Economics of Education Review 94 (2023) 102408

Contents lists available at ScienceDirect

Economics of Education Review

journal homepage: www.elsevier.com/locate/econedurev

Genetic risk, childhood obesity, and educational achievements

Maoyong Fan [a], Yanhong Jin [b], Man Zhang [c,#,*]

[a] Department of Economics, Ball State University, Muncie, IN, United States of America
[b] Department of Agricultural, Food and Resource Economics, Rutgers, The State University of New Jersey, United States of America
[c] School of Agricultural Economics and Rural Development, Renmin University of China, Beijing, China

ARTICLE INFO

JEL Classification:
I24
I25
I28
I18

Keywords:
Childhood obesity
Academic performance
Educational attainment
Genetic variants

ABSTRACT

We use the genetic risk exclusively related to body mass index as an instrumental variable to examine the causal effects of childhood obesity on educational achievements. We find that childhood obesity decreases high school GPA by 0.92 grade points (33.0%), GPAs of different subjects by 0.72–1.11 grade points (21.7–42.6%), the probability of college enrollment by 0.37, the probability of college completion by 0.65, and years of schooling by 2.19 years (14.8%). Additionally, we explore potential underlying mechanisms through which childhood obesity adversely influences educational outcomes. Our results indicate that childhood obesity does not have a statistically significant influence on cognitive abilities. Nevertheless, it negatively affects educational achievements via health factors (overall health status, health-damaging behaviors, and psychological well-being), school absenteeism and aggression, college aspirations and expectations, and family dynamics. This research provides evidence that childhood obesity can hinder children's educational progress, potentially affecting adult outcomes and exacerbating economic inequality.

1. Introduction

Childhood obesity stands out as the most widespread health issue affecting children in the United States.[1] The prevalence of childhood obesity reached 19.7% in 2017–2020, affecting 14.7 million children.[2] Childhood obesity is known to have profound effects on children's physical health, social, emotional wellbeing, and self-esteem (Cawley, 2010; Wang & Veugelers, 2008). It can have lasting effects into adulthood – obese children are likely to stay obese into adulthood (Simmonds et al., 2016) and develop non-communicable disease like diabetes at a younger age (Sahoo et al., 2015).

Several studies have investigated the potential impact of childhood obesity on academic performance and educational attainment (Almond et al., 2018; Jackson, 2009). However, the literature offers mixed results. Some studies suggest that childhood obesity has no effect on school performance (Kaestner & Grossman, 2009; Scholder et al., 2012; Von Hinke et al., 2016). Others show that it may have adverse effects (Ding et al., 2009; Rouse & Hunziker, 2020; Joseph J Sabia, 2007; J. J. Sabia & Rees, 2015). Furthermore, most previous studies have focused on the association between childhood obesity and school performance, with few exploring causal relationships (Martin et al., 2017; Santana et al., 2017). As such, further research is needed to advance the literature and identify the causal effects of childhood obesity on academic performance and educational attainment.

The key challenge in investigating the causal link between childhood obesity and educational achievements is that a naïve comparison of either academic performance or educational attainment between obese and normal-weight children is unlikely to establish causal effects. This is because childhood obesity is influenced by various confounding characteristics at the individual (e.g., cognition and self-control abilities), family (e.g., parents' income and education), and neighborhood (e.g., neighborhood safety) levels; and these characteristics may also affect education. To address this endogeneity issue, several studies have employed an instrumental variable (IV) approach. For example, Sabia and Rees (2015) use sibling BMI and maternal obesity status as an IV. They find that being overweight or obese in adolescence has statistically

* Corresponding author.
E-mail address: zhang_man@ruc.edu.cn (M. Zhang).
Authors are listed alphabetically, and the seniority of authorship is equally shared.

[1] Childhood obesity has been listed as one of the top ten child health concerns among parents. See the National Poll on Children's Health at https://mottpoll.org/tags/top-10-child-health-concerns. Last access on June 13th, 2022.

[2] Notably, the prevalence varied by age, with rates of 12.7% among 2- to 5-year-olds, 20.7% among 6- to 11-year-olds, and 22.2% among 12- to 19-year-olds in 2017-2020 (CDC, 2022).

https://doi.org/10.1016/j.econedurev.2023.102408
Received 10 January 2023; Received in revised form 22 March 2023; Accepted 22 April 2023
Available online 30 April 2023
0272-7757/

significant and adverse effects on high school GPA and college completion. However, using such variables as an IV may violate IV validity assumptions due to their associations with home environment and parental behavioral factors that are known to affect academic performance. Alternatively, researchers have suggested using child fixed effect (FE) models to address endogeneity (Palermo & Dowd, 2012; Scholder et al., 2012), which is useful only if the unobserved variables affecting both childhood obesity and educational outcomes are time-constant, not time-varying (Scholder et al., 2012). More research is needed to identify the causal effects of childhood obesity on educational achievements while adequately controlling for the confounding variables that impact both childhood obesity and education outcomes.

This study leverages genetic information to create a novel IV for children's weight status to address the endogeneity issue. The recent advancements in measuring genetic information within a large population at low cost have made genetic factors accessible to social science researchers (Belsky & Israel, 2014). As genetic information is predetermined before birth and remains stable throughout one's life, using genetic variants as an IV has gained increasing attention in economic and social studies (Belsky & Israel, 2014; Von Hinke et al., 2016). Previous studies have proposed using body mass index (BMI) genetic variants or BMI polygenic score (PGS)[3] as an IV for weight status (Böckerman et al., 2019; Ding et al., 2009; Scholder et al., 2012; Von Hinke et al., 2016). However, BMI PGS could compromise the validity of the IV due to the horizontal pleiotropy problem, where the pathways of BMI PGS correlate with the confounding factors between weight status and educational outcomes. For example, BMI PGS has been found to share genes with education PGS (Boardman et al., 2015; Okbay et al., 2016), thereby violating the independent and exclusion assumption of a valid IV as it correlates with intelligence-related control variables such as children's cognitive endowment or parents' education. To address this problem, we employ a novel two-stage regression approach to construct an IV based on BMI-related genetic information. This approach purifies the genetic information exclusively for BMI by residualizing the BMI PGS using a linear regression controlling for the confounding PGSs that share biological pathways with the BMI PGS, such as PGSs relating to education and cognitive ability, substance use, and mental health. By purifying the genetic information for BMI, our IV is less likely to be correlated with socioeconomic determinants of childhood obesity and better satisfies the assumptions of a valid IV.

Using the U.S. National Longitudinal Study of Adolescent to Adult Health (Add Health) data, we investigate the impact of childhood obesity on academic performance in high school and educational attainment in early adulthood. Our findings show that childhood obesity has adverse effects on both academic performance in high school and educational attainment in early adulthood. Childhood obesity reduced the overall high school GPA by 0.92 (33%) grade points and the subject GPAs of English, history, social science, math, science, and physical education by 0.72–1.11 (21.7–42.6%) grade points. Moreover, childhood obesity decreases educational attainment by 2.19 years (14.8%) and lowers the probability of college enrollment and completion by 0.37 and 0.49, respectively. As educational attainment has been shown to be a critical factor in determining earnings (Heckman & Mosso, 2014) and income inequality (Hoffmann et al., 2020; Lemieux, 2006), our study highlights that childhood obesity could significantly impact economic well-being and income equality through its influence on education. We further investigate the possible mechanisms through which childhood obesity could negatively affect educational outcomes. Our analysis rules out the possibility that childhood obesity impacted cognitive abilities and confirms that it had a negative effect on general health, health-compromising behaviors, psychological well-being, school absenteeism and aggression, college aspiration and expectation, and family dynamics. Our results are robust to various tests and placebo checks. The findings not only highlight the causal effects of childhood obesity on educational achievements, but also suggest that preventive measures for childhood obesity could improve high school academic performance in the short run and educational attainments in the long run, and ultimately reducing economic inequality.

This study contributes to three strands of literature. First, it adds to the general body of knowledge concerning the relationship between education and health. It is well established that education and health are highly interdependent, with a bi-directional, positive relationship between two (Currie, 2009; Cutler & Lleras-Muney, 2006). Given the impact of education on labor market outcomes, this study highlights how childhood obesity may have contributed to income inequality in the United States by lowering educational levels. Second, this research enriches the literature on human capital development by shedding light on the relationship between childhood obesity and the development of neurocognitive abilities, an area that has hitherto been poorly understood (Liang et al., 2014). A key challenge in this area is to measure cognitive endowment and development, which are the fundamental factors in children's educational outcomes (Benner & Mistry, 2007; A.K. Cohen, Rehkopf et al., 2013; Hossler & Stage, 1992). Employing a gene-based IV model and controlling for one's intelligence PGS as a covariate representing the cognitive endowment, we have identified the mechanisms through which childhood obesity may affect educational outcomes. Third, this study contributes to the literature using genetic information in social science studies. We develop a novel two-stage strategy to construct an IV more exclusively related to BMI and obesity than the raw BMI PGS. This strategy can be readily adapted to address the pleiotropy problem and accommodate new discoveries in PGSs and underlying biological pathways. The constructed IV allows for the separation of the reciprocal relationship between health and education and emphasizes the causal effects of childhood obesity on educational outcomes.

The rest of this study proceeds as follows. Section 2 provides an overview of genetics and its use in social science. Section 3 describes the data and key variables. Section 4 describes the empirical strategies and the construction of IV. The IV validation and main results are presented in Section 5. Section 6 validates our main results by providing robustness checks and placebo test. Section 7 discusses mechanisms through which childhood obesity can potentially affect educational achievements. Section 8 concludes.

2. Genetic information, polygenic scores, and their use in social science

The following section offers a comprehensive exposition of genetic information, particularly the construction of polygenic scores, and a concise review of how genetic data have been incorporated into social science and economic research in recent times. To facilitate the readers' comprehension, Table A1 contains definitions of scientific terminologies and acronyms used in this study.

2.1. BMI genes and polygenic score

The Human genome consists of 23 pairs of Deoxyribonucleic acid (DNA) molecules, which are also called chromosomes. DNA is a molecule containing genetic instructions for enzymes or functional molecules, and is responsible the development and functioning of organisms. DNA is composed of two polynucleotide chains with repeating units called nucleotides. Nucleotides are composed of three distinctive chemical units, and one important unit is called nucleobase (or nucleotide base). There are four types of nucleotide bases in DNA: adenine (A), cytosine (C), guanine (G), and thymine (T). The base pairs on the two polynucleotide chains, namely, adenine-thymine (AT) and guanine-cytosine(GC), contribute to the DNA double helix structure. Single-nucleotide polymorphism (SNP) is a term for genomic variation at a

[3] PGS is the weighted sum of genetic variants. Detailed explanation can be found in section 2.1.

single nucleotide base that occurs at a specific position in the genome. For example, the guanine (G) nucleotide may appear at a specific position in the genome for most individuals, whereas the adenine (A) nucleotide may occupy that position for a minority of individuals. To be considered as a SNP rather than a rare mutation, the variation must be presented in more than 1% of the population. More than 600 million SNPs have been found in populations worldwide.[4] Some SNPs are associated with susceptibility to a wide range of diseases. Owing to technological advances and reducing costs of DNA sequencing, several human genome projects have recruited thousands or even millions of participants. Such programs allow researchers to identify genetic contributors to a specific disease by comparing the SNP patterns between people with and without the disease. Gene-wide association studies (GWASs) screen the entire genome of a large number of individuals and identify SNPs associated with a specific phenotype[5] or disease.

Since an individual has two sets of chromosomes and inherits one from each parent, one has a score of risk allele[6] frequency (0, 1, or 2) at each SNP position. GWASs usually estimate the contributions of SNPs to the interested outcome (Y_{ij}) by gathering J observable SNPs and estimating J linear regressions separately using an equation similar to the following (Barth et al., 2020):

$$Y_{ij} = \beta_j SNP_{ij} + \alpha X_i + \varepsilon_{ij} \text{ such that } \forall\, j \in J \tag{1}$$

Where Y_{ij} is an outcome variable (e.g., BMI, years of schooling, substance use behaviors) for individual i on J observable SNPs; X_i is a vector of control variables, usually including individual characteristics (e.g., sex and age) and other factors such as recruitment center and genotyping batches (Locke et al., 2015; Yengo et al., 2018). SNP_{ij} is the risk allele frequency score $\in (0, 1,\ 2)$ at SNP_j for individual i. After checking the significance level for each SNP and subjecting to quality control measures, K key significant SNPs out of J SNPs for the outcome variable are identified in the GWAS. For example, Locke et al. (2015) report 97 BMI-associated SNPs, and Yengo et al. (2018) report 947 BMI-associated SNPs after involving more participants in their GWAS. PGSs are calculated as a weighted sum of associations between allele frequencies at individual SNPs and the specific phenotype (e.g., BMI) according to the summary statistics (beta-coefficients) from GWASs:

$$PGS_{BMI_i} = \sum_{k=1}^{K} \beta_k SNP_{ik} \text{ such that } \forall\, k,\ \beta_k \in \{\beta_J\} \tag{2}$$

In this study, the BMI PGS of the Add Health participants is estimated based on a two-sample prediction. A vector of coefficients $(\beta_1, \beta_2 \ldots \beta_K)$ in Eq. (2) is obtained from Yengo et al. (2018)'s GWAS as demonstrated via Eq. (1), because the sample size (~700,000 individuals) of Yengo et al. (2018) is larger and more representative than the Add Health cohort (~9000 participants). The BMI PGS is obtained by multiplying the coefficients identified by Yengo et al. (2018) by the corresponding allele frequencies of specific SNPs of Add Health participants according to their DNA sequencing results (Braudt & Harris, 2020). In addition, PGSs for education and cognitive ability, substance use, and mental health are constructed in a similar way and provided by Add Health. The raw PGSs are then standardized within each ancestry group.[7] We use the standardized PGSs in this study. To account for the population stratification, the first ten ancestry-specific principal components are always controlled for in our analyses.

2.2. Leveraging genetic information in social science

The advances in low-cost sequencing techniques and the explosion of GWASs in recent years make it possible for researchers to access genetic risks of health and diseases that are previously unobservable within large populations (Belsky & Israel, 2014). Epidemiology studies have applied genetic variants as instrumental variables for environmentally modifiable exposures (e.g., smoking, drinking, and serum iron levels) to generate causal inferences of medical interventions (Davey Smith & Hemani, 2014; Lawlor et al., 2008; Richmond & Smith, 2022). This method, also known as Mendelian Randomization (MR), has been applied to identify causal relationships in medical and epidemiological studies in recent decades (Koellinger & De Vlaming, 2019; Lawlor et al., 2008). On the other hand, integrating genetics into social science fields has received increasing interest, and is believed to deliver more precise and realistic answers to classical questions and theories in social science (Belsky & Israel, 2014; Harden & Koellinger, 2020).

Social science studies leverage genetic information, PGSs, in two different ways. First, some utilize PGSs as individual's genetic endowments and investigate their relationships with economic performance and life outcomes (Barth et al., 2020; Belsky et al., 2018; Cawley et al., 2019; Domingue et al., 2014). For example, Barth et al. (2020) find that educational attainment PGS is linked to wealth inequality at retirement, and the gene-wealth gradients could be partially explained by factors such as inheritances, risk preference, mortality, and portfolio decisions. Other researchers have found that the genetic environment, combined with the neighborhood and peer effects, could affect one's behaviors or socioeconomic outcomes (Belsky et al., 2019; Brunello et al., 2020). For example, Brunello et al. (2020) report that the BMI PGS of peers positively affects an individual's BMI PGS and obesity risk, especially for females. Belsky et al. (2019) find that compared with children in advantaged neighborhoods, children in disadvantaged neighborhoods have a genetic selection and concentration with higher teenage pregnancy PGS and lower educational attainment PGS, while no significant difference is found for BMI PGS and schizophrenia PGS for children growing up in different neighborhoods.

Second, empirical economic studies generally assume that genetic factors are represented in an error term, either averaged out with certain fixed effects or left as an unobserved heterogeneity (Papageorge & Thom, 2020). Recent empirical studies have used genetic variants as an IV for specific phenotypes (e.g., weight status) to deal with endogeneity. The genetic variants could be a single SNP, multiple SNPs, or PGS as unweighted or weighted sum of risk alleles scores for given SNPs (Richmond & Smith, 2022). The earliest studies using BMI genetic variants can be traced back to Ding et al. (2009) and Norton and Han (2008), who used no more than five SNPs as IVs for obesity, including dopamine transporter (*DAT*) and tryptophan hydroxylase locus (*TPH*). These genetic variants are called "candidate genes" for weight status, because their associations with BMI are found to suffer from limited robustness and poor reproducibility across populations (Benjamin et al., 2012; J. M. Fletcher, 2018). Speliotes et al. (2010) make a breakthrough by identifying 32 BMI-related SNPs in their GWAS with around 120,000 individuals and provide an unweighted sum of these SNPs as the BMI genetic score. Add Health has released a BMI genetic score for the first time for its 1886 participants based on Speliotes et al. (2010). Willage (2018) analyzes the Add Health cohort and reports a negative causal impact of weight status on mental health using the unweighted sum of SNPs identified by Speliotes et al. (2010) as IV for BMI. Von Hinke et al. (2016) use the BMI genetic score based on Speliotes et al. (2010) as the IV for children's fat mass and find no causal effects of children's fat mass on their academic performance in a British cohort.

