

DISCUSSION PAPER SERIES

IZA DP No. 13055

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

Vikesh Amin
Carlos A. Flores
Alfonso Flores-Lagunes

MARCH 2020

DISCUSSION PAPER SERIES

IZA DP No. 13055

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

Vikesh Amin

Central Michigan University

Carlos A. Flores

California Polytechnic State University at San Luis

Alfonso Flores-Lagunes

Syracuse University and IZA

MARCH 2020

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

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

We estimate the effect of BMI on mental health for young adults and elderly individuals using data from the National Longitudinal Study of Adolescent Health and the Health & Retirement Study. To tackle confounding due to unobserved factors, we exploit variation in a polygenic score (PGS) for BMI within two complementary econometric methods that differ in the assumptions they employ. First, we use the BMI PGS as an IV and adjust for PGSs for other factors (depression and educational attainment) that may invalidate this IV. We find a large statistically significant effect of BMI on mental health for the elderly: a 5 kg/m² increase in BMI (a difference equivalent to moving from overweight to obese) increases the probability of depression by 29%. In contrast, for young adults the IV estimates are statistically and economically insignificant. We show that IV estimates likely have to be interpreted as identifying a weighted average of effects of BMI on mental health mostly for compliers on the upper quantiles of the BMI distribution. Second, we use the BMI PGS as an “imperfect” IV and estimate an upper bound on the average treatment effect for the population. The estimated upper bounds reinforce the conclusions from the IV estimates.

JEL Classification: I10, I12

Keywords: BMI, depression, genetics, instrumental variables

Corresponding author:

Vikesh Amin
Department of Economics
Central Michigan University
Sloan Hall 312
Mount Pleasant, MI 48859
USA

E-mail: amin1v@cmich.edu

1. Introduction

Mental health is an important public health issue in the US due to its high prevalence and associated economic and societal costs. For instance, according to the National Institute of Mental Health, in 2017 approximately 19% of US adults suffered from a mental illness (about 46.6 million adults), 4.5% experienced a serious mental illness (i.e., one that substantially interfered with major life activities), and 7.1% had at least one major depressive episode. In 2014, \$186 billion was spent on mental health treatment, representing 6.4% of all health spending (Substance Abuse and Mental Health Services Administration, 2016). This figure does not include the cost coming from lost earnings, which can also be substantial—for example, Insel (2008) estimates such cost to be \$193.2 billion for serious mental illnesses. A factor that has been found to be associated to mental illness is obesity, usually defined as having a Body Mass Index (BMI) greater than or equal to 30. Several studies have shown that having a higher BMI or being obese is associated with higher odds of experiencing poor mental health and depression (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). In spite of this empirical evidence, it is unclear whether there is a causal relationship because of the inherent difficulty in estimating causal effects. A widely used econometric method to empirically uncover causal relationships exploits exogenous variation in an instrumental variable (IV) to estimate the effect of the variable of interest or “treatment” (BMI) on an outcome (mental health). In this paper, we study the effect of BMI on mental health for European-ancestry young adults (aged 25-34) and the elderly (aged 50-89) using a genetic index of risk for high BMI as an IV for BMI.

There are theoretical reasons why BMI and obesity may affect mental health. Biologically, 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). It has also been hypothesized by Markowitz et al. (2008) that BMI and obesity are associated with mental health through behavioral (functional impairment), cognitive (body image dissatisfaction, poor self-reported health), and social (stigma) mechanisms. 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, as individuals who believe their health to be poor may also hold other depressive beliefs. In terms of functional impairments, 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). Thus, it is theoretically possible for BMI to have a causal effect on mental health. Moreover, we hypothesize that the importance of the mechanisms above may be different for young adults and the elderly (e.g., Himes 2000; Willage 2018), which motivates us to analyze these two groups separately.

Although there are theoretical reasons for BMI affecting mental health, drawing empirical causal conclusions is difficult since BMI is not randomly assigned but rather is determined by several factors (some unobserved to researchers). In particular, it is hard to interpret standard estimated associations from Ordinary Least Squares (OLS) regressions as causal effects because they are likely to be confounded by unobserved factors (e.g. time preference, genetic endowments, innate ability) that affect both BMI and mental health, and because of reverse causality insofar poor mental health may increase the likelihood of being obese. To tackle these

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).

issues, some recent studies have employed genetic information in the form of a polygenic score (PGS) for BMI as an IV.^{2,3} 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.⁴ For 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 ($p < 5 \times 10^{-8}$), which explain approximately 1.45% of the variation in BMI. A PGS for BMI is constructed by aggregating the SNPs identified in Speliotes et al. (2010) and weighting them by the strength of their association with BMI.⁵ Under the assumptions employed by IV methods (to be discussed in section 3.1), the IV estimates represent causal effects.

Previous studies using a PGS for BMI as an IV in the context of the US indicate that there is no statistically significant effect of BMI on mental health. Hung et al. (2014) estimate the effect of BMI on the probability of having a major depressive disorder in young adults (average age of 26 years) using data from RADIANT (a case control study of major depression); 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; and 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.⁶ In all three studies, the OLS estimates showed a statistically positive association between BMI and poor mental health, whereas the IV estimates were statistically insignificant.⁷ In contrast to US results, Jokela et al. (2012) find that BMI has a statistically significant negative effect on mental health in adolescents and adults in Finland, while Tyrrell et al. (2018) find a statistically significant positive effect of BMI on depression for individuals (average age of 57 years) in the UK.

We contribute to this literature by estimating the effect of BMI on mental health (measured using the Center for Epidemiologic Studies Depression Scale; CES-D) for European ancestry individuals from 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

²The 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.

³In the health 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).

⁴A 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.

⁵The PGS for individual i is a weighted average across the number of SNPs (m) of the number of reference alleles A (0, 1 or 2) at that SNP multiplied by the corresponding beta estimate from the GWAS analysis: $PGS_i = \sum_{j=1}^m \beta_j A_{ij}$. 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 PGSs.

⁶One important difference between our analysis of young adults and Willage (2018) is that we use a more general sample from the National Longitudinal Study of Adolescent Health, described in section 2.1.

⁷Willage (2018) does however find a statistically positive effect of BMI on suicidal ideation at the 10% level.

Health & Retirement Study).^{8,9} Estimation of the effect of BMI on mental health for different age groups is important because this effect may differ over the life cycle as different mechanisms may be relevant for different age groups, and the impacts of such mechanisms on mental health may also differ.¹⁰ As a result, policies aimed at reducing BMI (and obesity) may have different effects on mental health for different age groups.

To estimate the effects of interest, we exploit variation in a PGS for BMI employing two complementary econometric methods that differ in the assumptions they employ. By providing results from two different methods, we offer a more complete analysis of these effects. The first econometric method builds upon the existing literature using genetic markers as IVs. There are three innovations in the way we perform our analysis relative to previous work. First, the PGS we employ is based on the more recent GWAS by Locke et al. (2015), which includes all SNPs and thus has more predictive power for BMI. Second, IV estimates represent causal effects under certain assumptions (discussed in section 3.1) that include the critical requirement that the BMI PGS affects mental health only through its effect on BMI (the “exclusion restriction” assumption). One concern is that the BMI PGS may violate the exclusion restriction because of pleiotropy (von Hinke et al. 2016; van Kippersluis & Rietveld 2018; DiPrete et al. 2018; Fletcher 2018). That is, 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. Unlike previous studies, we attempt to lessen this concern by using PGSs for depression and educational attainment as additional control variables. Using a PGS for depression as an additional control variable is in the spirit of the genetic instrumental variables (GIVs) framework developed by DiPrete et al. (2018), who argue that is possible to block pleiotropic pathways by including the PGS for the outcome conditional on treatment as a control. The educational attainment PGS is used to proxy for innate ability, which is an unobservable factor that can confound estimates.¹¹ Third, we empirically characterize the subpopulation for which the IV estimates identify the effects of interest; this characterization has been missing in the literature. In the presence of heterogeneous effects, IV methods identify a local average treatment effect (LATE) for those individuals whose treatment (BMI) is affected by the IV (PGS for BMI)—the so-called “compliers” (Imbens & Angrist 1994). This is in contrast to OLS estimates, which under strong assumptions such as unconfoundedness identify the population average treatment effect (ATE). Using unconditional quantile treatment effects estimates of the effect of the BMI PGS on BMI, we document that IV estimates using a BMI PGS as an IV largely identify the effects of interest for compliers on the upper quantiles of the BMI distribution. This characterization of the relevant

⁸ The CES-D is one common measure of mental health used in the literature and available in our data sets. Another common measure is the Beck’s Depression Inventory (BDI). The BDI was designed to measure depressive severity for individuals already diagnosed as clinically depressed (Beck & Steer 1993), while the CES-D was initially intended to measure depressive severity of adults in the general population (Radloff 1977). Another difference between the two measures is that the CES-D emphasizes the affective component of depression (Radloff 1977), whereas the BDI has a much stronger cognitive component (Beck et al. 1961). Given these differences, it is not straightforward to compare our results to those of studies that use the BDI. For reference, the correlation between these two important measures has been found to range between 0.58 to 0.82 (Skorikov & Vandervoort, 2003).

