The Impact of Poor Health on Education:
New Evidence Using Genetic Markers

Weili Ding
Queen's University

Steven F. Lehrer
Queen's University and NBER

J. Niels Rosenquist
University of Pennsylvania and Massachusetts General Hospital

Janet Audrain-McGovern
University of Pennsylvania

October 2006

MOTIVATION

There is a large and positive correlation between education and health.
  This result is extremely robust

It remains unclear whether better health actually causes higher education levels, or whether it is a spurious correlation

Two Major Challenges in Identifying the Role/Impact of Health on Academic Outcomes:
  Endogeneity and Measurement Error.

Substantial Public Policy Implications
  -Rationale for Medicaid in the US and Introduction of Medicare in Canada

KEY POINTS

We will discuss how we can use a set of genetic markers to identify the impact of adverse health on education.

We will have some clear answers but propose more questions:
- -Large impact of poor health on academic performance.
- -Substantial heterogeneity exists across gender; Girls really suffer from poor health.
- -Comorbidity of health disorders and health behavior must be accounted for.
- -Genes show great promise as good instruments.

SCIENCE

Lots of evidence suggest that the role of genetic factors leading to poor health outcomes is substantial.

As a result, pharmaceutical companies regularly develop drugs that target specific genetic markers.
  Bupropion (a.k.a. zyban) affects dopamine production.
 Ritalin blocks the reuptake of dopamine.

Different regions of the brain engage in different processes.

Many correlates of risky behaviors as well as mental and physical disorders are found in the brain's reward system.

This system regularly operates by sending messages via neurotransmitters from the VTA to the prefrontal cortex.

Some food and drugs have properties that hijack this system and give individuals a feeling of pleasure.

