CENTER FOR POLICY RESEARCH

THE MAXWELL SCHOOL

WORKING PAPER SERIES

The Impact of BMI on Mental Health: Further Evidence from Genetic Markers

Vikesh Amin, Carlos A. Flores, and Alfonso Flores-Lagunes

Paper No. 216
August 2019

ISSN: 1252-3066

SYRACUSE UNIVERSITY

Maxwell School
Center for Policy Research

426 Eggers Hall
Syracuse University
Syracuse, NY 13244-1020
(315) 443-3114/email: ctrpol@syr.edu
http://www.maxwell.syr.edu/CPR_Working_Papers.aspx

CENTER FOR POLICY RESEARCH –Fall 2019

Leonard M. Lopoo, Director
Professor of Public Administration and International Affairs (PAIA)

Associate Directors

Margaret Austin
Associate Director, Budget and Administration

John Yinger
Trustee Professor of Economics (ECON) and Public Administration and International Affairs (PAIA)
Associate Director, Center for Policy Research

SENIOR RESEARCH ASSOCIATES

Badi Baltagi, ECON
Robert Bifulco, PAIA
Leonard Burman, PAIA
Carmen Carrión-Flores, ECON
Alfonso Flores-Lagunes**,** ECON
Sarah Hamersma, PAIA
Madonna Harrington Meyer, SOC
Colleen Heflin, PAIA
William Horrace, ECON
Yilin Hou, PAIA
Hugo Jales, ECON
Jeffrey Kubik, ECON
Yoonseok Lee, ECON
Amy Lutz, SOC
Yingyi Ma, SOC
Katherine Michelmore, PAIA
Jerry Miner, ECON
Shannon Monnat, SOC
Jan Ondrich, ECON
David Popp, PAIA
Stuart Rosenthal, ECON
Michah Rothbart, PAIA
Alexander Rothenberg, ECON
Rebecca Schewe, SOC
Amy Ellen Schwartz, PAIA/ECON
Ying Shi, PAIA
Saba Siddiki, PAIA
Perry Singleton, ECON
Yulong Wang, ECON
Michael Wasylenko, ECON
Peter Wilcoxen, PAIA
Maria Zhu. ECON

GRADUATE ASSOCIATES

Rhea Acuña, PAIA
Mariah Brennan, SOC. SCI.
Jun Cai, ECON
Ziqiao Chen, PAIA
Yoon Jung Choi, PAIA
Dahae Choo, PAIA
Stephanie Coffey, ECON
Brandon De Bruhl, PAIA
Giuseppe Germinario, ECON
Myriam Gregoire-Zawilski, PAIA
Emily Gutierrez, PAIA
Jeehee Han, PAIA
Mary Helander, SOC. SCI.
Amra Kandic, SOC
Hyoung Kwon, PAIA
Mattie Mackenzie-Liu, PAIA
Maeve Maloney, ECON
Austin McNeil Brown, SOC. SCI.
Qasim Mehdi, PAIA
Claire Pendergrast, SOC
Jonathan Presler, ECON
Krushna Ranaware, SOC
Christopher Rick, PAIA
David Schwegman, PAIA
Saied Toossi, PAIA
Huong Tran, ECON
Joaquin Urrego, ECON
Yao Wang, ECN
Yi Yang, ECON
Xiaoyan Zhang, ECN
Bo Zheng, PAIA
Dongmei Zhu, SOC. SCI.

STAFF

Joanna Bailey, Research Associate
Joseph Boskovski, Maxwell X Lab
Katrina Fiacchi, Administrative Specialist
Emily Minnoe, Administrative Assistant
Candi Patterson, Computer Consultant
Laura Walsh, Administrative Assistant

Abstract

We examine the relationship between body mass index (BMI) and mental health for young adults and elderly individuals using data from the National Longitudinal Study of Adolescent Health and the Health & Retirement Study. While ordinary least squares (OLS) estimates show that BMI is significantly associated with worse mental health in both young adulthood and old age, they are likely to be confounded by (i) unobserved factors that affect both BMI and mental health and (ii) reverse causality. To tackle confounding, we take two complementary approaches. First, we use a polygenic score for BMI as an instrumental variable (IV) and adjust for polygenic scores for other factors that may invalidate this IV. The IV estimates indicate that there is no statistically significant relationship between BMI and mental health for young adults, whereas there is a positive and statistically significant relationship for the elderly. Moreover, we show that IV estimates likely have to be interpreted as identifying a weighted average of effects of BMI on mental health mostly for individuals on the upper quantiles of the BMI distribution. Given potential remaining concerns about the validity of the IV, our second approach is to consider it an "imperfect" IV and estimate an upper bound on the average treatment effect for the corresponding population following Nevo & Rosen (2012). The estimated upper bounds reinforce the conclusions from the IV estimates: they show little evidence of a detrimental effect of BMI on mental health for young adults while being consistent with an economically meaningful effect for elderly individuals. Lastly, we explore some of the potential channels through which BMI may affect mental health for the elderly.

JEL No. I10, I12

Keywords: Obesity, Mental Health, Depression, Polygenetic Scores

Authors: Vikesh Amin, Central Michigan University; Carlos A. Flores, California Polytechnic State University at San Luis Obispo; Alfonso Flores-Lagunes, Syracuse University, Center for Policy Research, Institute of Labor Economics (IZA), Global Labor Organization (GLO).

Acknowledgement: The authors thank research assistance from Giuseppe Germinario. This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due to Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis. The Health and Retirement Study (HRS accession number 0925-0670) is sponsored by the National Institute on Aging (grant numbers NIA U01AG009740, RC2AG036495, and RC4AG039029) and is conducted by the University of Michigan. Additional funding support for genotyping and analysis were provided by NIH/NICHD R01 HD060726. The authors acknowledge research funding from NIH grant number 1R01HD094011-01.

1. Introduction

Mental health is an important public health issue in the US because of the high prevalence of mental illnesses, and the associated economic and societal costs. In any given year, approximately 18.5% of US adults suffer from a mental illness (National Alliance on Mental Illness). Mental illnesses hasten mortality by 10-20 years, (Chesney et al. 2014), and are linked to diabetes, heart disease, strokes, and suicide, which in turn led to the decrease in US life expectancy in 2016 (Kochanek et al. 2017).

Poor mental health is a multifactorial problem. Although there is no single solution, one well-known risk factor is body mass. Several studies have shown that individuals who have a high BMI or are obese are more likely to have poor mental health and be depressed (Carpenter et al. 2000; Luppino et al. 2010; Ha et al. 2017; Roberts et al. 2000; Rosen-Reynoso 2011; Scott et al. 2008; Simon et al. 2006; Zhao et al. 2009). Biologically, there is a relationship between BMI and mental health, because obesity is associated with chronic low-grade inflammation in peripheral tissues and blood circulation (Gregor & Hotamisligil 2011). Inflammation in turn affects brain physiology, and alters mood and behavior leading to depression (Miller & Raison 2016). Markowitz et al. (2008) suggest that the relationship may also be due to behavioral mechanisms (functional impairment), cognitive mechanisms (body image dissatisfaction, poor self-reported health), and social mechanisms (stigma). There is a stigma associated with obesity. Obese individuals are viewed as being lazy, unintelligent, unsuccessful, and lacking self-discipline (Puhl & Heuer 2010). These stereotypes can lead to actual and/or perceived discrimination, low self-esteem and depression (Kessler et al. 1999). Discrimination stemming from obesity can also lead to depression through the labor market, to the extent that obesity is associated with worse labor market

outcomes, which in turn can lead to financial stress and depression[1]. Poor self-reported health can contribute to depression through a cognitive mechanism, because individuals who believe their health to be poor may also hold other depressive beliefs. The relationship can also run through functional impairments and exercise. Obesity is associated with a higher probability of experiencing limitations in carrying out activities of daily living (Himes 2000). Functional limitations affect one's ability to undertake physical exercise, which is associated with better mental health (Paolucci et al. 2018).

Although there are empirical associations, and theoretical reasons for BMI affecting mental health, it is unclear whether there is a causal relationship. First, the causal relationship could be confounded by unobserved factors (e.g. time preference, genetic endowments, innate ability) that affect both BMI and mental health. Second, there is a problem of reverse causality insofar as poor mental health is associated with a higher likelihood of being obese. To tackle confounding, some recent studies have employed a polygenic score (PGS) for BMI as an instrumental variable (IV). A PGS is a summary measure of an individual's genetic predisposition for a given trait, and is constructed using results from Genome Wide Association Studies (GWAS). In a GWAS, hundreds of thousands of single nucleotide polymorphisms (SNPs) are tested for associations with an outcome[2]. As an example, Speliotes et al. (2010) conducted a GWAS on a sample of 123,864 individuals, where they examined associations between 2.8 million SNPs with BMI. They identified 32 SNPs that reached genome-wide significance ($p<5x10^{-8}$), which explain approximately 1.45% of the variation in BMI. A PGS for BMI is constructed

[1] Studies have found that a higher BMI/obesity is associated with a lower probability of being employed and lower wages (Brunello & D'Hombres 2007; Greve 2008; Morris 2007).

[2] A SNP is a DNA sequence variation occurring when a single nucleotide (A, T, C or G) in the genome differs at a single position among individuals.

by aggregating the SNPs identified in Speliotes et al. (2010) and weighting them by the strength of their association.[3]

Previous studies have used as an IV a PGS consisting of 32 SNPs identified as genome-wide significant predictors of BMI in the GWAS by Speliotes et al. (2010)[4]. IV estimates for US based studies show that there is no statistically significant relationship between BMI and mental health. Hung et al. (2014) estimate the effect of BMI on the probability of major depressive disorder in young adults (average age of 26 years) using data from RADIANT. Walter et al. (2015) estimate the effect of BMI on depression for women (with an average age of 56 years) using data from the Nurses' Health Study. Willage (2018) estimates the effect of BMI on depression in young adults (18-34 years of age) using the siblings sample in the National Longitudinal Study of Adolescent Health. In all three studies, the OLS estimates showed a positive statistically significant association between BMI and mental health, whereas IV estimates were statistically insignificant. Willage (2018) does however find that BMI is associated with suicidal ideation. In contrast to US results, Jokela et al. (2012) find that a higher BMI is associated with worse mental health in adolescents and adults in Finland. Tyrrell et al. (2018) use a PGS consisting of 73 SNPs from those identified in the more recent GWAS by Locke et al. (2015)[5]. The IV estimates in Tyrrell et al. (2018) show a significant positive association between BMI and depression for individuals (average age of 57 years) in the UK.

[3] The PGS for individual *i* is a weighted average across the number of SNPs (n) of the number of reference alleles A (0, 1 or 2) at that SNP multiplied by the corresponding beta estimate from the GWAS analysis: $PGS_i = \sum_{j=1}^{n} \beta_j \ A_{ij}$. However, in practice constructing PGSs involves several complex decisions such as whether the PGS should be limited to genome-wide significant SNPs or whether a series of more liberal thresholds should be applied (e.g., p<0.001, 0.01). See Ware et al. (2017) for a detailed overview of the decisions involved in constructing GRSs.

