This is the post print version of the article, which has been published in Health economics. 2019, 28 (1),65-77.https://doi.org/10.1002/hec.3828.

The Effect of Weight on Labor Market Outcomes: an Application of Genetic Instrumental Variables

Petri Böckerman,* John Cawley,** Jutta Viinikainen,*** Terho Lehtimäki,**** Suvi Rovio,# Ilkka Seppälä,## Jaakko Pehkonen### and Olli Raitakari####

* University of Jyväskylä, Labour Institute for Economic Research, Helsinki, Finland and IZA. E-mail: petri.bockerman@labour.fi
** Department of Policy Analysis and Management, and Department of Economics, Cornell University, Ithaca NY 14853 USA. E-mail: JHC38@cornell.edu
*** University of Jyväskylä, Jyväskylä, Finland. E-mail: jutta.viinikainen@jyu.fi
**** Department of Clinical Chemistry, Fimlab Laboratories and Finnish Cardiovascular Research Center Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Finland. E-mail: terho.lehtimaki@uta.fi
# Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland. E-mail: suvi.rovio@utu.fi
Department of Clinical Chemistry, Fimlab Laboratories and Finnish Cardiovascular Research Center Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Finland. E-mail: ilkka.seppala@uta.fi
University of Jyväskylä, Jyväskylä, Finland. E-mail: jaakko.k.pehkonen@jyu.fi
Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku and Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland. E-mail: olli.raitakari@utu.fi

July 13, 2018

ABSTRACT

This paper contributes to the literature on the labor market consequences of obesity by using a novel instrument: genetic risk score, which reflects the predisposition to higher body mass index across many genetic loci. We estimate IV models of the effect of BMI on labor market outcomes using Finnish data that have many strengths: e.g. body mass index that is measured rather than self-reported, and data on earnings and social income transfers that are from administrative tax records and are thus free of the problems associated with non-response, reporting error or top-coding.

The empirical results are sensitive to whether we use a narrower or broader genetic risk score, and to model specification. For example, models using the narrower genetic risk score as an instrument imply that a one-unit increase in BMI is associated with 6.9% lower wages, 1.8% fewer years employed, and a 3 percentage point higher probability of receiving any social income transfers. However, when we use a newer, broader, genetic risk score, we cannot reject the null hypothesis of no effect. Future research using genetic risk scores should examine the sensitivity of their results to the risk score used.

KEYWORDS: Obesity, Overweight, Earnings, Employment, Social income transfers, Genetics

JEL CLASSIFICATION: D62, I1, I12, J01, J24, J3, J7

Acknowledgements

The Young Finns Study has been financially supported by the Academy of Finland: grants 286284, 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117787 (Gendi), and 41071 (Skidi); the Social Insurance Institution of Finland; Competitive State Research Financing of the Expert Responsibility area of Kuopio, Tampere and Turku University Hospitals (grant X51001); Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation for Cardiovascular Research ; Finnish Cultural Foundation; Tampere Tuberculosis Foundation; Emil Aaltonen Foundation; Yrjö Jahnsson Foundation; Signe and Ane Gyllenberg Foundation; Diabetes Research Foundation of Finnish Diabetes Association; and EU Horizon 2020 (grant 755320 for TAXINOMISIS). The use of linked data is supported by the Palkansaaja Foundation. Jutta Viinikainen and Jaakko Pehkonen acknowledge the financial support from theYrjö Jahnsson foundation (grants 6646 and 6664). Petri Böckerman thanks the Strategic Research Council funding for the project Work, Inequality and Public Policy (293120). John Cawley thanks the Robert Wood Johnson Foundation for an Investigator Award in Health Policy Research. We thank Barton Willage for research assistance. Finally, we also thank two anonymous reviewers, the seminar participants at the EALE, Oulu Business School and the VATT seminar and Jani-Petri Laamanen for helpful comments.

Data availability and disclosure statement

The data used in this study are confidential but other researchers can independently obtain access to it for replication purposes by the permission from Statistics Finland. To obtain access to the data, please contact Statistics Finland, FI-00022, Helsinki, Finland. The specific instructions to obtain access to the data are available at https://tilastokeskus.fi/tup/mikroaineistot/hakumenettely_en.html

The authors will provide guidance about acquiring the data upon request.

All participants of the Young Finns Study (YFS) provided written informed consent, and the study was approved by local institutional review boards (ethics committees of the participating universities). Parents or guardians provided written informed consent on behalf of the under aged children enrolled in the study.

The study does not disclose information concerning individual persons. The final linked YFS-FLEED data have been approved for research purposes by Statistics Finland, under the ethical guidelines of the institution which comply with the national standards.

All authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

Running head

The Effect of Weight on Labor Market Outcomes

The Effect of Weight on Labor Market Outcomes:
an Application of Genetic Instrumental Variables

July 13, 2018

Abstract

This paper contributes to the literature on the labor market consequences of obesity by using a novel instrument: genetic risk score, which reflects the predisposition to higher body mass index across many genetic loci. We estimate IV models of the effect of BMI on labor market outcomes using Finnish data that have many strengths: e.g. body mass index that is measured rather than self-reported, and data on earnings and social income transfers that are from administrative tax records and are thus free of the problems associated with non-response, reporting error or top-coding.

The empirical results are sensitive to whether we use a narrower or broader genetic risk score, and to model specification. For example, models using the narrower genetic risk score as an instrument imply that a one-unit increase in BMI is associated with 6.9% lower wages, 1.8% fewer years employed, and a 3 percentage point higher probability of receiving any social income transfers. However, when we use a newer, broader, genetic risk score, we cannot reject the null hypothesis of no effect. Future research using genetic risk scores should examine the sensitivity of their results to the risk score used.

KEYWORDS: Obesity, Overweight, Earnings, Employment, Social income transfers, Genetics

JEL CLASSIFICATION: D62, I1, I12, J01, J24, J3, J7

1. INTRODUCTION

The prevalence of obesity, defined as a body mass index (BMI) of 30 or higher,[1] has risen dramatically in many countries in the past several decades (GBD 2015 Obesity Collaborators, 2017). Economists have extensively studied the economic consequences of obesity, in particular whether obesity lowers wages or reduces the probability of employment (see the reviews in Averett, 2011; Cawley, 2015). Obesity could result in worse labor market performance for several reasons; e.g. obesity worsens health (Hu, 2008) which may lower productivity and thus wages, and there may be obesity-related discrimination in the labor market (Puhl, 2011; Rooth, 2009).

Correlations between weight and labor market outcomes are difficult to interpret. They reflect not only any impact of weight on earnings, but also any reverse causality that would arise if a low income results in weight gain (see e.g. Schmeiser, 2009), and the influence of any omitted variables such as rate of time preference (Komlos *et al.,* 2004). For this reason, numerous studies have sought to estimate the causal effect of weight on labor market outcomes. Most have instrumented for respondent weight using the weight of a biological relative; e.g. Cawley (2004), Brunello and D'Hombres (2007), Kline and Tobias (2008), and Lindeboom et al. (2010). This approach takes advantage of the substantial genetic variation in weight; genetics studies estimate a strong heritable component of BMI, roughly 40-70% (Barsh *et al.,* 2000; Pietiläinen *et al.,* 1999; Locke *et al.,* 2015). A potential concern with the approach is that unobserved characteristics may be correlated with both a person's own BMI and their relative's BMI.

[1] BMI is calculated as a person's weight in kilograms divided by height in meters squared.

This paper contributes to the literature by using a novel instrument – genetic risk score for high BMI – to estimate the causal effect of weight on labor market outcomes. This instrument takes advantage of the natural experiment known as Mendelian Randomization, which refers to the draw of an individual's genotype at conception (Conley, 2016; Haycock *et al.,* 2016; von Hinke *et al.,* 2016; Tyrrell *et al.,* 2016; Davey Smith *et al.,* 2017).[2] We utilize two different genetic risk scores for high BMI; a narrower one based on 32 SNPs[3] and a broader one based on 97 SNPs that have robustly and significantly been found to influence obesity in genome-wide association studies involving very large samples. We discuss the power and validity of these instruments in detail in the Methods section.