NEUROANATOMY
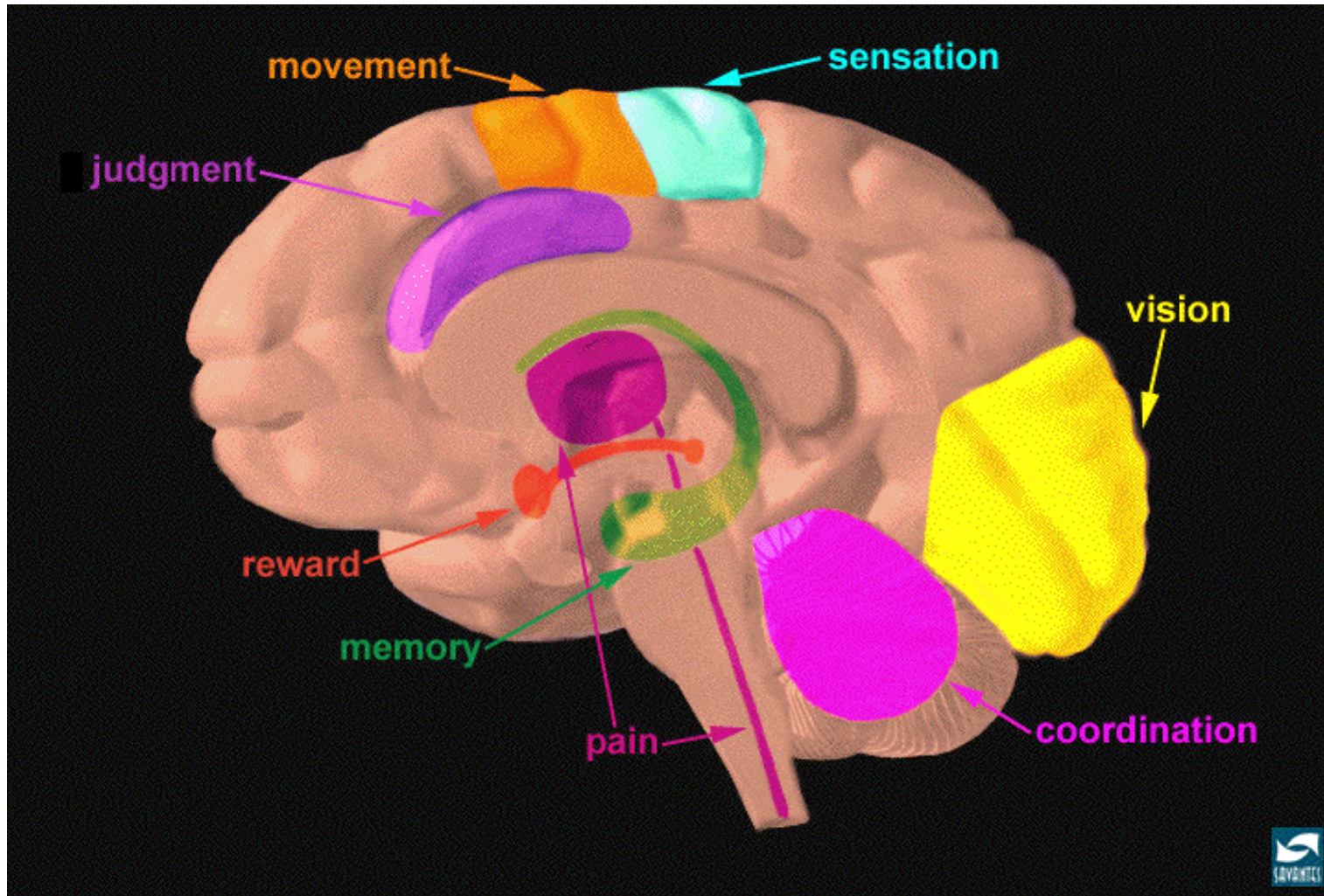
GENETIC MARKERS

Each person inherits an allele of a gene from each parent at conception. The two alleles combine to form a marker.

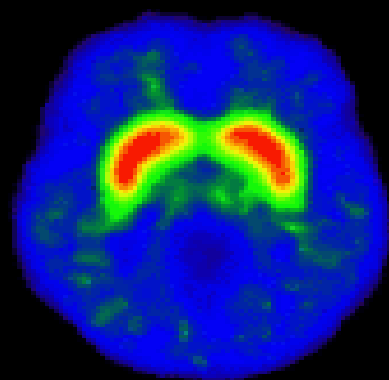**Terminology**
   Polygenic
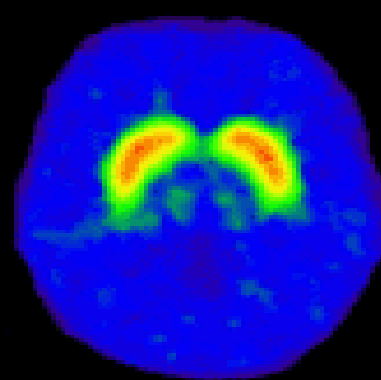   **Homozygous vs. Heterozygous**
   Polymorphism

Example: the Dopamine Receptor D2 locus. People in general either have A1/A1 or A1/A2 or A2/A2.
A1 alleles code for reduced density of dopamine receptors.

Control subjects   Obese subjects

These composite brain scan images show that obese individuals have significantly fewer dopamine receptors in the outlined area than control subjects. These receptors transmit pleasurable feelings from basic activities such as eating and sex. Low levels of these receptors also have been found in people addicted to drugs of abuse. The reduced reward experienced by people with this deficiency may make them more likely to engage in addictive behaviors.

http://www.drugabuse.gov/NIDA_Notes/NNVol16N4/pathological.html

Our data contains information on three other genes

    - SLC6A3 codes for the dopamine transporter protein involved with reuptake. (DAT 0, 1, 2)

    - Trypthohan Hydroxylase Gene is involved with mood and impulse as it synthesizes serotonin. (TPH AA, AC, CC)

    - CYP2B6 gene metabolises drugs and toxins in the liver. (CYP TT, CT, CC)

Interactions of these markers could have powerful effects.

DATA

Our primary data source is the Georgetown Adolescent TObacco Research Study

GATOR Study

Information on Genetic Markers:
    -Buccal swabs with standard techniques. Four markers were collected and double-checked at 20%.

Four follow-up surveys with the students were conducted.

Each survey conducted at school contains standard demographic information and detailed smoking information.

Academic performance collected in the last three surveys.

Very little attrition -- Follow up is 95%, 96%, 93% and 89%.

HEALTH MEASURES
(Table 1)

Depression assessed by the CES-D scale. Well-established 20 item questionnaire. Use adolescent cutoffs that vary by age and gender.

To assess ADHD as well as AD and HD separately, the Current Symptoms Scale Self-Report Form was used. Only known to the researchers; AD/HD; reduce measurement error.

Obesity is constructed from self-reported height and weight collected in the fifth survey. Not ideal.

All means fall within normal ranges for adolescent samples.

We matched by school identifiers to CCD and census records by zip codes for additional controls.

EMPIRICAL FRAMEWORK

Focus is to construct a model from health to education that allows for adolescent making some decisions.