Although the BMI genetic score of Speliotes et al. (2010) has been applied in several economic studies (Böckerman et al., 2019; Von Hinke et al., 2016; Willage, 2018), researchers are concerned that 32 BMI-related SNPs have limited power in explaining the variation of BMI, and the unweighted sum of 32 SNPs dismisses heterogeneous

[4] See details at https://medlineplus.gov/genetics/understanding/genomicresearch/snp/. Last access on June 14th, 2022.

[5] Phenotype refers to an individual's observable trait, such as height, BMI, educational attainment in years, etc. An individual's phenotype depends on his/her genetic endowment and environmental factors.

[6] An allele refers to the variant form of SNPs (Lawlor et al., 2008).

[7] The genotyped sample of Add Health participants was restricted into European, Hispanic, African, and East Asian ancestries (Braudt & Harris, 2020).

Table 1
Summary Statistics of Key Variables.

Variables	Full sample		Normal weight sample		Overweight sample (BMI percentile ≥ 0.85)		Obese sample (BMI percentile ≥ 0.95)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Polygenic scores (PGSs) measured in Wave IV								
BMI PGS	−0.048	0.993	−0.172	0.982	0.349	0.925	0.500	0.921
Confounding PGSs								
Intelligence	0.054	0.985	0.056	0.986	0.048	0.982	0.023	0.974
Cognitive function	0.060	0.995	0.047	0.984	0.104	1.029	0.084	1.062
Educational attainment	0.081	0.989	0.112	0.993	−0.017	0.971	−0.034	0.983
Smoking	0.039	0.996	0.042	0.994	0.033	1.003	−0.000	1.053
Drinking	−0.006	1.015	−0.041	1.002	0.105	1.047	0.111	1.056
Major depressive disorder	−0.040	1.016	−0.025	1.010	−0.087	1.034	−0.087	1.043
Depressive symptom	−0.024	1.009	−0.021	1.010	−0.031	1.006	−0.033	1.043
Mental health cross disorder	0.024	0.966	0.037	0.965	−0.021	0.966	0.014	0.704
High school GPA on a four-point scale from official transcripts obtained in Wave III								
Overall high school GPA	2.737	0.778	2.794	0.775	2.553	0.760	2.496	0.762
Math	2.360	0.921	2.421	0.924	2.166	0.885	2.105	0.839
Science	2.437	0.946	2.495	0.945	2.251	0.927	2.201	0.910
English	2.537	0.915	2.605	0.913	2.319	0.889	2.237	0.883
History and social science	2.628	0.956	2.690	0.943	2.429	0.972	2.350	0.963
Physical education	3.280	0.865	3.335	0.838	3.106	0.927	3.046	0.969
Educational attainment reported in Wave IV								
College enrollment	0.733	0.442	0.761	0.427	0.644	0.479	0.601	0.491
Received college degree	0.376	0.484	0.411	0.492	0.263	0.441	0.234	0.424
Years of schooling (years)	14.67	2.104	14.83	2.131	14.15	1.924	14.02	1.959
No. of observations	2585		1970		615		248	

Notes: The Add Health PGS data are provided in the sub-dataset, called "Polygenic Scores (PGSs) in Add Health-Release 2." PGSs for BMI, intelligence, cognitive function, and educational attainment are constructed by an out-of-sample prediction according to Genome-wide association studies (GWAS) of Yengo et al. (2018), Savage et al. (2018), Davies et al. (2018) and Lee et al. (2018), respectively. In addition, two substance-use PGSs and three mental health related PGSs provided in Add Health, namely smoking PGS (Liu et al., 2019), drinking PGS (Liu et al., 2019), major depressive disorder PGS (Howard et al., 2019), depressive symptom PGS (Baselmans et al., 2019), and mental health cross disorder PGS (Consortium, 2013).

contributions of each SNP on BMI across ancestries. Locke et al. (2015) further identify additional BMI-related SNPs and construct a BMI PGS as a weighted sum of 97 BMI-related SNPs. Emerging studies start to instrument weight using SNPs or PGS identified by Locke et al. (2015) to explore the consequence of weight status on income, medical costs, and socioeconomic status (Dixon et al., 2020; Edwards et al., 2021; Tyrrell et al., 2016). As more participants were involved in GWAS (~700,000 individuals), Yengo et al. (2018) expand BMI PGS by involving 947 BMI-related SNPs, leading to higher explanatory power for BMI than the previous PGSs using fewer SNPs. While the increasing number of BMI-related SNPs improves the prediction power of BMI or obesity status, it makes the pleiotropy problem a legitimate concern when using the PGS score as an IV. Given the potential overlapping of genetic variants between BMI PGS and educational attainment PGS (Boardman et al., 2015; Okbay et al., 2016), the BMI PGS may fail to be a valid IV as it is likely correlated to covariates (e.g., cognition and education) or the error term in the second stage of the IV model. Likewise, BMI PGS may also share confounding biological pathways with self-control abilities or mental health, as previous studies have suggested significant associations between obesity and self-control and mental health problems (Avila et al., 2015; Maoyong Fan & Jin, 2014b). Manually excluding the confounding SNPs from the BMI PGS is not scientifically realistic, as biological pathways of many SNPs are not fully understood. To address the pleiotropy problem between height PGS and educational attainment PGS, DiPrete et al. (2018) propose to divide the original GWAS sample into nonoverlap subsamples (i.e., by different education levels) and then construct height PGSs for each subgroup.[8] However, gaining access to the GWAS sample and calculating the PGS for each subgroup are less likely to be feasible for most social science researchers.[9] We propose a practical, convenient, and robust approach to deal with the pleiotropy problem, which does not require access to the GWAS sample and can be readily used by other researchers and adapted when new confounding SNPs and genetic overlapping pathways are discovered (See the discussion in Section 4).

3. Data

Add Health is a U.S. national longitudinal study starting from 1994 to 1995 (Wave I) and has been followed by four waves of interviews in 1996 (Wave II), 2001–2002 (Wave III), 2008–2009 (Wave IV), and 2016–2018 (Wave V). Around 20,000 students in grades 7–12 and 17,000 parents participated in-school and in-home interviews in Wave I. In Wave III, 91.7% of the Add Health participants ($N = 13{,}901$) signed a Transcript Release Form and authorized Add Health to request their official high school transcripts. In Wave IV, 9130 participants gave their consent to provide their saliva to collect genetic information through DNA sequencing techniques. Add Health has also collected a rich set of information about the participants' sociodemographic, psychological, cognitive, health conditions, and contextual information about their resided neighborhoods in each wave.

We restrict the sample to adolescents aged 11–19 years old in Wave I when they were asked to self-report weight and height and exclude underweight children (~4%). We assume that the participants in Wave IV had completed college if they ever attended, as their youngest age was 24 years old with a mean of 28.3 years old. This study uses individual, family, and neighborhood characteristics in childhood collected in Wave I, high school academic performance provided in Wave III, and educational attainment and genetic information collected in Wave IV. To avoid the population stratification issues of genetics, this study focuses on white participants. A total of 2585 observations is used in this study.

Table 1 summarizes the key variables. Approximately 10% of children in Wave I were obese (BMI percentile ≥ 0.95), and 24% were overweight (BMI percentile ≥ 0.85). The average high school GPA was 2.74 based on a four-point scale, and the average GPAs by subject were

[8] In this case, leveraging height PGS for populations with different education levels as an IV is able to yield reasonable estimates for the causal impact of body height on educational attainment (DiPrete et al., 2018).

[9] It requires several demanding steps, including gaining access to the original genetic data, combining genetic data from different cohorts, and identifying key SNPs to estimate the genetic scores for each subgroup.

2.36 in math, 2.44 in science, 2.54 in English, 2.63 in history and social science, and 3.28 in physical education. In Wave IV, 73.3% had ever attended college, 37.6% received a four-year college degree, and the average years of schooling was 14.67. Table 1 shows that obese children had the worst academic performance and educational attainment, followed by overweight children, whereas normal-weight children had the best outcomes. The definitions of the dependent variables are presented in Table A2, and the summary statistics of all covariates and mediating factors are shown in Table A3.

4. Empirical strategies

4.1. Constructing the IV exclusively related to BMI-related genetic risk

Genetic variants have several advantages in controlling for endogeneity. In general, an individual's DNA sequence is predetermined at conception and expected to remain the same across human tissues and lifetimes (Belsky & Israel, 2014). More importantly, genetic information is most likely independent of changes in one's social environment. Therefore, compared with traditional IVs such as the weight status of biological relatives (Joseph J Sabia, 2007; Scholder et al., 2012), using BMI genetic information as an IV for weight status could better deal with the endogeneity problem of obesity.

Several recent studies have used BMI PGS as an IV for either BMI or obesity (Böckerman et al., 2019; Edwards et al., 2021; Scholder et al., 2012; Von Hinke et al., 2016; Willage, 2018). However, the validity of BMI PGS as an IV may be compromised because of the pleiotropy problem (Baiocchi et al., 2014; Davey Smith & Hemani, 2014; Von Hinke et al., 2016). Given that the pathways of genes involved in BMI PGSs are reported as synaptic function, insulin action, energy metabolism, lipid metabolism, and others, some BMI-associated genes are also involved in the development of the central nervous system, which may affect intelligence, cognition, and educational outcomes (Locke et al., 2015; Yengo et al., 2018). Locke et al. (2015) identify the tissues and cell types where BMI-associated genes are highly expressed. They find that the enrichment sites in the brain also include the hippocampus and limbic system, which have a vital role in memory, learning, and cognition. *FTO,* an important BMI-associated gene involved in BMI PGS (Locke et al., 2015; Speliotes et al., 2010), is found to regulate dopaminergic signaling in the brain and affect the reward learning[10] in humans (Sevgi et al., 2015). Indeed, we find significant correlations between BMI PGSs and PGSs for educational attainment, cognition, and intelligence in the Add Health cohort, suggesting the overlap of biological pathways of SNPs involved in BMI and educational attainment, intelligence, and cognition. The IV estimates would be biased if we ignore the pleiotropy problem.

To deal with the pleiotropy problem, we create a genetic predictor of BMI by netting out the confounding PGSs from the raw BMI PGS using a two-stage regression approach:

$$BMIPGS_i = \gamma_0 + \sum_{j=1}^{J} \gamma_j ConfoudingPGS_{ij} + \omega_i \tag{3a}$$

$$GPBMI_i = BMIPGS_i - \widehat{BMIPGS}_i \tag{3b}$$

where $BMIPGS_i$ is the latest version of BMI PGS consisting of 947 key SNPs according to Yengo et al. (2018) for an individual *i*. $ConfoudingPGS_{ij}$ is a set of *J* confounding PGSs, including PGSs for three education related PGS for educational attainment, cognitive function, and intelligence; two PGSs for substance use; and three PGSs for mental health.

Lack of self-control in childhood is associated with a wide range of adverse adult outcomes, including poor education, poor physical and mental health, substance dependence, sexual risk-taking, criminal misdemeanor, and poor personal finances even after controlling for gender, socioeconomic status, and childhood IQ (Fergusson et al., 2013; Moffitt et al., 2011). Additionally, lack of self-control is more predominant for obese children than for their normal-weight peers (Fan & Jin, 2014b). The overlapping pathways between self-control and BMI PGSs would bias the IV results. Add Health does not provide measures of self-control ability. We use PGSs for substance use (e.g., smoking and drinking) based on Liu et al. (2019)'s GWAS as a proxy for self-control capability and incorporate them in Eq. (3a) as confounding PGSs.

Mental health is found to be associated with both childhood obesity (Avila et al., 2015), educational attainment (Amin et al., 2023; Cornaglia et al., 2015), and labor market outcomes (Fletcher, 2013a). The overlapping pathways between mental health and BMI PGSs might lead to estimation bias. We therefore incorporate three mental health related PGSs provided by Add Health in Eq. (3a), namely major depressive disorder PGS (Howard et al., 2019), depressive symptom PGS (Baselmans et al., 2019), and mental health cross disorder PGS (Consortium, 2013), as confounding PGSs.

Genetic predictor of BMI denoted by $GPBMI_i$ is the difference of the raw BMI ($BMIPGS_i$) and the predicted one ($\widehat{BMIPGS_i}$) from Eq. (3a). Intuitively, this procedure nets out the portion of the variation in BMI PGS we could have predicted using the confounding PGSs, leaving the residual genetic information that is exclusively related to weight status. We use the genetic predictor of BMI, denoted by *GPBMI* in Eq. (3b), as the IV for childhood obesity.

4.2. The IV model

There are three main threats to interpreting the ordinary least squares (OLS) estimates as causal effects, as BMI is not randomly assigned across individuals. First, OLS estimates may suffer from omitted variable bias because children's weight status is correlated with many individual, family, and neighborhood characteristics, many of which are unobservable to researchers. Second, OLS estimates may suffer from reverse causality bias because education (or more broadly knowledge) may improve decision-making ability and lead to better health (Cohen et al., 2013; Kenkel et al., 2006). Finally, OLS estimates may suffer attenuation bias because of measurement errors (e.g., people tend to over-reported height and underreport weight, especially for females (Fan & Jin, 2015)).

We use an IV approach to isolate random variation in BMI and estimate the causal effects of childhood obesity on educational achievements. We estimate the following IV model:

$$Obese_{ik} = \alpha_0 + \alpha_1 GPBMI_{ik} + \beta IntPGS_{ik} + X_{ik}^{'}\alpha + PC_{ik}^{'}\eta + \pi_k + \zeta_{ik} \tag{4a}$$

$$EDU_{ik} = \theta_0 + \theta_1 \widehat{Obese}_{ik} + \delta IntPGS_{ik} + X_{ik}^{'}\theta + PC_{ik}^{'}\psi + \tau_k + \mu_{ik} \tag{4b}$$

where EDU_{ik} is an educational outcome (i.e., high school GPA, college enrollment, college completion, and years of schooling) for individual *i* at school *k*. The variable of interest, $Obese_{ik}$, indicates whether individual *i* is obese in Wave I. $GPBMI_{ik}$ is the IV based on personal BMI-related genetic risk, which is constructed based on Eqs. (3a) and (3b). $PC_{ik}^{'}$ consists of the first ten ancestry-specific principal components for PGS (Braudt & Harris, 2020). Equations(4a) and (4b) also include the school fixed effects denoted by μ_k and τ_k and error terms denoted by ζ_{ik} and μ_{ik}.

The model also controls for individual, family, and neighborhood characteristics denoted by $X_{is}^{'}$. Individual characteristics consist of children's age, age squared, gender, and school grade. Since birth weight has been found to have long-term effects on adulthood outcomes, including educational outcomes (Oreopoulos et al., 2008; Rosenzweig & Zhang, 2013), we include an indicator for low birth weight. To account

[10] Reward learning is a learning process by which behaviors are trained by the received information (stimuli, actions, or contexts). https://www.nimh.nih.gov/research/research-funded-by-nimh/rdoc/constructs/reward-learning. Last access on June 7th, 2023.

Table 2
The Effect of Genetic Predictor of BMI (*GPBMI*) on Childhood Obesity and Childhood Overweight (First Stage of the IV estimation).