[4] In the economics literature, a PGS for BMI has also been used as an IV to estimate effects of BMI on labor market outcomes (Böckerman et al. 2019) and on academic performance and blood pressure (von Hinke et al. 2016).

[5] Locke et al. (2015) identified 97 SNPs (of which 56 are novel) as genome-wide significant predictors of BMI, which explain about 2.7% of the variation in BMI.

We contribute by examining the relationship between BMI and mental health (measured using the Center for Epidemiologic Studies Depression Scale; CES-D) for two important groups: young adults (aged 25-34 years from the National Longitudinal Study of Adolescent Health) and elderly individuals (aged 50-89 years from the Health & Retirement Study), using genetic data to form the IV. The comparison between young adults and the elderly is interesting because some studies (Crisp et al. 1980; Palinkas et al. 1996) have found a protective effect (often termed the Jolly Fat Hypothesis) of BMI/obesity on mental health. Like previous studies, notably Willage (2018) who also analyzed a sample of young adults from the National Longitudinal Study of Adolescent Health, we also use genetic data to form an IV, but there are two key differences in our study and approach[6]. First, our PGS is more powerful as it is based on the more recent Locke et al. (2015) GWAS, and includes all SNPs. A PGS that includes all SNPs has more predictive power than one that is just based on genome wide significant SNPs (Ware et al. 2017). Second, one concern with using a PGS as an IV is that it may violate the exclusion restriction because of pleiotropy (von Hinke et al., 2016; van Kippersluis and Rietveld, 2017; DiPrete et al, 2018); that is, that the PGS for BMI may affect mental health independently of its effect on BMI because genes associated with BMI may be related to other genes or traits that also affect mental health. We address this potential concern in two separate ways. First, by using a PGS for depression to control for the genetic predisposition to poor mental health, and a PGS for educational attainment to control for innate ability. We find that the IV estimates for the elderly are likely overestimated if we do not control for these PGSs. Second, we regard the PGS for BMI as a potentially "imperfect IV" (Nevo & Rosen, 2012)

[6] An additional important difference between our analysis of young adults and Willage (2018) is that he uses the siblings sample from the National Longitudinal Study of Adolescent Health, while we use a more general sample described in section 2.1.

in the sense that it is allowed to violate the exclusion restriction, and estimate an upper bound for the effect of BMI on mental health in the corresponding population.

Overall, the IV estimates indicate that there is no statistically significant relationship between body mass and mental health for young adults, whereas there is a positive, substantial, and statistically significant relationship for the elderly. For the latter group, a 5 kg/m^2 increase in BMI is related to a 20% increase in their CES-D score, and a 29% increase in the likelihood of depression. We stress that IV estimates are not directly comparable to OLS estimates. In the presence of heterogeneous effects, IV methods identify a local average treatment effect (LATE) for those individuals whose BMI is affected by the PGS for BMI (the so-called "compliers"; Imbens & Angrist, 1994). In contrast, OLS estimates, under strong assumptions such as unconfoundedness, identify the average treatment effect for the corresponding population (young adults or the elderly). Using unconditional quantile treatment effects estimates of the relationship between the BMI PGS and BMI, we document that using a BMI PGS as an instrument likely identifies a weighted average of effects for those individuals whose BMI is affected by their BMI PGS (compliers), with the weights being higher for individuals in the upper quantiles of the BMI distribution.

To complement the IV analysis, we estimate an upper bound on the average treatment effect for the corresponding population following Nevo & Rosen (2012). Their approach combines the use of an "imperfect" IV (in our case, the BMI PGS) with relatively weak assumptions about the direction (and in some cases relative magnitude) of the correlation between the IV and unobserved factors in the OLS model for mental health, and the correlation between those same unobserved factors and the endogenous regressor (BMI). The estimated upper bounds are largely consistent with the conclusions gathered using the IV estimates. They are consistent with the effects of BMI on mental health for young

adults being at most small and economically insignificant, and they are consistent with economically significant effects for the elderly population.

The paper is organized as follows. We describe the datasets and outline the econometric approaches in sections 2 and 3, respectively. The results are presented and discussed in section 4. Finally, section 5 concludes.

2. Data

We employ data from two sources: the National Longitudinal Study of Adolescent Health (Add Health) and the Health & Retirement Study (HRS). The first data focuses on young adults, while the second focuses on the elderly.

2.1 The National Longitudinal Study of Adolescent Health (Add Health)

Add Health is a nationally-representative sample of 20,745 students in grades 7 through 12 (aged 12-21) in 1994-95 (wave 1). Adolescents were surveyed from 132 schools that were selected to ensure representativeness with respect to region, urbanicity, school size and type, and ethnicity. In wave 1, data were collected from adolescents, their parents, siblings, friends, relationship partners, fellow students, and school administrators. The adolescents have been followed 1 year (wave 2, 1996), 6 years (wave 3, 2001-2002), and 13 years (wave 4, 2008) later.

We examine the relationship between BMI and mental health at wave 4, which corresponds to the wave when genetic data were collected. BMI (weight kg/ height m^2) is based on measurement taken by field interviewers as part of data collection. Mental health is measured using the 10 item CES-D. The CES-D score is created by summing responses (ranging from 0 to 3) from questions that asked respondents how often in the last week they (1) were bothered by things not normally bothersome; (2) could not shake the blues; (3) felt like they were not as good as others; (4) had trouble focusing; (5) were

depressed; (6) were too tired to do things they enjoyed; (7) felt sad; (8) felt happy; (9) enjoyed life, and (10) felt disliked. Hence, the CES-D score has ranges from 0 to 30, with higher values corresponding to poorer mental health. Depression is defined as having a score of 11 or higher (Suglia et al. 2016).

At wave 4 96% of participants consented to providing saliva samples. Approximately 12,200 (80% of those participants) consented to long-term archiving and were consequently eligible for genome-wide genotyping. Genotyping was done on two Illumina platforms, with approximately 80% of the sample genotyping performed with the Illumina Omni1-Quad BeadChip and 20% genotyped with the Illumina Omni2.5-Quad BeadChip. After quality control procedures, genotyped data are available for 9,974 individuals (7,917 from the Omni1 chip and 2,057 from the Omni2 chip) on 609,130 SNPs common across both genotyping platforms. Using this data, Add Health has released PGSs for 9,129 individuals. Of these 9129 individuals, 63% (5728 individuals) are of European ancestry. We concentrate on individuals of European ancestry because the GWAS we employ is for this population, and the PGSs for other ethnic groups may not have the same predictive power (Martin et al. 2017). Our instrument is a PGS for BMI, which is based on the GWAS by Locke et al. (2015). We use as controls a PGS for educational attainment, which is based on the GWAS by Lee et al. (2018), and a PGS for major depressive disorder based on the GWAS by Wray et al. (2018). We do this to increase the plausibility of the exclusion restriction assumption of the IV by controlling for factors that have the potential to be simultaneously related to mental health and the PGS for BMI, such as innate ability and genetic predisposition of depression. All PGSs provided by Add Health are standardized to have a mean of 0 and a standard deviation of 1.

The final sample consists of 4,928 European ancestry respondents with non-missing information on BMI, mental health, the PGSs, and a set of basic control variables (age, gender, birth order, mother's education, picture vocabulary score, PGS for education and PGS for depression).

2.2 Health & Retirement Study (HRS)

The HRS is a nationally-representative longitudinal survey of more than 37,000 individuals in 23,000 households over age 50 in the US. The HRS started in 1992 and data is collected every 2 years on income and wealth, health, cognition, use of health care services; work and retirement, and family connections. The initial HRS cohort consisted of persons born 1931-41 (then aged 51-61) and their spouses of any age. A second study, Asset and Health Dynamics Among the Oldest (AHEAD) was fielded the next year to capture an older birth cohort, those born 1890-1923. In 1998, the 2 studies merged, and, in order to make the sample fully representative of the older US population, two new cohorts were enrolled, the Children of the Depression (CODA), born 1924-1930, and the War babies, born 1942-1947. The HRS now employs a steady state design, replenishing the sample every six years with younger cohorts to continue making it fully representative of the population over age 50.

Although the HRS began in 1992, collection of genetic data only started in 2006. Genotype data on over 19,000 HRS participants was obtained using the Illumina HumanOmni2.5 BeadChips. The HRS recently released publicly available constructed PGSs for 12,090 European-ancestry individuals. We merge this genetic data to the RAND HRS dataset (version p), which is a cleaned and streamlined version of the HRS. We focus on the 2006 wave of the HRS, which consist of 18,469 individuals, as the genetic data was first collected in 2006. The RAND HRS contains a cleaned BMI variable, which is based on self-reported height and weight. Mental health is based on an 8 item CES-D score, which is based on the following questions with "yes/no" response options: much of the time during the last week: (1) I felt depressed; (2) everything I did was an effort; (3) my sleep was restless; (4) I felt lonely; (5) I felt sad; (6) I felt happy; (7) I enjoyed life, and (8) I could not get going. The total number of "yes" responses are summed to calculate the CES-D score. Hence, the range of the CES-D measure in HRS is from 0 to 8, with higher values of the variable corresponding to poorer mental health. Individuals with a CES-D score of 4

or more are classified as being depressed (Steffick 2000). Like in Add Health, our instrument is a PGS for BMI based on the GWAS by Locke et al. (2015), and the PGS for educational attainment is based on the GWAS by Lee et al. (2018). The PGS for depression in the HRS is different from that in Add Health. The HRS provides a PGS for depressive symptoms, which is based on an auxiliary GWAS conducted by Okbay et al. (2016), as part of their subjective wellbeing GWAS. All PGSs are standardized to have a mean of 0 and a standard deviation of 1. Our final sample consist of 8,867 European ancestry individuals with non-missing information on BMI, mental health, the PGSs, and mother's education.

3. Methodology

3.1 OLS and IV Estimation

We first estimate associations between BMI and mental health through OLS regressions where the mental health of individual *i* (MH_i) is modelled as a linear function of BMI (BMI_i), a vector of covariates (X_i), and a stochastic error term (u_i).

$$(1)\ MH_i = \beta_0 + \beta BMI_i + X_i'\delta + u_i$$

Our parameter of interest in this case is β, which represents the association between BMI and mental health. However, this association is likely to be confounded by unobserved factors that are correlated with both BMI and mental health. If no such unobserved factors exist, the causal effect estimated is the population (from which the sample is drawn) average treatment effect. To circumvent the likely influence of unobserved factors (confounders), the first approach we employ is IV estimation using a PGS for BMI as an instrument for BMI[7]. IV estimation is equivalent to two stage least squares. In the first stage

[7] The use of genetic variants as IVs is also known as Mendelian Randomization in the medical literature (von Hinke et al. 2016). The method exploits Mendel's law of independent assortment according to which a trait is inherited independently from other traits at conception.