⁹ We limit the analysis to European ancestry individuals due to issues relating to population stratification and the predictive power of genetic markers. These issues are discussed in sections 2 and 3.

¹⁰ For instance, although BMI is hypothesized to worsen mental health, 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 for elderly individuals.

¹¹ Our use of a more comprehensive PGS based on Locke et al. (2015) could arguably lead to additional concerns of pleiotropy. This provides an additional reason to control for other traits’ PGSs and employ a second, complementary approach that does not impose the validity of the exclusion restriction.

subpopulation for which the effect is estimated is critical to understand the external validity of the estimated effects and the possible incidence of policies.

The second econometric method we use exploits the PGS for BMI as a potentially “imperfect IV”—one that is allowed to violate the exclusion restriction—to estimate an upper bound on the ATE of BMI on mental health for the corresponding population (young adults or the elderly), following Nevo & Rosen (2012). While controlling for the PGSs for depression and education can block pleiotropic pathways (DiPrete et al., 2018), it is not possible to guarantee that the exclusion restriction is satisfied (Angrist et al. 1996; von Hinke et al. 2016). By not relying on this assumption, this second method is not subject to the criticism that pleiotropy may be present—at the cost of providing only a bound rather than a point estimate of the effect of interest. Nevo & Rosen (2012)’s 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 (1) the correlation between the IV and unobserved factors in the OLS model for mental health, and (2) the correlation between those same unobserved factors and the endogenous regressor (BMI). Moreover, under their corresponding assumptions, each method makes inference about a different parameter: IV methods estimate a LATE on compliers, while Nevo & Rosen (2012) methods bound the population’s ATE. By providing results from two different approaches, we offer a more complete picture of the effects of interest and allow the reader to put more weight on the estimates she deems more credible if she has concerns about the assumptions used by them.

The results from the two different econometric methods are consistent with each other. For young adults, the IV estimates indicate that there is no statistically significant effect of BMI on mental health, while the estimated upper bound for the ATE is consistent with, at most, a small economically insignificant effect. In contrast, the IV estimates for elderly individuals indicate a positive and statistically significant effect. In particular, the effect of a 5 kg/m² (a difference equivalent to moving from overweight to obese) increases the CES-D score by 20% and the likelihood of depression by 29%. Again, the estimated upper bounds on the ATE for the elderly population are consistent with the magnitude of the IV estimated effects. As previously indicated, our results also suggest that the (LATE) effect estimated in the IV method likely reflects a weighted average of effects for compliers on the upper quantiles of the BMI distribution. Lastly, we find no statistically significant gender differences using the IV estimates.

Our paper contributes to the literature on the effects of BMI on mental health in several ways. First, we provide estimates of these effects for two important groups, young adults and the elderly, which is relevant as it can increase our understanding of the effects and of policies aimed at decreasing BMI/obesity. Second, we employ a more recent PGS for BMI with more predictive power relative to the previous literature (an exception is Tyrrell et al. 2018). Third, we add as controls in the IV regressions the PGSs for depression and education to increase the plausibility of the exclusion restriction assumption in this setting (DiPrete et al 2018; Fletcher 2018), which is important since otherwise IV estimates may overestimate the effect of interest. Indeed, we document that the IV estimates for the elderly are likely overestimated if we do not control for those PGSs. Fourth, we analyze the first-stage effects of the BMI PGS on BMI to inform the characterization of the subpopulation for which the IV methods estimate an effect. Within this literature, we appear to be the first to illustrate how this can be done. We document that IV estimates likely reflect a weighted average of effects of BMI on mental health for complier individuals on the upper quantiles of the BMI distribution. From a policy perspective, having such a characterization of the subpopulation for which the effect is estimated is critical. Finally, we complement our IV estimates by using the bounding methods in Nevo & Rosen (2012). This econometric method uses the PGS for BMI as an “imperfect IV” that does not need to satisfy the exclusion restriction assumption, thus addressing a common critique of using genetic markers as IVs (von Hinke et al. 2016; Fletcher 2018). To our knowledge, we are the first to illustrate the use of the Nevo & Rosen (2012) bounds in this context.

The paper is organized as follows. We describe the datasets and outline the econometric approaches we use 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 estimate the effect of BMI on mental health at wave 4, which corresponds to the wave when genetic data were collected. BMI (weight kg/ height m²) 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% (5,728 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). As noted before, 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).¹²

¹² Most of the observations with missing values (732 out of 800) correspond to the variables mother's education and the picture vocabulary score. We checked that our results below hold when (1) not

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 two 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. Our analysis sample from the HRS includes individuals from the different cohorts.

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.

In the 2006 wave of the HRS there are 9,738 European ancestry individuals aged 50 or older with genetic data. Our final sample consists of 8,867 European ancestry individuals with non-missing information on BMI, mental health, the PGSs, and mother’s education.¹³

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).

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

controlling for these two variables and (2) imputing their missing values with the sample mean. These exercises are available from the authors upon request.

¹³ Most of the observations with missing values (675 out of 871) are due to the variable mother’s education. Our findings using the HRS remain largely unchanged if we (1) do not control for mother’s education, and (2) impute mother’s education with the sample mean. These results are available from the authors upon request.

Our parameter of interest in this case is β , which in the absence of unobserved factors affecting both BMI and mental health—an assumption usually referred to as unconfoundedness or selection on observables—can be interpreted as the causal ATE of BMI on mental health for the population from which the sample is drawn (Imbens 2004). However, in our context such assumption seems implausible as there are several important factors that are not included in X_i that are almost surely correlated with both BMI and mental health (e.g., unobserved health endowments and environmental factors), in which case OLS provides a biased estimate of the ATE. For this reason, we refer to such OLS estimates as estimating “associations” rather than (causal) effects.

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. IV estimation is equivalent to two stage least squares. In the first stage (equation 2), the BMI of individual i is related to the PGS for BMI (PGS_i) and a vector of exogenous control variables (X_i):

$$(2) \text{ BMI}_i = \alpha_0 + \gamma PGS_i + X_i' \psi + v_i$$

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:

$$(3) \text{ MH}_i = c_0 + \beta^{IV} \widehat{BMI}_i + X_i' p + \epsilon_i$$

Under heterogeneous effects, the IV estimates represent a 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 (A4) 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.¹⁴ 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 section 4. 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, 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

¹⁴ The HRS only provides the first 10 principal components of the genetic data, whereas Add Health provides the first 20 principal components.

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 exclusion restriction requires that the PGS for BMI affects mental health only through its effect on BMI. One important reason why the exclusion restriction may be violated is because of pleiotropy (von Hinke et al. 2016; van Kippersluis and Rietveld 2018; 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.¹⁵ For example, if the FTO gene (that is associated with obesity) is also related to intelligence, which affects mental health, then the exclusion restriction would be violated. On a broad level, Locke et al. (2015) found that genes affect 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.¹⁶ If these traits also affect mental health, then they represent paths— independent of BMI—through which the exclusion restriction may be violated. Our strategy to deal with pleiotropy is in the same spirit of the GIV framework developed by DiPrete et al. (2018). 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. Thus, 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. Controlling for these factors increases the plausibility of the exclusion restriction assumption in our setting.