	Models			
	(1)	(2)	(3)	(4)
Panel A: Obese (BMI percentile ≥ 0.95)				
GPBMI	0.066***	0.067***	0.067***	0.067***
	(0.007)	(0.007)	(0.007)	(0.007)
Cragg-Donald F-stat	87.27	90.57	91.00	90.58
Root MSE	0.306	0.303	0.303	0.304
No. of Observations	2218	2218	2218	2218
Panel B: Overweight (BMI percentile ≥ 0.85)				
GPBMI	0.092***	0.094***	0.093***	0.093***
	(0.008)	(0.008)	(0.008)	(0.008)
Cragg-Donald F-stat	110.6	115.6	115.5	114.9
Root MSE	0.411	0.408	0.407	0.407
No. of Observations	2585	2585	2585	2585
Personal characteristics	N	Y	Y	Y
Family characteristics	N	N	Y	Y
Neighborhood characteristics	N	N	N	Y
School fixed effects	Y	Y	Y	Y

Notes: This table presents the first-stage results of IV estimation, where childhood obesity or childhood overweight is the dependent variable. The instrumental variable is the genetic predictor of BMI constructed based on genetic information. Standard errors in parentheses are clustered by school reported in Wave I. * $p < .1$, ** $p < .05$, *** $p < .01$.
Models (1)-(4) stepwise add the respondents' personal, family, and neighborhood characteristics. Model (1) only controls for the school fixed effects and ten ancestry-specific principal components for PGSs. Model (2) additionally control for personal characteristics, including the respondents' age, age squared, gender, intelligence PGS used as the proxy for learning ability, school grade, and a dummy variable for low birth weight. Model (3) additionally controls for family characteristics reported in Wave I, including the number of children of the respondent's biological parents, logarithmic parental household income in Wave I, the ratio of parental household income to the tract-level per capita income in 1989, whether parents having education beyond high school, single-parent family or not, and participation in the parent-teacher organization in their children's school or not. Model (4) further includes neighborhood characteristics reported in Wave I. Neighborhood characteristics at the census tract level are the share of people aged 16–19 years old not in school or armed forces and not high school graduate, the share of people aged at least 25 years old having college degree or above, per capita income in 1989, unemployment rate, the share of female civilian labor force, the share of married-couple households, and poverty rate. Neighborhood characteristics at the county level consist of crime rate and per capita government budget spent on education, health and hospitals, and public welfare, respectively.

for individual's differences in learning ability, we include the intelligence PGS in the model, denoted by $IntPGS_{ik}$.

Family and neighborhood characteristics at childhood were collected in Wave I. Family characteristics include the number of children of their biological parents, the logarithm of household income of parents, the ratio of parental household income to per capita income at the census tract level, being a single-parent family or not, parents having post-secondary education after high school or not, and parent's participation in parent-teacher association in their children's school or not. We also include a set of neighborhood variables at either the census tract level or the county level, given that childhood neighborhood is a critical predictor for both obesity status and educational outcomes in adulthood (Chetty & Hendren, 2018a, 2018b; Maoyong Fan & Jin, 2014a). Neighborhood characteristics at the census tract level are the share of the population aged 16–19 not in school or armed forces and not high school graduate, the share of the population aged 25 years and above having a college degree or above, per capita income in 1989, unemployment rate, the share of female civilian labor force, the share of married-couple households, and poverty rate. County-level neighborhood characteristics consist of crime rate and per capita government budget spent on education, health and hospitals, and public welfare, respectively.

5. Results

5.1. The validity of IV

A valid IV needs to satisfy the assumptions of the non-zero effect of IV on the treatment, conditional independence, exclusion restriction, and monotonicity (Von Hinke et al., 2016). The non-zero effect of IV on treatment requires that IV has enough power to explain much of the variation of childhood obesity. Table 2 presents the first stage results of the IV models. *GPBMI* is highly associated with both childhood obesity and childhood overweight and its coefficients are statistically significant at the 1% level. The F statistics in the first stage ranges from 87 to 116 across all estimations, which is much larger than the traditional cut-off value for weak IV (Stock & Yogo, 2005). Thus, *GPBMI* is a powerful predictor of weight status.

Since the conditional independence and exclusion assumptions are impossible to be directly tested, we use several approaches to indirectly validate these two assumptions. First, Table 2 shows the first-stage estimation results by gradually adding control variables. Model 1 only includes the school fixed effects and the first ten ancestry-specific principal components of PGS. Models 2–4 stepwise add personal, family, and neighborhood characteristics. The coefficients of *GPBMI* remain extremely stable across the four models in both statistical significance and magnitude: 0.066–0.067 for childhood obesity (Panel A) and 0.092–0.094 for childhood overweight (Panel B). The results suggest that *GPBMI* is less likely to correlate with other confounding factors. Second, we regress *GPBMI* directly on all the covariates and find that it is not statistically significantly correlated with any covariate (Columns 1 of Table 3). Therefore, we find little evidence that our IV violates the conditional independence and exclusion assumptions.

We also compare our IV with two commonly used IVs in the obesity literature, namely raw BMI PGS and obesity status of parents. Columns 2 and 3 of Table 3 show the regression results where BMI PGS and obesity status of parents are regressed on the same set of covariates, respectively. We find that BMI PGS is statistically correlated with intelligence PGS[11] that is used as a proxy for cognitive ability and the share of female

[11] The significant correlations between BMI PGS and educational outcomes are also reported in Willage (2018), where BMI genetic risk is found to be associated with Picture Vocabulary Test (PVT) score and self-reported high school GPA. Overlaps or correlations of biological pathways between BMI PGS and education-related PGS are also found in two GWASs (Boardman et al., 2015; Okbay et al., 2016).

Table 3
Associations between the Generic Predictor for BMI (*GPBMI*) and each of the Control Variables.

Variables	*GPBMI*		BMI PGS		Parental Obesity Status	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
Personal Characteristics						
Age in Wave I	−0.240	0.331	−0.090	0.351	0.070	0.180
Age square	0.010	0.010	0.005	0.011	−0.002	0.006
Female (=1, male=0)	0.035	0.038	0.037	0.039	0.011	0.024
Low birth weight (=1, otherwise=0)	0.053	0.080	0.028	0.085	−0.060	0.038
Intelligence PGS	0.029	0.021	**−0.063*** **	0.022	0.006	0.011
Family Characteristics						
Number of children of the respondent's biological parents	−0.006	0.016	−0.008	0.017	0.013	0.010
Log parent self-reported household income in Wave I	−0.017	0.014	−0.024	0.015	−0.011	0.011
Ratio of parental household income to the tract-level per capita income in 1989	−0.002	0.003	−0.003	0.004	−0.004	0.002
Single-parent family (Yes=1, No=0)	−0.048	0.051	−0.016	0.054	**−0.062*** **	0.024
Parent having education beyond high school (Yes=1, No=0)	0.004	0.039	−0.054	0.039	**0.039***	0.021
Participated in the parent-teacher association (Yes=1, No=0)	−0.032	0.043	−0.034	0.045	0.012	0.020
Neighborhood characteristics at the tract level						
Share of people aged 16–19 not in school or armed forces and not high school graduate	−0.045	0.331	−0.013	0.339	−0.308	0.217
Share of people aged 25 years/over and with college degree/more	−0.020	0.488	−0.049	0.488	−0.311	0.213
Per capita income in 1989 (1000 dollars)	−0.011	0.013	−0.011	0.013	−0.000	0.006
Unemployment rate	−0.983	1.529	−0.799	1.602	−0.037	0.712
Share of female civilian labor force	−1.374	1.019	**−1.875***	1.043	0.346	0.469
Share of married-couple family households	0.215	0.641	0.016	0.680	0.083	0.357
Share of families with income in 1989 below poverty level	−0.074	0.821	−0.325	0.897	**0.646***	0.333
Neighborhood characteristics at the county level						
Per capita government budget spent on education ($1000)	−0.138	0.647	0.056	0.711	0.010	0.242
Per capita government budget spent on health and hospitals ($1000)	−0.612	0.712	−0.401	0.783	**−0.718*** **	0.238
Per capita government budget spent on public welfares ($1000)	0.699	1.697	0.709	1.668	0.179	0.855
Crime rate (*1000) per 100,000 population	0.011	0.041	0.025	0.048	0.014	0.019
Ancestry-specific ten principal components	Yes		Yes		Yes	
School and grade fixed effects	Yes		Yes		Yes	
No. of observations	2585		2585		2507	

Notes: This table presents the correlation between the genetic predictor of BMI (Column 1), the raw BMI PGS (Column 2) and parental obesity status (Column 3) and each of the control variables for education outcomes. Standard errors in parentheses are clustered by school reported in Wave I. * $p < .1$, ** $p < .05$, *** $p < .01$.

civilian labor force. Parental obesity status is statistically correlated with being a single-parent family or not, parental education, the share of families with income below the poverty level at the tract level, per capita government budget spent on health and hospitals at the county level. Therefore, using raw PGS or parental obesity status as an IV fails to satisfy the exclusion assumption. It is clear that *GPBMI*, as an IV for childhood obesity, outperforms the two alternative IVs.

We also check the monotonic assumption. We divide the Add Health participants into ten groups with equal sample sizes according to their *GPBMI*, and then plot their BMI, BMI percentile, overweight, and obesity on the *GPBMI* for each group. As shown in Fig. A1, we observe almost strictly positive relationships between *GPBMI* and each of the four weight measures, indicating that *GPBMI* is almost monotonically associated with all measures of weight status.

To summarize, the proposed genetic predictor of BMI (*GPBMI*) passes various validity tests of IV and is likely to be a valid IV to identify the causal effects of childhood obesity on educational achievements.

Table 4
The Effect of Childhood Obesity and Childhood Overweight on Academic Performance in High School.

	Overall high school GPA	GPA by Subject				
		Math	Science	English	History and social science	Physical education
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Obese (BMI percentile ≥ 0.95)						
Obese	−0.922***	−0.890***	−0.968***	−1.109***	−1.097***	−0.723***
	(0.234)	(0.286)	(0.300)	(0.265)	(0.329)	(0.257)
Root MSE	0.693	0.835	0.855	0.814	0.845	0.749
No. of observations	2218	2218	2218	2218	2218	2218
Panel B: Overweight (BMI percentile ≥ 0.85)						
Overweight	−0.544***	−0.586***	−0.577***	−0.681***	−0.598***	−0.416***
	(0.141)	(0.183)	(0.181)	(0.158)	(0.209)	(0.159)
Root MSE	0.680	0.832	0.847	0.802	0.836	0.751
No. of observations	2585	2585	2585	2585	2585	2585
Control	Y	Y	Y	Y	Y	Y
School fixed effects	Y	Y	Y	Y	Y	Y

Notes: This table presents the IV estimation of the effects of childhood obesity and childhood overweight on academic performance in high school. The dependent variables are the accumulative GPAs from the respondent's high school official transcript obtained in Wave III, including the overall GPA and GPAs for different subjects (math, science, English, history and social science, and physical education). The instrumental variable is the genetic predictor of BMI constructed based on genetic information. The covariates are characteristics of individual, family and childhood neighborhood, ten ancestry-specific principal components of PGSs, and school fixed effects. Covariates are the same with those in Model (4) in Table 2. Standard errors in parentheses are clustered by childhood school in Wave I. * $p < .1$, ** $p < .05$, *** $p < .01$.

5.2. Academic performance in high school

Table 4 presents the IV estimates of the effects of childhood obesity (Panel A) and childhood overweight (Panel B) on academic performance in high school measured by official overall and subject GPAs. In Panel A, all estimates are negative and statistically significant at the 1% level. Being obese as an adolescent significantly decreases the overall high school GPA by 0.922 (33%) grade points, English GPA by 1.109 (42.6%), history and social science GPA by 1.097 (40.8%), science GPA by 0.968 (38.8%), math GPA by 0.890 (36.8%), and physical education GPA by 0.723 (21.7%). The magnitudes of the obesity effects on GPAs differ across the five subjects – lowest on physical education (21.7% less) while similar magnitudes for English, history, and social science, and science (38.8%−42.6% less). Similarly, being overweight as an adolescent has statistically significant and adverse effects on both overall and subject GPAs. The magnitudes of the overweight effects are smaller, ranging from 12 to 26% reduction in GPAs. The results suggest a non-linear effect of body weight on academic performance. As a comparison, Table A4 presents the OLS estimates for childhood obesity and overweight. Notably, the IV estimates are about 4–6 times larger than the OLS estimates, suggesting that the OLS estimations would underestimate the effects of childhood obesity and overweight on high school GPAs.

Previous studies investigating the effects of childhood obesity on academic performance have mixed results (Kaestner & Grossman, 2009; Rouse & Hunziker, 2020; Sabia, 2007; Von Hinke et al., 2016). Santana et al. (2017) and Martin et al. (2017) provide systemic reviews of related studies and conclude that the link between childhood obesity and academic achievement is still not well established. Our study contributes to the literature by providing evidence of the adverse effects of childhood obesity on the official high school GPA among American youth using the gene-based IV method.

5.3. Educational attainments in adulthood

According to the U.S. Social Security Administration, individuals with a bachelor's degree are estimated to have higher median lifetime earnings than high school graduates, by approximately $630,000 for females and $900,000 for males.[12] College enrollment or completion, however, is influenced by a multitude of factors beyond the high school GPA, such as family support and peer effects (Lessard & Juvonen, 2022; Roksa & Kinsley, 2019). Given the significant return of post-secondary education on earnings, we examine whether childhood obesity has long-term effects on educational attainment in this section.

We measure educational outcomes in Wave IV, where the ADD Health participants have completed their formal education.[13] The IV estimations are summarized in Table 5. We find that childhood obesity significantly decreases the likelihood of college enrollment by 0.366, college completion by 0.649, and years of schooling by 2.2 years (14.8%). Similarly, adolescent overweight significantly reduces the probability of college enrollment by 0.202, college completion by 0.403, and years of schooling by 1.3 years (8.9%). Comparing our results with Logit/OLS models presented in Table A4, our IV estimates of childhood obesity (overweight) are 4.46 (3.11) times greater than the Logit estimates for college enrollment, 7.82 (5.60) times greater than the logit estimates for college completion, and 5.80 (3.68) times greater than the OLS estimates for years of schooling. The comparisons indicate that association studies could underestimate the effects of childhood obesity and overweight on educational attainment due to the endogeneity and reverse causality issues.

Table 5
The Effect of Childhood Obesity and Childhood Overweight on Educational Attainment.

	College Enrollment (1)	College Completion (2)	Years of Schooling (3)
Panel A: Obese (BMI percentile ≥ 0.95)			
Obese	−0.366**	−0.649***	−2.192***
	(0.159)	(0.167)	(0.644)
Root MSE	0.397	0.458	1.877
No. of observations	2218	2218	2218
Panel B: Obese and Overweight (BMI percentile ≥ 0.85)			
Obese and Overweight	−0.202**	−0.403***	−1.319***
	(0.101)	(0.101)	(0.390)
Root MSE	0.396	0.443	1.816
No. of observations	2585	2585	2585
Controls	Y	Y	Y
School Fixed Effects	Y	Y	Y

Notes: This table presents the IV estimation of the effects of childhood obesity and childhood overweight on educational attainment in early adulthood. The dependent variables are a dummy variable indicating ever attending college or not, a dummy variable indicating completing college or not, and years of schooling. The IV and the control variables are the same as defined in Table 4. Standard errors in parentheses are clustered by school in Wave I. * $p < .1$, ** $p < .05$, *** $p < .01$.

The negative impacts of childhood obesity on educational attainment are consistent with the literature (Cohen et al., 2013; Sabia & Rees, 2015). Sabia and Rees (2015) find that childhood overweight decreases the probability of completing college by 0.25 for females and 0.28 for males using maternal obesity and sibling's BMI as an IV. As the return of an additional schooling year on wage is approximately 8–10% (Card, 1999), we estimate that childhood obesity would cause a 17–22% decrease in one's wage through lower education.

Finally, we show the link between the IV and educational outcomes by regressing educational outcomes on *GPBMI* and the full set of control variables specified in Eq. (4b). Table 6 presents the reduced form estimation results. We find that *GPBMI* significantly and negatively correlates with educational outcomes in all regressions. One standard deviation increases in *GPBMI* is associated with a decrease in GPA of 0.062 grade points and the probability of college enrollment by 0.024 and college completion by 0.043. The findings imply that an individual with a higher *GPBMI* is likely to underperform in academic performance and educational attainment, compared with their peers with a lower *GPBMI*.

6. Robustness checks and placebo tests

6.1. Parental BMI genes

Intergenerational transmission of BMI-related genetic markers could bias the IV results as they correlate with parental socioeconomic status.[14] While Add Health does not collect genetic information of parents, we use the obesity status of parents as a proxy for parental BMI-related genetics. We create an alternative IV by adding the obesity status of parents as a covariate in Equation (3a) to exclude the potential influences of the parental obesity status from the IV. In addition, we also

[12] https://www.ssa.gov/policy/docs/research-summaries/education-earnings.html. Last access on June 7th, 2022.

[13] The youngest age of the Add Health participants in Wave IV was 24 years old, and by then, most had already completed their college education if they chose to attend. The results are almost identical if we use educational outcomes reported in Wave V, but the sample size is smaller due to attrition.

[14] Willage (2018) includes the obesity status of parents as a covariate and explores the causal impact of obesity on mental health using BMI genetic risk score as an IV. However, the independence assumption of IV is not likely to be satisfied as the obesity status of parents is highly correlated with children's BMI PGS.

Table 6
Reduced-Form Regression Results for Education Outcomes.