(equation 2), the BMI of individual *i* is related to the PGS for BMI (PGS_i, which is standardized to have a mean of 0 and standard deviation of 1) and a vector of exogenous control variables (X_i):

$$(2)\ BMI_i = \alpha_0 + \gamma PGS_i + X_i'\psi + v_i$$

To obtain the IV estimate of β, the second stage (equation 3) is performed by regressing mental health on the predicted BMI ($\widehat{BMI_i}$) from equation (2) and a vector of control variables:

$$(3)\ MH_i = c_0 + \beta^{IV}\widehat{BMI_i} + X_i'\mathrm{p} + \epsilon_i$$

Under heterogeneous effects, the IV estimates represent a Local Average Treatment Effect (LATE): the causal effect of the treatment on the outcome for "compliers"—those individuals whose BMI (treatment) is affected by the PGS for BMI (IV)—if the following assumptions hold (Imbens & Angrist, 1994; Angrist et al., 1996): (A1) Random assignment of the instrument; (A2) Non-zero average effect of the instrument on the endogenous variable (BMI); (A3) monotonicity: that the instrument affects the endogenous variable in the same direction for all individuals; and (ER) The exclusion restriction: that the instrument affects the outcome only through its effect on the endogenous variable.

Although genes are randomly inherited at conception conditional on parental genotype, assumption (A1) may fail because of population stratification. Population stratification refers to a situation where the distribution of genes systematically differs by population subgroups (e.g. by ethnicity/race). If these subpopulations also systematically have different health outcomes that are not due to genetic make-up, then this could lead to a spurious correlation between genetic risk and health. Population stratification can be controlled for by limiting analyses to ethnically homogenous samples (Cardon & Palmer 2003) and by including principal components from genome-wide SNP data as control variables, which account for genetic differences across ethnic groups (Price et al. 2006). We include the first 20 (10, respectively) principal components when using the Add Health (HRS) data to control for

population stratification and limit our analyses to individuals of European-ancestry[8]. Therefore, our analysis assumes the PGS is exogenous conditional on population stratification. Assumption (A2) can be empirically verified. Previous studies (e.g., Willage, 2018) have shown that the PGS for BMI is statistically significantly associated with BMI, and we show this is also the case in our samples in the next section. Assumption (A3) requires that increasing the number of risk alleles for an individual increases the exposure (BMI) or leaves it constant, but does not decrease it. In other words, this means that an individual who has genes related to BMI should have at least as high a BMI compared to if he/she did not have those genes. This assumption is untestable, but von Hinke et al. (2016) argue that it is likely to hold because genes are randomly assigned and individuals do not know their genotype, so they cannot act on knowledge of their genes in a way so as to violate the monotonicity assumption (e.g., being careless about their BMI if they do not have the gene and being over-reactive otherwise). Moreover, von Hinke et al. (2016) note that although the counterfactual is unobserved, studies have shown, at the population level, that individuals who have risk alleles have a higher BMI than those who do not, which is consistent with (A3).

The ER assumption requires that the PGS for BMI affects mental health only through its effect on BMI. One important reason why the ER may be violated is because of pleiotropy (von Hinke et al., 2016; van Kippersluis and Rietveld, 2017; DiPrete et al, 2018). Genes have multiple functions (pleiotropy), and genes related to BMI may be related to other traits (e.g. smoking, education) that also affect mental health. A gene could also be related to another gene (through linkage disequilibrium) that directly affects mental health[9]. For example, if the FTO gene (that is associated with obesity) is also related to

[8] The HRS only provides the first 10 principal components of the genetic data, whereas Add Health provides the first 20 principal components.

[9] Linkage disequilibrium is the non-random association of alleles at different loci in a given population.

intelligence, which affects mental health, then the ER would be violated. Locke et al. (2015) manually reviewed the literature related to 405 genes within 500 kb and a r^2>0.2. On a broad level, they found that genes effect BMI through hypothalamic function, energy homeostasis, and neuronal transmission and development. However, Locke et al. (2015) do report that some genes are associated with schizophrenia, smoking, and type 2 diabetes[10]. If these traits also affect mental health, then they represent paths—independent of BMI—through which the ER is violated. Our first strategy to deal with pleiotropy is quite straightforward. Two broad sources—other than BMI—through which the BMI PGS might affect mental health are through its relation to other genes that affect mental health or intelligence/cognitive ability. We use a PGS for depression to control for the genetic risk of having poor mental health, and a PGS for educational attainment to control for innate ability. The PGSs for mental health and educational attainment are correlated with the BMI PGS and mental health[11]. Hence, those two PGSs may also be related to mental health unobserved factors, in which case the ER assumption would be violated if they were not included as controls. This first strategy we employ is in the same spirit as the recent work by DiPrete et al. (2018), although it was not motivated by it.

3.2 Nevo & Rosen (2012) Bounds

The second strategy we employ to deal with potential violations of the ER—from pleiotropy or other sources—consists on estimating an upper bound for the corresponding population effect using the BMI PGS as a potentially "imperfect IV". Nevo & Rosen (2012) derive bounds on the average treatment

[10] The genes rs38888190, rs7903146, rs1558902, rs2176040, rs1558902 are associated with type 2 diabetes. rs11191560 is associated with schizophrenia, and rs11030104 is associated with smoking initiation. See Locke et al. (2015) supplementary table 17B.

[11] In Add Health: (1) the correlation (standard error) between the BMI and education PGSs is -0.1731 (0.014), and (2) the correlation (standard error) between the BMI and depression PGSs is 0.0322 (0.014). In the HRS: (1) the correlation (standard error) between the BMI and education PGSs is -0.1707 (0.011), and (2) the correlation (standard error) between the BMI and depression PGSs is -0.0152 (0.011). All correlations are statistically significant except for the correlation between the BMI and depression PGSs in the HRS.

effect in the original OLS model (β equation in (1)) using an IV that may or may not satisfy the ER. Instead, they maintain an assumption about the direction of the correlation between (i) the instrument (PGS for BMI) and the error term representing the outcome's (mental health's) unobserved factors, and (ii) the endogenous variable (BMI) and the same error term. Specifically, let ρ_{ab} denote the correlation between any two variables *A* and *B*, σ_{ab} denote their covariance, and σ_a denote the standard deviation of *A*. Nevo & Rosen (2012) assume that: (A4) $\rho_{wu}\rho_{zu} \geq 0$ (their "assumption 3"); that is, that the instrument (z) has (weakly) the same direction of correlation with the error term u as the endogenous variable w (e.g. BMI). In our context, we presume that the leading mental health unobserved factors present in the error term are unobserved health endowments and environmental factors (e.g., unobserved family environment). It is also important to keep in mind that lower values of the CESD score measure better mental health (the same applies to the depression indicator). Thus, "better" mental health unobservables lead to lower values of CES-D and, correspondingly, lower values of the error term. Then, since we would expect individuals with better unobserved health endowments and/or environmental factors (and thus lower values of the error term) to have a lower BMI on average, we expect that BMI is positively correlated with the error term. Similarly, we also expect our instrument—genetic risk for high BMI—to be positively correlated with the error term under a similar reasoning. Nevo & Rosen (2012) show that if the instrument is positively correlated with the endogenous variable ($\sigma_{zw} > 0$), and if the instrument and endogenous variable are positively correlated with the error term ($\sigma_{wu}, \sigma_{zu} \geq 0$), then their bounds on β only provide an upper bound:

$$(4) \quad \beta \leq \min\{\beta^{OLS}, \beta_z^{IV}\}$$

where β^{OLS} and β_z^{IV} denote the probability limits of the standard OLS and IV estimators for β, respectively.

The bounds in (4) can be tightened by assuming that: (A5) the instrument (PGS for BMI) is less correlated with the error term in equation (1) than is the endogenous variable (BMI), i.e., $|\rho_{wu}| \geq |\rho_{zu}|$ ("assumption 4" in Nevo & Rosen 2012). Although this assumption is untestable, we believe that it is likely to hold in our context for two main reasons. First, our instrument comes from a genetic lottery, whereas BMI is a choice variable that is likely to be affected by many unobserved factors from birth until BMI is measured. Second, our instrument could be correlated with the error term because it could be correlated with other unobserved genes associated to mental health. However, we control for PGSs for education and depression, which should imply that our instrument is less correlated with the error term than if we did not control for these PGSs, making this assumption more plausible.

The two previous assumptions give the definition of an "imperfect IV" in Nevo & Rosen (2012), as an IV that has the same direction of correlation with the unobserved error term as the "treatment" variable of interest (in our case BMI), but is less endogenous than the treatment variable. They define the function $V(\lambda) = \sigma_w Z - \lambda \sigma_z W$, which, when evaluated at $\lambda^* = \frac{\rho_{zu}}{\rho_{wu}}$, generates a variable that is uncorrelated with the error term, and thus serves as a valid instrument. Although λ^* is unknown, it is bounded between 0 and 1 under the assumptions above. Nevo & Rosen (2012) employ the bounds on λ^* to bound β in equation (1).

For the linear regression model, Nevo & Rosen (2012) use $V(1) = \sigma_w Z - 1\sigma_z W$ as the instrument. As it was the case before, if the instrument is positively correlated with the endogenous variable, and if the instrument and endogenous variables are positively correlated with the error term, then the Nevo & Rosen (2012) bounds on β in equation (1) only provide an upper bound. This upper bound is given by:

$$(5)\ \beta \leq \min\left\{\beta^{IV}_{V(1)}, \beta^{IV}_{Z}\right\}$$

where $\beta_{V(1)}^{IV}$ is the probability limit of the standard IV estimator for β when $V(1)$ is used as an instrument for w.

Being able to estimate an upper bound on β in equation (1) is useful. The estimated upper bound provides a benchmark magnitude for the population average effect while relaxing the potentially troublesome ER. This magnitude can be directly compared to the estimates obtained from OLS, which identifies the same effect under strong conditions. It is also important to keep in mind that the effect that is bounded differs from the LATE identified by IV methods under the validity of the exclusion restriction. Therefore, it is misleading to interpret differences in the two approaches as violations of the ER, since the differences may be due to effect heterogeneity. Still, the comparison of the estimated upper bound with the estimated LATE can be informative about treatment effect heterogeneity if the ER assumption is maintained. Lastly, the upper bound allows ruling out plausible magnitudes for the true effect if they fall outside the bound and corresponding confidence interval. Unfortunately, in our setting the estimated upper bound is not useful for ruling out zero effects.

4. Results

Table 1 provides summary statistics for Add Health and HRS respondents. In Add Health, the average age at wave 4 is 28.94 years, and over half (54%) of respondents are female. The average BMI is 28.56 kg/m^2. The mean CES-D score is 5.79, and 15% are classified as being depressed. Examining the summary statistics by gender reveals that women have substantially a higher CES-D score and incidence of depression, but there are no differences in BMI. The HRS respondents are 68.18 years old on average in 2006 and 58% are women. Average BMI is 27.69 kg/m^2. The average CES-D score is 1.26, and 12%

are depressed[12]. In comparison to young adults, there are large gender differences in both BMI and mental health. On average, women have a slightly lower BMI but much higher CES-D scores and incidence of depression compared to men.