In addition to the genetic IVs, the data we use has three noteworthy strengths. First, it includes measurements, as opposed to self-reports, of weight and height; thus, we avoid the problems arising from reporting error in weight such as inefficiency and bias (see Cawley *et al.,* 2015; Courtemanche *et al.,* 2015). Second, we utilize administrative data on earnings, which avoids problems associated with refusal to report, and reporting error in wages and salaries. Third, the data allow us to examine a novel outcome – social income transfers, taken from administrative records, which allows us to examine one potential negative externality related to obesity. Studying the existence and magnitude of such external costs is important because they may be associated with deadweight loss to society and thus represent an economic rationale for government intervention to prevent and reduce obesity (Cawley, 2015). Social

[2] Norton and Han (2008) used genetic information as an instrument for weight to estimate the effect of weight on labor market outcomes, although the specific genetic IVs were later called into question as relatively weak and likely invalid (Cawley, Han, and Norton, 2011; von Hinke *et al.,* 2016).
[3] Places where DNA differ between people are called polymorphisms, and a single nucleotide polymorphism or SNP is a single base-pair variation in DNA. Humans have two copies of each chromosome, so they have two alleles, or versions, of each SNP that may be the same (homozygous) or different (heterozygous). See Appendix A of von Hinke Kessler Scholder et al. (2012).

income transfers are also of interest because they tend to be a substantial item in government budgets, especially in countries like Finland with a comprehensive social safety net.

One study similar to this one is Tyrrell et al. (2016), which estimated the effect of obesity on labor market outcomes using genetic information as an instrument for BMI, and found that higher BMI lowered annual household income for women. Our study differs in significant ways from theirs. We use a measure of income that is more accurate for several reasons. Tyrrell et al. (2016) used a self-reported categorical income; in contrast, our paper uses administrative income data from national registers which avoids problems of refusal and reporting error. In addition, our measure of income is continuous rather than categorical. Moreover, the self-reported income in Tyrrell et al. (2016) is for the household, whereas our administrative information measures earnings specific to the individual that are more relevant for studying the consequences of individual weight. Furthermore, Tyrell et al. (2016) examined outcomes in a single year whereas we examine outcomes over 12 years. We also examine additional outcomes – employment, and receipt of social income transfers – and use data from Finland instead of the U.K.

2. DATA

We link data from three sources: 1) the Cardiovascular Risk in Young Finns Study (YFS); 2) the Finnish Longitudinal Employer-Employee Data (FLEED) of Statistics Finland (SF); and 3) the Longitudinal Population Census (LPC) of SF. The merge is executed using unique personal identifiers, which is exact matching; i.e. there are no misreported ID codes.

The YFS is an on-going epidemiological study that began in 1980 with the goal of examining how childhood cardiovascular risk factors and health behaviors, as

well as biological and psychological factors, contribute to the risk of cardiovascular diseases in adulthood.[4] Subjects in six age cohorts (aged 3, 6, 9, 12, 15 and 18 years) were randomly chosen from the five university hospital districts of Finland using the national population register (Raitakari *et al.,* 2008).[5] The sample is relatively small – 3,596 persons participated in the study at baseline – but the richness of the data are an offsetting advantage. Eight waves of data have been collected in 3-9 year intervals, starting with baseline in 1980 and most recently in 2011-12, with response rates between 60% and 80%. We use data from the 2001, 2007 and 2011 waves, because we have, from another source, labor market data for the years 1990-2012 (as we explain below).

The YFS data are collected through questionnaires, physical measurements, and blood tests. In all waves of the YFS, weight and height were measured to the nearest 0.1 kg by medical professionals at local health centers.

In 2009, genome-wide association studies (GWAS) were performed for YFS subjects using the 670K Illumina platform (Sanger Institute, UK). Variation in over 670,000 known single nucleotide polymorphisms (SNPs) were measured from 2,450 study subjects. Imputation for up to 2.5 million SNPs was performed using information on Hapmap 2 by using MACH. All the SNPs were imputed with excellent imputation quality (MACH $r^2>0.8$). These genetic data were used to construct the genetic risk scores, which will be explained in detail in the Methods section.

The second dataset that we use, the Finnish Longitudinal Employer-Employee Data (FLEED), is the source for data on employment status, salary, and other income,

[4] YFS is the largest running follow-up study in Europe that evaluates cardiovascular risk factors from childhood to adulthood; see http://youngfinnsstudy.utu.fi/studydesign.html

[5] Finland is divided into 20 hospital districts, five of which are university hospital districts.

for 2001 to 2012. FLEED data come directly from tax and other administrative registers that are collected and/or maintained by SF. Such register-based data have much less measurement error than self-reports from surveys; e.g. the income data in FLEED do not suffer from underreporting or recall error, nor are they top coded. This accuracy increases the efficiency of the estimates, which is particularly important for relatively small samples such as the YFS. The third dataset that we use, the Longitudinal Population Census (LPC), is the source of information on parental education.

3. METHODS

We estimate regressions of the following form:

$$Y_i = \alpha + W_i\beta + X_i\gamma + \varepsilon_i \quad (1)$$

We examine four labor market outcomes *Y* for an individual *i*. Out of a concern that short-term cross-sectional measures, such as yearly earnings or current employment status, contain idiosyncratic components that diminish the precision of the estimates (Dahl *et al.*, 2011), in this paper, *Y* is usually the average of the values over 2001 to 2012, which is the period that most respondents were of working age (i.e. between 24 and 50). Because the sample size is relatively small, reduction in variance and precision gains from averaging over several periods are important.

The first outcome we examine is the logarithm of the average of the individual's annual wage and salary earnings over 2001-2012. The second dependent variable is labor market attachment; specifically: the share of years employed during 2001-2012, with employment status in a year classified by the individual's status in the last week of each year in FLEED. Retirement is not an issue for this sample; the YFS participants are between 35 and 50 years old in 2012.

The third dependent variable we examine is an indicator variable for whether the respondent received any social income transfers between 2001 and 2012, and the fourth and final dependent variable is the logarithm of the average of the individual's annual social income transfers over the period 2001-2012, conditional on receiving any. These last two outcomes jointly represent a two-part model, in which the first part concerns whether the respondent received any social income transfers, and the second part concerns the average annual amount received, conditional on receiving any. Social income transfers include unemployment benefits, housing benefits and disability payments.[6] We examine their receipt for two reasons: first, social income transfers are an important component of total income in Finland, a welfare state, and second, the amount of social income transfers is informative about adverse labor market consequences and negative externalities related to obesity. For both wages/earnings and social income transfers, the values in each year are converted to constant (inflation-adjusted) values using the consumer price index (base year 2000) before the average is calculated.

The regressor of interest is weight *W*. In our primary models, we use BMI measured in 2001. Thus, the main estimates are based on cross-sectional variation across individuals in the value of BMI in 2001. BMI has only limited variation for each individual over the relatively short observation window (2001, 2007 and 2011) and thus it is not feasible to estimate individual fixed effects models.

[6] We exclude parental leave benefits from social income transfers because they are not a "negative indicator" in the same way as unemployment benefits and disability payments. Parental leave benefits are also strongly earnings-related in Finland. We focus on social income transfers that are indicators of poor labor market success and markers of negative externalities related to obesity. We have estimated the baseline models also by including parental leave benefits to the measure of social income transfers. The conclusions remain intact. Persons being on parental leave are coded as employed according to Statistics Finland, because parental leave does not dissolve the legal status of employment contract.

The vector of controls *X* includes indicator variables for birth month, birth year and gender. Typically, wage equations include education as a regressor (Mincer, 1974). In this case that practice is questionable, as there is some evidence that youth obesity may reduce academic performance and educational attainment (Sabia, 2007; von Hinke Kessler Scholder *et al.,* 2012). For this reason, we omit respondent education from the set of regressors and instead control for parental education: specifically, whether each parent has completed at least bachelor's degree (based on LPC data from 1980). Controlling for parental education also accounts for assortative mating within educational groups that could potentially violate the independence assumption of the IV estimation; i.e. it accounts for the possibility that the allele distribution differs according to parental education (Von Hinke *et al.,* 2016). Parental education is also a convenient control for family environment and resources.

Wage equations sometimes include controls for cognitive performance, when the data are available. In this context, however, that is questionable because there is evidence that obesity lowers scores on tests of cognitive achievement (Sabia, 2007; Averett and Stifel, 2010).[7]

Given the modest sample size, the main models in our paper are estimated for men and women pooled, and thus represent the average effect across both sexes. However, previous studies of the impact of weight on earnings have often found differences by gender (e.g. Cawley, 2004), so as an extension in the Appendix we also estimate models separately by gender.[8]

[7] See Appendix 1 for description of the measures for cognitive performance.
[8] Finland also exhibits a gender difference in wage penalty of obesity (Sarlio-Lähteenkorva *et al.,* 2004; Johansson *et al.,* 2009).