	Wave III	Wave IV		
	High school GPA (1)	College Enrollment (2)	College Completion (3)	Education (Years) (4)
Panel A: Obese (BMI percentile ≥ 0.95)				
GPBMI	−0.062***	−0.024**	−0.043***	−0.147***
	(0.015)	(0.011)	(0.010)	(0.041)
Root MSE	0.680	0.403	0.442	1.867
No. of Observations	2218	2218	2218	2218
Panel B: Obese and Overweight (BMI percentile ≥ 0.85)				
GPBMI	−0.147***	−0.019*	−0.038***	−0.123***
	(0.041)	(0.010)	(0.009)	(0.037)
Root MSE	1.867	0.406	0.436	1.837
No. of Observations	2585	2585	2585	2585
Controls	Y	Y	Y	Y
School Fixed Effects	Y	Y	Y	Y

Notes: The table presents the reduced-form estimation results of the impact of childhood obesity and childhood overweight on education outcomes. The dependent variable of Column 1 is defined in Table 4. The dependent variables of Columns 2–4 are defined in Table 5. The covariates are the same as defined in Table 4. Standard errors in parentheses are clustered by school. * $p < .1$, ** $p < .05$, *** $p < .01$.

incorporate the obesity status of parents as an additional covariate in the IV models formulated in Equations (4a) and (4b). As shown in Column 1 of Tables A5 and A6, the IV estimates of obesity and overweight effects on academic performance and educational attainment are almost identical to the main results presented in Tables 4 and 5. This robustness check alleviates the concerns that the IV estimates might be biased by the obesity-related genes transmitted from parents to their children.

6.2. Using the alternative BMI PGS based 97 SNPs

The BMI PGS based on 97 key SNPs identified by Locke et al. (2015) has been used in recent economic studies (Böckerman et al., 2019; Brunello et al., 2020; Edwards et al., 2021). As a robustness check, we construct an alternative *GPBMI* based on the 97-SNP BMI PGS from the GWAS in Locke et al. (2015). As shown in Column 2 of Tables A5 and A6, the re-estimated results also indicate that childhood obesity and overweight significantly and negatively impacts high school GPA, college enrollment, college completion, and years of schooling. Our main results are quite robust. Compared with the main results, the first-stage F statistics of re-estimations and the magnitude of second-stage coefficients are much smaller. This phenomenon could be resulted from the lower explanatory power of 97-SNP BMI PGS on weight status than the 947-SNP BMI PGS. Therefore, we rely on the estimations using the latest version of BMI PGS (Yengo et al., 2018) for the main results.

6.3. Controlling for learning ability using various PGSs

In the main analysis, we use the intelligence PGS as a proxy for children's intrinsic learning ability, a potentially critical omitted variable in the returns to education literature. Here, we experiment with alternative proxies for learning ability: the cognitive function PGS and educational attainment PGS. In this robustness check, we first replace the intelligence PGS with either the cognitive function PGS or educational attainment PGS in Eq. (4a) and (4b). Subsequently, we control for all three education-related PGSs in the model. As shown in Columns 3–5 in Tables A5 and A6, the IV estimates using alternative proxies for the learning ability are similar to our main results, suggesting that the original specification works well.

6.4. Using BMI and BMI percentil as alternative weight measures

We re-estimate the results using BMI and BMI percentile in one's adolescence as alternative weight measures for a robustness check. As shown in Table A7, a higher BMI (Panel A) or BMI percentile (Panel B) causes statistically significant, adverse effects on educational outcomes that are qualitatively robust compared with the main results. However, we note that the associations between education outcomes and BMI percentile or BMI are not necessarily linear, as we discussed in Section 5.2. A linear model may incorrectly assume a linear relationship betwen BMI and academic performance. However, in reality, the relationship exhibits diminishing returns, where the negative impact of higher BMI on academic performance becomes more significant as the BMI increases beyond certain thresholds (85 and 95 percentiles). The highly non-linear associations between BMI and educational outcomes suggest that a dichotomized measure of weight status, such as obesity or overweight, would be more appropriate to examine the impact of childhood weight status on educational outcomes.

6.5. Controlling for PGSs for substance use and mental health

We control for intelligence PGS that is a proxy for cognitive capacity and learning ability in Eqs. (4a) and (4b), but other PGSs for mental health and substance use are not incorporated due to a concern that education outcomes may have bi-directional relationships with mental health and substance use. Yet, as a robustness check, we control for the three PGSs for mental health and the two PGSs for substance uses in Eqs. (4a) and (4b). First, we check for the association between these additional PGSs and childhood weight status. As shown in Table A8, none of these additional PGSs is statistically correlated with *GPBMI*. Second, we re-estimate the IV models controlling for the PGSs for mental health and substance as covariates in Eqs. (4a) an (4b). The new IV results presented in Table A9 are similar as the main results in both statistical significance and magnitude. These results suggest that the novel IV is valid for the independence assumption and the main results are quite robust.

6.6. Placebo tests

To further assess the validity of the IV estimates, we conduct a placebo test by randomly assigning *GPBMI* to the Add Health participants 1000 times. We then re-estimate the effects of childhood obesity and overweight using the randomized *GPBMI*. For each outcome variable, we create 1000 IV estimates. If childhood obesity is primarily driven by *GPBMI* rather than other confounding factors, the IV estimates in the placebo tests should converge towards zero and remain statistically insignificant. Fig. 1 shows the IV estimates for both childhood obesity (Panel A) and childhood overweight (Panel B), which are centered at zero. Additionally, all T-statistics in the placebo tests are substantially lower than two and the corresponding real estimates in the main results (depicted as red dots), except for one outlier (college enrollment in Panel B). The distributions of P-values of 8000 IV estimates associated with the four outcome variables are presented in Fig. A2. Only four estimates with P-values less than 0.05 are identified among these 8000 IV estimates. The results of the placebo tests offer compelling evidence that childhood obesity has a causal impact on academic performance and educational attainment.

7. Mechanisms

In this section, we explore potential mechanisms through which childhood obesity may affect human capital formation relating to cognitive ability, health, and noncognitive ability. The evidence we provide here is suggestive. First, we show that there is no difference in cognitive ability between obese and normal-weight children. Second, we examine how childhood obesity relates to health conditions and psychological well-being. Third, we examine school behaviors that are

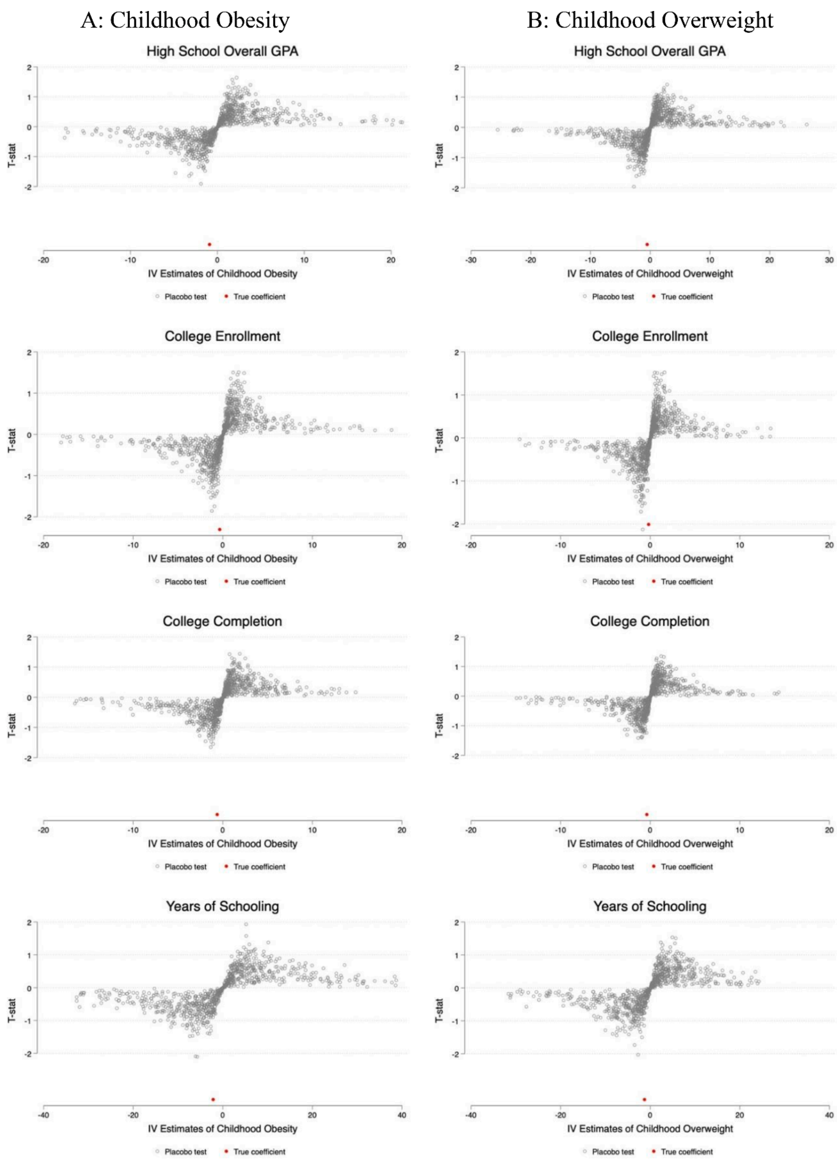

Fig. 1. Distributions of Placebo IV Estimates for Childhood Obesity and Childhood Overweight
Notes: For each of the four main education outcomes (high school GPA, college enrollment, college completion, and years of schooling), we repeat the placebo test 1000 times and plot the distribution of placebo IV estimates of childhood obesity (Panel A) and childhood overweight (Panel B). Fig. A2 present the corresponding p-values.

related to school performance. Finally, we examine the effects of childhood obesity on children's college aspiration/expectation and family dynamics.

7.1. *Cognitive abilities*

We investigate the most direct mechanism: whether childhood obesity affects cognitive ability and hence educational outcomes. Cognitive ability plays a critical role in educational choices (Hanushek & Woessmann, 2020), educational attainment (Heckman, 2007), and labor market outcomes (Heckman, 2007; Heckman et al., 2006). However, the direct link between childhood obesity and cognitive ability remains ambiguous. Liang et al. (2014) review 67 studies focusing on associations between obesity and obesity-related behaviors with neurocognitive functioning in youth. They find mixed evidence regarding the relationship between obesity and general cognitive functioning, learning and memory, language, and academic performance.

Commonly used tests for cognitive ability are the Peabody Picture Vocabulary Test (PPVT) and memory recall tests. The PPVT score measures verbal IQ and cognitive skills (Fletcher, 2013a; Fletcher, 2013b). The participants' PPVT percentile rank ranging from 1 to 100 is reported in Waves I and III. This measure has a sample mean of 61.2 and a standard deviation of 25.6 in Wave I and a sample mean of 60.3 and a standard deviation of 25.8 in Wave III, suggesting that cognitive ability stays stable from adolescence to early adulthood. Memory is a key component of cognitive functioning, reflecting one's skills for acquiring, encoding, storing, and retrieving information (Liang et al., 2014). Memory recall tests, namely, Delayed Word Recall and Counting Backward, are conducted in Wave IV. Based on the available information in Add Health, we construct three variables to measure memory-related cognitive abilities: 1) the number of words the participant correctly recalls within 60 s after having heard a list of 15 words; 2) the number of words the participant correctly recalls within 90 s after having heard a list of 15 words; and 3) an indicator whether the respondent can correctly count backward the six numbers. The Delayed Word Recall has a sample mean of 5.6 (SD = 2.0) for the 60-second assessment and 7.0 (SD = 1.9) for the 90-second assessment. The backward memory recall test has an average passing ratio of 44.5%. In addition to cognitive ability tests, we also examine the participant's self-rated intelligence level. The Add health participants were asked how intelligent they were compared with other people of a similar age and whether they had trouble paying attention daily at school in Wave I. More than half (60.3%) of the respondents reported that they were above the average intelligent level, and 12.5% reported having attention problems at school.

Table 7
The Effect of Childhood Obesity and Childhood Overweight on Cognitive Abilities.

	PPVT percentile		Working Memory Capacity in Wave IV			Self-rated cognition in Wave I	
	Wave I (1)	Wave III (2)	Words recall within 90s (3)	Words recall within 60s (4)	Backward digit recall (5)	Self-rated intelligence (6)	Having attention problems in school (7)
Panel A: Obese (BMI percentile ≥ 0.95)							
Obese	8.327	−1.952	−0.000	−0.302	0.053	0.009	−0.017
	(8.629)	(7.071)	(0.581)	(0.660)	(0.153)	(0.161)	(0.109)
Root MSE	22.42	22.55	1.816	1.841	0.466	0.454	0.311
No. of observations	2104	2104	2214	2214	2214	2214	2218
Panel B: Overweight (BMI percentile ≥ 0.85)							
Overweight	5.407	1.804	0.210	−0.162	0.096	0.037	−0.016
	(5.871)	(4.957)	(0.398)	(0.433)	(0.099)	(0.100)	(0.076)
Root MSE	22.43	22.69	1.828	1.853	0.468	0.459	0.318
No. of observations	2448	2448	2581	2581	2581	2581	2585
Controls	Y	Y	Y	Y	Y	Y	Y
School Fixed Effects	Y	Y	Y	Y	Y	Y	Y

Notes: This table presents the IV estimation of the effects of childhood obesity and overweight on cognitive abilities. The dependent variables are the percentile ranking of the Peabody Picture Vocabulary Test (PPVT) score reported in Waves I and III (Columns 1–2), working memory capacity test performance in Wave IV (Columns 3–5), and self-rated cognition in Wave I (Columns 6–7). The IV and covariates are the same as defined in Table 4. Standard errors in parentheses are clustered by school in Wave I. * $p < .1$, ** $p < .05$, *** $p < .01$.

Table 8
The Effect of Childhood Obesity and Childhood Overweight on Self-reported Health Condition, Health-Compromising Behaviors, and Psychological Well-being.

	Self-rated very good/excellent health condition		Substance use in Wave I			Mental health conditions in Wave I	
	Wave I (1)	Wave III (2)	Weekly drink alcohol (3)	Ever daily smoking (4)	Feel moody daily (5)	Standardized CES-D score (6)	Suicide ideation (7)
Panel A: Obese (BMI percentile ≥ 0.95)							
Obese	−0.594***	−0.405***	0.221**	0.218*	0.218**	0.586*	0.093
	(0.156)	(0.145)	(0.105)	(0.121)	(0.101)	(0.343)	(0.100)
Root MSE	0.420	0.401	0.263	0.380	0.298	0.889	0.327
No. of observations	2218	2218	2210	2211	2216	2215	2208
Panel B: Overweight (BMI percentile ≥ 0.85)							
Overweight	−0.397***	−0.305***	0.173***	0.176**	0.189***	0.311	0.091
	(0.115)	(0.096)	(0.062)	(0.076)	(0.066)	(0.199)	(0.061)
Root MSE	0.427	0.412	0.269	0.383	0.308	0.885	0.332
No. of observations	2585	2585	2577	2577	2583	2582	2575
Controls	Y	Y	Y	Y	Y	Y	Y
School Fixed Effects	Y	Y	Y	Y	Y	Y	Y

Notes: This table presents the IV estimation of the effects of childhood obesity and overweight on health, including self-rated health condition, health-compromising behaviors, and psychological wellbeing. The outcome variables are dummies for self-rated good health condition reported in Waves I and III (very good/excellent = 1; zero otherwise) (Columns 1–2) and dummies indicating whether the respondent drink alcohol weekly and ever smoke daily in Wave I (Columns 3–4). The outcome variables for psychological wellbeing are whether feeling moody daily, the general depressive (CES-D) score, and whether having suicide ideation (Columns 5–7). The IV and covariates are the same as defined in Table 4. Standard errors in parentheses are clustered by school in Wave I. * $p < .1$, ** $p < .05$, *** $p < .01$.

We estimate the IV models using cognitive test scores (PPVT rankings and memory test scores) and self-rated intelligence as dependent variables and present the results in Table 7. None of the fourteen IV estimates is statistically significant at the 10% level, suggesting no statistically significant effect of childhood obesity or overweight on cognitive ability measured by either cognitive test scores or self-rated intelligence. Palermo and Dowd (2012) find similar results using Child Development Supplement (CDS) data from the Panel Study of Income Dynamics (PSID). Our results show that cognition is not the mechanism through which childhood obesity affects academic achievements, which motivates us to explore noncognitive channels.

7.2. Health outcomes, health behaviors, and psychological well-being

Child health affects academic performance and educational attainment (Suhrcke & de Paz Nieves, 2011). Childhood obesity has been found to adversely affect both physical and mental health (Crosnoe, 2007; Litsfeldt et al., 2020; Martin et al., 2017; Sahoo et al., 2015). This section focuses on three health-related channels through which childhood obesity could affect education, including general self-reported health status, health-compromising behaviors, and psychological well-being.