3.2 Nevo & Rosen (2012) Bounds

While controlling for the PGSs for depression and education increases the plausibility of the exclusion restriction assumption, it is not possible to guarantee that it is satisfied (Angrist et al. 1996; von Hinke et al. 2016). The second econometric strategy we employ to investigate the effect of BMI on mental health exploits the PGS for BMI as a potentially “imperfect instrument”, which is one that is allowed to violate the exclusion restriction. Therefore, this method is not subject to the common critique that genetic markers may violate this critical assumption of IV methods because of pleiotropy or other reasons. In particular, we use the methods in Nevo & Rosen (2012), who introduce the concept of an imperfect IV to derive bounds on the ATE in the original OLS model in equation (1), β . To motivate their approach, note that in the context of the linear model in equations (1) and (2), the exclusion restriction can be interpreted as the assumption that the instrument Z is uncorrelated with the error term u in the outcome equation. Instead of assuming such correlation is exactly zero, Nevo & Rosen (2012) impose assumptions on the direction of this correlation. Specifically, 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. In this section, we present the Nevo & Rosen (2012) bounds we employ and discuss their assumptions in the present context.¹⁷

To describe the bounds, let ρ_{ab} denote the correlation between any two random variables A and B , σ_{ab} denote their covariance, and σ_a denote the standard deviation of A . Also, to simplify

¹⁵ Linkage disequilibrium is the non-random association of alleles at different loci in a given population.

¹⁶ 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.

¹⁷ For further details on the methodology, see Nevo & Rosen (2012).

notation in this section, let W denote the treatment (or endogenous) variable, in our case BMI. Nevo & Rosen (2012) assume that: (A4) $\rho_{wu}\rho_{zu} \geq 0$; that is, that the instrument Z (in our case, PGS for BMI) has (weakly) the same direction of correlation with the error term u as the endogenous variable W (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 CES-D 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 (W) is positively correlated with the error term u , so that $\rho_{wu} \geq 0$. Following a similar reasoning, we also expect our instrument Z —genetic risk for high BMI—to be positively correlated with the error term u , implying that $\rho_{zu} \geq 0$, and hence $\rho_{wu}\rho_{zu} \geq 0$.

Nevo & Rosen (2012) show that (i) if $\sigma_{zw} > 0$, that is, if the instrument (PGS for BMI) is positively correlated with the endogenous variable (BMI), which is the case in our application, and (ii) if $\sigma_{wu}, \sigma_{zu} \geq 0$, that is, if the instrument and endogenous variable are positively correlated with the error term u (which we assume, as discussed above), then in the linear IV model in equations (1) and (2) with fixed coefficients, an upper bound for β is given by:

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

where β^{OLS} and β_z^{IV} denote the probability limits of the standard OLS and IV estimators for β , respectively. In other words, in this case β (which represents the ATE) is bounded from above by the minimum of the OLS estimate from equation (1) and its corresponding standard IV estimate computed assuming the validity of the exclusion restriction assumption (i.e., assuming $\rho_{zu} = 0$). Note that in our setting and under our assumptions on ρ_{wu} and ρ_{zu} , the Nevo & Rosen (2012) methods provide only an upper bound on β .¹⁸

The bounds in (4) can be tightened by assuming that: (A5) $|\rho_{wu}| \geq |\rho_{zu}|$; that is, that the instrument Z (PGS for BMI) is less correlated with the error term u in equation (1) than is the endogenous variable W (BMI). 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, thus making BMI likely to be more correlated with u than the BMI PGS. Second, our instrument could be correlated with the error term u 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 u than if we did not control for these PGSs, making the assumption that $|\rho_{wu}| \geq |\rho_{zu}|$ more plausible.

The two previous assumptions give the definition of an “imperfect IV” in Nevo & Rosen (2012), as an IV that: (A4) has the same direction of correlation with the unobserved error term u in the OLS equation for the outcome (in our case, mental health) as the “treatment” variable of interest (BMI), and (A5) is less endogenous than the treatment variable.

¹⁸ Intuitively, as discussed in Nevo & Rosen (2012), in the linear model $MH = \beta_0 + \beta W + u$, the probability limits of the standard OLS and IV estimators for β are, respectively, $\beta^{OLS} = \beta + (\sigma_{wu}/\sigma_w^2)$ and $\beta_z^{IV} = \beta + (\sigma_{zu}/\sigma_{wz})$. Given our assumptions that $\sigma_{wu} > 0$, $\sigma_{zu} > 0$, and the fact that $\sigma_{wz} > 0$, β^{OLS} and β_z^{IV} provide upper bounds for β , with the minimum of the two being a tighter bound and thus preferred. For the model with additional covariates X , the intuition is similar after appropriately partialling out the covariates (see Nevo & Rosen (2012) for details).

To derive their bounds on β under assumptions (A4) and (A5), Nevo & Rosen (2012) define the function $V(\lambda) = \sigma_w Z - \lambda \sigma_z W$, which, when evaluated at $\lambda^* = \frac{\rho_{zu}}{\rho_{wu}}$ (a measure of how much “less invalid” is the IV assumption that $\rho_{zu} = 0$ relative to the OLS assumption that $\rho_{wu} = 0$), generates a variable that by construction is uncorrelated with the error term u , and thus serves as a valid instrument. Although λ^* is unknown, it is bounded between 0 and 1 under the assumptions above (as they imply $\rho_{wu} \geq \rho_{zu} \geq 0$). Nevo & Rosen (2012) employ the bounds on λ^* to bound β in equation (1).

For the linear regression model we employ, 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 (PGS for BMI) is positively correlated with the endogenous variable (BMI), and if the instrument and the endogenous variable are positively correlated with the error term u , then Nevo & Rosen (2012) show that an upper bound on β in equation (1) is given by:

$$(5) \beta \leq \min\{\beta_{V(1)}^{IV}, \beta_z^{IV}\}$$

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 . As before, the Nevo & Rosen (2012) methods provide only an upper bound on β in our setting, which is given by the minimum of two IV estimators.¹⁹ Note that, while two IV estimates are compared to estimate an upper bound for β , this does not imply that the method assumes that the instrument used (PGS for BMI) satisfies the exclusion restriction (although this assumption holds for the “auxiliary” instrument $V(1)$ by construction). Moreover, it does not imply that the parameters of interest are now the local average treatment effects that would be estimated by each instrument when the exclusion restriction holds. Instead, those two IV estimates are used to obtain an upper bound for β (the population average treatment effect, or ATE).

Being able to estimate an upper bound on β in equation (1)—the ATE—is useful. First, the estimated upper bound provides a benchmark magnitude for the population ATE while relaxing the potentially troublesome exclusion restriction. Second, 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. Lastly, note that the magnitude of the upper bound can be directly compared to the estimates obtained from OLS, which identifies the same effect under the strong condition of selection on observables. It is also crucial to keep in mind that direct comparisons between the estimated upper bound and the IV estimate are not accurate because the effect that is bounded (ATE) 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 exclusion restriction, since the differences may be due to effect heterogeneity.

4. Results

Table 1 provides summary statistics for our Add Health and HRS samples. 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/m². 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 a substantially higher CES-D score and incidence of depression, but there are no noticeable differences in BMI.

¹⁹ The intuition for the form of the bound in this case is similar to that discussed for the bound in equation (4), now using the IV estimator $\beta_{V(1)}^{IV}$. In this case, Nevo & Rosen (2012) show that under our assumptions that $\sigma_{wu} > 0$, $\sigma_{zu} > 0$, and the fact that $\sigma_{wz} > 0$, having $\beta^{OLS} \leq \beta_z^{IV}$ implies $\beta_{V(1)}^{IV} \leq \beta^{OLS}$, and hence, $\min\{\beta^{OLS}, \beta_z^{IV}, \beta_{V(1)}^{IV}\} = \min\{\beta_z^{IV}, \beta_{V(1)}^{IV}\}$.