4.1 OLS Estimates

The OLS estimates for young adults are presented in Table 2 columns 1-3. Panel A shows estimates for the CES-D score and panel B for the depression indicator. All regressions control for a PGS for depression, PGS for educational attainment, age, age squared, gender, birth order, mother's education, picture vocabulary score, and the first 20 ancestry-specific principal components of the genetic data. There are three important points. First, the associations between BMI and mental health are relatively small. Column 1 shows that there is a small statistically significant positive association between BMI and the CES-D score. The coefficient on BMI indicates that a 5 kg/m^2 increase in BMI is associated with a 0.14 unit increase in the CES-D score[13,14]. This represents an increase of 2.4%, given the mean CES-D score of 5.79. This means that BMI would have to increase by 10 kg/m^2 (e.g. going from a BMI of 20 and being classified as "normal weight" to a BMI of 30 and being classified as "obese") for the CES-D score to increase by 5%. Similarly, the OLS estimate in panel B column 1 shows that a 5 kg/m^2 increase in BMI is associated with a 1 percentage point increase (6.67% off the mean) in the likelihood of depression. Second, there are statistically significant differences by gender. There is a positive association between BMI and the CES-D score for women. A 5 kg/m^2 increase in BMI is associated with

[12] Recall that the scale of the CES-D score in Add Health and HRS differ, with the first going from 0 to 30 and the second from 0 to 8.

[13] A 5 kg/m^2 increase in BMI would move an individual's classification from overweight to obese. For example, for a man (women, respectively) with the average height for men (women) of 5 foot 9 inches (5 foot 3 inches), the weight corresponding to a 25 BMI would be 169 lbs (141 lbs). For him (her), a 5 kg/m^2 increase in BMI would correspond to a 43 lbs (28 lbs) increase in weight. The average height of men and women are taken from Fryar et al. (2018).

[14] According to the Centers for Disease Control and Prevention, an adult person is classified as being underweight if his/her BMI is less than 18.5, of normal weight if his/her BMI is in the interval [18.5, 25), overweight if his/her BMI is in the interval [25, 30), and obese if his/her BMI is 30 or higher. (https://www.cdc.gov/obesity/adult/defining.html).

a 0.31 unit (5%) increase in the CES-D score. In contrast, for men a 5 kg/m^2 increase in BMI is associated with a decrease in the CES-D score by 0.014 units (2.5%). Third, the estimates show that genetic predisposition has a large impact on mental health, but perhaps surprisingly, innate ability (as measured by the education PGS) is not significantly associated with mental health. Specifically, a 1 standard deviation increase in the PGS for depression is associated with a 0.359 unit (6.2%) increase in the CES-D score, and a 2.3 percentage point (15.33%) increase in the incidence of depression. In contrast, the education PGS is negatively associated with mental health, but it is statistically insignificant.

Estimates for elderly individuals from the HRS are given in Table 2 columns 4-6. The regressions control for a PGS for depression, a PGS for educational attainment, gender, age, age squared, mother's education, and the first 10 ancestry-specific principal components of the genome wide data. The association between BMI and mental health is larger for the elderly than for young adults. For example, in column 4, a 5 kg/m^2 increase in BMI is associated with a 0.145 unit increase in the CES-D score, and a 1.5 percentage point increase in the probability of depression. This is equivalent to a 11.5% and 12.5% increase in CES-D score and incidence of depression, respectively, relative to the corresponding sample means. The association between BMI and mental health does not differ substantially between elderly men and women. Estimates in columns 5 and 6 show that a 5 kg/m^2 increase in BMI is associated with a 10.49% (12.62%) increase in the CES-D score and a 14.29% (16.67%) increase in the probability of depression score for women (men). Finally, it is interesting to note that compared to young adults (i) the impact of genetic predisposition for depression in old age is larger, and (ii) innate ability as measured by the education PGS is significantly associated with better mental health for the elderly (except for men in panel B). For example, the coefficient on the depression (education) PGS in column 4 of panel A indicates

that, a 1 standard deviation increase in the PGS for depression (education) is associated with a 10.31% (6.74%) increase (decrease) in the CES-D score relative to the sample mean[15].

4.2 IV Estimates

The OLS estimates of β could be biased because of unobserved factors that affect both BMI and mental health. We now turn to IV estimates, using the PGS for BMI as the instrument. Recall that IV estimates identify a different parameter from OLS. Under heterogeneous effects, IV estimates represent a LATE for individuals whose treatment (BMI) is affected by the instrument (PGS for BMI).

Table 3 provides the first stage estimates. Column 1 presents results from an OLS regression of BMI on PGS for BMI and controls. It shows that a 1 standard deviation increase in the PGS for BMI is associated with a 1.803 (1.403) kg/m^2 increase in the BMI of young adults (elderly individuals, respectively), with a first stage F-statistic of 289 (529). The PGS for BMI thus satisfies assumption A2 (non-zero first stage effect). Columns 2-6 of Table 3 give estimates from conditional quantile regressions of BMI on PGS for BMI and controls. These estimates serve a number of purposes. First, they allow us to learn about the heterogeneity of the effect of our instrument (PGS for BMI) on our treatment variable (BMI). Second, they can provide indirect evidence about the monotonicity assumption (A3). Lastly, they provide our first suggestive evidence to characterize the effect that the IV estimates identify in our application. We observe that the effect of the PGS for BMI on all the selected quantiles of the conditional BMI distribution is positive and highly statistically significant. This is consistent with assumption A3, which states that the effect of the BMI PGS on BMI is non-negative for all individuals, thus providing

[15] The OLS estimates without the depression and education PGS controls are similar to those in Table 2 (see appendix table A1).

indirect evidence in its favor (although not implying it holds, as the assumption is not directly testable)[16]. Also observed is a pattern of increasing conditional quantile estimates as one moves to higher quantiles, indicating that the IV has stronger effects on higher quantiles of the conditional BMI distribution.

To shed additional light on how to interpret the effect identified by the IV estimates in our setting, we conduct an exercise similar in spirit to that in Angrist & Imbens (1995). In a setting with a binary IV and a treatment with variable intensity, they show that IV estimates identify a weighted average of treatment effects for those individuals whose treatment intensity is affected by the instrument (the compliers), where the weights are proportional to the (unconditional) quantile treatment effects (QTEs) of the IV on the treatment[17]. Intuitively, such weights allow to characterize the parameter identified by IV estimates, since the effects for those individuals (or compliers) with larger weights contribute more to the effect that is estimated with the binary IV. To implement this exercise using our continuous IV, we create a binary IV equal to 1 if the (standardized) PGS for BMI is greater than 0. We then estimate unconditional QTEs using the approach by Firpo (2007). First, we estimate a propensity score by running a logit regression of the IV on the control variables. Next, we perform a weighted quantile regression using the propensity score to form individual weights[18]. The unconditional QTEs for Add Health and the HRS are shown in Figure 1. The unconditional QTEs are clearly larger at higher quantiles. For example, in the HRS there is a 2.20 kg/m^2 difference in BMI between treated (PGS for BMI>0) and

[16] In particular, note that the monotonicity assumption in A3 is imposed on the individual-level effect of the PGS for BMI on BMI, while quantile regression provides effects on the quantiles of the conditional distribution of BMI (rather than on the distribution of individual effects).

[17] Technically, the weights defined in Angrist & Imbens (1995) are proportional to the difference between the CDF of the potential treatment intensities in the absence of the IV (i.e., the treatment intensities that individuals would receive if Z=0) and the CDF of the potential treatment intensities under receipt of the IV (i.e., the treatment intensities that individuals would receive if Z=1). The result in Angrist & Imbens (1995) can be written in such a way that the weights are defined in terms of the difference in quantiles of those CDFs, which equal the unconditional QTEs we employ here.

[18] The weights=(z/phat) + ((1-z)/(1-phat)) where z is the binary instrument and phat is the predicted probability from the estimated propensity score model.

untreated individuals at the 60th percentile of the BMI distribution. At the 70th (80th) percentile the difference in BMI is 2.60 kg/m^2 (2.80 kg/m^2). This suggests that the groups of individuals that contribute the most to the IV estimate are those in the upper quantiles of the BMI distribution for both young adults and the elderly, implying that for both demographic groups the IV estimate reflects the effect of BMI on mental health mostly for individuals in the upper quantiles of the BMI distribution. It is important to keep this information present when the IV estimates are interpreted, since they likely identify the effect—under the IV assumptions—for a subpopulation in which individuals on the upper quantiles of the BMI distribution have considerably more weight.

The IV estimates are presented in Table 4. All the IV estimates for the effect of BMI for young adults are statistically insignificant and relatively small in magnitude, suggesting that there is no causal relationship between BMI and mental health. In comparison, the IV estimates for elderly individuals are statistically significant and large in magnitude. For example, IV estimates in column 4 indicate that a 5 kg/m^2 increase in BMI is associated with a 0.255 (20%) increase in the CES-D score, and a 3.57 percentage point (29%) increase in the probability of depression. There are no statistically significant gender differences. For completeness, appendix Table 2 presents IV estimates using a dummy variable equal to 1 if the BMI PGS is greater than 0 as the instrument. The IV estimates using the binary instrument are not substantially different from those in Table 4.

It is important to note that the IV estimates appear to be overestimated when models do not control for the education and depression PGSs, which is consistent with DiPrete et al.'s (2018) argument that the inclusion of the outcome's PGS may reduce the consequences of pleiotropy. More concretely, the change in the IV estimates occurs because the education and depression PGSs are correlated with the BMI PGS (see footnote 11), which means that the instrument picks up some of the effects of these two PGSs when they are omitted. This point is illustrated in Table 5, which presents IV estimates without

controlling for the education and depression PGSs. The IV estimates are considerably larger in the HRS without adjusting for the education and depression PGSs, although the differences in coefficients are not statistically significant. For example, the effect of BMI on CES-D score (depression) when controlling for the PGSs in Table 4 column 4 is 0.051 (0.007). In contrast, the effect of BMI on CES-D (depression) when not controlling for them in Table 5 column 4 increases to 0.065 (0.009), giving a percentage increase of 27% (29%). In Add Health, the IV estimates appear overestimated to a lesser extent. The effect of BMI on CES-D score when adjusting for the education and depression PGSs is 0.020 in Table 4 column 1. The corresponding IV estimate without controlling for the education and depression PGSs in Table 5 column 1 is 0.024, giving a percentage increase of 20%. However, the IV estimates for depression in Tables 4 and 5 are the same.

Finally, we note that even with the inclusion of the depression and education PGSs, the BMI PGS may still violate the ER assumption through other channels. To complement the evidence presented thus far, we employ the alternative approach of obtaining an upper bound for the average effect of BMI on mental health for the corresponding population.

4.3 Nevo & Rosen (2012) Bound Estimates

In this subsection, we present evidence about the effect for the population parameter β in equation (1) by using the BMI PGS as a potentially imperfect IV based on the approach by Nevo & Rosen discussed in section 3.2. Recall, that since (i) our IV (BMI PGS) is positively correlated with the endogenous variable (BMI), and (ii) the IV and endogenous variable are assumed to be positively correlated with the error term in equation (1) (see discussion in section 3.2), the Nevo & Rosen (2012) procedure yields an upper bound on β.