The Add Health participants were asked to rank their health condition as excellent, very good, good, fair, or poor in Waves I and III. We create an indicator for good health condition if an individual chooses "excellent" or "very good" as an answer. Approximately 71.2% and 75.6% of the sample indicate very good to excellent health conditions in Waves I and III, respectively. We estimate the IV model with the good health indicator as the dependent variable and present the results in Columns 1–2 of Table 8. We find that childhood obesity decreases the probability of having very good or excellent health by 0.594 in Wave I and 0.405 in Wave III (approximately six years later). The results indicate that childhood obesity could affect education outcomes via its effect on physical health, consistent with the literature (Suhrcke & de Paz Nieves, 2011).

Second, we consider health behaviors. Lack of self-control in childhood is associated with substance use and dependence (Fergusson et al., 2013; Moffitt et al., 2011). Obese children are found to have lower self-control than their normal weight peers (Fan & Jin, 2014b). Health-compromising behaviors like drinking and smoking can be a potential channel for childhood obesity to affect educational achievements. Based on the questions on substance use in Wave I, we create a dummy variable for alcohol drinking – it equals one if the participants reported drinking at least once a week in the last 12 months; and zero otherwise. We also create a smoking dummy by assigning a value of one if the participants reported ever smoking at least one cigarette every day for a month; and zero otherwise. We estimate the IV model with smoking and drinking behaviors as dependent variables and present the results in Columns 3–4 of Table 8. We find that childhood obesity increases the probability of being a weekly alcohol drinker by 0.221 and a daily smoker by 0.218. The IV estimates for childhood overweight are also statistically significant but smaller in magnitude. Our results are consistent with the literature, where frequent smoking and binge drinking are found to be negatively associated with poor academic performance in high school (Balsa et al., 2011; Cox et al., 2007) and educational attainment (Staff et al., 2008; Yan & Brocksen, 2013). Therefore, frequent substance use of alcohol and cigarettes is likely to be a facilitating factor for childhood obesity to adversely affect educational achievements.

Third, we examine psychological well-being. We create several variables measuring psychological well-being. The first one is a dummy variable indicating whether an individual felt moody daily. The second one is the Center for Epidemiologic Studies Depression Scale (CES-D) score developed by Radloff (1977). CES-D is widely used to evaluate depressive symptom severity, with a higher score indicating increased severity. The participants in Wave I were asked to answer 19 CES-D questions. Following the suggestion in the literature (Willage, 2018), we standardize the sum of the CES-D score with a mean of zero and a standard deviation of one. The third variable of psychological well-being is a dummy variable indicating whether an individual ever had suicide ideation. We estimate the IV model with these measures of psychological well-being as the dependent variables and present the results in Columns 5–7 of Table 8. We find that childhood obesity has a significant effect on the depressive CES-D score and feeling moody daily. Childhood obesity increases the probability of feeling moody daily by 0.218 and the depressive CES-D score by 0.586. Both depressive CES-D score and feeling moody are often recognized as representative symptoms of internalizing and externalizing behavior problem index, which are also measures of noncognitive ability in literature (Palermo & Dowd, 2012). Therefore, our results on psychological well-being suggest that childhood obesity has adverse effects on one's mental health and noncognitive ability.

To summarize, we find that childhood obesity potentially affects educational achievements through three health-related channels, general physical health condition, healthy-compromising behaviors, and psychological well-being.

7.3. School absenteeism and aggressive behaviors

Academic performance for U.S. youth is found to be negatively associated with school absenteeism (Gershenson et al., 2017; Ruff et al., 2019) and aggressive behaviors (Glew et al., 2005; Lepore & Kliewer, 2013). In this section, we examine the effects of childhood obesity on school absenteeism and aggressive behaviors through which childhood obesity may affect academic performance.

First, we create two dummy variables for having an excused school absence of at least three days and an unexcused school absence of at least one day during the school year.[15] We estimate the IV models with these two dummy variables for school absenteeism as the dependent variables and present the IV estimation results in Table 9. We find that childhood obesity and overweight significantly increase the probability of having excused and unexcused school absences. The results may be explained by psychological stress caused by obesity and obesity-related illness (An et al., 2017; Fuxa & Fulkerson, 2011). Our results are consistent with the findings in the literature. For example, An et al. (2017) find that the odds of being absent from school are 54% higher for obese children and 27% higher for overweight children than their normal-weight peers.

Add Health provides information on physical fights in Wave I. About 27.3% report ever getting into a physical fight during the past 12

Table 9
The Effect of Childhood Obesity and Childhood Overweight on School Absenteeism, Aggressive Behaviors, and Feeling of being Safe in school in Wave I.

	School absence		Aggression: Physical fight		Feeling unsafe in school
	with excuses	without excuses	at least one time in last 12 months	with friends or known ones last time	
	(1)	(2)	(3)	(4)	(5)
Panel A: Obese (BMI percentile ≥ 0.95)					
Obese	0.446***	0.254**	0.444***	0.444*	0.245**
	(0.156)	(0.126)	(0.153)	(0.230)	(0.111)
Root MSE	0.464	0.389	0.419	0.423	0.302
No. of observations	2217	2216	2211	1637	2218
Panel B: Obese and Overweight (BMI percentile ≥ 0.85)					
Overweight	0.266**	0.201**	0.270***	0.205*	0.174**
	(0.113)	(0.091)	(0.095)	(0.120)	(0.076)
Root MSE	0.463	0.396	0.419	0.417	0.306
No. of observations	2584	2583	2578	1887	2585
Controls	Y	Y	Y	Y	Y
School Fixed Effects	Y	Y	Y	Y	Y

Notes: This table presents the IV estimation of the effects of childhood obesity and childhood overweight on school absenteeism, aggressive behaviors, and feeling of being safe in school reported in Wave I. The outcome variables are all dummies, indicating whether the respondent has at least 3-day excused school absence and ever skips school at least 1 day during school year (Columns 1–2), whether the respondent ever gets into physical fight in the last 12 months (Column 3), ever fights with friends or known ones in the last physical fight (Column 4), and feels unsafe in school (Column 5), respectively. The IV and covariates are the same as defined in Table 4. Standard errors in parentheses are clustered by school in Wave I. * $p < .1$, ** $p < .05$, *** $p < .01$.

[15] In Wave I, Add Health participants were asked the following question: "how many times were you absent from school for a full day with an excuse during the school year (never, 1-2 times, 3-10 times, and 10+ times)?" About 64.0% of the participants reported having at least three times of excused school absences. The participants were also asked to report the number of times they had skipped school for a full day without an excuse during the school year. About 25.0% of the participants reported skipping school at least once during the school year.

Table 10
The Effect of Childhood Obesity and Childhood Overweight on College Aspiration, College Expectation, and Family Dynamics in Wave I.

	Personal perspective		Family dynamics		
	College aspiration (1)	College expectation (2)	Parental college expectation (3)	Perceived family attentions (4)	Feeling understood at home (5)
Panel A: Obese (BMI percentile ≥ 0.95)					
Obese	−0.315**	−0.274***	−0.478***	−0.245*	−0.409**
	(0.128)	(0.103)	(0.171)	(0.132)	(0.182)
Root MSE	0.422	0.356	0.469	0.420	0.486
No. of observations	2216	2215	2216	2213	2216
Panel B: Overweight (BMI percentile ≥ 0.85)					
Overweight	−0.157*	−0.148**	−0.307***	−0.173*	−0.244**
	(0.088)	(0.066)	(0.114)	(0.094)	(0.120)
Root MSE	0.420	0.361	0.468	0.425	0.483
No. of observations	2583	2582	2583	2580	2583
Controls	Y	Y	Y	Y	Y
School Fixed Effects	Y	Y	Y	Y	Y

Notes: This table presents the IV estimation of the effects of childhood obesity and childhood overweight on college aspiration, college expectation, and family dynamics reported in Wave I. The dependent variables are children's college aspiration (Column 1), children's college expectation (Column 2), and parents' expectation of their children to complete college (Column 3) in Wave I. Two indicators for feelings at home measure: whether the participants agree/strongly agree the statements "you receive family attention" (Column 4) and "you feel understood at home" (Column 5), respectively. The IV and the covariates are the same as defined in Table 4. Standard errors in parentheses are clustered by school in Wave I. * $p < .1$, ** $p < .05$, *** $p < .01$.

months, and 35.1% ever fight with friends or known ones in the last physical fight,[16] and 11.3% feel unsafe at school.[17] The IV estimates using whether getting into a physical fight, ever fighting with friends or known ones in the last physical fight, and feeling unsafe in school as the dependent variables are shown in Columns 3–5 of Table 9. We find that obese and overweight children are more likely to be involved in a physical fight, fight with friends or known ones, and feel unsafe at school. Our results are consistent with the findings in the literature as it documents a positive association between obesity/overweight and aggressive behaviors in youth (Farhat et al., 2010; Hasler et al., 2004). Another possible explanation is that obese children might be more likely to be targets of bullying or teasing, leading them into physical fights (Lumeng et al., 2010). Fuxa and Fulkerson (2011) report that overweight and obese adolescents are more likely to have negative perceptions about the safety and social environment at school. Overall, our results suggest that childhood obesity would increase school absenteeism and aggressive behaviors, potentially leading to poor academic outcomes.

7.4. College aspiration/expectation and family dynamics

College aspiration or expectation is a critical determinant of college attendance and completion. College aspiration reflects students' internal motivation and perceived benefits of attending college, peer effects, school environment, and social norms (Cohen et al., 2013; Dyce et al., 2012). College expectation is a relatively realistic aspiration, primarily shaped by structural forces such as socioeconomic status, social background, family support, school academic performance, and social norms (Cohen et al., 2013). We investigate whether childhood obesity affects college aspiration and expectation in this section.

The Add Health participants were asked in Wave I: "how much do you want to go to college on a scale from 1 (low) to 5 (high)?" Almost two-thirds of the participants chose a scale of 5. We construct a dummy variable for high college aspiration by assigning one if the respondent chose 5; and zero otherwise. The participants were also asked: "how likely will you go to college on a scale from 1 (least likely) to 5 (most likely)?" Approximately 80.4% of the participants chose a scale of 4 or 5. We construct a dummy variable for high college expectation by assigning one if the respondent chose 4 or 5; and zero otherwise. Columns 1–2 of Table 10 show the IV estimates of childhood obesity when adolescents' own college aspirations and expectation dummies are dependent variables. We find that childhood obesity reduces the probability of high college aspiration by 0.315 (p<.05) and the probability of high college expectation by 0.274 (p<.01). Being overweight as an adolescent also reduces the probability of high college aspiration and expectation, while the magnitudes of the effects are smaller than childhood obesity.

Family environment is important for children to develop human capacities and could affect children's educational outcomes (Forehand et al., 1986; Martin et al., 2015; Rivers et al., 2012). Parental expectation has been found to be the most vital factor for a child's college plan, rather than family background and characteristics and the child's high school GPA and activities (Hossler & Stage, 1992). Parental expectations of their children's college attendance and completion, along with their own educational expectations, significantly affect the child's academic performance and competence (Benner & Mistry, 2007; Cohen et al., 2013).

We investigate the impact of childhood obesity on parental college expectations of their children and children's feelings at home documented in Wave I. The Add Health parent participants were asked: "how disappointed would you be if your child did not graduate from college?" Approximately 38.4% of the participants chose "very disappointed." The Add Health participating children were asked how much they felt on a scale of 1 (not at all) to 5 (very much) for "your family pays attention to you" and "people in your family understand you." We create three dummy variables to indicate: 1) high expectation of parents for their children to complete college, which equals to one if choosing "very disappointed," and zero otherwise; 2) family attention perceived by children, which equals to one if choosing 4 or 5, and zero otherwise; and 3) children feeling understood at home, which equals to one if choosing 4 or 5, and zero otherwise. We estimate the IV model using parental college expectations and adolescents' feelings at home as dependent variables and present the results in Columns 3–5 of Table 10. Childhood obesity is estimated to reduce the probability of parents' college expectations of their children by 0.478. Childhood obesity also reduces the

[16] People who fight with strangers or family members in the last time of physical fight are excluded. Thus, this dummy variable equals one if people fight with friends or unknown ones in the last physical fight, and zero if they never fight.

[17] Participants were asked: "how much do you agree or disagree: you feel safe in your school (strongly agree, agree, neither agree nor disagree, disagree, and strongly disagree)." The dummy for unsafe feeling in school equals one if participants chose to disagree or strongly disagree, and zero otherwise.

probability of receiving sufficient attention and feeling understood at home by 0.245 and 0.409. The estimates for childhood overweight are smaller in magnitude. The results suggest that childhood obesity may send a negative signal to parents about children's abilities and aspirations and reduce family care and attention children receive. Family dynamics could be a potential pathway for childhood obesity to affect academic performance and educational attainment.

Overall, we find that childhood obesity has no statistically significant effect on cognitive ability. It is found to exacerbate general self-reported physical health condition, health-compromising behaviors (e.g., substance use), and psychological well-being, indicating that obesity affects educational achievements through these health-related channels. Being obese as an adolescent is also statistically linked with increased probabilities of school absenteeism and aggressive behaviors (e.g., physical fights). Furthermore, childhood obesity is found to negatively affect one's noncognitive abilities, indicated by typical internalizing and externalizing behavior problems (i.e., depressed scale and feeling moody), feeling unsafe at school, and having negative perceptions of family dynamics. Childhood obesity and overweight significantly decrease college aspirations and expectations of adolescents and their parents. Overall, our findings suggest interventions for childhood obesity would help mitigate the adverse effect on educational outcomes by improving health and noncognitive ability rather than cognitive ability.

8. Discussion and conclusion

This study examines the causal effects of childhood obesity on academic performance in high school and educational attainment in adulthood, employing a gene-based IV approach. The results show that childhood obesity has negative consequences on educational achievements in both short-term and long term. This study also examines the mechanisms through which obesity could adversely affect educational achievements. The results indicate that childhood obesity has an impact on education outcomes not through its effect on cognitive abilities, but rather through its influence on factors such as general physical health condition, health-compromising behaviors, and psychological well-being, as well as school absenteeism and aggression, college aspiration/expectation, and family dynamics. The mechanism analyses imply that improving health and non-cognitive ability rather than cognitive ability can help mitigate adverse effects of childhood obesity on educational outcomes. Overall, our findings suggest that public policies or interventions targeting childhood obesity have the potential to affect human capital formation and ultimately impact economic well-being and income inequality through its effects on education.

This study has several limitations. First, the absence of genetic information about parents in Add Health makes it impossible to conduct within-family analysis and avoid dynastic effects. Sibling analysis is also not plausible due to the limited number of sibling participants with genetic information ($N = 451$). Second, the majority of Add Health participants are white (>73%), leading to the insufficient observation of ethnic minorities with genetic information and high school transcript ($N = 940$). In addition, the PGSs for BMI, education, substance use, and mental health used in this study are mainly calculated based on the European GWAS ancestry group. Therefore, those PGSs have quite lower prediction accuracy for corresponding phenotypes among ethnic minorities than individuals of European ancestry (Braudt & Harris, 2020). Third, although the instrumental variable constructed in this study controls for various confounding effects of intelligence, substance use, and mental health PGSs, the undiscovered pleiotropy problems of many SNPs might still affect the results. Future research is encouraged to re-examine this topic with new pleiotropic factors identified and better data available. Finally, this study explores potential mechanisms through which childhood obesity would affect educational outcomes. Potentially, there are three possibilities for the potential pathways: a) childhood obesity affects education outcomes through its effect on mechanism outcomes; b) childhood obesity affects academic outcomes, and the latter affect mechanism outcomes; and c) childhood obesity affects both education and mechanism outcomes. Add Health collected mechanism outcomes in Wave I (1994–1995), high school transcripts in Wave III (2001–2002), and education outcomes in Wave IV (2008–2009). The order of the timing is less supportive for the second pathway. In order to exclude the possibility of the third pathway, more research is needed to reveal the causal relationship between education outcomes and mechanism outcomes on health and non-cognitive skills.

CRediT authorship contribution statement

Maoyong Fan: Conceptualization, Methodology, Validation, Formal analysis, Writing – review & editing. **Yanhong Jin:** Conceptualization, Methodology, Validation, Formal analysis, Writing – review & editing. **Man Zhang:** Conceptualization, Software, Formal analysis, Writing – original draft, Writing – review & editing.

Declarations of Competing Interest

None.

Data availability

The data used in this study can be requested on the official website of Add Health by signing restricted-use contract. We will provide the STATA code upon request.