The HRS respondents are 68.18 years old on average in 2006 and 58% are women. Average BMI is 27.69 kg/m². The average CES-D score is 1.26, and 12% are depressed.²⁰ In comparison to young adults, there are noticeable gender differences in both BMI and mental health. On average, women have a slightly lower BMI but considerably 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 estimated associations 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/m² increase in BMI is associated with a 0.14 (0.028 x 5) unit increase in the CES-D score.^{21,22} This represents an increase of 2.4% relative to the mean CES-D score of 5.79. Similarly, the OLS estimate in panel B column 1 shows that a 5 kg/m² increase in BMI is associated with a 1 (0.002 x 5) percentage point increase (7% off the mean) in the likelihood of depression. Second, there are statistically significant differences in the estimated associations by gender. Looking at the CES-D score, for women, a 5 kg/m² increase in BMI is associated with a 0.31 unit (5%) increase in the CES-D score. In contrast, for men, a 5 kg/m² increase in BMI is associated with a decrease in the CES-D score by 0.014 units (3%). A similar pattern is seen for the depression indicator. Third, the estimated associations show that genetic predisposition to depression 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%) increase in the CES-D score, and a 2.3 percentage point (15%) increase in the incidence of depression. In contrast, the education PGS is negatively associated with mental health, but it is statistically insignificant.

Estimated associations 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 estimated association between BMI and mental health is larger for the elderly than for young adults. For example, in column 4, a 5 kg/m² increase in BMI is associated with a 0.145 (0.029 x 5) unit increase in the CES-D score, and a 1.5 (0.003 x 5) percentage point increase in the probability of depression. This is equivalent to a 12% and 13% increase in CES-D score and incidence of depression, respectively, relative to the corresponding sample means. In contrast to young adults, the estimated 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/m² increase in BMI is associated with a 10% (13%) increase in the CES-D score and a 14% (17%) increase in the probability of depression score for women (men),

²⁰ 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.

²¹ 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>).

²² A 5 kg/m² 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/m² 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).

respectively. Finally, it is interesting to note that, compared to young adults, innate ability—as measured by the education PGS—is statistically significantly associated with better mental health for the elderly (except for men in panel B). For example, the coefficient on the education PGS in column 4 of panel A indicates that, a 1 standard deviation increase in the education PGS is associated with a 7% decrease in the CES-D score relative to the sample mean.²³

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 that use 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 from equation (2). Column 1 presents results from an OLS regression of BMI on PGS for BMI and the 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 in Panel A (elderly individuals in Panel B, respectively), with a first stage F-statistic of 289 (529). The PGS for BMI thus satisfies assumption (A2) of a non-zero first stage effect. Columns 2-6 of Table 3 show estimates from conditional quantile regressions of BMI on PGS for BMI and the 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 provide indirect evidence about the monotonicity assumption (A3). Lastly, they provide our first suggestive characterization of the latent subpopulation for which the IV estimates identify a causal effect. We observe that the effect of the PGS for BMI on all the selected quantiles of the conditional BMI distribution shown in the table 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 indirect evidence in its favor (although not implying it holds, as the assumption is not directly testable).²⁴ Also observed is a suggestive pattern of increasing conditional quantile estimates from lower to higher quantiles, indicating that the IV has stronger effects on higher quantiles of the conditional BMI distribution.

Next, we conduct a more formal exercise to characterize the latent subpopulation for which the effect is estimated by IV methods, which is 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.²⁵ Intuitively, such weights allow to characterize the latent subpopulation for which the LATE parameter identified by IV estimates applies, since the effects for those complier individuals with larger weights contribute more to the effect that is estimated using 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

²³ The OLS estimates without the depression and education PGS controls are similar to those in Table 2 (see Appendix Table A1).

²⁴ 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).

²⁵ 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 $Z=1$) at a given level of the treatment. 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.

0.²⁶ 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.²⁷ 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/m² difference in BMI between treated (PGS for BMI>0) and untreated individuals at the 60th percentile of the BMI distribution. At the 70th (80th) percentile the difference in BMI is 2.60 kg/m² (2.80 kg/m²). This suggests that the groups of complier 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. This implies that, for both demographic groups, the IV estimate reflects the effect of BMI on mental health mostly for complier individuals in the upper quantiles of the BMI distribution. This interpretation is critical to understand the external validity of the estimates and the possible incidence of policies.

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 effect of BMI on mental health. In comparison, the IV estimates for elderly individuals are statistically significant and comparably large in magnitude. For example, IV estimates in column 4 indicate that the effect of a 5 kg/m² increase in BMI is 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 comparison, 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, which means that the instrument picks up some of the influence of these two PGSs when they are omitted.²⁸ This point is illustrated in Table 5, which presents IV estimates that do not control for the education and depression PGSs. The IV estimates are considerably larger in the HRS when not 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 the PGSs in Table 5 column 4 are 0.065 (0.009), reflecting an 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 20% larger at 0.024. 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 exclusion restriction assumption through other channels. Thus, to

²⁶ The IV estimates employing this binary IV are similar to the IV estimates employing the continuous PGS for BMI as the IV, as can be seen from Appendix Table A2.

²⁷ The weights = $(z/\text{phat}) + ((1-z)/(1-\text{phat}))$ where z is the binary instrument and phat is the predicted probability from the estimated propensity score model.

²⁸ In Add Health: the correlation (standard error) between (1) the BMI and education PGSs is -0.1731 (0.014), (2) the BMI and depression PGSs is 0.0322 (0.014), and (3) the depression and education PGSs is -0.1521 (0.014). In the HRS: the correlation (standard error) between (1) the BMI and education PGSs is -0.1707 (0.011), (2) the BMI and depression PGSs is -0.0152 (0.011), and (3) the depression and education PGSs is -0.1136 (0.011).

complement the evidence from the IV estimates, 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 section, we present evidence about the population ATE of BMI on mental health given by the parameter β in equation (1), by using the BMI PGS as a potentially imperfect IV based on the approach by Nevo & Rosen (2012) 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 u in equation (1) (see discussion in section 3.2), the bounding procedure yields an upper bound on the ATE (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 show the OLS and IV estimates from Tables 2 and 4, respectively. 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 on β under (A4)—that the instrument (PGS for BMI) has weakly the same direction of correlation with the error term u in equation (1) as the endogenous variable (BMI)—is given in column 5, and corresponds to the upper bound given in equation (4), which equals the min of the OLS and IV estimates. The estimated upper bound, while not able to rule out a null effect, indicates that the largest value of the ATE of BMI on the CES-D score for the population of young adults is 0.02. Based on this largest possible value for the effect, a 5 kg/m² increase in BMI would increase the average CES-D score in the population of young adults by at most a 0.1 unit (2%), suggesting that the average effect for young adults is, at most, small. Hence, in this case the estimated upper bound provides valuable information as it rules out plausible effects. The upper endpoint of the 95% confidence interval on the bounded parameter is 0.052.²⁹ Looking at the OLS estimate of β in equation (1), we see that it falls below the upper endpoint of the 95% confidence interval of the bounded parameter β . Since they both undertake statistical inference on the same parameter β (ATE), this implies that we cannot statistically rule out that the OLS estimate is unbiased for the true population average effect (β) using the estimated upper bound. The estimated upper bounds on β also suggest that 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 on β is consistent with a 5 kg/m² increase in BMI increasing the average CES-D score of young women by at most 2%. For men, the estimated upper bound is negative, although the upper endpoint of the 95% confidence interval on the bounded parameter β straddles zero, implying that a null effect is possible. Column 6 gives the estimated upper bounds on β when we add (A5)—that the instrument (PGI for BMI) is less correlated with the error term u in equation (1) than the endogenous variable (BMI). These bounds correspond to that in equation (5), which equals the min of the IV estimators using either PGS for BMI or $V(1)$ as an instrument (see section 3.2 for details). These estimated upper bounds on β turn out to be not much different from those in column 5 for young adults, the main difference being that the upper endpoints of the 95% confidence interval on the bounded parameter β are somewhat larger.

