-0.479	-1.110	-1.388
	(0.18)	(0.18)	(0.29)	(0.29)	(2.26)	(2.72)	(1.62)	(1.87)
Overweight	-0.317***	-0.320***	-0.109	-0.110	-1.401	-0.343	0.690	1.768
	(0.11)	(0.11)	(0.19)	(0.19)	(1.24)	(1.21)	(1.03)	(1.60)
Age	0.251***	0.249***	0.231***	0.229***	0.254***	0.245***	0.316***	0.299***
	(0.025)	(0.025)	(0.084)	(0.084)	(0.039)	(0.038)	(0.084)	(0.094)
Male	-0.366***	-0.368***	-0.413***	-0.411***	-0.435***	-0.462***	-0.450**	-0.345
	(0.086)	(0.086)	(0.15)	(0.16)	(0.13)	(0.16)	(0.18)	(0.22)
Black	-0.0204	-0.0151			-0.0207	0.0335		
	(0.12)	(0.12)			(0.16)	(0.16)		
Hispanic	-0.0272	-0.0266			0.0620	0.0218		
	(0.13)	(0.13)			(0.25)	(0.27)		
Sibling Sample	0.0320	0.0338			0.0348	0.0300		
	(0.089)	(0.089)			(0.15)	(0.13)		
Birth Order	-0.144***	-0.144***	-0.328*	-0.330*	-0.158***	-0.150***	-0.266	-0.241
	(0.036)	(0.036)	(0.17)	(0.17)	(0.048)	(0.045)	(0.18)	(0.22)
Family Income	0.00579***	0.00583***			0.00504***	0.00594***		
	(0.0014)	(0.0014)			(0.0018)	(0.0020)		
Maternal Education	0.215***	0.215***			0.180***	0.209***		
	(0.022)	(0.022)			(0.054)	(0.058)		
Parent Age	0.0592***	0.0596***			0.0600***	0.0619***		
	(0.0099)	(0.0098)			(0.013)	(0.013)		
Married Parents	0.176	0.182*			0.251	0.206		
	(0.11)	(0.11)			(0.15)	(0.15)		
Missing Information	-0.281***	-0.283***			-0.242*	-0.313**		
	(0.10)	(0.10)			(0.14)	(0.13)		
Constant	4.015***	4.023***	10.29***	10.33***	4.550***	3.989***		
	(0.59)	(0.59)	(1.74)	(1.73)	(1.01)	(0.97)		
Observations	1618	1618	1618	1618	1618	1618	1618	1618

Note: Robust standard errors clustered at the sibling-pair level. ***1*, **5%, *10%

Table 4
The Effects of Health Status on Years of Schooling Completed
Results Using Only Same-Sex Twins

Estimation Method	OLS	OLS	Family Fixed Effects	Family Fixed Effects	Instrumental Variables	Instrumental Variables	Genetic Lottery IV Fixed Effects	Genetic Lottery IV Fixed Effects
AD	-0.549		-0.119		-4.716***		-2.385**	
	(0.45)		(0.50)		(0.86)		(1.12)	
HD	0.0694		0.230		3.139**		0.253	
	(0.47)		(0.79)		(1.38)		(1.00)	
ADHD		-0.347		-0.0655		-0.866		-3.770***
		(0.35)		(0.50)		(2.26)		(0.62)
Depressed	-0.602	-0.599	-0.418	-0.429	0.716	-2.610	-0.140	-1.578
	(0.44)	(0.44)	(0.51)	(0.51)	(2.96)	(2.41)	(1.41)	(2.26)
Overweight	-0.415**	-0.417**	-0.0871	-0.0723	-0.854	0.219	0.306	0.446
	(0.18)	(0.18)	(0.30)	(0.30)	(1.16)	(1.17)	(1.03)	(1.08)

Note: Robust standard errors clustered at the sibling-pair level. ***1*, **5%, *10%. Each set of results contains the same controls as previous tables.