We first estimate equation (1) using ordinary least squares (OLS) in order to estimate the conditional correlation of weight with labor market outcomes.[9] These correlations reflect not only any causal effect of weight on wages, but also potentially reverse causality and the influence of omitted variables that may be correlated with both weight and the outcomes.

3.1. Method of instrumental variables: Genetic risk score for obesity

In order to estimate the causal effect of body weight on these outcomes, we estimate models of instrumental variables (IV) in which our IV is one of the two genetic risk scores for BMI.[10] It is estimated that 40-70% of inter-individual variability in BMI is due to genetic factors (e.g. Locke *et al.*, 2015), so the genetic risk factor score has the potential to be a powerful instrument.

As described in the data section, subjects in the YFS contributed DNA samples; results of the analysis of these samples are used to construct genetic risk scores (GRS) for high BMI. We use two different measures of GRS for high BMI. The first is based on the 32 SNPs that were found to be significantly ($p<1.0 \times 10^{-8}$) associated with high BMI by Speliotes et al. (2010) and which is used as an example of a powerful and likely valid application of genes as instruments by von Hinke *et al.* (2016). The second genetic risk score is based on 97 SNPs identified as associated with high BMI by Locke et al. (2015). This second, broader GRS includes all of the 32 SNPs included in the first, narrower, GRS. While the 32 SNP score was used by von Hinke Kessler Scholder (2016), the 97 SNP score has not, to our knowledge, been used in any study examining the economic consequences of obesity.

[9] For the binary outcome of receiving any social income transfers we estimate linear probability models. We prefer the use of linear probability models, because they facilitate easy interpretation of coefficients and are less sensitive to distributional assumptions. The results remain intact using a Tobit specification, where the social income transfers are left censored at zero.
[10] For all outcomes, the IV model is two-stage least squares.

The first GRS is equal to the sum of the alleles in the 32 SNPs that put one at elevated risk of high BMI. A person's risk score is equal to the number of alleles they have that are associated with an elevated risk of high BMI; because each person has either 0, 1, or 2 alleles for each of the relevant SNPs, the first GRS (based on 32 SNPs) ranges from 0 to 64. The second GRS (based on 97 SNPs) is available to us only in weighted form; the weights are based on the contribution of each SNP to high BMI in a meta-analysis. This difference in weighting explains the difference in means between the two risk scores shown in Appendix Table A2. The weighting may not be ideal in this context because the weights are based on all of the international data used in the meta-analysis, but the YFS represented only 0.8% of these observations, and thus the weights are not necessarily appropriate or best for the Finnish sample we study.

The genetic risk scores have two advantages as an instrument: first, they are more powerful (explain more variation in weight) than any of the SNPs individually; and second, they may be more valid because they reduce the risk that any alternative biological pathway (pleiotropy) in any individual SNP will bias the IV results (Davey Smith, 2011; Palmer *et al.,* 2012); the reason is that the instrument is a count of number of alleles associated with high BMI instead of indicator variables for having specific alleles of specific SNPs.[11]

Speliotes et al. (2010) report that the mechanisms by which these SNPs affect weight are through: 1) regulators of appetite or energy balance; and 2) insulin secretion or response. It is estimated that the 32 loci that constitute the first risk score explain 1.45% of the variation in BMI (Speliotes et al., 2010) and the 97 SNPs in the

[11] Using simulations comparing different methods, Palmer et al. (2012, p. 239) provide evidence that supports the use of genetic scores over indicator variables for individual SNPs.

second risk score explain 2.7% of the variation in BMI (Locke et al., 2015).[12] (Even though it is estimated that 40-70% of inter-individual variation in BMI is due to genetic factors, all currently-identified SNPs explain several percentage points of the variation; in other words, the vast majority of genetic variability in BMI remains unexplained; see, e.g., Locke *et al.*, 2015.) Each one-unit increase in the first genetic risk factor score was associated with an increase in BMI of 0.17 units, or roughly one-half of a kilogram of weight for an average-sized adult (Speliotes *et al.*, 2010). This same instrument (an unweighted risk score based on the 32 SNPs associated with obesity by Speliotes *et al.*, 2010) was also used in von Hinke (2016) as an example of a valid and powerful application of genes as instruments; they used the IV to estimate the effect of fat mass on academic achievement and blood pressure.

A threat to the validity of genetic instruments is pleiotropy – genes having more than one function (see, e.g., von Hinke *et al.*, 2016; Cawley, Han, and Norton 2011). For example, if the same genes associated with higher weight were also associated with unrelated traits or conditions that affect employment or earnings, then those genes are invalid instruments because the exclusion restriction is violated.

There is a possible tradeoff between power and validity associated with using a broader SNP risk score (i.e. one based on more SNPs). The advantage is that it a broader risk score may be more powerful (explains more variation in BMI), given that it is based on additional SNPs.[13] However, there is also a risk that some of those additional SNPs will also be correlated with other traits that affect labor market

[12] Speliotes et al. (2010) reports that the SNP with the greatest explanatory power for BMI is FTO (which explains 0.34% of variation in BMI), and that having the risky allele for FTO is associated with 20.3% greater odds of obesity. We have also estimated models in which the IV is a genetic risk score in which each SNP is weighted based on their effect size in the meta-analysis; this was no more powerful in the first stage than the unweighted genetic risk score.

[13] When the 97 SNP score is used as an instrument, the F-statistics of that instrument in the first stage of IV range from 52 to 63, depending the specification. Using the 32 SNP GRS the first-stage F-statistics varied between 23 and 40.

outcomes (pleiotropy) and thus there may be a greater risk of bias in the IV estimates. The two genetic risk scores we use are correlated but far from perfectly (0.64, which is statistically significant at the 1% level). Moreover, the SNPs most strongly correlated with the trait are usually identified first and thus are likely to already be included in the 32 SNP score.

We investigate the possibility of pleiotropy two ways. First, we check whether, in the genetics literature, the genes significantly associated with BMI are also significantly associated with other possible determinants of labor market outcomes. Speliotes et al. (2010) and Locke et al. (2015) search the genetics literature for evidence of any pleiotropy of the BMI-related SNPs. Of the SNPs linked to BMI, some have been linked to waist circumference and waist-to-hip ratio, but these are clearly related to weight. Some are associated with height, a component of BMI (Speliotes *et al.*, 2010).[14]

Some SNPs are linked to obesity-related illnesses; these could be either downstream effects of a high BMI, but it is also possible that coincidentally the SNPs affect these illnesses through pathways other than obesity. Specifically, some SNPs in the risk score are associated with either Type 2 diabetes, fasting glucose, fasting insulin, or insulin resistance, which is not surprising given that excess fat (by secreting the hormone resistin) causes insulin resistance and thus diabetes (Hu, 2008). Some are linked to serum cholesterol levels and one to blood pressure; both of these conditions are strongly associated with obesity (e.g. Hu, 2008). Some are associated with age of onset of menstruation (menarche), but this too is related to fatness (Wang, 2002; Kaplowitz, 2008). In summary, the other phenotypes that the obesity-related

[14] We report the results additionally controlling for height in the working paper version. The conclusions remain intact.

SNPs are associated with tend to be obesity-related comorbidities. We assume that the associations with obesity-related conditions occur because of the SNPs association with high BMI, but acknowledge that it could be through other pathways, which could threaten the validity of the instrument. It is noteworthy that the searches of Speliotes et al. (2010) and Locke et al. (2015) did not yield evidence that the SNPs associated with high BMI are associated with characteristics unrelated to obesity that might directly affect labor market outcomes, such as intelligence.

As a second check of the validity of the genetic instruments, we follow McClellan et al. (1994) and divide our sample into those with an above-average and below-average value of the instrument, and test whether the two groups significantly differ in their observable characteristics that are likely correlated with the second-stage outcome. It is impossible to confirm the null hypothesis that the instrument is uncorrelated with the second-stage error term, but a lack of correlation between the instrument and observed variables would be consistent with the exclusion restriction.[15] These comparisons will be discussed in the Empirical Results section.

An additional assumption is that the allele distribution does not vary systematically in different population subgroups (von Hinke *et al.,* 2016). There are two key facts that support the independence assumption in our setting. First, our data originate from Finland, which is ethnically very homogeneous. Second, following von Hinke et al. (2016), we have tested whether the distribution of our covariates is the same across the instrument distribution by regressing each of the covariates on the instrument. In Appendix Table A1, we report p-values associated with a joint test based on regressing a covariate on each of the 32 SNPs and then testing whether the

[15] Comparing distribution of observables between above- versus below-average genetic score does not address the potential concern about the remaining endogeneity stemming from unobservables.