Results for the CES-D score are presented in Table 6. Panel A shows results for young adults in Add Health. Columns 2 and 3 give the OLS and IV estimates from Tables 2 and 4. Column 4 shows the IV estimate using $V(1) = \sigma_{BMI}PGS\_BMI - 1\sigma_{PGS\_BMI}BMI$ as the instrument, which is an intermediate step in the Nevo & Rosen procedure. The estimated upper bound (the min of the OLS and IV estimates; see equation (4)) under (A4)—that the instrument has weakly the same direction of correlation with the error term as the endogenous variable—is given in column 5. The estimated upper bound, while not able to rule out a null effect, indicates that the largest value of the average effect of BMI on the CES-D score for the population of young adults (β) is 0.02 (the IV estimate). Based on this largest possible value for the effect, a 5 kg/m^2 increase in BMI would be associated with at most a 0.1 unit (1.7%) increase in the CES-D score. The upper endpoint of the 95% confidence interval on the bounded parameter is 0.052[19]. Looking at the OLS estimate, we see that it falls within the 95% confidence interval of the bounded parameter. Since they both undertake statistical inference on the same parameter, this implies that we cannot statistically rule out that the OLS estimate is unbiased for the true population average effect (β). The estimated upper bounds also suggest there could be a gender difference in the effect of BMI on the CES-D score among young adults. In particular, for women, the estimated upper bound, which corresponds to the IV estimate, is consistent with a 5 kg/m^2 increase in BMI increasing the CES-D score by at most 1.5%. For men, the estimated upper bound, which corresponds to the OLS estimate, is negative, although the 95% confidence interval on the bounded parameter straddles zero, implying that a null effect is possible. Column 6 gives the estimated upper bounds (the min of the IV and IV(1) estimates; see equation 5) when we add (A5)—that the instrument is less correlated with the error term

[19] The estimation of the upper bound and the construction of confidence intervals on the bounded parameter in the Nevo & Rosen approach necessitates a non-standard procedure given that the bounds' expression contains a minimum operator. The stata program we use (imperfectiv) follows Nevo & Rosen (2012), who employ a variant of the Chernozhukov et al. (2013) method.

than the endogenous variable. These estimated upper bounds turn out to be not much different from those in column 5 for young adults, the main difference being that the 95% confidence intervals on the bounded parameter are somewhat larger.

Results for elderly individuals in the HRS are given in panel B of Table 6. The estimated upper bound on the average effect for the elderly population in column 5 corresponds to the OLS estimate (0.029), implying that a 5 kg/m^2 increase in BMI would be associated with, at most, an 11.5% increase in the CES-D score. Thus, this estimated upper bound provides valuable information, as it rules out plausible larger effects. Compared to the IV estimate (0.051), the estimated upper bound is about half as large. Indeed, looking at the 95% confidence interval (on the bounded parameter) of 0.037, the IV point estimate is considerably larger. As mentioned before, this difference has implications for the ER assumption only under the overly restrictive assumption of homogeneous effects. Instead, under heterogeneous effects and maintaining the ER assumption, the implication is that the average effect of BMI on the CES-D score for compliers (within the elderly population) is larger than the average effect on the elderly population. This interpretation is consistent with the previously presented evidence that the IV estimate likely captures the effect mostly for individuals at the upper part of the BMI distribution, who likely experience larger effects of their BMI on their CES-D score. Contrary to the young adult population, the estimated upper bounds on the corresponding average effect for elderly women and men are very similar. Lastly, the estimated upper bounds that add (A5), shown in column 6, yield very similar conclusions to the results in column 5, although with the estimated upper bounds in column 6 being somewhat smaller.

Results for depression are shown in Table 7. The patterns observed are similar to those documented in Table 6, although the differences in estimated upper bounds between young adults and the elderly are slightly less pronounced. Looking at the estimated bounds under (A4) in column 5, the

largest possible value for the (population) effect indicates that a 5 kg/m^2 increase in BMI would be associated with at most a 6.6% increase in the probability of experiencing depression for the population of young adults. For the elderly population, the estimated upper bound is related to a corresponding increase of 12.5% in the probability of experiencing depression. Importantly, as before, these estimated upper bounds are able to rule out plausible larger effects.

Overall, the results from employing the Nevo & Rosen (2012) bounds—which yield an upper bound on β in this setting—can be summarized as follows. First, the estimated bounds for the population of young adults reinforce the IV-based finding that their effect of BMI on mental health is at most small and arguably economically insignificant. Second, the estimated upper bounds are suggestive of economically significant effects of BMI on mental health for the elderly population. Third, for the elderly population, when we compare the estimated upper bound on β with the IV estimate that maintains the ER assumption (and identifies a LATE), the former is about half the magnitude of the latter. If effects are heterogeneous, this is consistent with the IV estimate reflecting the effect for individuals (or compliers) mostly from the upper part of the BMI distribution, as documented in section 4.2, who likely experience larger effects of BMI on mental health. Fourth, since in general the OLS estimates are below the estimated upper bounds (and/or below the 95% confidence intervals on the bounded parameter), we cannot rule out the possibility that the former estimates are unbiased for the population average effect of BMI on mental health. Lastly, the estimated upper bounds for both young adults and the elderly are able to rule out plausible effects, and are thus informative.

4.4 Other Analyses

4.4.1 Obesity

Our main analyses focus on BMI rather than obesity (BMI$\geq$30) because the instrument affects mental health through variation in BMI across the BMI distribution, not just incidence in obesity.

Nevertheless, we have analyzed the relationship between obesity and mental health in Appendix Table A3. The results for obesity are qualitatively similar to the BMI results. For young adults, the IV estimates show no statistically significant relationship between obesity and mental health. For elderly individuals, the IV estimates indicate that there is a large positive relationship between obesity and mental health. Obese individuals have a higher CES-D score by 0.705 units (56% relative to the sample mean), and are 10.2 percentage points (85% relative to the sample mean) more likely to be depressed than non-obese individuals. From the estimated Nevo & Rosen (2012) upper bounds shown in columns 3 and 6, the patterns already documented emerge. Noteworthy is that, when considering obesity, the difference between the estimated upper bounds and the IV estimates is generally larger, which is in line with the notion that the effect identified by the IV estimates presented in section 4.2 capture the effect of BMI on mental health mostly for individuals from the upper part of the BMI distribution, since here the treatment is defined as having a more extreme value of BMI ($\geq$30).

4.4.2 Reverse Causality and Mortality Attrition

We have also examined whether the findings for elderly individuals are influenced by (i) reverse causality, and (ii) mortality attrition. To assess the influence of reverse causality, we estimated the effect of mental health on BMI using the PGS for depression as the IV. Unlike the PGS for BMI, which predicts both BMI (first-stage effect) and mental health (reduced-form effect), the PGS for depression only predicts mental health (first-stage effect). Appendix Table A4 shows OLS, IV, and upper bound estimates using the HRS sample. While OLS estimates are positive and statistically significant, IV estimates are statistically insignificant. The estimated upper bounds are negative with the 95% confidence interval on the bounded parameter straddling zero. The IV estimates and estimated upper bounds provide suggestive evidence that reverse causality is unlikely to be driving our results. A second

potential concern is that our estimates could be affected by mortality attrition, as genetic data is only available for HRS respondents who have survived until 2006. As a robustness check, we weighted the IV regressions with the inverse of the probability of living until 2006. The probabilities were the predicted values from a logit regression where the dependent variable was an indicator for being alive in 2006 on basic demographics (year of birth, gender), health (ever smoked, ever had diabetes, ever had heart disease, average BMI up to 2004, average CES-D score up to 2004, average self-reported health up to 2004), and socioeconomic variables (years of education, mother's education). As shown in Appendix Table A5, the IV estimates are qualitatively similar to those presented in Table 4. The one exception is that the effect of BMI on depression for women becomes statistically insignificant.

4.4.3 Exploration of Some Mechanisms

As discussed in the introduction, BMI is hypothesized to affect mental health through various channels, including health concerns and functional limitations (Himes 2000; Markowitz et al. 2008; Paolucci et al. 2018). A useful discussion of different mechanisms through which BMI could affect mental health is provided in Willage (2018). Appendix Table A6 presents estimates (OLS, IV and bounds) of the effect of BMI on the following self-reported outcomes for the elderly: (i) probability of being in very good/excellent health, (ii) probability of engaging in vigorous exercise more than once per week, and (iii) probability of reporting that health limits at least 1 daily activity. Note that, as the outcome changes in these exercises, the conditions to obtain a lower or upper bound using the Nevo & Rosen (2012) approach may change. The rationale for justifying the direction of correlation between the error term corresponding to each of the outcomes in turn and the endogenous variable (BMI) and instrument (PGS for BMI) follows the same logic outlined in section 3.2. In panel A, we find evidence that BMI is linked to poor self-reported health. The IV estimate in column 3 suggest that a 5 kg/m^2 increase in BMI is related

to a 9.5 percentage point (20%) decrease in the likelihood of reporting being in very good or excellent health. The estimated lower bound in column 4 (using (A4) and (A5)) is somewhat smaller (in absolute value) than the IV estimate. Unfortunately, the estimated lower bound does not allow ruling out a null effect since it is negative, but it allows ruling out effects from a 5 kg/m^2 increase in BMI greater than 16%. This estimated effect is present in both women and men. In panel B, we see that, according to the IV estimate, a 5 kg/m^2 increase in BMI is related to a 6.5 percentage point (25%) decrease in the likelihood of engaging in vigorous exercise more than once per week. The estimated lower bound is smaller than the IV estimate (in absolute value), and cannot be used to rule out a null effect. This estimated effect is also present in both women and men. Looking at panel C, the IV estimate suggests that a 5 kg/m^2 increase in BMI is related to a 5 percentage point (42%) increase in the likelihood of reporting that health limits at least 1 daily activity. For this outcome, we obtain an upper bound, which is estimated to be just slightly smaller than the IV estimated effect. According to the estimated upper bound, the effect is at most 38% for the elderly population. Based on these results, it appears that health concerns are plausible mechanisms through which BMI has an impact on mental health for elderly individuals.

5. Conclusion

We examine the relationship between BMI and mental health for young adults in Add Health and elderly individuals in the HRS. To account for confounding due to unobserved factors and reverse causality, we first take the same approach as previous studies (Hung et al. 2014; Jokel et al. 2012; Tyrrell et al. 2018; Walter et al. 2015; Willage 2018) and use a PGS for BMI as an IV. However, unlike the previous studies, our PGS is more powerful because it is based on the more recent GWAS by Locke et al. (2015) and includes all SNPs, rather than just genome-wide significant SNPs. Moreover, we attempt to address the possible violation of the exclusion restriction (ER) assumption due to pleiotropy by using

PGSs for depression and education to control for the genetic predisposition to poor mental health and innate ability. We also contribute to this literature by analyzing effects of the IV on the BMI quantiles. We document that IV estimates, which identify a weighted average of effects for individuals whose BMI is affected by their BMI PGS (the compliers), are likely to capture the effect of BMI on mental health mostly for individuals in the upper quantiles of the BMI distribution. As a second approach to account for unobserved confounding and reverse causality, we estimate an upper bound on the average treatment effect for the corresponding population (i.e., on the parameter identified by OLS, β in equation (1)). We use the approach of Nevo & Rosen (2012), which has not been used in this context as far as we are aware. The attractiveness of this approach is to employ the PGS for BMI as a potentially "imperfect" instrument that may not satisfy the ER assumption, instead imposing relatively weaker conditions about the direction of certain correlations.