Table 5
Within Family Associations Between Child Investments and Poor Health Status

Outcome	Feel Close to Mother	Feel Close to Peers	Feel Close to School	College Expectations
AD	-0.111	-0.258	-0.209	-0.057
	(0.193)	(0.298)	(0.271)	(0.269)
HD	-0.165	-0.053	-0.206	-0.134
	(0.189)	(0.264)	(0.220)	(0.253)
Depressed	-0.280	-0.476*	-0.602***	-0.319
	(0.204)	(0.266)	(0.229)	(0.230)
Obese	-0.022	0.139	0.116	-0.105
	(0.098)	(0.144)	(0.132)	(0.143)
Age	0.003	-0.018	-0.010	-0.039
	(0.047)	(0.064)	(0.055)	(0.066)
Male	0.200**	-0.012	-0.056	-0.338***
	(0.082)	(0.115)	(0.102)	(0.113)
Birth Order	0.114	0.021	-0.017	-0.185
	(0.112)	(0.154)	(0.114)	(0.138)
Constant	4.180***	3.011**	3.132***	5.398***
	(0.961)	(1.311)	(1.071)	(1.300)
Observations	1584	1622	1622	1654
R-squared	0.66	0.56	0.64	0.68

Note: Robust standard errors clustered at the sibling-pair level. ***1*, **5%, *10%. . Each outcome is a categorical variable with higher values indicating higher investments

References

Almond, D., K. Y. Chay, and D. S. Lee. (2005). "The Costs of Low Birth Weight." *Quarterly Journal of Economics* 120(3), 1031-1083.

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.

Becker, G. S. (1962). Human Capital: A Theoretical and Empirical Analysis. Columbia University Press.

Becker, G. S. (1964). Human Capital: A Theoretical and Empirical Analysis, With Special Reference to Higher Education. Columbia University Press.

Bleakley, H. C. (2007). "Disease and Development: Evidence from Hookworm Eradication in the American South." *Quarterly Journal of Economics*, 122(1), 73-117.

Bound, J., C. Brown, and N. Mathiowetz. (2001). "Measurement Error in Survey Data." in Handbook of Econometrics, Volume 5, J.J. Heckmand and E.E. Leamer (eds.). Amsterdam: Elsevier.

Card, D. (1999). "The Causal Effect of Education on Earnings". in O. Ashenfelter and D. Card, (eds.). Handbook of Labor Economics Volume 3. Amsterdam: Elsevier.

Case, A. and C. Paxson. (2008). "Stature and Status: Height, Ability and Labor Market Outcomes." *Journal of Political Economy* 116(3), 499-532.

Case, A., D. Lubotsky and C. Paxson. (2002). "Economic Status and Health in Childhood: The Origins of the Gradient." *American Economic Review* 92(5), 1308-1334.

Case, A., A. Fertig, and C. Paxson. (2005). "The Lasting Impact of Childhood Health and Circumstance." *Journal of Health Economics* 24, 365-389.

Catts, H.W. (1993). The relationship between speech-language impairments and reading disabilities." *Journal of Speech and Hearing Research* 36, 948-958.

Centers for Disease Control (2004). National Center on Birth Defects and Developmental Disabilities, Vision Impairment Factsheet, http://www.cdc.gov/ncbddd/dd/ddvi.htm.

Conley, T., C. Hansen and P. E. Rossi. (2007). "Plausibly Exogenous." *mimeo*, University of Chicago.

Costante, C.C. (2002). Healthy learners: The link between health and student achievement. *American School Board Journal* 103(1), 1-3.

Council Of Chief State School Officers (1998). State Education Accountability Reports and Indicator Reports: Status of Reports across the States. Washington, D. C.: Council of Chief State School Officers.

Cragg, J. G., and S. G. Donald. (1993). "Testing Identifiability and Specification in Instrumental Variables Models." *Econometric Theory* 9, 222-240.

Cunha, F. and J.J. Heckman. (2007). "Technology of Skill Formation." *American Economic Review Papers and Proceedings* 97(2), 31-47.

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.

Cutler, D. and A. Lleras-Muney (2006). "Education and Health: Evaluating Theories and Evidence." NBER Working Paper 12352.

Datar, A. and R. Sturm. 2006. "Childhood Overweight and Elementary School Outcomes." *International Journal of Obesity* 30, 1449-1460.

Datar, A., R. Sturm, and J. L. Magnabosco. (2004). "Childhood Overweight and Academic Performance: National Study of Kindergartners and First-Grade." *Obesity Research* 12, 58-68.

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.

Deutsch, C. (2000). "Common cause: School health and school reform." *Educational Leadership*, March.

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.

Ding, W., S. F. Lehrer, J. N. Rosenquist and J. Audrain-McGovern. (in press). "The Impact of Poor Health on Academic Performance: New Evidence Using Genetic Markers." *Journal of Health Economics*

Fletcher , J. M. (2008a). "Adolescent Depression: Diagnosis, Treatment, and Educational Attainment." *Health Economics* 17, 1215-1235.