Results for elderly individuals in the HRS are given in panel B of Table 6. The estimated upper bound on the ATE (β) for the elderly population in column 5 equals 0.029, implying that for the elderly the average effect of a 5 kg/m² increase in BMI is, at most, an 12% increase in the CES-D score. Thus, this estimated upper bound on β is informative as it rules out plausible larger

²⁹ 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 statistical 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.

effects. Compared to the IV estimate (0.051), the estimated upper bound on β is about half as large. However, the lower endpoint of the 95% confidence interval of the IV estimate (note that its standard error is 0.015) is below the upper endpoint of the 95% confidence interval of the upper bound of β , making their difference non-statistically significant. Contrary to the young adult population, the estimated upper bounds on the corresponding ATE (β) for elderly women and men are very similar. Lastly, the estimated upper bounds on β 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 in magnitude.

Results for depression are shown in Table 7. In general, the patterns observed are similar to those documented in Table 6 for the CES-D score, although the differences in estimated upper bounds on β between young adults and the elderly are less pronounced. Looking at the estimated bounds on β under (A4) in column 5, for the population of young adults the largest possible value for the ATE indicates that a 5 kg/m² increase in BMI increases the probability of experiencing depression by at most 7%. For the elderly population, the estimated upper bound on β indicates a corresponding increase of at most 13% in the probability of experiencing depression. As before, the estimated upper bounds on β are able to rule out plausible larger effects.

Overall, the results from employing the Nevo & Rosen (2012) bounds can be summarized as follows. First, the estimated upper bounds on β for the population of young adults reinforce the general finding based on the IV estimates that the effect of BMI on mental health is at most small and arguably economically insignificant for young adults. Second, the estimated upper bounds on β are suggestive of economically significant effects of BMI on mental health for the elderly population. Lastly, the estimated upper bounds on β for both young adults and the elderly rule out plausible large effects (especially for young adults), and are thus informative.

4.4 Other Analyses

4.4.1 Obesity and Extreme 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, since the effect of BMI on mental health could be nonlinear, we have analyzed the effect of obesity and extreme obesity ($BMI \geq 40$) on mental health in Appendix Table A3. The results for obesity and extreme obesity are qualitatively similar to the BMI results. For young adults, the IV estimates show no statistically significant difference in mental health by obesity or extreme obesity status. We note that the IV estimates in Appendix Table A3 are somewhat large (although smaller than the OLS associations) but imprecisely estimated. This is suggestive of the effect of BMI on mental health for young adults possibly being nonlinear; however, there is not enough precision in this sample to have statistical confidence that the effects are non-zero. For elderly individuals, the IV estimates indicate that there are large statistically significant differences in mental health by obesity and extreme obesity status. For instance, 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. Noteworthy is that, for elderly individuals, when considering obesity and extreme obesity status, the IV estimates are generally larger and remain statistically significant, which is consistent 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 complier individuals on the upper part of the BMI distribution.

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 (second-stage effect). Appendix Table A4 shows OLS and IV estimates using the HRS sample. While OLS estimates are positive and statistically significant, IV estimates are statistically insignificant, which provides 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. A key concern is that mortality attrition changes the composition of the elderly sample in a way that is related to body mass and mental health. The direction of the potential resulting bias is not entirely clear. There is evidence documenting that mortality attrition in the HRS results in a sample with better self-reported health, less reported medical conditions, and higher socioeconomic status (e.g., Zajacova & Burgard, 2013). In light of this sample selection caused by mortality attrition, we conjecture that the potential resulting bias can be an underestimation of the true effect if the causal link between BMI and mental health is muted for generally healthier individuals. Therefore, as a robustness check, we weighted the IV regressions with the inverse of the probability of surviving until 2006. The probabilities are the predicted values from a logit regression where the dependent variable is 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). This exercise assumes that, conditional on the previous covariates, the potential effect of mortality attrition on our estimates can be ignored. As shown in Appendix Table A5, the IV estimates are qualitatively similar to those presented in Table 4.³⁰ The one exception is that the effect of BMI on depression for women becomes statistically insignificant. While not definitive, this exercise suggests that our results are not dramatically affected by mortality attrition.

4.4.3 Exploration of Some Mechanisms

In general, our analysis in sections 4.2 and 4.3 suggests that there are economically and statistically significant effects of BMI on mental health (as measured by the CES-D score and depression) for the elderly, but not for young adults. For the latter group, this result is consistent with previous findings in the literature. For example, Hung et al. (2014) find no effect of BMI on the probability of having a major depressive disorder using data from RADIANT, while Willage (2018) also finds no effect of BMI on the CES-D score using the sibling sample of the Add Health.³¹

As discussed in the introduction, BMI is hypothesized to affect mental health through various channels, including social stigma (which can lead to perceived or actual discrimination), body-image dissatisfaction, poor health, functional limitations, and biological processes (among others). To understand our findings, in this section we analyze the effect of BMI on some of these potential mechanisms through which BMI may affect mental health, using information available in the Add Health and the HRS. Importantly, the relevance of different mechanisms—in terms of both how much BMI affects a mechanism and how much that mechanism in turn affects mental health—may be different for young adults and the elderly. Based on prior literature (e.g., Himes 2000; Willage 2018), we hypothesize that the more relevant mechanisms for young adults are

³⁰ Interestingly, the estimated effects for males appear in line with the conjecture that mortality attrition may result in an underestimation of the true effect.

³¹ As mentioned in the introduction, Willage (2018) finds a statistically significant (at the 10 percent level) positive effect on suicidal ideation. However, when we estimate the effect of BMI on suicidal ideation using our more general sample from Add Health (rather than using the siblings sample), we find no statistically significant effect. Both our estimated effect and its standard error are substantially smaller than those in Willage (2018) (we omit the results for brevity, but they are available from the authors upon request). Recall that besides using a different sample, we also use a different PGS and we control for the PGSs for depression and education.

social stigma and self-image, while for the elderly are functional limitations and poor health, among others.

At the outset, it is important to note that the analysis below only looks at the effect of BMI on potential mechanisms. Thus, finding a statistically significant effect for a potential mechanism does not imply that such variable is indeed a mechanism. To statistically show that such a variable is a mechanism, one would also need to show that the effect of BMI on the potential mechanism in turn affects mental health, which would require a formal causal mediation analysis (e.g., Huber 2019). Given that causal mediation typically requires considerably stronger assumptions than those necessary to estimate “total” average effects (e.g., Flores & Flores-Lagunes 2009; Huber 2019), we follow prior literature and focus on estimating the effect of the treatment (in our case BMI) on potential mechanisms (e.g., Currie and Moretti 2003; Willage 2008).

Appendix Table A6 presents estimates (OLS and IV) of the effect of BMI on the following potential mechanisms for young adults: (i) the probability of being in very good/excellent health, (ii) the probability that the interviewer rates the respondent as being very attractive (which is likely to capture social stigma), and (iii) the probability that the respondent rates him/herself as being very attractive. In panel A, we find evidence that BMI decreases the probability of self-reporting very good/excellent health. The IV estimate in column 3 suggests that a 5 kg/m² increase in BMI decreases the likelihood of reporting being in very good or excellent health by 7 percentage points (11% relative to the sample mean). This estimated effect is present in both women and men. The estimates in panel B show that there is a relatively large effect of BMI on perceived attractiveness by the interviewer for women. For example, according to the IV estimate in column 3, the effect of a 5 kg/m² increase in BMI decreases the probability of women being viewed very attractive by 8 percentage points (73%). In panel C, the IV estimates show that there is no strong effect of BMI on self-reported attractiveness. In fact, the IV estimate for the sample implies a surprising positive relationship between BMI and self-reported attractiveness—which is only marginally statistically significant—that is driven by men. For men, the IV estimate is statistically significant and implies that a 5 kg/m² increase in BMI increases the probability of judging oneself very attractive by 5 percentage points (50%). Overall, for young adults, BMI has a statistically significant effect on potential mechanisms for the effect of BMI on mental health. However, given our results from the previous section that BMI does not statistically significantly affect mental health for young adults, we interpret these findings as suggesting that those potential mechanisms are not strong enough to produce a detectable effect of BMI on mental health, as measured using CES-D score, depression, and suicidal ideation (see footnote 31).