Acknowledgments

The authors acknowledge that the Add Health Restricted-Use Contract (#24021901) established between Rutgers University and Add Health at the University of North Caroline at Chapel Hill made this research possible. Add Health is directed by Robert A. Hummer and funded by the National Institute on Aging cooperative agreements U01 AG071448 (Hummer) and U01 AG071450 (Aiello and Hummer) at the University of North Carolina at Chapel Hill. Waves I-V data are from the Add Health Program Project, grant P01 HD31921 (Harris) from Eunice Kenney Shriver National Institute of Child Health and Human Development (NICHD), with cooperative funding from 23 other federal agencies and foundations. Add Health was designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill.

Man Zhang acknowledges the support of National Natural Science Foundation of China (Grant No. 72203223). The authors also thank seminar participants in the Department of Economics at Ball State University, Rutgers University, and Renmin University of China for their feedback. Any errors that inevitably remain are entirely our responsibility.

Appendix

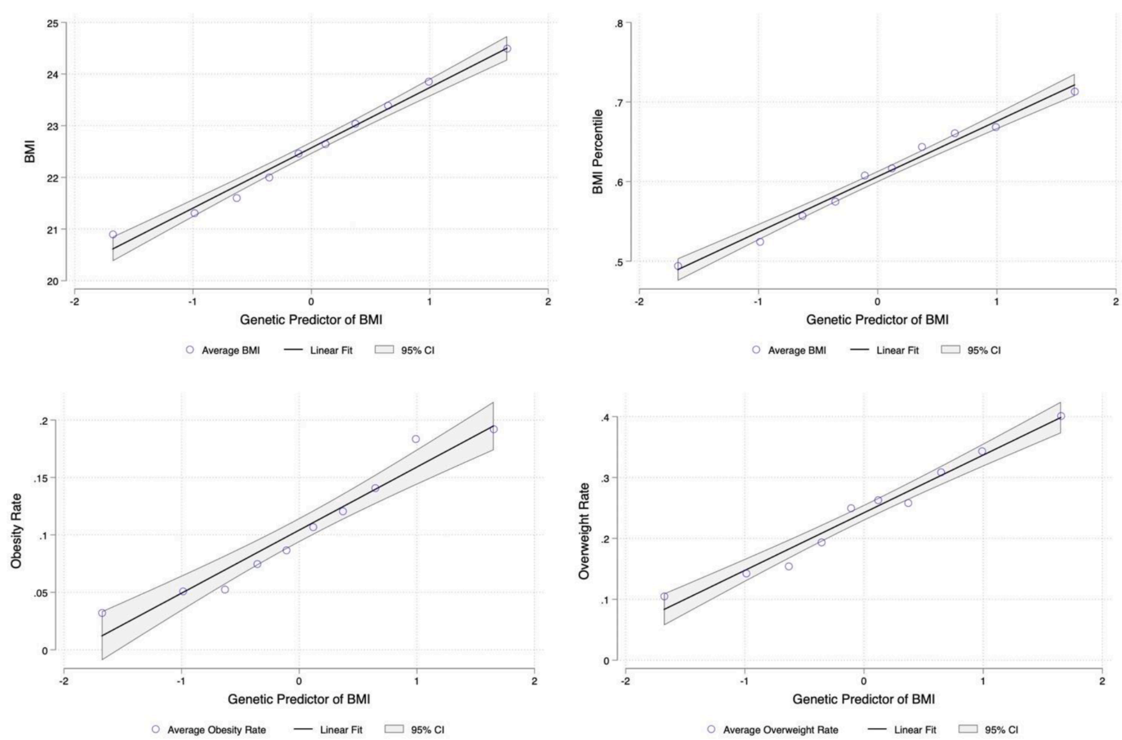

Fig. A1. Relationship between the Genetic Predictor of BMI and Weight Measures in Wave I for Add Health Participants with Available Genetic Information in Wave IV and Weight Information in Wave I

Notes: The Add Health participants are binned into ten groups of equal size based on their *GPBMI* scores. The x-axis is the mean value of *GPBMI* scores from the ten groups. The y-axis is average weight information (BMI, BMI percentile, overweight rate, and obese rate) in Wave I for the ten groups.

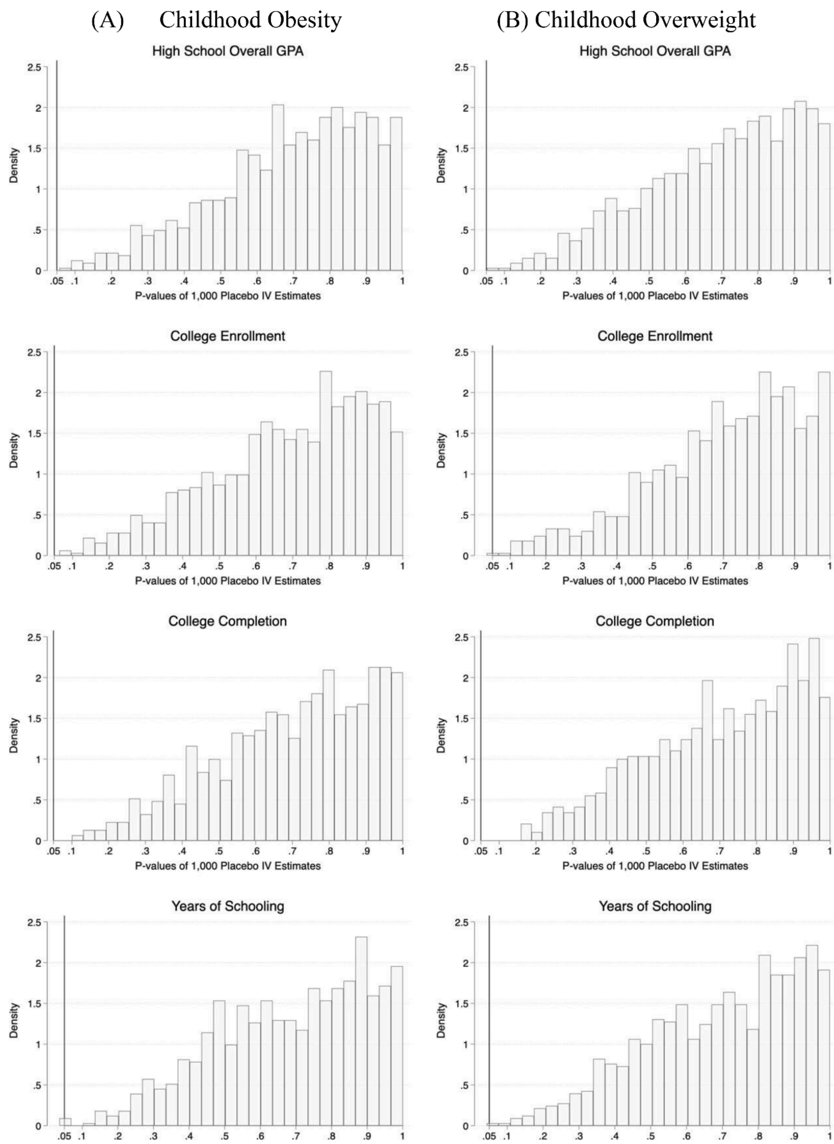

Fig. A2. Distributions of P-values of Placebo IV Estimates of Childhood Obesity and Childhood Overweight on Educational Outcomes
Notes: For each of four educational outcome variables (high school GPA, college enrollment, college completion, and years of schooling), we create 1000 IV estimates using the randomized genetic predictor of BMI as the IV for childhood obesity (Panel A) and childhood overweight (Panel B).

Table A1
Definition of scientific terms.

Variables	Definitions
deoxyribonucleic acid (DNA)	DNA is a molecule containing genetic instruction for particular enzymes or functional molecules in organisms. DNA is composed of two polynucleotide chains with double helix structure.
Nucleotide	Nucleotide is composed of three subunits: nucleobase (nucleotide base), sugar molecule, and phosphate group. DNA has four types of nucleotide bases: has four types of nucleotide bases: namely adenine (A), cytosine (C), guanine (G), and thymine (T).
Single-nucleotide polymorphism (SNP)	Genomic variation at a single nucleotide base that occurs at a specific position in the genome
Phenotype	Phenotype refers to an individual's observable trait, such as height, BMI, educational attainment in years, etc. An individual's phenotype depends on his/her genetic endowment and environmental factors.
Genotype	Genotype is a scoring or the type of genetic variants at a specific location in the genome
Gene-wide association study (GWAS)	GWAS is a research approach exploring associations between SNPs and risks of a disease or phenotype.
Allele	An allele refers to the variant form of SNPs.
Polygenic scores (PGS)	PGS is calculated as a weighted sum of associations between allele frequencies at individual SNPs and the specific phenotype (e.g., BMI) according to summary statistics from GWASs

Table A2
Definition of dependent variables .

Variables	Wave	Definition
Key dependent variables		
High school GPA	III	The overall high school GPA and GPAs of different subjects were provided from the official high school transcripts with the scale of 0–4.
Education enrollment	IV	Participants reported that their highest level of education achieved to date are "some college" or above.
Education completion	IV	Participants reported that their highest level of education achieved to date are "completed college (bachelor's degree)" or above.
Years of schooling (years)	IV	Years of schooling was calculated from the self-reported highest level of education.
Mechanisms-cognitive ability		
PPVT score ranking	I/III	Participant's Peabody Picture Vocabulary Test (PPVT) score as the percentile rank ranging from 1 to 100.
Words recall within 90 s	IV	the number of words the participant correctly recalled within 90 s after having heard a list of 15 words.
Words recall within 60 s	IV	the number of words the participant correctly recalled within 60 s after having heard a list of 15 words.
Backwards digit recall	IV	An indicator whether the respondent could correctly count backward the six numbers or more
Self-rated intelligence (above average=1, Otherwise=0)	I	An indicator whether respondents rate their intelligence is above the average compared with other people of a similar age.
Having troubles of paying attention in school daily	I	An indicator whether respondents reported to have troubles paying attention (almost) every day or not.
Mechanisms-Health condition, risky behaviors and psychological wellbeing		
Self-rated very good or excellent health condition	I/III	An indicator whether participants rated their health condition "excellent" or "very good".
Drinking alcohol weekly	I	An indicator whether participants reported drinking at least once a week in the last 12 months.
Ever smoking daily	I	An indicator whether participants reported ever smoking at least one cigarette every day for a month.
Feeling moody daily	I	An indicator whether participants felt moody (almost) every day.
Standardized CES-D score	I	The standardized sum of answers of 19 Epidemiologic Studies Depression Scale (CES-D) questions.
Suicide ideation	I	An indicator whether participants ever had suicide ideation.
Mechanisms-Academic performance related behaviors and feeling in school		
Having excused school absence at least 3 times during school year	I	An indicator whether participants reported excused school absence at least three days during the school year.
Ever skipping school without excuses during school year	I	An indicator whether participants reported unexcused school absence at least one day during the school year.
Ever getting into physical fight in the past 12 months	I	An indicator whether participants ever got into a physical fight in the past 12 months.
Ever fighting with friends or known ones in the last physical fight	I	An indicator whether participants ever fought with friends or known ones (excluding family members) in the last physical fight.
Feeling unsafe in school	I	An indicator whether participants disagreed or very disagreed with "you feel safe in your school".
Mechanisms-Feelings at home and college inspiration and expectation		
Personal college aspiration	I	An indicator of participant's choice of 5 for the question "how much do you want to go to college on a scale from 1 (low) to 5 (high)?".
Personal college expectation	I	An indicator of participant's choice of 4/5 for the question "how likely will you go to college on a scale from 1 (least likely) to 5 (most likely)?".
Parent's college expectation	I	An indicator of parental choice of "very disappointed" for the question "how disappointed would you be if your child did not graduate from college?".
Perceiving receiving family attention	I	An indicator of participant's choice of "quite a bit" or "very much" for the question "how much do you feel that your family pays attention to you".
Feeling been understood in family	I	An indicator of participant's choice of "quite a bit" or "very much" for the question "How much do you feel that people in your family understand you".

Table A3
Summary statistics for covariates and mediating factors.

Variables	Wave	Full sample		Normal Weight sample		Overweight sample (BMI percentile > 0.85)		Obese sample (BMI percentile > 0.95)	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
Personal characteristics									
Obesity status (BMI percentile≧0.95)	I	0.112	0.315	0.000	0.000	N/A	N/A	1.000	0.000
Overweight or obesity (BMI percentile≧0.85)	I	0.238	0.426	0.000	0.000	1.000	0.000	1.000	0.000
Age	I	15.43	1.671	15.46	1.679	15.35	1.645	15.43	1.618
Female (=1, male=0)	I	0.518	0.500	0.550	0.498	0.415	0.493	0.347	0.477
School grade	I	9.532	1.606	9.564	1.610	9.426	1.587	9.484	1.530
Low birth weight	I	0.065	0.247	0.068	0.252	0.055	0.229	0.056	0.231
Family characteristics									
Logarithmic parental household income ($1000)[α]	I	3.658	1.139	3.693	1.197	3.548	0.922	3.573	0.937
Ratio of parental income to the tract-level per capita income in 1989	I	3.861	3.508	3.975	3.860	3.496	1.962	3.616	1.855
Parents having education beyond high school (Yes=1, No=0)	I	0.590	0.492	0.613	0.487	0.515	0.500	0.508	0.501
Single-parent family (Yes=1, No=0)	I	0.188	0.391	0.178	0.383	0.221	0.415	0.206	0.405
Number of children of respondent's biological parents	I	2.497	1.178	2.519	1.193	2.428	1.126	2.452	1.165
Being member of parent-teacher association (Yes=1, No=0)	I	0.378	0.485	0.391	0.488	0.338	0.473	0.351	0.478
Neighborhood characteristics -at census tract level									
Share of families living under the 1989 poverty level	I	0.085	0.078	0.080	0.074	0.100	0.089	0.107	0.095
Per capita income in 1989 ($1000)	I	13.69	4.857	13.96	5.013	12.81	4.206	12.47	4.153
Share of married-couple family households	I	0.828	0.078	0.831	0.076	0.816	0.082	0.814	0.077
Share of people aged 25+ years and with at least college degree	I	0.236	0.125	0.243	0.128	0.213	0.112	0.202	0.107
Share of people aged 16–19 not in school or armed forces and not high school graduate	I	0.101	0.073	0.100	0.072	0.106	0.075	0.112	0.078
Share of female civilian labor force	I	0.447	0.034	0.448	0.032	0.444	0.039	0.440	0.040
Unemployment rate	I	0.061	0.035	0.058	0.032	0.068	0.040	0.069	0.038
Neighborhood characteristics-at county level									
Crime rate (*1000) per 100,000 population	I	4.728	2.485	4.743	2.466	4.679	2.547	4.643	2.575
Per capita government budget spent on education ($1000)	I	0.692	0.143	0.692	0.142	0.690	0.146	0.686	0.132
Per capita government budget spent on health and hospitals ($1000)	I	0.149	0.151	0.147	0.149	0.155	0.159	0.154	0.162
Per capita government budget spent on public welfares ($1000)	I	0.059	0.075	0.058	0.074	0.060	0.080	0.055	0.074
Cognitive abilities									
PPVT score ranking	I	61.22	25.59	61.69	25.54	59.70	25.70	59.33	26.76
	III	60.34	25.78	61.13	25.42	57.77	26.76	57.53	26.65
Words recall within 90 s	IV	6.998	1.940	7.049	1.964	6.834	1.852	6.790	1.830
Words recall within 60 s	IV	5.571	1.974	5.629	1.996	5.385	1.894	5.423	1.791
Backwards digit recall	IV	0.445	0.497	0.454	0.498	0.415	0.493	0.395	0.490
Self-rated intelligence (above average=1, Otherwise=0)	I	0.603	0.489	0.621	0.485	0.545	0.498	0.524	0.500
Having troubles of paying attention in school daily	I	0.125	0.331	0.121	0.326	0.138	0.345	0.117	0.322
Health condition, risky behaviors and psychological wellbeing									
Self-rated very good or excellent health condition	I	0.712	0.453	0.762	0.426	0.553	0.498	0.448	0.498
	III	0.756	0.430	0.795	0.403	0.628	0.484	0.548	0.499
Drinking alcohol weekly	I	0.085	0.278	0.078	0.268	0.106	0.308	0.101	0.302
Ever smoking daily	I	0.210	0.407	0.200	0.400	0.241	0.428	0.258	0.438
Feeling moody daily	I	0.117	0.321	0.110	0.313	0.138	0.345	0.137	0.345
Standardized CES-D score	I	−0.155	0.943	−0.172	0.945	−0.101	0.933	−0.137	0.909
Suicide ideation	I	0.139	0.346	0.134	0.341	0.155	0.362	0.138	0.345
Academic performance related behaviors and feeling in school									
Having excused school absence at least 3 times during school year	I	0.640	0.480	0.629	0.483	0.676	0.468	0.725	0.448
Ever skipping school without excuses during school year	I	0.250	0.433	0.246	0.431	0.261	0.439	0.279	0.450
Ever getting into physical fight in the past 12 months	I	0.273	0.445	0.247	0.431	0.356	0.479	0.379	0.486
Ever fighting with friends or known ones in the last physical fight	I	0.351	0.478	0.320	0.467	0.460	0.499	0.471	0.501
Feeling unsafe in school	I	0.113	0.317	0.107	0.309	0.133	0.340	0.113	0.317
Feelings at home and college inspiration and expectation									
Perceiving receiving family attention	I	0.726	0.446	0.744	0.437	0.667	0.472	0.677	0.468
Feeling been understood in family	I	0.534	0.499	0.544	0.498	0.502	0.500	0.484	0.501
Personal college aspiration	I	0.720	0.449	0.731	0.443	0.685	0.465	0.673	0.470
Personal college expectation	I	0.804	0.397	0.821	0.384	0.750	0.434	0.714	0.453
Parent's college expectation	I	0.384	0.486	0.389	0.488	0.369	0.483	0.367	0.483

Notes: After restricting the Add Health participants to those with the available information of weight status (excluding underweight sample), educational outcomes, PGSs, and key personal, family and neighborhood characteristics, we have 2585 observations for our study. The observations for mediating factors (i.e. cognitive ability, health outcomes, behaviors and feelings, etc.) may vary due to data availability.