32 coefficients on the SNPs are jointly equal to zero. In Column 2 of Appendix Table A1, the p-values indicate that, for each covariate, we cannot reject the null hypothesis that the coefficients on the 32 SNPs in the first score are jointly equal to zero. However, in Column 4, the p-values indicate that we can reject the null for the 97 SNPs for the following covariates: female, age, marital status, and father's income in 1980. Thus, the additional SNPs in the broader risk score may vary systematically in population subgroups. The evidence on this point is more supportive of the identifying assumptions for the narrower 32 SNP score than for the broader 97 SNP score.

4. EMPIRICAL RESULTS

4.1. Baseline OLS estimates

OLS estimates are presented in Column 1 of Table I.[16] In OLS regressions, a one-unit increase in BMI is associated with: 0.7% lower average earnings, 0.2% fewer years spent employed, 0.1 percentage point lower probability of any social income transfers, and 0.5% lower social income transfers, none of which are statistically significant.[17]

4.2. Power and validity of the IV Model

Because we seek to estimate the causal effect of BMI on these outcomes, we next estimate IV models. The genetic risk score for BMI is a powerful instrument for BMI. In the first stage of IV, the F-statistic on the instrument varies by outcome, but ranges between 23.5 and 39.9 for the 32 SNP genetic risk score, and between 52.0 and

[16] Appendix Table A2 reports summary statistics for our regression sample.
[17] We estimate several additional models to assess the robustness of the baseline results using the narrower score (Appendix 3).

62.7 for the 97 SNP genetic risk score; all of these far exceed the minimum standard of F=10 suggested in Staiger and Stock (1997).[18]

In order to examine the validity of the genetic instruments, Appendix Table A3 presents differences in means of the observed variables for those with above- and below-average values of the BMI genetic risk score, and tests for equality of the means. As expected, those with above-average genetic risk factor scores have a significantly higher BMI (by 0.92 units for the 32 SNP GRS and by 0.96 units for the 97 SNP GRS); this is consistent with the instrument being powerful. The table also shows that those with a higher genetic 32 SNP risk score for obesity have significantly lower earnings, which is consistent with BMI worsening labor market outcomes. The difference in earnings for those with a high and low value of the 97 SNP score is smaller and not statistically significant. While the difference in the probability of social income transfers for those with a high and low value of the instrument is not statistically significant for the 32 SNP score, it is statistically significant for the 97 SNP score.

Lower rows in Appendix Table A3 shed some light on the criteria of validity. The 32 SNP score is associated with differences in two, and the 97 SNP score is associated with a difference in one, cognitive test score. There is also some evidence that obesity worsens academic test scores (Sabia, 2007; Averett and Stifel, 2010), although von Hinke Kessler Scholder et al. (2012) could not reject the null of no effect. The literature searches of Speliotes et al. (2010) and Locke et al. (2015) did not turn up evidence of a link between the BMI-related SNPs and intelligence. The 32 SNP score is not associated with other covariates, but the 97 SNP score is associated

[18] We have also estimated IV models that use both the genetic risk score and its square as instruments. The first-stage F-statistics in these models are lower (roughly 18), and the results are similar, so we continue to estimate models that simply use the level of the risk score as the IV.

with father's income. This is a second piece of evidence that the 97 SNP score may be less valid than the 32 SNP score. (The first was the evidence that the 97 SNP score varies by subgroup, as seen in Appendix Table A1.)

We also assessed the validity of our IV setting by examining potential heterogeneity between the variant-specific estimates. If all SNPs were valid instruments, their Mendelian randomization estimates should only vary by chance so that larger between-instrument heterogeneity would indicate a violation of IV assumptions, most likely due to pleiotropy (Greco *et al.*, 2015). To visually illustrate the potential heterogeneity Figure A1 plots the genetic associations with log earnings (vertical axis) against genetic associations with the BMI (horizontal axis) for each of the 32 SNPs. (Figure A2 plots the same associations for the 97 SNPs.) Each point in Figure A1 stands for a genetic variant. The points should be compatible with a straight line through the origin under the null hypothesis of homogeneity and any point that substantially deviates from this horizontal line from the origin should be investigated for potential pleiotropy (Burgess *et al.*, 2017, p. 35). Based on visual inspection the estimates do seem clustered along the horizontal at zero; the one vertical outlier is not significantly different from zero.

In Figure A2, for the 97 SNP score, some estimates are statistically significantly different from zero. This could be the result of different SNPs having different Local Average Treatment Effects (LATE) by operating through different biological mechanisms; however, it is also consistent with pleiotropy. This is a third piece of evidence that the 97 SNP score may be less valid than the 32 SNP score.

A formal statistical test of pleiotropy can be conducted using the Sargan's over-identification test. To perform this test we estimated our main models using individual SNPs (both 32 and 97) as instruments for BMI. In all but one case, the

Sargan's test supported the null hypothesis that all instruments yield the same Mendelian randomization estimate and thus provided support to the validity of our instrument. The exception is that when an indicator variable for social income transfers was used as the outcome variable, the null hypothesis was rejected at the 10% level (p = 0.041) when 32 individual SNPs were used as instruments.

4.3. IV estimates

The coefficients from the IV models are presented in Columns 2 and 3 of Table I. The estimates based on using the 32 SNP GRS as an instrument (Column 2) indicate that a one-unit increase in BMI is associated with 6.9% lower wages[19] and 1.8 percentage point (2.1%) fewer years employed, both of which are statistically significant at the 5% level. A one-unit increase in BMI is also associated with a 3.0 percentage point (3.7%) higher probability of receiving any social income transfers, which is also statistically significant at the 10% level.[20] There is no statistically significant effect on the amount of social income transfers, conditional on receiving any.

Expressed another way, the results imply that a one-standard-deviation increase in BMI (of 4.3 units) lowers wages by 29.7%, lowers years employed by 9.0%, and raises the probability of receiving any social income transfers by 15.9%.

Interestingly, when we use the broader risk score as an instrument in IV models (Column 3) we cannot reject the null hypothesis of no effect of BMI on any outcome. The IV coefficients on BMI in the regressions for employment and probability of social income transfers have the same sign as the earlier IV coefficients

[19] Earnings and social income transfers have been log-transferred, so to interpret the coefficients on the GRS as a percent change, one must raise e to the power of the coefficient and then subtract one.
[20] If we estimate a specification for the whole sample setting zeros to 1 Euro and then using the logarithmic transformation of social income transfers as the outcome, using IV models we find that higher BMI leads to a significant overall increase in transfers (Appendix Table A4).

based on the 32 SNP score, but both are smaller and neither is statistically significant. In the Discussion, we consider explanations for the differences in results between the 32 SNP score and the 97 SNP score.

Reduced-Form Estimates

Table II presents results of reduced-form models that regress outcomes on the instrument (BMI genetic risk score) directly, controlling for the same set of regressors as earlier. The results are consistent with those of the IV models. Raising the 32 SNP genetic risk score by one (meaning that an individual has one additional allele that raises their risk of high BMI) is associated with 1.2% lower earnings, 0.3 percentage points (0.3%) fewer years of employment, and a 0.5 percentage point (0.6%) higher probability of receiving any social income transfers, all of which are statistically significant at the 5% level (Column 1 of Table II).

Again, the choice of GRS makes a difference in the results. Column 2 of Table II shows that the reduced-form estimates for the 97 SNP score are typically not statistically significant. The exception is that an additional risky allele is associated with a 8.7 percentage point (10.6%) increase in the probability of receiving any social income transfers.[21]

5. CONCLUSION

Much of the evidence about causal effects of obesity is based on IV models in which the instrument for respondent weight is the weight of a biological relative. This paper contributes to the literature by using a novel instrument: genetic risk score for obesity based on many SNPs that are robustly associated with high BMI.

[21] We also regressed the 97 SNP score on the 32 SNP score and used the residual as a predictor in the reduced-form model along with 32 SNP GRS. In earnings equation, there was a significant difference between 32 SNP GRS and residual coefficients. This suggests that the newly-added SNPs in the larger score may have a different relationship to earnings than the SNPs in the narrower score. This might indicate that the newly added SNPs are less exogenous or less powerful.