For young adults, the IV estimates indicate that there is no statistically significant relationship between BMI and mental health. The estimated upper bounds are consistent with the notion that the effect of BMI on mental health for young adults is at most small and economically insignificant. The results for young adults are consistent with those in Willage (2018), who also used Add Health, but with a smaller sample and less powerful PGS. In contrast, the IV estimates for elderly individuals are statistically and economically significant. They indicate that a 5kg/m^2 increase in BMI (a difference equivalent to moving from overweight to obese) is related to a 20% increase in the CES-D score, and a 29% increase in the likelihood of depression. We find no statistically significant gender differences in the IV estimates. The estimated upper bounds are consistent with the IV estimates in that they are also suggestive of large and economically significant effects of BMI on mental health for the elderly population. These results do not support the "jolly fat" hypotheses (Crisp et al., 1980; Palinkas et al., 1996), which argues that BMI

has a protective effect on mental health for elderly individuals. Interestingly, for the elderly population, the IV estimates (which identify a weighted average of effects for compliers) are about twice as large as our estimated upper bounds on β. Under heterogeneous effects, this speaks to the notion that the effect for compliers is larger than the effect for the overall population. This same notion is supported by our estimates of the quantile treatment effects of the PGS for BMI on BMI, which suggest that the IV estimates identify the effect of BMI on mental health mostly for individuals from the upper part of the BMI distribution, who thus are likely to experience large effects of BMI on mental health.

Overall, our results suggest that policy interventions aimed at reducing obesity (e.g. the healthy food financing initiative) could have indirect benefits of improving mental health among the elderly. To put our estimates in perspective, Baicker et al. (2013) report that the Oregon Health Insurance Experiment (which randomly provided low-income uninsured adults access to Medicaid) reduced the probability of depression by 30%. Among individuals between age 50 and 64, this effect is larger at 58%[20]. Our preferred results indicate that a reduction of 5kg/m^2 in BMI—admittedly a substantial reduction—could have an impact on the incidence of depression of between one-seventh (based on our bounds estimates) and one-half (based on our IV estimates) of the effect of having access to health insurance.

There are some limitations with our paper. First, we focus on European ancestry individuals because GWASs are mostly conducted on European descent populations. This means that PGSs for other ethnic groups may not have the same predictive power (Martin et al. 2017). We also do not look at the effect of BMI on mental health in midlife, which may differ from the effect found for young adults and elderly individuals.

[20] Author's own calculations with the public use data provided by Baicker et al. (2013), using their same specification to obtain LATE effects. The percentage effect is relative to the corresponding control mean.

References

Angrist, J. D., & Imbens, G. W. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association*, *90*(430), 431-442.

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, *91*(434), 444-455.

Baicker, K., Taubman, S. L., Allen, H. L., Bernstein, M., Gruber, J. H., Newhouse, J. P., ... & Finkelstein, A. N. (2013). The Oregon experiment—effects of Medicaid on clinical outcomes. *New England Journal of Medicine*, *368*(18), 1713-1722.

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

Brunello, G., & d'Hombres, B. (2007). Does body weight affect wages?: Evidence from Europe. *Economics & Human Biology*, *5*(1), 1-19.

Cardon, L. R., & Palmer, L. J. (2003). Population stratification and spurious allelic association. *The Lancet*, *361*(9357), 598-604.

Carpenter, K. M., Hasin, D. S., Allison, D. B., & Faith, M. S. (2000). Relationships between obesity and DSM-IV major depressive disorder, suicide ideation, and suicide attempts: results from a general population study. *American Journal of Public Health*, *90*(2), 251.

Chernozhukov, V., Lee, S., & Rosen, A. (2013). Intersection bounds: Estimation and inference. *Econometrica*, *81*(2), 667-737.

Chesney, E., Goodwin, G. M., & Fazel, S. (2014). Risks of all-cause and suicide mortality in mental disorders: a meta-review. *World Psychiatry*, *13*(2), 153-160.

Crisp, A. H., Queenan, M., Sittampaln, Y., & Harris, G. (1980). 'Jolly fat' revisited. *Journal of Psychosomatic Research*, *24*(5), 233-241.

DiPrete, T., Burik, CAP & Koellinger, PD. (2018). Genetic Instrumental Variable Regression: Explaining Socioeconomic and Health Outcomes in Nonexperimental Data. *Proceedings of the National Academy of Sciences*, *115*, E4970-E4979.

Firpo, S. (2007). Efficient semiparametric estimation of quantile treatment effects. *Econometrica*, *75*(1), 259-276.

Fryar CD, Kruszon-Moran D, Gu Q, Ogden CL (2018). Mean body weight, height, waist circumference, and body mass index among adults: United States, 1999-2000 through 2015-2016. National Health Statistics Reports; no 122. Hyattsville, MD: National Center for Health Statistics.

Gregor, M. F., & Hotamisligil, G. S. (2011). Inflammatory mechanisms in obesity. *Annual Review of Immunology*, *29*, 415-445.

Greve, Jane. "Obesity and labor market outcomes in Denmark." *Economics & Human Biology* 6, no. 3 (2008): 350-362.

Ha, H., Han, C., & Kim, B. (2017). Can Obesity Cause Depression? A Pseudo-panel Analysis. *Journal of Preventive Medicine and Public Health*, *50*(4), 262.

Himes, C. L. (2000). Obesity, disease, and functional limitation in later life. *Demography*, *37*(1), 73-82.

Hung, C. F., Rivera, M., Craddock, N., Owen, M. J., Gill, M., Korszun, A., ... & Rietschel, M. (2014). Relationship between obesity and the risk of clinically significant depression: Mendelian randomisation study. *The British Journal of Psychiatry*, *205*(1), 24-28.

Imbens, G., & Angrist, J. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica, 62(2)*, 46475.

Jokela, M., Elovainio, M., Keltikangas-Järvinen, L., Batty, G. D., Hintsanen, M., Seppälä, I., ... & Kivimäki, M. (2012). Body mass index and depressive symptoms: instrumental-variables regression with genetic risk score. *Genes, Brain and Behavior*, *11*(8), 942-948.

Kessler, R.C., Mickelson, K. D., & Williams, D. R. (1999). The prevalence, distribution, and mental health correlates of perceived discrimination in the United States. *Journal of Health and Social Behavior*, *40*, 208-230.

Kochanek KD, Murphy SL, Xu JQ, and E. Arias (2017). Mortality in the United States, 2016. NCHS Data Brief, no 293. Hyattsville, MD: National Center for Health Statistics.

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., ... & Fontana, M. A. (2018). Gen

Locke, A. E., Kahali, B., Berndt, S. I., Justice, A. E., Pers, T. H., Day, F. R., ... & Croteau-Chonka, D. C. (2015). Genetic studies of body mass index yield new insights for obesity biology. *Nature*, *518*(7538), 197.

Luppino, F. S., de Wit, L. M., Bouvy, P. F., Stijnen, T., Cuijpers, P., Penninx, B. W., & Zitman, F. G. (2010). Overweight, obesity, and depression: a systematic review and meta-analysis of longitudinal studies. *Archives of General Psychiatry*, *67*(3), 220-229.

Markowitz, S., Friedman, M. A., & Arent, S. M. (2008). Understanding the relation between obesity and depression: causal mechanisms and implications for treatment. *Clinical Psychology: Science and Practice*, *15*(1), 1-20.

Martin, A. R., Gignoux, C. R., Walters, R. K., Wojcik, G. L., Neale, B. M., Gravel, S., ... & Kenny, E. E. (2017). Human demographic history impacts genetic risk prediction across diverse populations. *The American Journal of Human Genetics*, *100*(4), 635-649.

Miller, A. H., & Raison, C. L. (2016). The role of inflammation in depression: from evolutionary imperative to modern treatment target. *Nature Reviews Immunology*, *16*(1), 22.

Morris, S. (2007). The impact of obesity on employment. *Labour Economics*, *14*(3), 413-433.
National Alliance on Mental Illness "Mental Health By The Numbers". Retrieved from: https://www.nami.org/learn-more/mental-health-by-the-numbers

Nevo, A., & Rosen, A. M. (2012). Identification with imperfect instruments. *Review of Economics and Statistics*, *94*(3), 659-671.

Okbay, A., Baselmans, B. M., De Neve, J. E., Turley, P., Nivard, M. G., Fontana, M. A., ... & Gratten, J. (2016). Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nature Genetics*, *48*(6), 624.

Palinkas, L. A., Wingard, D. L., & Barrett-Connor, E. (1996). Depressive symptoms in overweight and obese older adults: a test of the "jolly fat" hypothesis. *Journal of Psychosomatic Research*, *40*(1), 59-66.

Paolucci, E. M., Loukov, D., Bowdish, D. M., & Heisz, J. J. (2018). Exercise reduces depression and inflammation but intensity matters. *Biological Psychology*, *133*, 79-84.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, *38*(8), 904-909.

Puhl, R. M., & Heuer, C. A. (2010). Obesity stigma: important considerations for public health. *American Journal of Public Health*, *100*(6), 1019-1028.

Roberts, R. E., Kaplan, G. A., Shema, S. J., & Strawbridge, W. J. (2000). Are the obese at greater risk for depression?. *American Journal of Epidemiology*, *152*(2), 163-170.

Rosen-Reynoso, M., Alegría, M., Chen, C. N., Laderman, M., & Roberts, R. (2011). The relationship between obesity and psychiatric disorders across ethnic and racial minority groups in the United States. *Eating Behaviors*, *12*(1), 1-8.

Scott, K. M., McGee, M. A., Wells, J. E., & Browne, M. A. O. (2008). Obesity and mental disorders in the adult general population. *Journal of Psychosomatic Research*, *64*(1), 97-105.

Simon, G. E., Von Korff, M., Saunders, K., Miglioretti, D. L., Crane, P. K., Van Belle, G., & Kessler, R. C. (2006). Association between obesity and psychiatric disorders in the US adult population. *Archives of General Psychiatry*, *63*(7), 824-830.

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

Steffick DE (2000). Documentation of Affective Functioning Measures in the Health and Retirement Study (http://hrsonline.isr.umich.edu/sitedocs/userg/dr-005.pdf).

Suglia, S. F., Demmer, R. T., Wahi, R., Keyes, K. M., & Koenen, K. C. (2016). Depressive symptoms during adolescence and young adulthood and the development of type 2 diabetes mellitus. *American Journal of Epidemiology*, *183*(4), 269-276.

Tyrrell, J., Mulugeta, A., Wood, A.R., Beaumont, R.N., Tuke, M.A., Jones, S.E., ... & Hyppönen, E., (2018). Using genetics to understand the causal influence of higher BMI on depression. *International Journal of Epidemiology*, dyy223, 1-15.

van Kippersluis, H., & Rietveld, C. (2018). Pleiotropy-robust Mendelian Randomization. *International Journal of Epidemiology*, *47*, 1279-1288.

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.