Step 1. Adolescent chooses whether or not to engage in a health behavior to maximize her utility:

$$k_{iT} = k\ (X_{1iT},\ p_k,\ G^k_i,\ H_{iT-1},\ \varepsilon^k_{iT})$$

Step 2. Altruistic parents observe a signal of their child's health behavior (ḱ) and make health input decisions (l_{iT}) that enter into a health production function for the adolescent:

$$H_{ijT} = g\ (X_{2iT}...X_{2i0},\ K_{iT}...K_{i0},\ l_{iT}...l_{i0},\ H_{i0},\ G^H_i,\ \varepsilon^H_{iT}...\varepsilon^H_{i0})$$

For simplicity, assume parents have a single mindedness in preference towards their child's health:

$$U(H_{iT1}, \cdot) \geq U(H_{iT2}, \cdot) \quad \text{if } H_{iT1} > H_{iT2}$$

Step 3. Altruistic parents choose education inputs such as the optimal school for the child, whether or not to hire tutor etc. that will maximize household utility.

Conditional on the selection of inputs in step 3, adolescent achievement is obtained through an education production function:

$$A_{ijT} = f\left(X^e_{iT}...X^e_{i0},\ Q_{jT}...Q_{j0},\ H_{iT},\ U_i,\ \varepsilon^e_{iT}...\varepsilon^e_{i0}\right)$$

HEALTH AS AN EDUCATION INPUT

Directly
- **Physical energy a child has for learning.**
- **Affects child's mental status: ability to concentrate.**

Indirectly
- **Child's health status may affect the response of peer, parents and teachers; thus influence the inputs in education.**

Possible Dynastic Effects
- **Potential proxy for the part of family inputs caused by the same symptoms in parent(s) and sibling(s).**

ESTIMATION

Linearize the education production function:

$$A_{ijT} = \beta_0 + \beta_1 X^e_{iT} + \beta_2 H_{iT} + \beta_3 Q_{jT} + \varepsilon_{ijT}$$

Similarly health production and decision to engage in risky behavior

$$H_{iT} = \gamma_0 + \gamma_1 X^H_{iT} + \gamma_2 k_{iT} + \gamma_3 G^H_i + \varepsilon^H_{iT}$$

$$k_{iT} = \delta_0 + \delta_1 X^k_{iT} + \delta_2 H_{iT-1} + \delta_3 G^k_i + \varepsilon^k_{iT}$$

This creates a system of equations.

IDENTIFICATION

Use genetic markers as instruments.

IDENTIFICATION DETAILS

Which genes to pick?

There does not appear to be a systematic relationship between the genes. (Table 2)

The raw data shows that unhealthy kids have significantly lower GPAs. (Table 5)

Genes are strongly related to health outcomes and behaviours. (Table 3)

Additional challenge is presented by comorbid conditions. (Table 4)

We consider two health vectors in our analysis.

Table 2: Number of Individuals with Each Genetic Marker

Genetic Markers	Genetic Marker Alleles	Number Of People with each allele	Interaction with the rare alleles			
			AA	TT	A1A1	DAT0
TPH	AA	120	****	4	5	16
	AC	393	****	15	20	39
	CC	380	****	12	27	65
CYP	TT	31	4	****	2	3
	CT	191	24	****	9	19
	CC	671	92	****	41	56
DRD2	A1A1	52	5	2	****	3
	A1A2	286	34	9	****	19
	A2A2	555	81	20	****	56
DAT	DAT0	72	16	3	3	****
	DAT1	317	38	13	17	****
	DAT2	498	65	15	32	****

Table 2b: Summary Statistics on GPA Performance by Health Disorder and Health Behavior

	Grade 10	Grade 11	Grade 12
Smokers	2.622 (0.633)	2.707 (0.768)	2.872 (0.660)
Non Smokers	3.211 (0.538)	3.198 (0.546)	3.236 (0.526)
Depressed	3.040 (0.640)	3.079 (0.643)	3.074 (0.624)
Non depressed	3.200 (0.536)	3.143 (0.598)	3.212 (0.539)
Obese	2.829 (0.615)	2.688 (0.721)	2.802 (0.619)
BMI <30	3.205 (0.555)	3.180 (0.570)	3.202 (0.554)
ADHD	2.841 (0.722)	2.931 (0.687)	2.923 (0.691)
No ADHD Diagnosis	3.169 (0.560)	3.145 (0.597)	3.192 (0.556)
AD	2.613 (0.722)	2.779 (0.730)	2.762 (0.734)
No AD Diagnosis	3.173 (0.558)	3.148 (0.594)	3.194 (0.554)
HD	3.017 (0.614)	3.030 (0.608)	3.047 (0.630)
No HD Diagnosis	3.154 (0.575)	3.136 (0.605)	3.181 (0.566)