Specifically, we use two such genetic risk scores, and find evidence that IV model estimates are sensitive to which risk score is used.

The estimates of the IV models that use the genetic risk score based on 32 SNPs confirm those of the previous literature that used a different instrument (the weight of a biological relative): weight lowers wages and the probability of employment. Specifically, our IV estimates indicate that an additional unit of BMI lowers wages by 6.9% and reduces the share of years employed by 2.1%. We also examine the novel outcome of social income transfers and find that an additional unit of BMI increases the probability of receiving social income transfers by 3.7%. This represents potential negative externalities of obesity – social costs of obesity paid by non-obese individuals – and thus an economic rationale for government intervention to prevent and reduce obesity. It is well-established that obesity imposes negative externalities through higher health care costs (e.g. Cawley and Meyerhoefer, 2012) but this paper offers the first evidence that there may also be negative externalities through social income transfers.

Reduced form models that regress outcomes directly on the genetic risk score based on 32 SNPs are also consistent with the hypothesis that additional weight worsens labor market outcomes; raising the genetic risk score by one (meaning that an individual has one additional allele that raises their risk of weight gain) is associated with 1.2% lower earnings, 0.3% fewer years employed, and a 0.6% higher probability of receiving social income transfers.

When we use a GRS based on 97 SNPs as the instrument in the IV model, however, the estimates are quite different. We cannot reject the null hypothesis of no effect of BMI on labor market outcomes. The 97 SNP score is typically not statistically significant in reduced-form models either; the exception is that an

additional risky allele is associated with a 10.6% increase in the probability of receiving any social income transfers.

There are several possible explanations for the sensitivity of the results to the genetic risk score used. First, the 97 SNP score is available for a slightly smaller sample than the 32 SNP score, so some statistical power is lost. This does not appear to be a critical factor, because when we re-estimate the IV model using the 32 SNP score but for the smaller sample with a valid 97 SNP score, the IV results are similar to those for the full sample. Second, it is possible that the weighting of the 97 SNP score makes a difference. To explore this we re-estimated the IV models using a weighted 32 SNP score, and we find that it does raise the standard errors to the point that the results are not statistically significant. Thus, the weights may be playing some role in the difference in results. The weights are based on each SNP's predictive power in a large international sample, of which the YFS constitutes less than 1%; as a result, the weights may not be optimal or appropriate for the YFS sample. A third reason why the results differ for the 97 SNP and 32 SNP scores is that the additional SNPs in the broader score may operate through different biological mechanisms and thus may have different Local Average Treatment Effects. A fourth explanation for the difference in results between the two risk scores is that the larger 97 SNP score may face a greater risk of bias because it includes SNPs less highly correlated with BMI but potentially correlated with other things that could affect labor market outcomes. We find two pieces of evidence that the 97 SNP score may not be as valid as the 32 SNP score: 1) the 97 SNP score but not the 32 SNP score varies by sex, age, marital status (Appendix Table A1), as well as father's income (Appendix Table A1 and Appendix Table A3); and 2) the 97 SNP score but not the 32 SNP score exhibits significant heterogeneity between variant-specific estimates, which could be due to

different SNPs having different LATEs through different biological mechanisms but may be due to pleiotropy (Appendix Figures A1 and A2). Because of this evidence casting some doubt on the validity of the 97 SNP score, the 32 SNP score is the preferred instrument in this study.

In general, it is noteworthy that the IV results are sensitive to the choice of genetic instrument. Future studies in this area may wish to test the robustness of their results to the use of alternate genetic risk scores, and to explore reasons for any differences that are found.

A strength of the paper is that the key variables are free of reporting error; i.e. weight and height are measured and information on employment, earnings, and social income transfers are taken from administrative records. This implies that the estimates are relatively free of the problems of bias and inflated standard errors that result from error in the dependent and independent variables (Bound *et al.,* 2001; Cawley *et al.,* 2015; Courtemanche *et al.,* 2015).

We acknowledge the limitations of this paper. The sample is relatively small (N=2,062), providing little statistical power to estimate models separately by gender or other subgroups. Despite being rich in other ways, the data do not allow us to further investigate the mechanisms by which BMI affects labor market outcomes. It is always important to stress when using the method of IV that important assumptions regarding the validity of the instruments are not testable. Although the SNPs that are used in the genetic risk score for BMI were generally not found to be linked to non-obesity-related outcomes, the failure to reject the null hypothesis of no effect is not the same as proving the null. It is also possible that the reason the SNPs are linked to obesity-related illness is because of some direct effect that does not operate through a high BMI. The exact function and mechanisms of these SNPs are not known with

certainty. Although the 32 SNP GRS was used as an example of a powerful and likely valid application of genes as IVs (von Hinke *et al.,* 2016), that study also pointed out the need for caution regarding instrument validity.

When considering the generalizability of these results, it should be noted that the Local Average Treatment Effect that we measure concerns the impact of genetic variation in weight; it is possible that variation in weight due to other sources could have a different impact on labor market outcomes. Moreover, our IVs measure only the genetic variation due to the specific SNPs included in the risk scores. Those in the 32 SNP score affect weight through regulators of appetite or energy balance, or insulin secretion or response (Speliotes *et al.,* 2010). It is possible that genetic variation in weight that operates through other mechanisms (e.g. resting metabolic rate, or propensity to add muscle mass) could exhibit a different relationship with labor market outcomes.

Our data are from Finland, a relatively small nation where the wage distribution is narrower than in the Anglo-Saxon countries, which may raise some issues of generalizability, but it is a highly economically developed country that is a member of the European Union and shares many labor market characteristics with the rest of Western Europe. The prevalence of obesity in Finland is 20.9% among adult men and 22.3% among adult women (Ng *et al.,* 2014), which is similar to that of other Western European countries. Despite these limitations, the strengths of the data, such as genetic information, measured weight and height, and comprehensive administrative data on wages, employment, and social income transfers that are measured without reporting error, make it well-suited to investigate this research question.

Acknowledgements

The Young Finns Study has been financially supported by the Academy of Finland: grants 286284, 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117787 (Gendi), and 41071 (Skidi); the Social Insurance Institution of Finland; Competitive State Research Financing of the Expert Responsibility area of Kuopio, Tampere and Turku University Hospitals (grant X51001); Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation for Cardiovascular Research ; Finnish Cultural Foundation; Tampere Tuberculosis Foundation; Emil Aaltonen Foundation; Yrjö Jahnsson Foundation; Signe and Ane Gyllenberg Foundation; Diabetes Research Foundation of Finnish Diabetes Association; and EU Horizon 2020 (grant 755320 for TAXINOMISIS). The use of linked data is supported by the Palkansaaja Foundation. Jutta Viinikainen and Jaakko Pehkonen acknowledge the financial support from theYrjö Jahnsson foundation (grants 6646 and 6664). Petri Böckerman thanks the Strategic Research Council funding for the project Work, Inequality and Public Policy (293120). John Cawley thanks the Robert Wood Johnson Foundation for an Investigator Award in Health Policy Research. We thank Barton Willage for research assistance. Finally, we also thank two anonymous reviewers, the seminar participants at the EALE, Oulu Business School and the VATT seminar and Jani-Petri Laamanen for helpful comments.

REFERENCES

Averett SL. 2011. Labor market consequences: Employment, wages, disability, and absenteeism. In *The Oxford Handbook of the Social Science of Obesity*, Cawley J (ed). Oxford University Press: New York, pp. 531–552.

Averett SL, Stifel DC. 2010. Race and gender differences in the cognitive effects of childhood overweight. *Applied Economics Letters* **17**: 1673–1679.

Barsh GS, Farooqi IS, O'Rahilly S. 2000. Genetics of body-weight regulation. *Nature* **404**: 644–651.

Bound J, Brown C, Mathiowetz N. 2001. Measurement error in survey data. In *Handbook of Econometrics*, Volume 5, Leamer E, Heckman J (eds). North Holland: Amsterdam, pp. 3705–3843.

Bowden J, Davey Smith G, Burgess S. 2015. Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression. *International Journal of Epidemiology* **44**: 512–525.

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

Burgess S, Bowden J, Fall T, Ingelsson E, Thompson SG 2017. Sensitivity analyses for robust causal inference from Mendelian randomization analyses with multiple genetic variants. *Epidemiology* **28**: 30–42.

Cawley J. 2004. The impact of obesity on wages. *Journal of Human Resources* **39**: 451–474.

Cawley J. 2015. An economy of scales: A selective review of obesity's economic causes, consequences, and solutions. *Journal of Health Economics* **43**: 244–268.