Fletcher, J. M. (2008b). "Adolescent Depression and Educational Attainment: Evidence from Sibling Fixed Effects." *mimeo*, Yale University.

Fletcher, J. M. and S. F. Lehrer. (2008). “Using Genetic Lotteries within Families to Examine the Causal Impact of Poor Health on Academic Achievement.” *mimeo*, Yale University.

Fletcher, J. M. and B. L. Wolfe. (2008a). “Child Mental Health and Human Capital Accumulation: The Case of ADHD Revisited.” *Journal of Health Economics* 27(3), 794-800.

Fletcher, J. M. and B. L. Wolfe. (2008b). "Long-term Consequences of Childhood ADHD on Criminal Activities." *mimeo*, Yale University.

Fuchs, V. R. (1982). “Time Preference and Health: An Exploratory Study”, in V. Fuchs (ed.) *Economic Aspects of Health* (Chicago: The University of Chicago Press).

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.

Grossman, M. (1972), "On the concept of health capital and the demand for health", *Journal of Political Economy* 80(2), 223-255.

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.

Hanson, T. L., G. Austin, and J. Lee-Bayha. (2004). “Ensuring that no child is left behind: How are student health risks and resilience related to the academic progress of schools?” San Francisco, CA: WestEd

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.

Heckman, J. J., M. I. Larenas, and S. Urzua. 2004. “Accounting for the Effect of Schooling and Abilities in the Analysis of Racial and Ethnic Disparities in Achievement Test Scores.” *mimeo*, University of Chicago, Department of Economics.

Heckman, J. J. and D. V. Masterov. (2007). The Productivity Argument for Investing in Young Children. NBER Working Paper No. 13016.

Kaester, R., M. Grossman. (2008). "Effects ofWeight on Children.s Educational Achievement." NBER Working Paper 13764.

Kalemli-Ozcan, S., H. E. Ryder and D. N. Weil. (2000). "Mortality decline, human capital investment, and economic growth," *Journal of Development Economics* 62(1), 1-23.

Kessler, R.C., L. A. Adler, R. Barkley, J. Biederman, C. K. Conners, S. V. Faraone, L. L. Greenhill, S. Jaeger, K. Secnik, T. Spencer, T. B. Ustun and A. M. Zaslavsky. (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.

Kremer M. and E. Miguel. (2004). "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." *Econometrica* 72, 159-217.

Mincer, J. (1974). Schooling, Experience and Earnings, Columbia University Press: New York.

National Institutes of Health (1993). "Early identification of hearing impairment in infants and young children." NIH Consens Statement Online. 11(1), 1-24.

Newacheck, P. W. (2000). "Prevalence, impact and trends in childhood disability due to Asthma". *Archives of Pediatric and Adolescent Medicine* 154, 287-293.

Norton, E. C. and E. Han. (2008). "Genetic Information, Obesity, and Labor Market Outcomes." *Health Economics* 17(9), 1089-104.

Perri, T. J. (1984). "Health Status and Schooling Decisions of Young Men." *Economics of Education Review* 3, 207-213.

Persico, N., A. Postlewaite, and D. Silverman. (2004). "The Effect of Adolescent Experience on Labor Market Outcomes: The Case of Height," *Journal of Political Economy* 112(5), 1019-1053, October.

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.

Plomin R., Y. Kovas and C.M.A. Haworth. (2007). Generalist genes: Genetic links between brain, mind and education, *Mind, Brain and Education* 1, 11–19.

Roberts, R. E., P. M. Lewinsohn, 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.

Sabia, J. J. (2007). "The effect of body weight on adolescent academic achievement." *Southern Economic Journal* 73(4), 871-900

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

Strauss, J. and D. Thomas. (1998). “Health, Nutrition, and Economic Development.” *Journal of Economic Literature* 36(2), 766-817.

Tamura, R. (2006). "Human capital and economic development," *Journal of Development Economics* 79(1), 26-72

Tefft, N. (2008). “Mental Health and Employment: The SAD Story.” *mimeo,* University of Wisconsin-Madison.

Weiss, Y. (1971). "Ability and the Investment in Schooling: A Theoretical Note on J. Mincer's 'Distribution of Labor Income'," *Journal of Economic Literature* 9(2), 459-61, June.