Table 3: Relationship Between Genetic Markers with Health Behaviors and Health Outcomes During Adolescence

Gene	Marker	Depressed	Smokes	Obesity	BMI	ADHD	AD	HD
TPH	AA	0.149	0.158*	0.108*	23.939*	0.067	0.033	0.033
		(0.357)	(0.365)	(0.312)	(4.516)	(0.250)	(0.180)	(0.180)
	AC	0.150	0.105	0.074	23.291	0.074*	0.048	0.043
		(0.357)	(0.306)	(0.262)	(4.140)	(0.262)	(0.215)	(0.204)
	CC	0.156	0.101	0.079	23.403	0.050	0.039	0.039
		(0.363)	(0.301)	(0.270)	(4.640)	(0.218)	(0.195)	(0.195)
CYP	TT	0.165	0.121	0.032*	22.536*	0.129*	0.129*	0.097*
		(0.373)	(0.328)	(0.180)	(3.283)	(0.341)	(0.341)	(0.301)
	CT	0.159	0.111	0.058*	23.082*	0.031*	0.010*	0.026
		(0.366)	(0.315)	(0.234)	(4.195)	(0.175)	(0.102)	(0.160)
	CC	0.150	0.109	0.089	23.565	0.069	0.048	0.042
		(0.357)	(0.312)	(0.286)	(4.508)	(0.253)	(0.213)	(0.200)
DRD2	A1A1	0.189**	0.122	0.096	23.562	0.058	0.038	0.038
		(0.393)	(0.328)	(0.298)	(5.998)	(0.235)	(0.194)	(0.194)
	A1A2	0.174*	0.100	0.115*	23.860*	0.049**	0.021)*	0.035
		(0.380)	(0.301)	(0.320)	(4.651)	(0.216)	(0.144)	(0.184)
	A2A2	0.138	0.114	0.061	23.189	0.070	0.054	0.043
		(0.345)	(0.318)	(0.240)	(4.088)	(0.256)	(0.226)	(0.204)
DAT	DAT0	0.155*	0.155	0.077	23.685	0.064	0.038	0.051
		(0.363)	(0.363)	(0.268)	(5.310)	(0.247)	(0.194)	(0.222)
	DAT1	0.109	0.122*	0.095**	23.775*	0.091*	0.063*	0.060*
		(0.311)	(0.327)	(0.293)	(4.749)	(0.289)	(0.244)	(0.238)
	DAT2	0.172	0.104	0.072	23.161	0.044	0.030	0.026
		(0.378)	(0.306)	(0.259)	(4.004)	(0.206)	(0.171)	(0.160)

Note: Each cell presents the conditional mean and standard deviation in parentheses. We include information on the full sample used in our analysis for GPA, depression and smoking. The remaining health outcomes present summary information where each only one observation per individual is included.

Table 4: Relationship Between Health Behaviors and Health Outcomes During Adolescence

Behavior	Total Number	Nothing Else[1]	Also Smokes	Also AD	Also HD	Also ADHD	Also Obese	Also Depressed
Wave 3, N=834								
Nothing	471	***	***	***	***	***	***	***
Smokes	73	36	***	7	4	8	7	16
AD	33	5	7	***	14	33	3	15
HD	30	8	4	14	***	29	2	10
ADHD	49	25	8	33	29	***	4	19
Obese	68	39	7	3	2	4	***	17
Depression	140	93	16	15	10	19	17	***
Wave 4, N=863								
Nothing	477	***	***	***	***	***	***	***
Smokes	82	42	***	9	5	10	10	21
AD	37	7	9	***	17	37	4	15
HD	34	9	5	17	***	33	3	9
ADHD	54	25	10	37	33	***	5	19
Obese	70	34	10	4	3	5	***	17
Depression	146	96	21	15	9	19	17	***
Wave 5, N=879								
Nothing	483	***	***	***	***	***	***	***
Smokes	129	60	***	15	11	18	15	20
AD	38	8	15	***	18	38	4	10
HD	36	8	11	18	***	36	3	9
ADHD	56	30	18	38	36	***	5	15
Obese	67	28	15	4	3	5	***	10
Depression	107	66	20	10	9	15	10	***

Note: Each cell contains the number of individuals diagnosed with the respective row and column combination.