Cawley J, Han E, Norton E. 2011. The validity of genes related to neurotransmitters as instrumental variables. *Health Economics* **20**: 884–888.

Cawley J, Maclean JC, Hammer M, Wintfeld N. 2015. Reporting error in weight and its implications for bias in economic models. *Economics and Human Biology* **19**: 27–44.

Cawley J, Meyerhoefer C. 2012. The medical care costs of obesity: An instrumental variables approach. *Journal of Health Economics* **31**: 219–230.

Conley D. 2016. Socio-genomic research using genome-wide molecular data. *Annual Review of Sociology* **42**: 275–299.

Courtemanche C, Pinkston JC, Stewart J. 2015. Adjusting body mass for measurement error with invalid validation data. *Economics and Human Biology* **19**: 275–293.

Dahl M, DeLeire T, Schwabish JA. 2011. Estimates of year-to-year volatility in earnings and in household incomes from administrative, survey, and matched data. *Journal of Human Resources* **46**: 750–774.

Davey Smith G. 2011. Random allocation in observational data: How small but robust effects could facilitate hypothesis-free causal inference. *Epidemiology* **22**: 460–463.

Davey Smith G, Paternoster L, Relton C. 2017. When will Mendelian randomization become relevant for clinical practice and public Health? *Journal of the American Medical Association* **317**: 589–591.

De Luca CR, Wood SJ, Anderson V, Buchanan J-A, Proffitt TM, Mahony K, Pantelis C. 2003. Normative data from the CANTAB. I: Development of executive function over the lifespan. *Journal of Clinical and Experimental Neuropsychology* **25**: 242–254.

GBD 2015 Obesity Collaborators. 2017. Health effects of overweight and obesity in 195 countries over 25 years. *New England Journal of Medicine* **377**: 13–27.
Greco FD, Minelli C, Sheehan NA, Thompson JR. 2015. Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. *Statistics in Medicine* **34**: 2926–2940.
Haycock PC, Burgess S, Wade KH, Bowden J, Relton C, Davey Smith G. 2016. Best (but oft-forgotten) practices: The design, analysis, and interpretation of Mendelian randomization studies. *The American Journal of Clinical Nutrition* **103**: 965–978.
Hu F. 2008. *Obesity Epidemiology.* Oxford University Press: New York.
International Consortium for Blood Pressure Genome-Wide Association Studies. 2011. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**: 103–109.
Johansson E, Böckerman P, Kiiskinen U, Heliövaara M. 2009. Obesity and labour market success in Finland: The difference between having a high BMI and being fat. *Economics and Human Biology* **7**: 36–45.
Kaplowitz PB. 2008. Link between body fat and the timing of puberty. *Pediatrics* **121** (Supplement 3): S208–S217.
Kline B, Tobias JL. 2008. The wages of BMI: Bayesian analysis of a skewed treatment–response model with nonparametric endogeneity. *Journal of Applied Econometrics* **23**: 767–793.
Komlos J, Smith PK, Bogin B. 2004. Obesity and the rate of time preference: Is there a connection? *Journal of Biosocial Science* **36**: 209–219.
Lindeboom M, Lundborg P, van der Klaauw B. 2010. Assessing the impact of obesity on labor market outcomes. *Economics and Human Biology* **8**: 309–319.
Locke AE, Kahali B, Berndt SI et al. 2015. Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**: 197–206.
McClellan M, McNeil BJ, Newhouse JP. 1994. Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables. *Journal of the American Medical Association* **272**: 859–866.
Mincer JA. 1974. *Schooling, Experience, and Earnings.* National Bureau of Economic Research: New York.
Ng M, Fleming T, Robinson M et al. 2014. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: A systematic analysis for the global burden of disease study 2013. *The Lancet* **384**: 766–781.
Norton EC, Han E. 2008. Genetic information, obesity and labor market outcomes. *Health Economics* **17**: 1089–1104.
Palmer TM, Lawlor DA, Harbord RM, Sheehan NA, Tobias JH, Timpson NJ, Davey Smith G, Sterne JAC. 2012. Using multiple genetic variants as instrumental variables for modifiable risk factors. *Statistical Methods in Medical Research* **21**: 223–242.
Pietiläinen KH, Kaprio J, Rissanen A, Winter T, Rimpelä A, Viken RJ, Rose RJ. 1999. Distribution and heritability of BMI in Finnish adolescents aged 16 y and 17 y: A study of 4884 twins and 2509 singletons. *International Journal of Obesity* **23**: 107–115.
Puhl RM. 2011. Bias, stigma, and discrimination. In *The Oxford Handbook of the Social Science of Obesity,* Cawley J (ed). Oxford University Press: New York, pp. 553–571.

Raitakari OT, Juonala M, Rönnemaa T et al. 2008. Cohort profile: The cardiovascular risk in Young Finns Study. *International Journal of Epidemiology* **37**: 1220–1226.

Robbins TW, James M, Owen AM, Sahakian, BJ, McInnes L, Rabbitt P. 1994. Cambridge Neuropsychological Test Automated Battery (CANTAB): A factor analytic study of a large sample of normal elderly volunteers. *Dementia* **5**: 266–281.

Rooth D-O. 2009. Obesity, attractiveness, and differential treatment in hiring: A field experiment. *Journal of Human Resources* **44**: 710–735.

Sabia JJ. 2007. The effect of body weight on adolescent academic performance. *Southern Economic Journal* **73**: 871–900.

Sarlio-Lähteenkorva S, Silventoinen K, Lahelma E. 2004. Relative weight and income at different levels of socioeconomic status. *American Journal of Public Health* **94**: 468–472.

Schmeiser MD. 2009. Expanding wallets and waistlines: The impact of family income on the BMI of women and men eligible for the earned income tax credit. *Health Economics* **18**: 1277–1294.

Speliotes EK, Willer CJ, Berndt SI et al. 2010. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature Genetics* **42**: 937–948.

Staiger D, Stock JH. 1997. Instrumental variables regression with weak instruments. *Econometrica* **65**: 447–486.

Teslovich TM, Musunuru K, Smith AV et al. 2010. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**: 707–713.

Tyrrell J, Jones SE, Beaumont R et al. 2016. Height, body mass index, and socioeconomic status: Mendelian randomisation study in UK Biobank. *British Medical Journal* **352**: i582.

von Hinke Kessler Scholder S, Davey Smith G, Lawlor DA, Propper C, Windmeijer F. 2012. The effect of fat mass on educational attainment: Examining the sensitivity to different identification strategies. *Economics and Human Biology* **10**: 405–418.

von Hinke S, Davey Smith G, Lawlor DA, Propper C, Windmeijer F. 2016. Genetic markers as instrumental variables. *Journal of Health Economics* **45**: 131–148.

Wang Y. 2002. Is obesity associated with early sexual maturation? A comparison of the association in American boys versus girls. *Pediatrics* **110**: 903–910.

Table I. The effect of BMI on average labor market outcomes, 2001-2012

Panel A: Log of Average Earnings, 2001-2012	**(1) OLS**	**(2) IV – 32 SNP score**	**(3) IV – 97 SNP score**
BMI	-0.007 (0.005)	-0.071** (0.036)	0.010 (0.027)
F-statistics	…	36.53	58.92
Mean outcome	9.863	9.863	9.866
N	2038	2038	1886
Panel B: Share of Years Employed, 2001-2012	**OLS**	**IV – 32 SNP score**	**IV – 97 SNP score**
BMI	-0.002 (0.001)	-0.018** (0.009)	-0.005 (0.007)
F-statistics	..	39.90	62.73
Mean outcome	0.857	0.857	0.859
N	2062	2062	1909
Panel C: Indicator for Social Income Transfers, 2001-2012 (Extensive margin)	**OLS**	**IV – 32 SNP score**	**IV – 97 SNP score**
BMI	-0.001 (0.002)	0.030* (0.016)	0.019 (0.012)
F-statistics	..	36.53	58.92
Mean outcome	0.821	0.821	0.819
N	2038	2038	1886
Panel D: Log of Average Social Income Transfers, 2001-2012 (Intensive margin)	**OLS**	**IV – 32 SNP score**	**IV – 97 SNP score**
BMI	-0.005 (0.009)	0.061 (0.068)	-0.020 (0.044)
F-statistics	..	23.49	51.97
Mean outcome	6.836	6.836	6.831
N	1673	1673	1545

Notes: Earnings are measured as the log of average earnings over the period 2001-2012. Employment is measured as the average share of employment years over the period 2001-2012. Indicator for social income transfers equals one for those who have received social security transfers at least once during 2001-2012. Social income transfers are measured as the log of average transfers over the period 2001-2012, conditional on obtaining a positive amount of transfers. The mean values for the dependent variables are reported. BMI is measured in 2001. All models include controls for the birth month and birth year effects. Gender and parental education (1980) are also controlled for in all models. The instrument used in the IV models is the BMI risk score, based on genetic markers. Angrist-Pischke multivariate F-tests of excluded instrument are reported for the IV models. Heteroscedasticity-robust standard errors are reported in parentheses: * statistically significant at the 0.10 level; ** at the 0.05 level; *** at the 0.01 level.