For the elderly, Appendix Table A7 shows results for the effect of BMI on the (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. Similar to young adults, we find that there are large effects of BMI on self-reported health for elderly individuals. The IV estimate in column 3 of panel A suggests that a 5 kg/m² increase in BMI decreases the likelihood of reporting being in very good or excellent health by 9.5 percentage points (20%). In panel B we see that, according to the IV estimate, the effect of a 5 kg/m² increase in BMI is a 6.5 percentage point (25%) decrease in the likelihood of engaging in vigorous exercise more than once per week. Looking at panel C, the IV estimate suggests that a 5 kg/m² increase in BMI increases the likelihood of reporting that health limits at least 1 daily activity by 5 percentage points (42%). We interpret these findings as suggesting that functional limitations and health concerns are among the mechanisms through which BMI may affect mental health for the elderly. Contrary to young adults, for the elderly these mechanisms appear strong enough to significantly impact their mental health given the reported statistically significant effects of BMI on mental health for this group in sections 4.2 and 4.3.

5. Conclusion

We study the effect of BMI on 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 previous studies, we attempt to address the possible violation of the exclusion restriction 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 to help characterize the latent subpopulation for which the IV estimates identify an effect. We document that IV estimates capture the effect of BMI on mental health for (complier) individuals in the upper quantiles of the BMI distribution. As a second, complementary approach, we estimate an upper bound on the ATE for the corresponding population of young adults or the elderly. We use the bounding 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 that it allows the PGS for BMI to be an “imperfect” instrument that may not satisfy the exclusion restriction.

For young adults, the IV estimates indicate that there is no statistically significant effect of BMI on mental health. The estimated upper bounds on the ATE 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). In contrast to young adults, the IV estimates for elderly individuals are statistically and economically significant. They indicate that a 5 kg/m² increase in BMI (a difference equivalent to moving from overweight to obese) increases the CES-D score by 20%, and the likelihood of depression by 29%. We find no statistically significant gender differences in the IV estimates. The estimated upper bounds on the ATE 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. Finally, our analysis of the effect of BMI on potential mechanisms through which BMI may affect mental health suggests that functional limitations and health concerns may be important mechanisms for the elderly. However, analyzing only the effect of BMI on potential mechanisms (a common practice in the literature) falls short on determining the causality of the mechanisms and quantifying their importance, as this would require the use of a formal causal mediation framework (e.g., Huber 2019). We leave the task of formally applying causal mediation in this setting for further research.

Overall, our paper provides new insights regarding the effect of BMI on mental health that can be valuable from a policy perspective, for example, when considering the costs of the ongoing obesity epidemic in the United States or the benefits of policies that have been implemented to fight it. First, our analysis suggests that the IV estimates of the kind used in this paper likely reflect a weighted average of effects of BMI on mental health for complier individuals on the upper quantiles of the BMI distribution. This is a critical consideration in using this kind of IV estimates for policy purposes. For example, since the compliers in the upper quantiles of the BMI distribution are a subset of individuals in those upper quantiles, our estimates suggest that policies targeting the upper quantiles of the BMI distribution (e.g., obesity or extreme obesity) would have significant effects on mental health at least for a subset of the individuals in those upper quantiles.³² Second, taken together, our results from the IV estimates and the estimated bounds on the ATE suggest that the effect of BMI on mental health is economically significant for the elderly, but not for young

³² There exists empirical evidence that some policies to reduce BMI have larger effects on the upper part of the BMI distribution. For example, Bonanno et al. (2018) find that reading food nutritional labels reduces BMI mostly at higher BMI quartiles. Han & Powell (2013) find that fiscal pricing policies, such as fast food taxes, are likely to have larger impacts on BMI for low income women or women with children on the upper tail of the conditional BMI distribution. Wehby & Courtemanche (2012) find that the negative effect of cigarette prices on BMI is more than three times as strong at high BMI levels.

adults.³³ Hence, for example, if a policy were to affect the BMI of different age groups in a similar way, our results would suggest that the indirect benefits of such policies in terms of improved mental health are likely larger for the elderly than for young adults. Third, our analysis of potential mechanisms through which BMI may affect mental health suggests that, for the elderly, functional limitations and health concerns are potentially important mechanisms. Taken at face value, this suggests, for example, that policies targeted at reducing the functional limitations of the elderly (e.g., by subsidizing appropriate medical equipment or human assistance) may ameliorate the negative effects of BMI on the mental health for them.

There are some limitations with our paper. For instance, 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.

6. Acknowledgements

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.

³³ We give this description in terms of the population of young adults and the elderly (and not only compliers) given that the Nevo & Rosen (2012) upper bounds we estimate are on the ATE for the corresponding population. See section 3.2.

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.
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, 4(6), 561-571.
- Beck, A. T., & Steer, R. A. (1993). *Beck Depression Inventory manual*. San Antonio, TX: Harcourt.
- Bonanno, A., Bimbo, F., Cleary, R., & Castellari, E. (2018). Food labels and adult BMI in Italy—An unconditional quantile regression approach. *Food Policy*, 74, 199-211.
- 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.
- Crisp, A. H., Queenan, M., Sittampaln, Y., & Harris, G. (1980). 'Jolly fat' revisited. *Journal of Psychosomatic Research*, 24(5), 233-241.
- Currie, J., & Moretti, E. (2003). Mother's education and the intergenerational transmission of human capital: Evidence from college openings. *Quarterly Journal of Economics*, 118(4), 1495-1532.
- 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.
- Fletcher, J. M. (2018). Economics and Genomics. In *Oxford Research Encyclopedia of Economics and Finance*.

- Flores, C. A. & Flores-Lagunes, A. (2009). Identification and estimation of causal mechanisms and net effects of a treatment under unconfoundedness. IZA DP No. 4237.
- 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.
- Han, E., & Powell, L. M. (2013). Fast food prices and adult body weight outcomes: evidence based on longitudinal quantile regression models. *Contemporary Economic Policy*, 31(3), 528-536.
- Himes, C. L. (2000). Obesity, disease, and functional limitation in later life. *Demography*, 37(1), 73-82.
- Huber, M. (2019). A review of causal mediation analysis for assessing direct and indirect treatment effects. SES Working Paper 500, University of Fribourg.
- 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. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and Statistics*, 86(1), 4-29.
- Imbens, G., & Angrist, J. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2), 467-475.
- Insel, T. R. (2008). Assessing the economic costs of serious mental illness. *The American Journal of Psychiatry*, 165(6), 663-665.
- 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.
- Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghziyan, O., Zacher, M., ... & Fontana, M. A. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics*, 50(8), 1112.

- 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>
- National Institute of Mental Health, "Mental health Statistics". Retrieved from: <https://www.nimh.nih.gov/health/statistics/index.shtml>
- 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.
- Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*(3), 385-401.

- 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.
- Skorikov, V. B. & Vandervoort, D. J. (2003). Relationships between the underlying constructs of the beck depression inventory and the center for epidemiological studies depression scale. *Educational and Psychological Measurement*, 63(2), 319-335.
- 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>).
- Substance Abuse and Mental Health Services Administration (2016). Behavioral Health Spending and Use Accounts, 1986–2014. HHS Publication No. SMA-16-4975.
- 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., ... & Kardina, 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.

Wehby, G. L., & Courtemanche, C. J. (2012). The heterogeneity of the cigarette price effect on body mass index. *Journal of Health Economics*, 31(5), 719-729.

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.

Zajacova, A. & Burgard, S. (2013). Healthier, wealthier, and wiser: A demonstration of compositional changes in aging cohorts due to selective mortality. *Population Research and Policy Review*, 32(3), 311-324.