Walter, S., Kubzansky, L. D., Koenen, K. C., Liang, L., Tchetgen Tchetgen, E. J., Cornelis, M. C., ... & Glymour, M. M. (2015). Revisiting mendelian randomization studies of the effect of body mass index on depression. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *168*(2), 108-115.

Ware, E. B., Schmitz, L. L., Faul, J. D., Gard, A., Mitchell, C., Smith, J. A., ... & Kardia, S. L. (2017). Heterogeneity in polygenic scores for common human traits. *BioRxiv*, 106062.

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

Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., ... & Bacanu, S. A. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics*, *50*(5), 668.

Zhao, G., Ford, E. S., Dhingra, S., Li, C., Strine, T. W., & Mokdad, A. H. (2009). Depression and anxiety among US adults: associations with body mass index. *International Journal of Obesity*, *33*(2), 257.

Table 1: Summary Statistics

	Add Health			HRS		
	All	Women	Men	All	Women	Men
	(1)	(2)	(3)	(4)	(5)	(6)
Age	28.94	28.82	29.01	68.18	68.05	68.36
	(1.73)	(1.75)	(1.72)	(9.84)	(10.06)	(9.54)
Female	0.54	---	---	0.58	---	---
	(0.49)			(0.49)		
CES-D	5.79	6.18	5.33	1.26	1.43	1.03
	(4.65)	(4.89)	(4.31)	(1.33)	(1.94)	(1.61)
Depressed	0.15	0.17	0.12	0.12	0.14	0.09
	(.35)	(.37)	(.32)	(.32)	(.35)	(.28)
BMI	28.56	28.48	28.67	27.69	27.42	28.05
	(7.14)	(7.66)	(6.49)	(5.51)	(5.96)	(4.83)
Obese	0.34	0.34	0.35	0.28	0.28	0.28
	(.47)	(.47)	(.48)	(.45)	(.45)	(.45)
N	4928	2643	2285	8867	5104	3763

Notes: Standard deviations in parentheses

Table 2: OLS Estimates of the Effect of BMI on Mental Health in Young and Old Adults

	Add Health			HRS		
	All (1)	Women (2)	Men (3)	All (4)	Women (5)	Men (6)
A: CESD-Score						
Outcome Mean	5.79	6.18	5.33	1.26	1.43	1.03
BMI	0.028***	0.062***	-0.028**	0.029***	0.030***	0.026***
	(.010)	(.014)	(.014)	(.004)	(.005)	(.006)
Depression PGS	0.359***	0.345***	0.389***	0.130***	0.121***	0.104***
	(.086)	(.126)	(.116)	(.020)	(.029)	(.027)
Education PGS	-0.101	-0.087	-0.010	-0.085***	-0.099***	-0.064***
	(.068)	(.098)	(.096)	(.068)	(.098)	(.096)
B: Depressed						
Outcome Mean	0.15	0.17	0.12	0.12	0.14	0.09
BMI	0.002***	0.004***	-0.002	0.003***	0.004***	0.003**
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
Depression PGS	0.023***	0.024**	0.022**	0.017***	0.019***	0.014***
	(.007)	(.010)	(.009)	(.004)	(.005)	(.005)
Education PGS	-0.005	-0.004	-0.004	-0.013***	-0.018***	-0.007
	(.005)	(.008)	(.007)	(.004)	(.005)	(.007)
N	4928	2643	2285	8867	5104	3763

Notes: Add Health regressions in columns 1-3 control for age, age squared, gender, birth order, mother's education, picture vocabulary score, and the first 20 ancestry-specific principal components of the genetic data. HRS regressions in columns 4-6 control for age, age squared, gender, mother's education and the first 10 ancestry-specific principal components of the genetic data. Robust standard errors in parentheses. ***significant at 1% **significant at 5% *significant at 10%

Table 3: First Stage Estimates of the Effect of the BMI PGS on BMI

	BMI	BMI 10th percentile	BMI 25th percentile	BMI 50th percentile	BMI 75th percentile	BMI 90th percentile
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Add Health						
BMI PGS	1.803***	0.861***	1.252***	1.698***	2.269***	2.788***
	(.101)	(.071)	(.074)	(.102)	(.149)	(.242)
Depression PGS	0.220*	-0.064	0.189*	0.341**	0.442**	0.252
	(.129)	(.094)	(.100)	(.135)	(.197)	(.310)
Education PGS	-0.119	0.124*	-0.029	-0.013	-0.224	-0.415
	(.104)	(.073)	(.078)	(.105)	(.154)	(.249)
F-Statistic	289	144	256	256	225	144
N	4928	4928	4928	4928	4928	4928
Panel B: HRS						
BMI PGS	1.403***	0.671***	0.936***	1.315***	1.699***	2.079***
	(.060)	(.062)	(.056)	(.060)	(.083)	(.129)
Depression PGS	-0.015	-0.033	0.022	-0.012	0.023	-0.035
	(.058)	(.062)	(.054)	(.061)	(.083)	(.126)
Education PGS	-0.089	-0.092	-0.089	-0.006	-0.002	-0.112
	(.058)	(.062)	(.056)	(.060)	(.083)	(.127)
F-Statistic	529	100	256	441	400	256
N	8867	8867	8867	8867	8867	8867

Notes: Add Health regressions control for age, age squared, gender, birth order, mother's education, picture vocabulary score and the first 20 ancestry-specific principal components of the genetic data. HRS regressions control for age, age squared, gender, mother's education and the first 10 ancestry-specific principal components of the genetic data. Robust standard errors in parentheses. ***significant at 1% **significant at 5% *significant at 10%

Figure 1: Unconditional Quantile Treatement Effects of the Binary IV on BMI

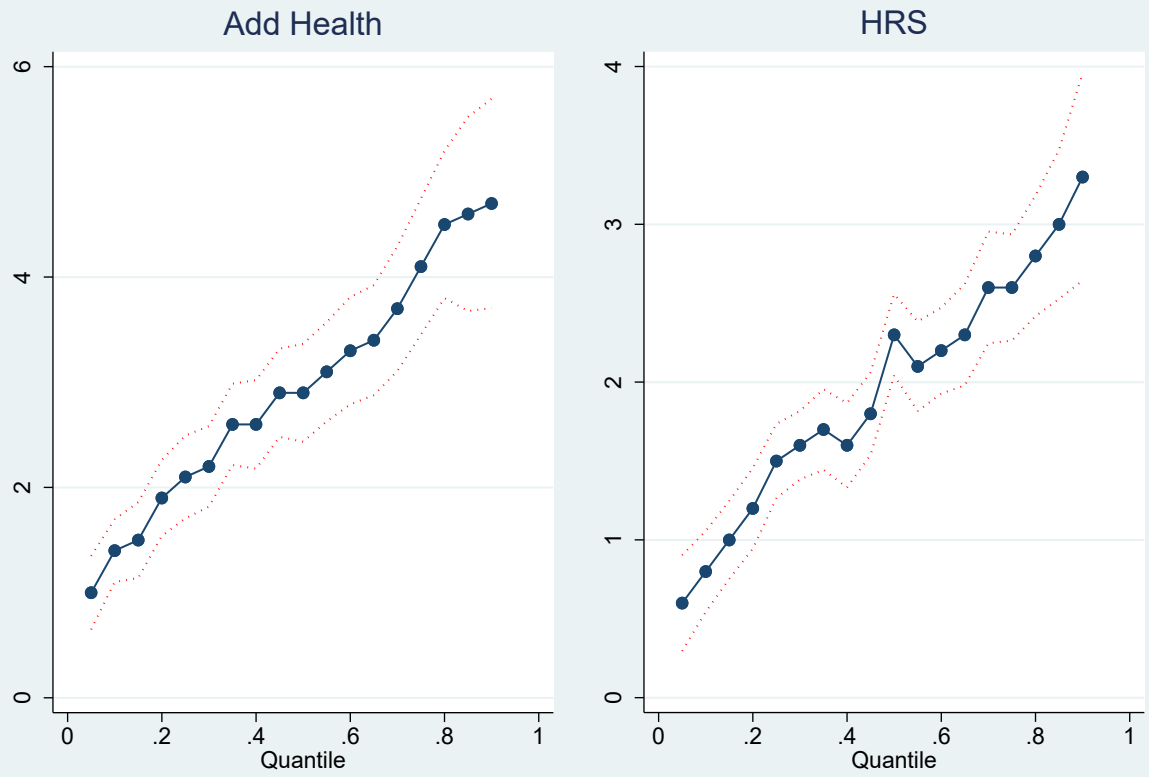

Table 4: IV Estimates of the Effect of BMI on Mental Health in Young and Old Adults

	Add Health			HRS		
	All (1)	Women (2)	Men (3)	All (4)	Women (5)	Men (6)
A: CESD-Score						
Outcome Mean	5.79	6.18	5.33	1.26	1.43	1.03
BMI	0.020	0.019	0.020	0.051***	0.053***	0.046***
	(.038)	(.054)	(.053)	(.015)	(.019)	(.023)
Depression PGS	0.360***	0.353***	0.386	0.128***	0.122***	0.138***
	(.086)	(.125)	(.117)	(.020)	(.029)	(.027)
Education PGS	-0.104	-0.111	-0.086	-0.078***	-0.088***	-0.062***
	(.076)	(.102)	(.096)	(.021)	(.031)	(.029)
B: Depressed						
Outcome Mean	0.15	0.17	0.12	0.12	0.14	0.09
BMI	0.003	0.004	0.001	0.007***	0.007**	0.007*
	(.003)	(.004)	(.004)	(.003)	(.004)	(.004)
Depression PGS	0.022***	0.024**	0.022	0.024***	0.019***	0.013**
	(.007)	(.010)	(.009)**	(.006)	(.005)	(.005)
Education PGS	-0.004	-0.003	-0.003	-0.010	-0.016***	-0.006
	(.005)	(.008)	(.007)	(.006)	(.006)	(.005)
N	4928	2643	2285	8867	5104	3763

Notes: Add Health regressions in columns 1-3 control for age, age squared, gender, birth order, mother's education, picture vocabulary score, and the first 20 ancestry-specific principal components of the genetic data. HRS regressions in columns 4-6 control for age, age squared, gender, mother's education and the first 10 ancestry-specific principal components of the genetic data. Robust standard errors in parentheses. ***significant at 1% **significant at 5% *significant at 10%

Table 5: IV Estimates of the Effect of BMI on Mental Health in Young and Old Adults without Controlling for PGSs for Depression and Educational Attainment

	Add Health			HRS		
	All	Women	Men	All	Women	Men
	(1)	(2)	(3)	(4)	(5)	(6)
A: CESD-Score						
Outcome Mean	5.79	6.18	5.33	1.26	1.43	1.03
BMI	0.024	0.022	0.026	0.065***	0.070***	0.056***
	(.037)	(.053)	(.052)	(.014)	(.019)	(.019)
B: Depressed						
Outcome Mean	0.15	0.17	0.12	0.12	0.14	0.09
BMI	0.003	0.004	0.001	0.009***	0.010***	0.008***
	(.003)	(.004)	(.004)	(.003)	(.003)	(.004)
N	4928	2643	2285	8867	5104	3763