[1] For ADHD nothing else excludes AD and HD.

RESULTS I

(Table 5) OLS estimates suggest
- **Each health disorder is negatively and significantly associated with academic performance for the full sample.**
- **Large Negative impact of obesity on GPA for females.**
- **Boys' GPAs have a strong negative association with AD.**

(Table 7) 2SLS estimates suggest
- **Depression and obesity is significant for the full sample.**
- **Large offsetting impacts of AD and HD**
- **Substantial heterogeneity across genders:**
    The negative impact of poor health is substantial for girls.
    Nothing is significant for boys.

(Appendix Table 7) 3SLS estimates results not different from 2SLS; little efficiency gains.

Hausman tests reject the exogeneity of the health vectors.

Table 5: Ordinary Least Squares Estimates of the Achievement Equation

	Full Sample	Females Only	Males Only	Full Sample	Females Only	Males Only
ADHD	-0.218*** (0.071)	-0.216** (0.106)	-0.230** (0.098)	N/A	N/A	N/A
AD	N/A	N/A	N/A	-0.408*** (0.099)	-0.358 (0.185)	-0.464*** (0.125)
HD	N/A	N/A	N/A	0.111 (0.084)	0.032 (0.107)	0.164 (0.124)
Depression	-0.130*** (0.029)	-0.059* (0.033)	-0.221*** (0.050)	-.125*** (0.029)	-0.056* (0.033)	-0.214*** (0.049)
Obesity	-0.341*** (0.071)	-0.469*** (0.088)	-0.191 (0.103)	-0.34*** (0.071)	-0.474*** (0.087)	-0.185* (0.101)
Smoker in Home	-0.159*** (0.037)	-0.110** (0.045)	-0.215*** (0.057)	-0.16*** (0.036)	-0.11** (0.045)	-0.21*** (0.056)
N	2576	1366	1210	2576	1366	1210
R squared	0.24	0.26	0.21	0.25	0.26	0.22

Table 7: 2SLS Estimates of the Achievement Equation

	Full Sample	Females Only	Males Only	Full Sample	Females Only	Males Only
ADHD	0.218 (0.288)	-0.053 (0.274)	0.503 (0.330)	N/A	N/A	N/A
AD	N/A	N/A	N/A	-0.513 (0.364)	-0.455 (0.395)	-0.111 (0.400)
HD	N/A	N/A	N/A	0.822 (0.512)	0.032 (0.374)	0.204 (0.569)
Depression	-0.452** (0.198)	-0.186 (0.192)	-0.322 (0.240)	-0.322** (0.161)	-0.353** (0.167)	-0.273 (0.197)
Obesity	-0.450** (0.222)	-0.500*** (0.190)	0.096 (0.300)	-0.460** (0.229)	-0.470** (0.199)	0.023 (0.295)
Smoker in Home	-0.161*** (0.032)	-0.111*** (0.033)	-0.253*** (0.050)	-0.157*** (0.030)	-0.099*** (0.033)	-0.224*** (0.046)
Parent HS Dropout	-0.145** (0.059)	-0.153** (0.065)	-0.204** (0.103)	-0.147** (0.060)	-0.124* (0.072)	-0.227** (0.096)

Note: Corrected standard errors in parentheses. Regressions include parental age, parental age squared, parental gender, indicator for whether the responding parent is a biological parent, school and time period indicators. ***,**, * denote statistical significance at 1%, 5%, 10% level respectively.

DISCUSSION OF 2SLS RESULTS

These are reduced form coefficients

The coefficient estimates may capture a dynastic and cumulative effect of the impact of health disorders

The dynastic effect may be desirable
- Assortative mating critique

The availability of genes as instruments makes it crystal clear the level of difficulty in obtaining structural parameters
Yet even if possible – may be uninteresting

Compare to within twins strategies (epigenic modication)

Robustness to parental smoking and inclusion of parental characteristics.

RESULTS II

First stage regressions
   For each subsample the F statistic is above current cutoffs for weak instruments for each health outcome and health behaviour.

Overidentification tests
   J tests provide little evidence against the overidentifying restrictions.

The results suggest that statistically genes are promising instruments.

ROBUSTNESS

Results are robust to alternative genetic instrument sets.

Some have better first stage properties for the two subsamples than others.

Appendix Table 2.