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 (1)	Women (2)	Men (3)	All (4)	Women (5)	Men (6)
Age	28.94 (1.73)	28.82 (1.75)	29.01 (1.72)	68.18 (9.84)	68.05 (10.06)	68.36 (9.54)
Female	0.54 (0.49)	---	---	0.58 (0.49)	---	---
CES-D	5.79 (4.65)	6.18 (4.89)	5.33 (4.31)	1.26 (1.33)	1.43 (1.94)	1.03 (1.61)
Depressed	0.15 (.35)	0.17 (.37)	0.12 (.32)	0.12 (.32)	0.14 (.35)	0.09 (.28)
BMI	28.56 (7.14)	28.48 (7.66)	28.67 (6.49)	27.69 (5.51)	27.42 (5.96)	28.05 (4.83)
Obese	0.34 (.47)	0.34 (.47)	0.35 (.48)	0.28 (.45)	0.28 (.45)	0.28 (.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						
<i>Outcome Mean</i>	5.79	6.18	5.33	1.26	1.43	1.03
BMI	0.028*** (.010)	0.062*** (.014)	-0.028** (.014)	0.029*** (.004)	0.030*** (.005)	0.026*** (.006)
Depression PGS	0.359*** (.086)	0.345*** (.126)	0.389*** (.116)	0.130*** (.020)	0.121*** (.029)	0.104*** (.027)
Education PGS	-0.101 (.068)	-0.087 (.098)	-0.010 (.096)	-0.085*** (.068)	-0.099*** (.098)	-0.064*** (.096)
B: Depressed						
<i>Outcome Mean</i>	0.15	0.17	0.12	0.12	0.14	0.09
BMI	0.002*** (.001)	0.004*** (.001)	-0.002 (.001)	0.003*** (.001)	0.004*** (.001)	0.003** (.001)
Depression PGS	0.023*** (.007)	0.024** (.010)	0.022** (.009)	0.017*** (.004)	0.019*** (.005)	0.014*** (.005)
Education PGS	-0.005 (.005)	-0.004 (.008)	-0.004 (.007)	-0.013*** (.004)	-0.018*** (.005)	-0.007 (.007)
N	4928	2643	2285	8867	5104	3763

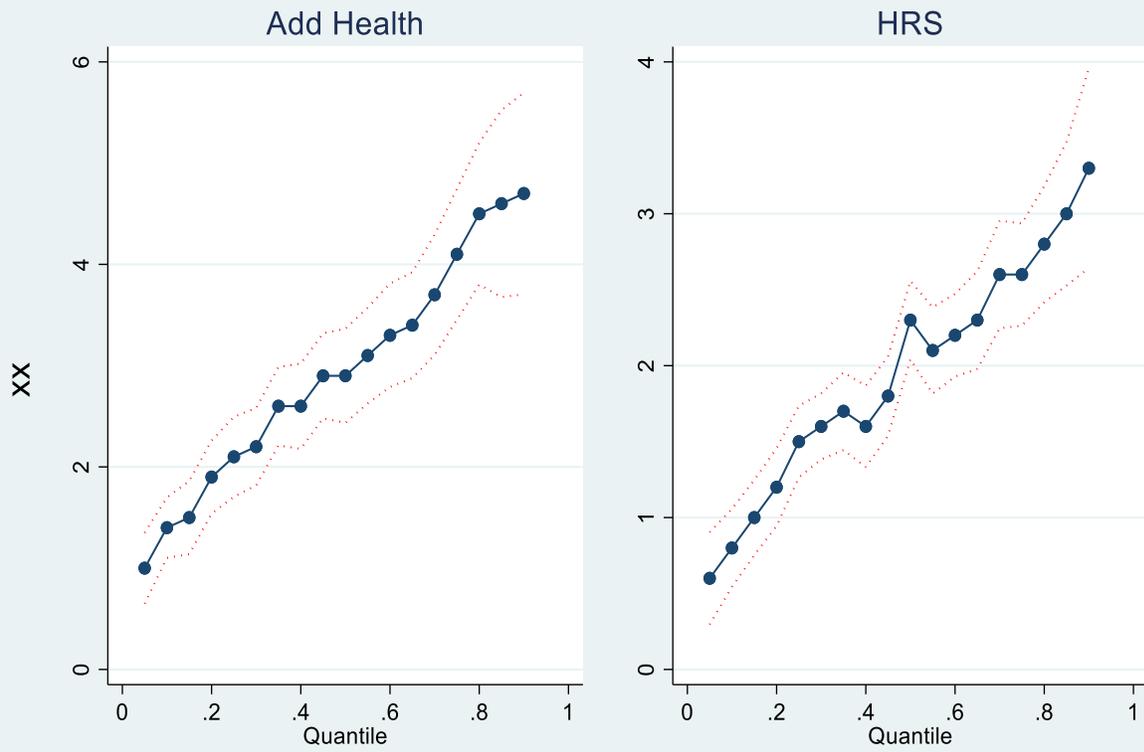
Notes: Add Health regressions in columns 1-3 control for age, age squared, gender, birth order, mother's education, picture vocabulary score, and the first 20 ancestry-specific principal components of the genetic data. HRS regressions in columns 4-6 control for age, age squared, gender, mother's education and the first 10 ancestry-specific principal components of the genetic data. Heteroscedasticity-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*** (.101)	0.861*** (.071)	1.252*** (.074)	1.698*** (.102)	2.269*** (.149)	2.788*** (.242)
Depression PGS	0.220* (.129)	-0.064 (.094)	0.189* (.100)	0.341** (.135)	0.442** (.197)	0.252 (.310)
Education PGS	-0.119 (.104)	0.124* (.073)	-0.029 (.078)	-0.013 (.105)	-0.224 (.154)	-0.415 (.249)
F-Statistic	289	144	256	256	225	144
N	4928	4928	4928	4928	4928	4928
Panel B:						
HRS						
BMI PGS	1.403*** (.060)	0.671*** (.062)	0.936*** (.056)	1.315*** (.060)	1.699*** (.083)	2.079*** (.129)
Depression PGS	-0.015 (.058)	-0.033 (.062)	0.022 (.054)	-0.012 (.061)	0.023 (.083)	-0.035 (.126)
Education PGS	-0.089 (.058)	-0.092 (.062)	-0.089 (.056)	-0.006 (.060)	-0.002 (.083)	-0.112 (.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. Heteroscedasticity-robust standard errors in parentheses. ***significant at 1% **significant at 5% *significant at 10%

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



Notes: The 95% confidence interval is given by the dotted red lines

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						
<i>Outcome Mean</i>	5.79	6.18	5.33	1.26	1.43	1.03
BMI	0.020 (.038)	0.019 (.054)	0.020 (.053)	0.051*** (.015)	0.053*** (.019)	0.046*** (.023)
Depression PGS	0.360*** (.086)	0.353*** (.125)	0.386 (.117)	0.128*** (.020)	0.122*** (.029)	0.138*** (.027)
Education PGS	-0.104 (.076)	-0.111 (.102)	-0.086 (.096)	-0.078*** (.021)	-0.088*** (.031)	-0.062*** (.029)
B: Depressed						
<i>Outcome Mean</i>	0.15	0.17	0.12	0.12	0.14	0.09
BMI	0.003 (.003)	0.004 (.004)	0.001 (.004)	0.007*** (.003)	0.007** (.004)	0.007* (.004)
Depression PGS	0.022*** (.007)	0.024** (.010)	0.022 (.009)**	0.024*** (.006)	0.019*** (.005)	0.013** (.005)
Education PGS	-0.004 (.005)	-0.003 (.008)	-0.003 (.007)	-0.010 (.006)	-0.016*** (.006)	-0.006 (.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. Heteroscedasticity-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 (1)	Women (2)	Men (3)	All (4)	Women (5)	Men (6)
A: CESD-Score						
<i>Outcome Mean</i>	5.79	6.18	5.33	1.26	1.43	1.03
BMI	0.024 (.037)	0.022 (.053)	0.026 (.052)	0.065*** (.014)	0.070*** (.019)	0.056*** (.019)
B: Depressed						
<i>Outcome Mean</i>	0.15	0.17	0.12	0.12	0.14	0.09
BMI	0.003 (.003)	0.004 (.004)	0.001 (.004)	0.009*** (.003)	0.010*** (.003)	0.008*** (.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. Heteroscedasticity-robust standard errors in parentheses. ***significant at 1% **significant at 5% *significant at 10%

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

	Mean	OLS	IV	IV(1)	UB of β under (A4)	UB of β under (A4 & A5)
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A:						
Add health						
All	5.79	0.028*** (.010)	0.020 (.038)	0.032** (.016)	0.020 [.052]	0.020 [.066]
Women	6.18	0.062*** (.014)	0.019 (.054)	0.075*** (.020)	0.019 [.094]	0.019 [.118]
Men	5.33	-0.028** (.014)	0.020 (.053)	-0.045* (.023)	-0.028 [-.003]	-0.027 [.010]
Panel B:						
HRS						
All	1.26	0.029*** (.004)	0.051*** (.015)	0.022*** (.006)	0.029 [.037]	0.022 [.035]
Women	1.43	0.030*** (.005)	0.053*** (.019)	0.023*** (.008)	0.030 [.041]	.023 [.040]
Men	1.03	0.026*** (.006)	0.046*** (.023)	0.019** (.010)	0.026 [.041]	0.018 [.042]

Notes: Add Health regressions 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 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. Heteroscedasticity-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 Estimates for the Effect of BMI on Depression

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

Notes: Add Health regressions 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 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. Heteroscedasticity-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 [.]