Notes: Add Health regressions in columns 1-3 control for age, age squared, gender, birth order, mother's education, picture vocabulary score, and the first 20 ancestry-specific principal components of the genetic data. HRS regressions in columns 4-6 control for age, age squared, gender, mother's education and the first 10 ancestry-specific principal components of the genetic data. Robust standard errors in parentheses. ***significant at 1% **significant at 5% *significant at 10%

Table 6: OLS, IV and Upper Bounds for the Effect of BMI on CES-D Score

	Mean (1)	OLS (2)	IV (3)	IV(1) (4)	UB (A4) (5)	UB (A4 & A5) (6)
Panel A: Add Health						
All	5.79	0.028***	0.020	0.032**	0.020	0.020
		(.010)	(.038)	(.016)	[.052]	[.066]
Women	6.18	0.062***	0.019	0.075***	0.019	0.019
		(.014)	(.054)	(.020)	[.094]	[.118]
Men	5.33	-0.028**	0.020	-0.045*	-0.028	-0.027
		(.014)	(.053)	(.023)	[-.003]	[.010]
Panel B: HRS						
All	1.26	0.029***	0.051***	0.022***	0.029	0.022
		(.004)	(.015)	(.006)	[.037]	[.035]
Women	1.43	0.300***	0.053***	0.023***	0.030	0.023
		(.030)	(.019)	(.008)	[.041]	[.040]
Men	1.03	0.026***	0.046***	0.019**	0.026	0.018
		(.006)	(.023)	(.010)	[.041]	[.042]

Notes: Add Health regressions control for age, age squared, gender, birth order, mother's education, picture vocabulary score, and the first 20 ancestry-specific principal components of the genetic data. HRS regressions control for age, age squared, gender, mother's education and the first 10 ancestry-specific principal components of the genetic data. Robust standard errors in (.) ***significant at 1% **significant at 5% *significant at 10%. The Nevo & Rosen (2012) approach is implemented using the imperfectiv command in Stata. The upper endpoint of the 95% confidence interval on the bounded parameter is given in [.].

Table 7: OLS, IV and Upper Bounds for the Effect of BMI on Depression

	Mean (1)	OLS (2)	IV (3)	IV(1) (4)	UB (A4) (5)	UB (A4 & A5) (6)
Panel A: Add Health						
All	0.15	0.002***	0.003	0.002	0.002	0.002
		(.001)	(.003)	(.001)	[.004]	[.004]
Women	0.17	0.004***	0.004	0.004**	0.004	0.004
		(.001)	(.004)	(.002)	[.006]	[.0007]
Men	0.12	-0.002	-0.001	-0.002	-0.002	-0.001
		(.0010)	(.004)	(.002)	[.001]	[.002]
Panel B: HRS						
All	0.12	0.003***	0.007***	0.002**	0.003	0.002
		(.001)	(.003)	(.001)	[.005]	[.005]
Women	0.14	0.004***	0.007**	0.003*	0.004	0.003
		(.001)	(.004)	(.001)	[.006]	[.006]
Men	0.09	0.003***	0.007*	0.002	0.003	0.001
		(.001)	(.004)	(.002)	[.005]	[.006]

Notes: Add Health regressions control for age, age squared, gender, birth order, mother's education, picture vocabulary score, and the first 20 ancestry-specific principal components of the genetic data. HRS regressions control for age, age squared, gender, mother's education and the first 10 ancestry-specific principal components of the genetic data. Robust standard errors in (.) ***significant at 1% **significant at 5% *significant at 10%. The Nevo & Rosen (2012) approach is implemented using the imperfectiv command in Stata. The upper endpoint of the 95% confidence interval on the bounded parameter is given in [.].

Appendix Table A1: OLS Estimates of the Effect of BMI on Mental Health in Young and Old Adults without Controlling for the Depression and Education PGSs

	Add Health			HRS		
	All	Women	Men	All	Women	Men
	(1)	(2)	(3)	(4)	(5)	(6)
A: CESD-Score						
Outcome Mean	5.79	6.18	5.33	1.06	1.43	1.03
BMI	0.030***	0.063***	-0.027*	0.030***	0.031***	0.027***
	(.010)	(.014)	(.014)	(.004)	(.005)	(.007)
B: Depressed						
Outcome Mean	0.15	0.17	0.12	0.12	0.14	0.09
BMI	0.002***	0.004***	-0.001	0.004***	0.004***	0.003**
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
N	4928	2643	2285	8867	5104	3763

Notes: Add Health regressions in columns 1-3 control for age, age squared, gender, birth order, mother's education, picture vocabulary score, and the first 20 ancestry-specific principal components of the genetic data. HRS regressions in columns 4-6 control for age, age squared, gender, mother's education and the first 10 ancestry-specific principal components of the genetic data. Robust standard errors in parentheses. ***significant at 1% **significant at 5% *significant at 10%

Appendix Table A2: IV Estimates of the Effect of BMI on Mental Health in Young and Old Using a Binary Instrument

	Add Health			HRS		
	All	Women	Men	All	Women	Men
	(1)	(2)	(3)	(4)	(5)	(6)
A: CESD-Score						
Outcome Mean	5.79	6.18	5.33	1.26	1.43	1.03
BMI	0.039	0.009	0.061	0.063***	0.071***	0.051*
	(.043)	(.060)	(.069)	(.018)	(.025)	(.023)
B: Depressed						
Outcome Mean	0.15	0.17	0.12	0.12	0.14	0.09
BMI	0.004	0.004	0.004	0.008***	0.009**	0.004
	(.003)	(.005)	(.005)	(.003)	(.004)	(.005)
N	4928	2643	2285	8867	5104	3763

Notes: The IV is a binary indicator=1 if the BMI PGS>0. Add Health regressions in columns 1-3 control for age, age squared, gender, birth order, mother's education, picture vocabulary score, PGSs for depression and education, and the first 20 ancestry-specific principal components of the genetic data. HRS regressions in columns 4-6 control for age, age squared, gender, mother's education, PGSs for depression and education, and the first 10 ancestry-specific principal components of the genetic data. Robust standard errors in parentheses. ***significant at 1% **significant at 5% *significant at 10%

Appendix Table A3: OLS, IV and Upper Bounds for the Effect of Obesity on Mental Health

	CES-D Score			Depression		
	OLS	IV	UB (A4 & A5)	OLS	IV	UB (A4 & A5)
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Add Health						
All	0.507***	0.334	0.334	0.027**	0.050	0.021
	(.142)	(.638)	[1.04]	(.011)	(.050)	[.059]
Women	0.913***	0.383	0.383	0.059***	0.086	0.052
	(.211)	(1.07)	[1.72]	(.016)	(.082)	[.108]
Men	0.008	0.284	-0.083	-0.010	0.010	-0.017
	(.189)	(.758)	[.615]	(.014)	(.060)	[.038]
Panel B: HRS						
All	0.311***	0.705***	0.213	0.036***	0.102***	0.019
	(.045)	(.206)	[.359]	(.008)	(.037)	[.046]
Women	0.332***	0.800***	0.219	0.037***	0.106***	0.020
	(.064)	(.293)	[.425]	(.012)	(.052)	[.059]
Men	0.276***	0.568**	0.201	0.033***	0.090*	0.018
	(.061)	(.284)	[.403]	(.011)	(.050)	[.055]

Notes: Add Health regressions control for age, age squared, gender, birth order, mother's education, picture vocabulary score, PGSs for education and depression, and the first 20 ancestry-specific principal components of the genetic data. HRS regressions control for age, age squared, gender, mother's education, PGSs for education and depression, and the first 10 ancestry-specific principal components of the genetic data. Robust standard errors in (.) ***significant at 1% **significant at 5% *significant at 10%. The Nevo & Rosen (2012) approach is implemented using the imperfectiv command in Stata. The upper bound is based on assumptions (A4 & A5). The upper endpoint of the 95% confidence interval on the bounded parameter is given in [.].

Appendix Table A4: The Effect of Mental Health on BMI for Old Adults

	All		
	OLS	IV	UB Ass 3&4
	(1)	(2)	(3)
Panel A:			
CES-D Score	0.231***	-0.117	-0.118
	(.036)	(.450)	[0.354]
Panel B:			
Depressed	0.846***	-0.914	-0.914
	(.206)	(3.56)	[1.52]

Notes: All regressions control for age, age squared, gender, mother's education, PGSs for BMI and education, and the first 10 ancestry-specific principal components of the genetic data. Robust standard errors in (.) ***significant at 1% **significant at 5% *significant at 10%. The 95% confidence interval of the estimated upper bound is given in [.].

Appendix Table A5: IV Estimates of the Effect of BMI on Mental Health for Old Adults, Accounting for Mortality Attrition

	All (1)	Women (2)	Men (3)
A: CESD-Score			
Outcome Mean	1.26	1.43	1.03
BMI	0.051***	0.043*	0.061*
	(.018)	(.023)	(.031)
B: Depressed			
Outcome Mean	0.12	0.14	0.09
BMI	0.008**	0.004	0.013**
	(.003)	(.004)	(.006)

Notes: All regressions control for age, age squared, gender, mother's education, PGSs for depression and education, and the first 10 ancestry-specific principal components of the genetic data. Robust standard errors in parentheses. ***significant at 1% **significant at 5% *significant at 10%.

Appendix Table A6: Effect of BMI on Health and Exercise for Older Adults

	(1)	(2)	(3)	(4)
Panel A: Self-reported Health	**Mean**	**OLS**	**IV**	**Lower Bound**
All [N=8861]	0.48	-0.016***	-0.019***	-0.015
		(.001)	(.004)	[-.018]
Women[N=5098]	0.49	-0.016***	-0.018***	-0.016
		(.001)	(.005)	[-.020]
Men [N=2225]	0.48	-0.014***	-0.021***	-0.011
		(.002)	(.007)	[-.019)
Panel B: Vigorous exercise More than once per week	**Mean**	**OLS**	**IV**	**Lower Bound**
All [N=8867]	0.26	-0.010***	-0.013***	-0.0059
		(.001)	(.003)	[-.012]
Women [N=5104]	0.23	-0.009***	-0.011**	-0.009
		(.001)	(.004)	[-.012]
Men [N=3763]	0.32	-0.012***	-0.016**	-0.011
		(.001)	(.006)	[-.017]
Panel C: Health Limits Daily activities	**Mean**	**OLS**	**IV**	**Upper Bound**
All [N=8867]	0.12	0.009***	0.010***	0.009
		(.001)	(.003)	[.012]
Women [N=5104]	0.13	0.009**	0.011***	0.009
		(.001)	(.003)	[.012]
Men [N=3763]	0.11	0.009***	0.007*	0.007
		(.001)	(.004)	[.015]

Notes: HRS regressions control for age, age squared, gender, mother's education, PGSs for education and BMI, and the first 10 ancestry-specific principal components of the genetic data. Robust standard errors in (.) ***significant at 1% **significant at 5% *significant at 10%. The Nevo & Rosen (2012) approach is implemented using the imperfectiv command in Stata. The bounds are estimated under assumptions (A4 & A5). The upper endpoint of the 95% confidence interval on the bounded parameter is given in [.].