[α] The household income ($1000) in Wave I is self-reported by parents. To obtain the logarithm of parental income, we replace parental income with one dollar if parents did not report income in Wave I.

Table A4
The effect of childhood obesity and overweight on education outcomes using OLS or Logit models.

	Wave III		Wave IV	
	High school GPA (1)	College Enrollment (2)	College Completion (3)	Education (Years) (4)
Panel A: Obese (BMI percentile ≥ 0.95)				
Obesity	−0.164***	−0.082**	−0.083***	−0.378***
	(0.053)	(0.032)	(0.030)	(0.131)
Root MSE	0.681	0.403	0.443	1.868
No. of Observations	2218	2218	2218	2218
Panel B: Obese and Overweight (BMI percentile ≥ 0.85)				
Overweight	−0.129***	−0.065***	−0.072***	−0.358***
	(0.034)	(0.023)	(0.021)	(0.093)
Root MSE	0.682	0.406	0.437	1.835
No. of Observations	2585	2585	2585	2585
Controls	Y	Y	Y	Y
School Fixed Effects	Y	Y	Y	Y

Notes: The table presents the OLS/logit estimates of childhood obesity and childhood overweight on education outcomes. The dependent variable of Column 1 is defined in Table 4. The dependent variables of Columns 2–4 are defined in Table 5. The covariates are the same as defined in Table 4. Standard errors in parentheses are clustered by school. * $p < .1$, ** $p < .05$, *** $p < .01$.

Table A5
Robustness Checks: different specifications of IV for childhood obesity and alternative controls for intelligence PGS.

	Different specifications of IV		Alternative control for intelligence PGS		
	Controlling for Parental obesity status (1)	Alternative *GPBMI* based on 97 SNPs (2)	Cognitive function PGS (3)	Educational attainment PGS (4)	Three education related PGS (5)
Panel A. High School GPA (Wave III)					
Obese	−0.918***	−0.595**	−0.911***	−0.937***	−0.925***
	(0.270)	(0.266)	(0.234)	(0.227)	(0.225)
First-stage F statistics	68.81	63.20	90.52	91.13	91.10
Root MSE	0.691	0.667	0.692	0.685	0.684
No. of Observations	2154	2218	2218	2218	2218
Panel B. College Enrollment (Wave IV)					
Obese	−0.374*	−0.285*	−0.357**	−0.367**	−0.371**
	(0.197)	(0.169)	(0.158)	(0.158)	(0.158)
First-stage F statistics	68.81	63.20	90.52	91.13	91.10
Root MSE	0.397	0.393	0.397	0.396	0.396
No. of Observations	2154	2218	2218	2218	2218
Panel C. College Completion (Wave IV)					
Obese	−0.635***	−0.377**	−0.641***	−0.654***	−0.651***
	(0.193)	(0.177)	(0.166)	(0.160)	(0.159)
First-stage F statistics	68.81	63.20	90.52	91.13	91.10
Root MSE	0.456	0.435	0.458	0.456	0.455
No. of Observations	2154	2218	2218	2218	2218
Panel D. Years of schooling (Wave IV)					
Obese	−2.140***	−1.206*	−2.151***	−2.221***	−2.224***
	(0.755)	(0.675)	(0.641)	(0.612)	(0.613)
First-stage F statistics	68.81	63.20	90.52	91.13	91.10
Root MSE	1.867	1.814	1.877	1.860	1.860
No. of Observations	2154	2218	2218	2218	2218

Notes: This table presents the IV estimates of the impact of childhood obesity on high school GPA (Panel A), college enrollment (Panel B), college completion (Panel C), and years of schooling (Panel D) using different specifications of IV and alternative controls for intelligence PGS. The estimates are obtained by additionally controlling for the obesity status of parents (Column 1) and using alternative *GPBMI* based on 97 SNPs (Locke et al., 2015)(Column 2). Instead of intelligence PGS, leveraging cognitive function PGS, educational attainment PGS, and all three education-related PGSs as the proxy for learning ability yields alternative IV estimates in Columns 3–5, respectively. The covariates are the same as defined in Table 4. Standard errors in parentheses are clustered by school reported in Wave I. * $p < .1$, ** $p < .05$, *** $p < .01$.

Table A6
Robustness Checks: different specifications of IV for childhood overweight and alternative controls for intelligence PGS.

	Different specifications of IV		Alternative control for intelligence PGS		
	Controlling for Parental obesity status (1)	Alternative *GPBMI* based on 97 SNPs (2)	Cognitive function PGS (3)	Educational attainment PGS (4)	Three education related PGS (5)
Panel A. High School GPA (Wave III)					
Overweight	−0.524***	−0.378**	−0.540***	−0.558***	−0.553***
	(0.159)	(0.161)	(0.141)	(0.139)	(0.138)
First-stage F statistics	94.02	84.18	114.6	116.1	115.9
Root MSE	0.679	0.667	0.679	0.673	0.672
No. of Observations	2507	2585	2585	2585	2585
Panel B. College Enrollment (Wave IV)					
Overweight	−0.209*	−0.198*	−0.196*	−0.202**	−0.206**
	(0.117)	(0.108)	(0.101)	(0.102)	(0.101)
First-stage F statistics	94.02	84.18	114.6	116.1	115.9
Root MSE	0.396	0.396	0.397	0.395	0.395
No. of Observations	2507	2585	2585	2585	2585
Panel C. College Completion (Wave IV)					
Overweight	−0.385***	−0.231**	−0.398***	−0.408***	−0.408***
	(0.113)	(0.104)	(0.100)	(0.099)	(0.098)
First-stage F statistics	94.02	84.18	114.6	116.1	115.9
Root MSE	0.440	0.427	0.442	0.440	0.440
align="center"No. of Observations	2507	2585	2585	2585	2585
Panel D. Years of schooling (Wave IV)					
Overweight	−1.295***	−0.753*	−1.292***	−1.344***	−1.348***
	(0.441)	(0.424)	(0.388)	(0.380)	(0.381)
First-stage F statistics	94.02	84.18	114.6	116.1	115.9
Root MSE	1.812	1.781	1.817	1.801	1.801
No. of Observations	2507	2585	2585	2585	2585

Notes: This table presents the IV estimates of the impact of childhood overweight on high school GPA (Panel A), college enrollment (Panel B), college completion (Panel C), and years of schooling (Panel D) using different specifications of IV and alternative controls for intelligence PGS. The estimates are obtained by additionally controlling for the obesity status of parents (Column 1) and using alternative *GPBMI* based on 97 SNPs (Locke et al., 2015) (Column 2). Instead of intelligence PGS, leveraging cognitive function PGS, educational attainment PGS, and all three education-related PGSs as the proxy for learning ability yields alternative IV estimates in Columns 3–5, respectively. The covariates are the same as defined in Table 4. Standard errors in parentheses are clustered by school reported in Wave I. * $p < .1$, ** $p < .05$, *** $p < .01$.

Table A7
IV estimates using BMI and BMI percentile as the treatment variable

	Wave III	Wave IV		
	High school GPA (1)	College Enrollment (2)	College Completion (3)	Years of schooling (4)
Panel A: BMI				
BMI	-0.046***	-0.017**	-0.034***	-0.110***
	(0.012)	(0.008)	(0.008)	(0.033)
Root MSE	0.668	0.393	0.432	1.792
Observations	2,585	2,585	2,585	2,585
Panel B: BMI percentile				
BMI percentile	-0.728***	-0.271**	-0.539***	-1.766***
	(0.185)	(0.135)	(0.133)	(0.525)
Root MSE	0.669	0.394	0.435	1.804
Observations	2,585	2,585	2,585	2,585
Controls	Y	Y	Y	Y
School Fixed Effects	Y	Y	Y	Y

Notes: The table presents the IV estimates of BMI and BMI percentile on education outcomes. The dependent variable of Column 1 and all the covariates are defined in Table 4. The dependent variables of Columns 2-4 are defined in Table 5. Standard errors in parentheses are clustered by school. * $p < .1$, ** $p < .05$, *** $p < .01$.

Table A8
The associations between *GPBMI* and each of the control variables.

Variables	Normal weight & obese sample		All sample	
	Coef.	S.E.	Coef.	S.E.
Personal Characteristics				
Age in Wave I	−0.398	0.375	−0.249	0.333
Age square	0.016	0.012	0.010	0.010
Female (=1, male=0)	0.064	0.040	0.035	0.038
Low birth weight (=1, otherwise=0)	0.030	0.087	0.049	0.082
Intelligence PGS	0.024	0.022	0.029	0.022
Smoking PGS	−0.009	0.022	−0.004	0.020
Drinking PGS	0.004	0.024	−0.002	0.022
Major depressive disorder PGS	−0.041	0.036	−0.036	0.036
Depressive symptom PGS	0.035	0.030	0.029	0.029
Mental health cross disorder PGS	−0.009	0.023	−0.025	0.022
Family Characteristics				
Number of children of the respondent's biological parents	0.009	0.019	−0.006	0.016
Log parent self-reported household income in Wave I	−0.019	0.013	−0.018	0.014
Ratio of parental household income to the tract-level per capita income in 1989	−0.001	0.003	−0.002	0.003
Single-parent family (Yes=1, No=0)	−0.073	0.056	−0.049	0.050
Parent having education beyond high school (Yes=1, No=0)	−0.007	0.042	0.004	0.039
Participated in the parent-teacher association (Yes=1, No=0)	−0.049	0.047	−0.033	0.042
Neighborhood characteristics at the tract level				
Share of people aged 16–19 not in school or armed forces and not high school graduate	−0.096	0.338	−0.042	0.334
Share of people aged 25 years/over and with college degree/more	0.131	0.529	−0.014	0.489
Per capita income in 1989 (1000 dollars)	−0.010	0.014	−0.011	0.013
Unemployment rate	−1.277	1.693	−0.981	1.544
Share of female civilian labor force	−0.518	1.187	−1.384	1.032
Share of married-couple family households	0.696	0.681	0.227	0.634
Share of families with income in 1989 below poverty level	0.718	0.899	−0.072	0.826
Neighborhood characteristics at the county level				
Per capita government budget spent on education ($1000)	−0.361	0.793	−0.145	0.655
Per capita government budget spent on health and hospitals ($1000)	−0.478	0.817	−0.589	0.721
Per capita government budget spent on public welfares ($1000)	0.024	2.845	0.750	1.735
Crime rate(*1000) per 100,000 population	0.005	0.047	0.008	0.040
Ancestry-specific ten principal components	Yes		Yes	
School and grade fixed effects	Yes		Yes	
No. of observations	2218		2585	

Notes: This table presents the correlation between *GPBMI* and each of the control variables for education outcomes for normal weight & obese sample(Column 1) and all sample(Column 2). Standard errors in parentheses are clustered by school reported in Wave I. * $p < .1$, ** $p < .05$, *** $p < .01$.

Table A9
The effect of childhood obesity and overweight on education outcomes after additionally controlling for PGSs of substance use and mental health.

	Wave III		Wave IV	
	High school GPA (1)	College Enrollment (2)	College Completion (3)	Education (Years) (4)
Panel A: Obese (BMI percentile ≥ 0.95)				
Obesity	−0.919***	−0.367**	−0.647***	−2.194***
	(0.236)	(0.161)	(0.168)	(0.652)
Root MSE	0.686	0.395	0.455	1.867
Observations	2218	2218	2218	2218
Panel B: Obese and Overweight (BMI percentile ≥ 0.85)				
Overweight	−0.551***	−0.208**	−0.408***	−1.348***
	(0.145)	(0.101)	(0.103)	(0.394)
Root MSE	0.675	0.395	0.441	1.809
Observations	2585	2585	2585	2585
Controls	Y	Y	Y	Y
School Fixed Effects	Y	Y	Y	Y

Notes: The table presents the IV estimates of childhood obesity and childhood overweight on education outcomes. The dependent variable of Column 1 is defined in Table 4. The dependent variables of Columns 2–4 are defined in Table 5. Besides of controlling for covariates defined in Table 4, we additionally control for PGSs for substance use (i.e., smoking and drinking) and mental health (i.e., major depressive disorder, depressive symptom, and mental health cross disorder). Standard errors in parentheses are clustered by school. * $p < .1$, ** $p < .05$, *** $p < .01$.

References

Almond, D., Currie, J., & Duque, V. (2018). Childhood circumstances and adult outcomes: Act II. *Journal of Economic Literature, 56*(4), 1360–1446.

Amin, V., Fletcher, J. M., Lu, Q., & Song, J. (2023). *Re*-examining the relationship between education and adult mental health in the UK: A research note. *Economics of Education Review, 93*, Article 102354.

An, R., Yan, H., Shi, X., & Yang, Y. (2017). Childhood obesity and school absenteeism: A systematic review and meta-analysis. *Obesity Reviews, 18*(12), 1412–1424.

Avila, C., Holloway, A. C., Hahn, M. K., Morrison, K. M., Restivo, M., Anglin, R., et al. (2015). An overview of links between obesity and mental health. *Current Obesity Reports, 4*(3), 303–310.

Baiocchi, M., Cheng, J., & Small, D. S. (2014). Instrumental variable methods for causal inference. *Statistics in Medicine, 33*(13), 2297–2340.

Balsa, A. I., Giuliano, L. M., & French, M. T. (2011). The effects of alcohol use on academic achievement in high school. *Economics of Education Review, 30*(1), 1–15.

Barth, D., Papageorge, N. W., & Thom, K. (2020). Genetic endowments and wealth inequality. *Journal of Political Economy, 128*(4), 1474–1522.

Baselmans, B. M., Jansen, R., Ip, H. F., van Dongen, J., Abdellaoui, A., van de Weijer, M. P., et al. (2019). Multivariate genome-wide analyses of the well-being spectrum. *Nature Genetics, 51*(3), 445–451.

Belsky, D. W., Caspi, A., Arseneault, L., Corcoran, D. L., Domingue, B. W., Harris, K. M., et al. (2019). Genetics and the geography of health, behaviour and attainment. *Nature Human Behaviour, 3*(6), 576–586.

Belsky, D. W., Domingue, B. W., Wedow, R., Arseneault, L., Boardman, J. D., Caspi, A., et al. (2018). Genetic analysis of social-class mobility in five longitudinal studies. *Proceedings of the National Academy of Sciences, 115*(31), E7275–E7284.

Belsky, D. W., & Israel, S. (2014). Integrating genetics and social science: Genetic risk scores. *Biodemography and Social Biology, 60*(2), 137–155.

Benjamin, D. J., Cesarini, D., Van Der Loos, M. J., Dawes, C. T., Koellinger, P. D., Magnusson, P. K., et al. (2012). The genetic architecture of economic and political preferences. *The Proceedings of the National Academy of Sciences, 109*(21), 8026–8031.

Benner, A. D., & Mistry, R. S. (2007). Congruence of mother and teacher educational expectations and low-income youth's academic competence. *Journal of Educational Psychology, 99*(1), 140.

Boardman, J. D., Domingue, B. W., & Daw, J. (2015). What can genes tell us about the relationship between education and health? *Social Science & Medicine, 127*, 171–180.