Table 6: Summary Information on the Performance of the Instruments

	Full Sample	Females Only	Males Only	Full Sample	Females Only	Males Only
First Stage F statistics						
Full System	19.01	12.03	9.25	19.14	12.84	9.33
ADHD	10.61	8.13	5.37	N/A	N/A	N/A
AD	N/A	N/A	N/A	14.37	8.19	10.20
HD	N/A	N/A	N/A	8.66	11.83	6.70
Depression	12.20	5.18	10.41	12.20	5.18	10.41
Obesity	10.16	11.32	11.39	10.16	11.32	11.39
Smoking	7.33	7.27	6.30	7.33	7.27	6.30
P-values from Overidentification Tests						
Full System	0.611	0.278	0.386	0.217	0.236	0.486
ADHD	0.553	0.420	0.236	N/A	N/A	N/A
AD	N/A	N/A	N/A	0.842	0982	0.440
HD	N/A	N/A	N/A	0.845	0.812	0.266
Depression	0.773	0.822	0.465	0.773	0.822	0.465
Obesity	0.216	0.232	0.817	0.216	0.232	0.817
Smoking	0.267	0.874	0.421	0.524	0.617	0.293

RESULTS III

What if we do not account for endogenous health behaviour?

- The impacts of all health outcomes on GPAs increase markedly. (Table 8)

- In addition, depression now comes in significantly for boys.

- Hausman tests reject the exogeneity of smoking decisions.

- The different outcomes may be a result of different smoking patterns. In particular, boys with AD, HD and Depression smoke more frequently than girls with same disorders, and with more tar and nicotine content.

- Compensating benefits from smoking?

Table 8: Two Stage Least Squares Estimates of the Achievement Equation where Years of Smoking is Treated as Exogenous

	Full Sample	Females Only	Males Only	Full Sample	Females Only	Males Only
ADHD	-0.646* (0.343)	-0.672** (0.287)	0.010 (0.345)	N/A	N/A	N/A
AD	N/A	N/A	N/A	-1.18*** (0.408)	-1.46*** (0.393)	-0.414 (0.428)
HD	N/A	N/A	N/A	0.611 (0.588)	0.220 (0.420)	0.108 (0.613)
Depression	-1.115*** (0.230)	-0.474** (0.208)	-0.94*** (0.229)	-0.76*** (0.176)	-0.495*** (0.186)	-0.753*** (0.192)
Obesity	-0.501* (0.287)	-0.659*** (0.209)	0.290 (0.324)	-0.627** (0.263)	-0.724*** (0.219)	0.192 (0.317)
Smoker in Home	-0.082** (0.040)	-0.073** (0.036)	-0.192*** (0.053)	-0.100*** (0.033)	-0.069* (0.037)	-0.182*** (0.049)

RESULTS IV

What if We Ignore Comorbid Conditions?

We consider 2SLS estimates that include one health variable at a time. (Table 9)

- The impact of ADHD becomes larger and significant. Approximately equal to size of depression and obesity impacts.

- The impact of depression is approximately 40% larger.

Results suggest that even with valid instruments such as genetic markers biases regarding measurement error is still a concern.

Proper measures of health require a rich vector.

Table 9: Two Stage Least Squares Estimates of the Achievement Equation Including A Subset of Health Outcomes

Include health behaviors	Full Sample	Girls	Boys
ADHD	-0.351 (0.319)	-0.319 (0.359)	0.284 (0.452)
AD	1.392 (0.669)***	0.648 (0.633)	0.615 (0.546)
HD	-1.966 (1.183)***	-1.040 (0.609)	0.237 (0.911)
AD	0.529 (0.304)	-0.124 (0.400)	0.766 (0.383)
HD	-0.144 (0.517)	-0.330 (0.445)	0.972 (0.766)
Depression	-0.713 (0.302)**	-1.250 (0.455)*	-0.032 (0.391)
Obesity	-0.331 (0.329)	-0.352 (0.235)	1.067 (0.738)
Observations	2576	1366	1210

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 samples. Regressions include the non-health inputs in Table 7, school and time period indicators.
*, **, *** denote statistical significance at 1%, 5%, 10% level respectively.

CONCLUSIONS

1. Genetic markers are promising as instruments.

2. The impact of poor health outcomes on academic achievement is large (one s.d.) with substantial heterogeneity across genders. Why?

3. When investigating the impact of health status on education, it is important to account for endogenous health enhancing or health deteriorating behaviors.

4. The presence of high comorbidity of adverse health outcomes is striking which presents a challenge to properly control them all.

DIRECTIONS FOR FUTURE RESEARCH

Understanding the heterogeneity across genders
    Differences in self-perception (?), discriminating inputs (?), impact of physical disorders on energy levels (?)