Table II. Reduced form IV estimates

	(1) **32 SNP score**	**(2)** **97 SNP score**
Panel A: Log of Average Earnings, 2001-2012	-0.012** (0.006)	0.046 (0.122)
Mean outcome	9.863	9.866
N	2038	1886
	32 SNP score	**97 SNP score**
Panel B: Share of Years Employed, 2001-2012	-0.003** (0.002)	-0.023 (0.035)
Mean outcome	0.857	0.859
N	2062	1909
	32 SNP score	**97 SNP score**
Panel C: Indicator for Social Income Transfers, 2001-2012 (Extensive margin)	0.005** (0.002)	0.087* (0.053)
Mean outcome	0.821	0.819
N	2038	1886
	32 SNP score	**97 SNP score**
Panel D: Log of Average Social Income Transfers, 2001-2012 (Intensive margin)	0.009 (0.010)	-0.095 (0.211)
Mean outcome	6.836	6.831
N	1673	1545

Notes: Earnings are measured as the log of average earnings over the period 2001-2012. Employment is measured as the average share of employment years over the period 2001-2012. Indicator for social income transfers equals one for those who have received social security transfers at least once during 2001-2012. Social income transfers are measured as the log of average transfers over the period 2001-2012, conditional on obtaining a positive amount of transfers. The mean values for the dependent variables are reported. All models include controls for the birth month and birth year effects. Gender and parental education (1980) are also controlled for in all models. Heteroscedasticity-robust standard errors are reported in parentheses: * statistically significant at the 0.10 level; ** at the 0.05 level; *** at the 0.01 level.

ONLINE SUPPLEMENTARY APPENDICES

Appendix 1:
Tests of cognitive performance

In the latest follow-up of YFS (2011-12), cognitive function was assessed with commercially available Cambridge Neuropsychological Test Automated Battery (CANTAB®). The CANTAB® is a computerized, predominantly non-linguistic and culturally neutral test performed using a validated touch-screen computer system. The full test battery includes 25 individual tests from which, five age sensitive tests (Robbins et al. 1994; De Luca et al. 2003) were selected for YFS. The tests measured several cognitive domains: 1) short term memory, 2) spatial working memory, 3) problem solving, 4) reaction time, 5) attention, 6) rapid visual processing, 7) visual memory, 8) episodic memory, and 9) visuospatial learning.

During cognitive testing the participants first conducted a motor screening test (MOT) measuring psychomotor speed and accuracy. In this study, the MOT test was considered as a training procedure in which the participants were introduced to the equipment used in the testing, and as a screening tool to point out any difficulties in vision, movement, comprehension or ability to follow simple instructions. Paired associates learning (PAL) test was used to assess visual and episodic memory as well as visuospatial associative learning containing aspects of both delayed response procedure and conditional learning. Spatial working memory (SWM) test was used to measure ability to retain spatial information and to manipulate items stored in the working memory, problem solving as well as the ability to conduct a self-organized search strategy. Reaction time (RTI) test assessed speed of response and movement on tasks where the stimulus was either predictable (simple location task) or

unpredictable (five-choice location task). Rapid visual information (RVP) test was used to assess visual processing, recognition and sustained attention.

Each of the CANTAB® tests produced several variables. Therefore, principal component analysis was conducted to reduce the number of variables and to identify components accounting for the majority of the variation within the cognition dataset. First, principal component analysis was conducted for the complete cognitive data and the resulting first component was considered as an indicator for overall cognitive performance. Second, principal component analyses were performed separately for all individual tests (MOT, PAL, SWM, RTI, RVP). The first components resulting from these analyses were considered to represent cognitive performance related to the particular domain. After creating the overall and testwise principal components their distributions were analyzed. The component for MOT test was excluded from further analyses because it did not discriminate the subjects indicating a ceiling effect. All other components were normalized based on the rank order normalization procedure resulting in five separate variables, each with mean value of 0 and standard deviation of 1.

Appendix 2: Additional tables and figures

Appendix Table A1: An indirect test of independence assumption: regressing the covariates on the BMI 32 SNP GRS

	(1) OLS 32 SNP GRS	(2) 32 independent variants p-value of F-statistics; instrument SNPs jointly equal to zero	(3) OLS 97 SNP GRS	(4) 97 independent variants p-value of F-statistics; instrument SNPs jointly equal to zero
Female	-0.005 (0.003)	0.529	-0.015 (0.071)	0.002
Age in 2001	-0.004 (0.033)	0.937	-0.739 (0.707)	0.012
Married in 2001	0.001 (0.003)	0.257	-0.100 (0.071)	0.080
Family income (1980), mother (euros)	-26.083 (22.346)	0.902	-511.666 (489.019)	0.144
Family income (1980), father (euros)	-68.203* (35.352)	0.270	-912.339 (868.100)	0.003
University education, mother	0.001 (0.002)	0.372	0.037 (0.037)	0.870
University education, father	0.000 (0.002)	0.352	0.005 (0.046)	0.164

Notes: * statistically significant at the 0.10 level; ** at the 0.05 level; *** at the 0.01 level.

Appendix Table A2:
Summary statistics

Variable	Mean (SD)	N
Average annual earnings (2001-2012), euros	24527.93 (15042.27)	2038
Share of years employed (2001-2012)	0.857 (0.245)	2062
Received social income transfers at any point (2001-2012) (extensive margin)	0.821 (0.384)	2038
Average annual social income transfers (2001-2012), euros (intensive margin)	2005.13 (2156.62)	1673
Earnings (2001), euros	18915.49 (14382.31)	2038
Earnings (2007), euros	26399.55 (19239.1)	1940
Earnings (2011), euros	29528.85 (19952.31)	1729
Indicator for being employed (2001)	0.807 (0.395)	2062
Indicator for being employed (2007)	0.885 (0.320)	1957
Indicator for being employed (2011)	0.906 (0.291)	1742
Indicator for social income transfers (2001) (extensive margin)	0.367 (0.482)	2038
Indicator for social income transfers (2007) (extensive margin)	0.334 (0.472)	1940
Indicator for social income transfers (2011) (extensive margin)	0.291 (0.454)	1729
Social income transfers (2001), euros (intensive margin)	4195.33 (3806.08)	748
Social income transfers (2007), euros (intensive margin)	4770.23 (4567.03)	647
Social income transfers (2011), euros (intensive margin)	5182.99 (5097.73)	503
BMI (2001)	25.052 (4.290)	2038
BMI (2007)	25.864 (4.432)	1940
BMI (2011)	26.338 (4.621)	1729
BMI risk score based on 32 SNPs (unweighted)	29.144 (3.358)	2038
BMI risk score based on 97 SNPs (weighted)	2.316 (0.161)	1886
BMI>=30 (2001)	0.124 (0.329)	2038
Weight (2001)	74.686 (15.854)	2038
University education (1980), mother	0.072 (0.258)	2038
University education (1980), father	0.102 (0.303)	2038
Income (1980), mother (euros)	4616.65 (3503.96)	2023
Income (1980), father (euros)	8739.78 (5775.44)	1931
Married (2001)	0.445 (0.497)	2038
Overall cognitive performance (2011-2012)	0.010 (0.996)	1334
Visual and episodic memory and visuospatial associative learning	0.013 (0.989)	1334
Reaction time	0.021 (0.996)	1334
Rapid visual information processing	0.042 (0.985)	1334
Spatial working memory	0.005 (0.974)	1334
Genetic risk score for blood pressure	30.449 (3.215)	2001
Genetic risk score for total cholesterol	27.462 (3.089)	2001
Genetic risk score for triglycerides	26.128 (2.875)	2001

Notes: Descriptive statistics are reported for the samples that are used in the estimations.