Böckerman, P., Cawley, J., Viinikainen, J., Lehtimäki, T., Rovio, S., Seppälä, I., et al. (2019). The effect of weight on labor market outcomes: An application of genetic instrumental variables. *Health Economics, 28*(1), 65–77.

Braudt, D., & Harris, K.M. (2020). Polygenic scores (PGSs) in the national longitudinal study of adolescent to adult health (add health)–release 2. In. https://addhealth.cpc.unc.edu/wp-content/uploads/docs/user_guides/WaveIVPGSRelease2UserGuide.pdf.

Brunello, G., Sanz-de-Galdeano, A., & Terskaya, A. (2020). Not only in my genes: The effects of peers' genotype on obesity. *Journal of Health Economics, 72*, Article 102349.

Card, D. (1999). The causal effect of education on earnings. *Handbook of Labor Economics, 3*, 1801–1863.

Cawley, J. (2010). The economics of childhood obesity. *Health Affairs, 29*(3), 364–371.

Cawley, J., Han, E., Kim, J., & Norton, E. C. (2019). Testing for family influences on obesity: The role of genetic nurture. *Health Economics, 28*(7), 937–952.

CDC. (2022). Childhood obesity facts. In (Vol. 2022). https://www.cdc.gov/obesity/data/childhood.html (last access date: 2022, Nov. 1st): CDC.

Chetty, R., & Hendren, N. (2018a). The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects. *The Quarterly Journal of Economics, 133*(3), 1107–1162.

Chetty, R., & Hendren, N. (2018b). The impacts of neighborhoods on intergenerational mobility II: County-level estimates. *The Quarterly Journal of Economics, 133*(3), 1163–1228.

Cohen, A. K., Rai, M., Rehkopf, D. H., & Abrams, B. (2013a). Educational attainment and obesity: A systematic review. *Obesity Reviews, 14*(12), 989–1005.

Cohen, A. K., Rehkopf, D. H., Deardorff, J., & Abrams, B. (2013b). Education and obesity at age 40 among American adults. *Social Science & Medicine, 78*, 34–41.

Consortium, C.-D. G.o.t. P. G. (2013). Identification of risk loci with shared effects on five major psychiatric disorders: A genome-wide analysis. *The Lancet, 381*(9875), 1371–1379.

Cornaglia, F., Crivellaro, E., & McNally, S. (2015). Mental health and education decisions. *Labour Economics, 33*, 1–12.

Cox, R. G., Zhang, L., Johnson, W. D., & Bender, D. R. (2007). Academic performance and substance use: Findings from a state survey of public high school students. *Journal of School Health, 77*(3), 109–115.

Crosnoe, R. (2007). Gender, obesity, and education. *Sociology of education, 80*(3), 241–260.

Currie, J. (2009). Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development. *Journal of Economic Literature, 47*(1), 87–122.

Cutler, D. M., & Lleras-Muney, A. (2006). *Education and health: Evaluating theories and evidence*. Cambridge, Mass., USA: National bureau of economic research.

Davey Smith, G., & Hemani, G (2014). Mendelian randomization: Genetic anchors for causal inference in epidemiological studies. *Human Molecular Genetics, 23*(R1), R89–R98.

Davies, G., Lam, M., Harris, S. E., Trampush, J. W., Luciano, M., Hill, W. D., et al. (2018). Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nature Communications, 9*(1), 1–16.

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

DiPrete, T. A., Burik, C. A., & Koellinger, P. D. (2018). Genetic instrumental variable regression: Explaining socioeconomic and health outcomes in nonexperimental data. *Proceedings of the National Academy of Sciences, 115*(22), E4970–E4979.

Dixon, P., Hollingworth, W., Harrison, S., Davies, N. M., & Smith, G. D. (2020). Mendelian Randomization analysis of the causal effect of adiposity on hospital costs. *Journal of Health Economics, 70*, Article 102300.

Domingue, B. W., Fletcher, J., Conley, D., & Boardman, J. D. (2014). Genetic and educational assortative mating among US adults. *Proceedings of the National Academy of Sciences, 111*(22), 7996–8000.

Dyce, C. M., Albold, C., & Long, D. (2012). Moving from college aspiration to attainment: Learning from one college access program. *The High School Journal*, 152–165.

Edwards, C. H., Bjørngaard, J. H., & Minet Kinge, J. (2021). The relationship between body mass index and income: Using genetic variants from HUNT as instrumental variables. *Health Economics, 30*(8), 1933–1949.

Fan, M., & Jin, Y. (2014a). Do neighborhood parks and playgrounds reduce childhood obesity? *American Journal of Agricultural Economics, 96*(1), 26–42.

Fan, M., & Jin, Y. (2014b). Obesity and Self-control: Food Consumption, Physical Activity, and Weight-loss Intention. *Applied Economic Perspectives and Policy, 36*(1), 125–145.

Fan, M., & Jin, Y. (2015). The effects of weight perception on adolescents' weight-loss intentions and behaviors: Evidence from the youth risk behavior surveillance survey. *International Journal of Environmental Research and Public Health, 12*(11), 14640–14668.

Farhat, T., Iannotti, R. J., & Simons-Morton, B. G. (2010). Overweight, obesity, youth, and health-risk behaviors. *American Journal of Preventive Medicine, 38*(3), 258–267.

Fergusson, D. M., Boden, J. M., & Horwood, L. J. (2013). Childhood self-control and adult outcomes: Results from a 30-year longitudinal study. *Journal of the American Academy of Child & Adolescent Psychiatry, 52*(7), 709–717. e701.

Fletcher, J. (2013a). Adolescent depression and adult labor market outcomes. *Southern Economic Journal, 80*(1), 26–49.

Fletcher, J. M. (2013b). The effects of personality traits on adult labor market outcomes: Evidence from siblings. *Journal of Economic Behavior and Organization, 89*, 122–135.

Fletcher, J. M. (2018). Economics and genetics. *Oxford research encyclopedia of economics and finance*.

Forehand, R., Long, N., Brody, G. H., & Fauber, R. (1986). Home predictors of young adolescents' school behavior and academic performance. *Child Development*, 1528–1533.

Fuxa, A. J., & Fulkerson, J. A. (2011). Adolescent obesity and school performance and perceptions of the school environment among Minnesota high school students. *School Mental Health, 3*(2), 102–110.

Gershenson, S., Jacknowitz, A., & Brannegan, A. (2017). Are student absences worth the worry in US primary schools? *Education Finance and Policy, 12*(2), 137–165.

Glew, G. M., Fan, M.-. Y., Katon, W., Rivara, F. P., & Kernic, M. A. (2005). Bullying, psychosocial adjustment, and academic performance in elementary school. *Archives of Pediatrics & Adolescent Medicine, 159*(11), 1026–1031.

Hanushek, E. A., & Woessmann, L. (2020). Education, knowledge capital, and economic growth. *The economics of education* (pp. 171–182).

Harden, K. P., & Koellinger, P. D. (2020). Using genetics for social science. *Nature Human Behaviour, 4*(6), 567–576.

Hasler, G., Pine, D., Gamma, A., Milos, G., Ajdacic, V., Eich, D., et al. (2004). The associations between psychopathology and being overweight: A 20-year prospective study. *Psychological Medicine, 34*(6), 1047–1057.

Heckman, J. J. (2007). The economics, technology, and neuroscience of human capability formation. *Proceedings of the National Academy of Sciences, 104*(33), 13250–13255.

Heckman, J. J., & Mosso, S. (2014). The economics of human development and social mobility. *Annual Review of Economics, 6*(1), 689–733.

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

Hoffmann, F., Lee, D. S., & Lemieux, T. (2020). Growing income inequality in the United States and other advanced economies. *Journal of Economic Perspectives, 34*(4), 52–78.

Hossler, D., & Stage, F. K. (1992). Family and high school experience influences on the postsecondary educational plans of ninth-grade students. *American Educational Research Journal, 29*(2), 425–451.

Howard, D. M., Adams, M. J., Clarke, T.-. K., Hafferty, J. D., Gibson, J., Shirali, M., et al. (2019). Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nature Neuroscience, 22*(3), 343–352.

Jackson, M. I. (2009). Understanding links between adolescent health and educational attainment. *Demography, 46*(4), 671–694.

Kaestner, R., & Grossman, M. (2009). Effects of weight on children's educational achievement. *Economics of Education Review, 28*(6), 651–661.

Kenkel, D., Lillard, D., & Mathios, A. (2006). The roles of high school completion and GED receipt in smoking and obesity. *Journal of Labor Economics, 24*(3), 635–660.

Koellinger, P. D., & De Vlaming, R. (2019). Mendelian randomization: The challenge of unobserved environmental confounds. *International Journal of Epidemiology, 48*(3), 665–671.

Lawlor, D. A., Harbord, R. M., Sterne, J. A., Timpson, N., & Davey Smith, G. (2008). Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Statistics in Medicine, 27*(8), 1133–1163.

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., et al. (2018). Gene discovery and polygenic prediction from a 1.1-million-person GWAS of educational attainment. *Nature Genetics, 50*(8), 1112.

Lemieux, T. (2006). Postsecondary education and increasing wage inequality. *American Economic Review, 96*(2), 195–199.

Lepore, S. J., & Kliewer, W. (2013). Violence exposure, sleep disturbance, and poor academic performance in middle school. *Journal of Abnormal Child Psychology, 41*(8), 1179–1189.

Lessard, L. M., & Juvonen, J. (2022). Developmental changes in the frequency and functions of school-related communication with friends and family across high school: Effects on college enrollment. *Developmental Psychology, 58*(3), 575.

Liang, J., Matheson, B., Kaye, W., & Boutelle, K. (2014). Neurocognitive correlates of obesity and obesity-related behaviors in children and adolescents. *International Journal of Obesity, 38*(4), 494–506.

Litsfeldt, S., Ward, T. M., Hagell, P., & Garmy, P. (2020). Association between sleep duration, obesity, and school failure among adolescents. *The Journal of School Nursing, 36*(6), 458–463.

Liu, M., Jiang, Y., Wedow, R., Li, Y., Brazel, D. M., Chen, F., et al. (2019). Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nature Genetics, 51*(2), 237–244.

Locke, A. E., Kahali, B., Berndt, S. I., Justice, A. E., Pers, T. H., Day, F. R., et al. (2015). Genetic studies of body mass index yield new insights for obesity biology. *Nature, 518*(7538), 197–206.

Lumeng, J. C., Forrest, P., Appugliese, D. P., Kaciroti, N., Corwyn, R. F., & Bradley, R. H. (2010). Weight status as a predictor of being bullied in third through sixth grades. *Pediatrics, 125*(6), e1301–e1307.

Martin, A., Booth, J. N., McGeown, S., Niven, A., Sproule, J., Saunders, D. H., et al. (2017). Longitudinal associations between childhood obesity and academic achievement: Systematic review with focus group data. *Current Obesity Reports, 6*(3), 297–313.

Martin, M. J., Conger, R. D., Sitnick, S. L., Masarik, A. S., Forbes, E. E., & Shaw, D. S. (2015). Reducing risk for substance use by economically disadvantaged young men: Positive family environments and pathways to educational attainment. *Child Development, 86*(6), 1719–1737.

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., et al. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences, 108*(7), 2693–2698.

Norton, E. C., & Han, E. (2008). Genetic information, obesity, and labor market outcomes. *Health Economics, 17*(9), 1089–1104.

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., et al. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature, 533*(7604), 539–542.

Oreopoulos, P., Stabile, M., Walld, R., & Roos, L. L. (2008). Short-, medium-, and long-term consequences of poor infant health an analysis using siblings and twins. *Journal of Human Resources, 43*(1), 88–138.

Palermo, T. M., & Dowd, J. B. (2012). Childhood obesity and human capital accumulation. *Social Science & Medicine, 75*(11), 1989–1998.

Papageorge, N. W., & Thom, K. (2020). Genes, education, and labor market outcomes: Evidence from the health and retirement study. *Journal of the European Economic Association, 18*(3), 1351–1399.

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*(3), 385–401.

Richmond, R. C., & Smith, G. D. (2022). Mendelian randomization: Concepts and scope. *Cold Spring Harbor Perspectives in Medicine, 12*(1), Article a040501.

Rivers, J., Mullis, A. K., Fortner, L. A., & Mullis, R. L. (2012). Relationships between parenting styles and the academic performance of adolescents. *Journal of Family Social Work, 15*(3), 202–216.

Roksa, J., & Kinsley, P. (2019). The role of family support in facilitating academic success of low-income students. *Research in Higher Education, 60*, 415–436.

Rosenzweig, M. R., & Zhang, J. (2013). Economic growth, comparative advantage, and gender differences in schooling outcomes: Evidence from the birthweight differences of Chinese twins. *Journal of Development Economics, 104*, 245–260.

Rouse, K., & Hunziker, B. (2020). Child bodyweight and human capital: Test scores, teacher assessments and noncognitive skills. *Economics of Education Review, 79*, Article 102042.

Ruff, R. R., Senthi, S., Susser, S. R., & Tsutsui, A. (2019). Oral health, academic performance, and school absenteeism in children and adolescents: A systematic review and meta-analysis. *The Journal of the American Dental Association, 150*(2), 111-121. e114.

Sabia, J. J. (2007). The effect of body weight on adolescent academic performance. *Southern Economic Journal*, 871–900.

Sabia, J. J., & Rees, D. I. (2015). Body weight, mental health capital, and academic achievement. *Review of Economics of the Household, 13*(3), 653–684.

Sahoo, K., Sahoo, B., Choudhury, A. K., Sofi, N. Y., Kumar, R., & Bhadoria, A. S. (2015). Childhood obesity: Causes and consequences. *Journal of Family Medicine and Primary Care, 4*(2), 187.

Santana, C. C. A., Hill, J. O., Azevedo, L. B., Gunnarsdottir, T., & Prado, W. L. (2017). The association between obesity and academic performance in youth: A systematic review. *Obesity Reviews, 18*(10), 1191–1199.

Savage, J. E., Jansen, P. R., Stringer, S., Watanabe, K., Bryois, J., De Leeuw, C. A., et al. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature Genetics, 50*(7), 912–919.

Scholder, S.v. H. K., Smith, G. D., Lawlor, D. A., Propper, C., & Windmeijer, F. (2012). The effect of fat mass on educational attainment: Examining the sensitivity to different identification strategies. *Economics & Human Biology, 10*(4), 405–418.

Sevgi, M., Rigoux, L., Kühn, A. B., Mauer, J., Schilbach, L., Hess, M. E., et al. (2015). An obesity-predisposing variant of the FTO gene regulates D2R-dependent reward learning. *Journal of Neuroscience, 35*(36), 12584–12592.

Simmonds, M., Llewellyn, A., Owen, C. G., & Woolacott, N. (2016). Predicting adult obesity from childhood obesity: A systematic review and meta-analysis. *Obesity Reviews, 17*(2), 95–107.

Speliotes, E. K., Willer, C. J., Berndt, S. I., Monda, K. L., Thorleifsson, G., Jackson, A. U., et al. (2010). Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature Genetics, 42*(11), 937–948.

Staff, J., Patrick, M. E., Loken, E., & Maggs, J. L. (2008). Teenage alcohol use and educational attainment. *Journal of Studies on Alcohol and Drugs, 69*(6), 848–858.

Stock, J. H., & Yogo, M. (2005). Testing for weak instruments in linear IV regression. *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg, 80*(4.2), 1.

Suhrcke, M., & de Paz Nieves, C (2011). *The impact of health and health behaviours on educational outcomes in high-income countries: A review of the evidence.* Copenhagen: World Health Organization. Regional Office for Europe.

Tyrrell, J., Jones, S. E., Beaumont, R., Astley, C. M., Lovell, R., Yaghootkar, H., et al. (2016). Height, body mass index, and socioeconomic status: Mendelian randomisation study in UK Biobank. *BMJ (Clinical research ed.)*, 352.

Von Hinke, S., Smith, G. D., Lawlor, D. A., Propper, C., & Windmeijer, F. (2016). Genetic markers as instrumental variables. *Journal of Health Economics, 45*, 131–148.

Wang, F., & Veugelers, P. (2008). Self-esteem and cognitive development in the era of the childhood obesity epidemic. *Obesity Reviews, 9*(6), 615–623.

Willage, B. (2018). The effect of weight on mental health: New evidence using genetic IVs. *Journal of Health Economics, 57*, 113–130.

Yan, J., & Brocksen, S. (2013). Adolescent risk perception, substance use, and educational attainment. *Journal of Risk Research, 16*(8), 1037–1055.

Yengo, L., Sidorenko, J., Kemper, K. E., Zheng, Z., Wood, A. R., Weedon, M. N., et al. (2018). Meta-analysis of genome-wide association studies for height and body mass index in∼ 700000 individuals of European ancestry. *Human Molecular Genetics, 27*(20), 3641–3649.