ONLINE APPENDIX –NOT INTENDED FOR PUBLICATION

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 (1)	Women (2)	Men (3)	All (4)	Women (5)	Men (6)
A: CESD-Score						
<i>Outcome Mean</i>	5.79	6.18	5.33	1.06	1.43	1.03
BMI	0.030*** (.010)	0.063*** (.014)	-0.027* (.014)	0.030*** (.004)	0.031*** (.005)	0.027*** (.007)
B: Depressed						
<i>Outcome Mean</i>	0.15	0.17	0.12	0.12	0.14	0.09
BMI	0.002*** (.001)	0.004*** (.001)	-0.001 (.001)	0.004*** (.001)	0.004*** (.001)	0.003** (.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. Heteroscedasticity-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 (1)	Women (2)	Men (3)	All (4)	Women (5)	Men (6)
A: CESD-Score						
<i>Outcome Mean</i>	5.79	6.18	5.33	1.26	1.43	1.03
BMI	0.039 (.043)	0.009 (.060)	0.061 (.069)	0.063*** (.018)	0.071*** (.025)	0.051* (.023)
B: Depressed						
<i>Outcome Mean</i>	0.15	0.17	0.12	0.12	0.14	0.09
BMI	0.004 (.003)	0.004 (.005)	0.004 (.005)	0.008*** (.003)	0.009** (.004)	0.004 (.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. Heteroscedasticity-robust standard errors in parentheses. ***significant at 1% **significant at 5% *significant at 10%

Appendix Table A3: Differences in Mental Health by Obesity Status

	CES-D OLS (1)	CES-D IV (2)	Depressed OLS (3)	Depressed IV (4)
Panel A: Add Health				
Obese				
All	0.507*** (.142)	0.334 (.638)	0.027** (.011)	0.050 (.050)
Women	0.913*** (.211)	0.383 (1.07)	0.059*** (.016)	0.086 (.082)
Men	0.008 (.189)	0.284 (.758)	-0.010 (.014)	0.010 (.060)
Panel B: Add Health, Severely Obese				
All	0.715*** (.275)	1.130 (2.17)	0.032 (.021)	0.170 (.169)
Women	1.040*** (.386)	0.961 (2.69)	0.053* (.029)	0.216 (.208)
Men	0.129 (.366)	1.402 (3.74)	-0.006 (.028)	0.050 (.297)
Panel C: HRS, Obese				
All	0.311*** (.045)	0.705*** (.206)	0.036*** (.008)	0.102*** (.037)
Women	0.332*** (.064)	0.800*** (.293)	0.037*** (.012)	0.106*** (.052)
Men	0.276*** (.061)	0.568** (.284)	0.033*** (.011)	0.090* (.050)
Panel D: HRS, Severely Obese				
All	0.750*** (.137)	3.592*** (1.087)	0.106*** (.026)	0.521*** (.192)
Women	0.754*** (.176)	3.648*** (1.395)	0.118*** (.033)	0.484* (.256)
Men	0.722*** (.212)	3.356*** (1.707)	0.080** (.040)	0.533* (.299)

Notes: Add Health regressions 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 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. Heteroscedasticity-robust standard errors in parentheses. ***significant at 1% **significant at 5% *significant at 10%

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

	OLS (1)	IV (2)
Panel A: CES-D Score		
All [N=8867]	0.231*** (.036)	-0.117 (.450)
Women [N=5104]	0.240*** (.045)	0.631 (.560)
Men [N=3763]	0.205*** (.056)	-0.716 (.745)
Panel B: Depressed		
All [N=8867]	0.846*** (.206)	-0.914 (3.56)
Women [N=5104]	0.878*** (.260)	6.459 (5.98)
Men [N=3763]	0.724** (.332)	-4.623 (4.85)

Notes: In panel A mental health is measured using the CES-D score. In panel B mental health is measured using the depression indicator. 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. Heteroscedasticity-robust standard errors in parentheses.

***significant at 1% **significant at 5% *significant at 10%

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)
Panel A: CES-D Score			
Outcome Mean	1.26	1.43	1.03
BMI	0.051*** (.018)	0.043* (.023)	0.061* (.031)
Panel B: Depressed			
Outcome Mean	0.12	0.14	0.09
BMI	0.008** (.003)	0.004 (.004)	0.013** (.006)
N	8,685	5,028	3,657

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. Heteroscedasticity-robust standard errors in parentheses. All regressions are weighted by the inverse probability of being alive in 2006. ***significant at 1% **significant at 5% *significant at 10%.

Appendix Table A6: Effect of BMI on Potential Mechanisms for Young Adults

	(1) Mean	(2) OLS	(3) IV
Panel A: Probability of being in very good/excellent health			
All [N=4928]	0.61	-0.018*** (.001)	-0.014*** (.004)
Women [N=2643]	0.61	-0.020*** (.001)	-0.025*** (.005)
Men [N=2285]	0.60	-0.017*** (.002)	-0.021*** (.006)
Panel B: Very attractive according to interviewer			
	Mean	OLS	IV
All [N=4925]	0.10	-0.006*** (.001)	-0.010*** (.002)
Women [N=2642]	0.11	-0.008*** (.001)	-0.016** (.004)
Men [N=2283]	0.07	-0.002*** (.001)	-0.002 (.003)
Panel C: Very attractive according to oneself			
	Mean	OLS	IV
All [N=4919]	0.08	-0.002*** (.001)	0.004* (.002)
Women [N=2638]	0.06	-0.001* (.001)	0.000 (.003)
Men [N=2281]	0.10	-0.002* (.001)	0.010** (.003)

Notes: All regressions control for age, age squared, gender, mother's education, picture vocabulary score, PGSs for education and depression, and the first 20 ancestry-specific principal components of the genetic data. Heteroscedasticity-robust standard errors in (.). ***significant at 1% **significant at 5% *significant at 10%.

Appendix Table A7: Effect of BMI on Health and Exercise for Older Adults

	(1) Mean	(2) OLS	(3) IV
Panel A: Self-reported health			
All [N=8861]	0.48	-0.016*** (.001)	-0.019*** (.004)
Women [N=5098]	0.49	-0.016*** (.001)	-0.018*** (.005)
Men [N=225]	0.48	-0.014*** (.002)	-0.021*** (.007)
Panel B: Vigorous exercise more than once per week			
All [N=8867]	0.26	-0.010*** (.001)	-0.013*** (.003)
Women [N=5104]	0.23	-0.009*** (.001)	-0.011** (.004)
Men [N=3763]	0.32	-0.012*** (.001)	-0.016 (.006)
Panel C: Health limits daily activities			
All [N=8867]	0.12	0.009*** (.001)	0.010*** (.003)
Women [N=5104]	0.13	0.009*** (.001)	0.011*** (.003)
Men [N=3763]	0.11	0.009*** (.001)	0.007* (.004)

Notes: All 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. Heteroscedasticity-robust standard errors in (.). ***significant at 1% **significant at 5% *significant at 10%.