Appendix Table A3:
Comparison of observables by value of instrument

	Difference (32 SNP GRS)	t-statistics (32 SNP GRS)	Difference (97 SNP GRS)	t-statistics (97 SNP GRS)
Earnings, 2001-2012 (euros)	1493.685	2.241**	-679.512	-0.981
Share of years employed, 2001-2012	0.017	1.575	-0.003	-0.227
Indicator for social income transfers, 2001-2012 (extensive margin)	-0.024	-1.431	-0.037	-2.105**
Social income transfers, 2001-2012 (euros) (intensive margin)	-103.129	-0.977	-2.105	-0.069
BMI (2001)	-0.920	-4.840***	-0.958	-4.864***
Married (2001)	0.001	0.029	0.017	0.758
Cognitive performance (2011-2012)				
Visual and episodic memory and visuospatial associative learning	0.061	1.114	0.038	0.664
Reaction time	-0.028	-0.517	0.059	1.042
Rapid visual information processing	0.187	3.481***	0.097	1.728*
Spatial working memory	0.124	2.321**	0.055	0.984
Family background (1980)				
Income, mother (euros)	247.164	1.591	143.705	0.893
Income, father (euros)	135.016	0.511	459.600	1.688*
University education, mother	-0.001	-0.112	-0.003	-0.270
University education, father	-0.004	-0.323	0.008	0.544

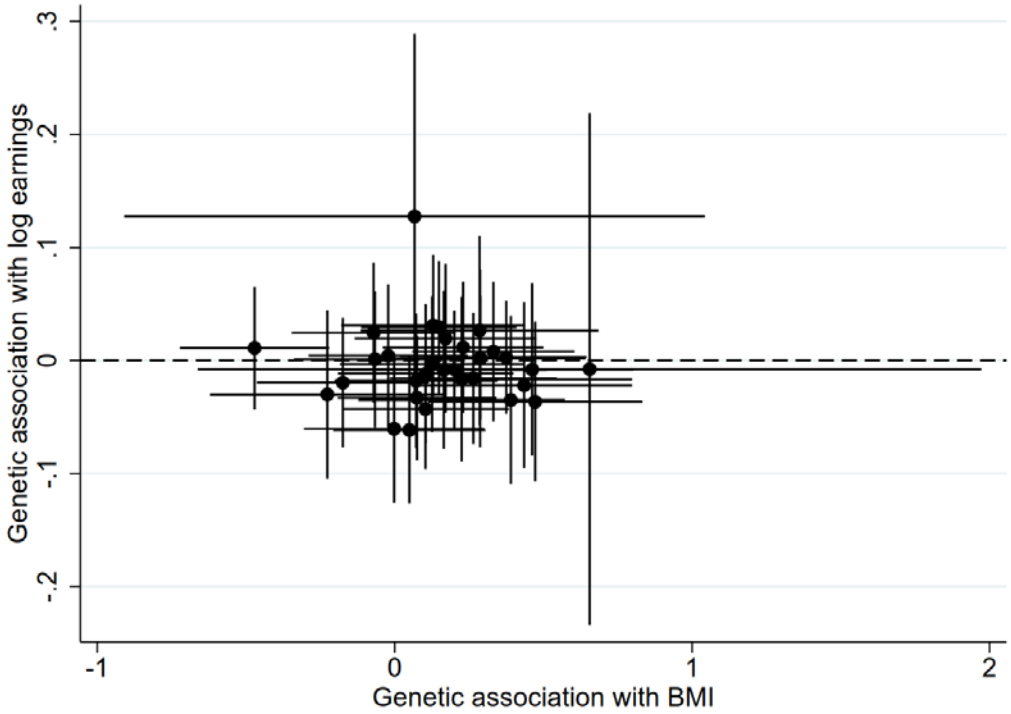
Notes: The difference is calculated by subtracting variable's mean value in the low GRS group from the mean value in the high GRS group. Low (high) GRS group consists of individuals whose GRS is below (above) the average value of the risk score.

Appendix Table A4. The effect of BMI on social income transfers

Log of Average Social Income Transfers, 2001-2012	**OLS**	**IV – 32 SNP score**	**IV – 97 SNP score**
BMI	-0.012 (0.015)	0.251** (0.120)	0.111 (0.088)
F-statistics	..	36.53	58.92
Mean outcome	5.603	5.603	5.587
N	2038	2038	1886

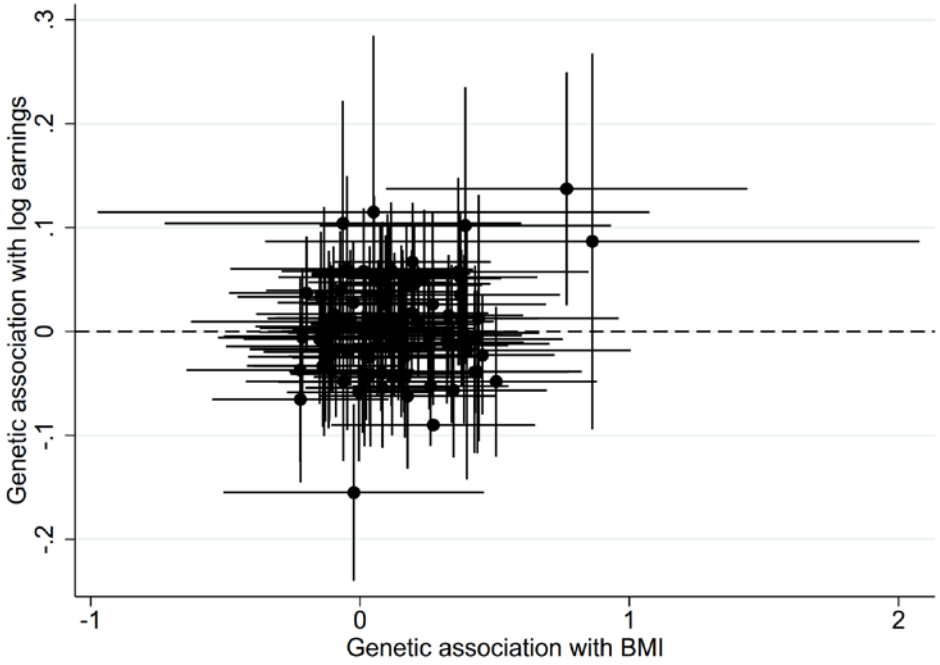
Notes: Social income transfers are measured as the log of average transfers over the period 2001-2012 for the whole sample, setting zeros to 1 Euro. The mean value for the dependent variable is reported. BMI is measured in 2001. All models include controls for the birth month and birth year effects. Gender and parental education (1980) are also controlled for in all models. The instrument used in the IV models is the BMI risk score, based on genetic markers. Angrist-Pischke multivariate F-tests of excluded instrument are reported for the IV models. Heteroscedasticity-robust standard errors are reported in parentheses: * statistically significant at the 0.10 level; ** at the 0.05 level; *** at the 0.01 level.

Figure A1. Scatter plot of genetic association with log earnings against genetic associations with BMI (32 SNP score)



Notes: Lines represent the 95% confidence intervals.

Figure A2. Scatter plot of genetic association with log earnings against genetic associations with BMI (97 SNP score)



Appendix 3:
Extensions

We estimate several additional models to assess the robustness of the results. Specifically, we estimate models separately by wave, control for genetic predisposition to obesity-related illness, control for cognitive performance, estimate models separately by sex, and estimate models for alternate measures of fatness other than BMI.

In brief, we find that when the sample size is reduced by estimating models separately by wave (instead of pooling all years), separately by sex (instead of pooling men and women), or by controlling for cognitive performance that is available for only a subset of the sample, the point estimates do not change significantly but the standard errors rise sufficiently that often the coefficients are not significantly different from zero.

We also find that the overall result that weight worsens labor market outcomes is robust to controlling for the genetic risk score for obesity-related illnesses such as high blood pressure, high cholesterol, and high triglycerides (section B); this is useful as it suggests that the mechanisms by which BMI affects labor market outcomes may not be those specific health conditions. We also find that the results are robust to using an indicator variable for obesity or weight in kg instead of BMI as the measure of fatness (section E).

A. EXTENSION 1: ESTIMATE MODELS SEPARATELY BY WAVE

In the main results we examine average labor market outcomes over the period 2001-2012 in order to minimize the influence of idiosyncratic variation that would diminish the precision of the estimates in our relatively small sample (cf. Dahl *et al.*, 2011). As an extension we estimate models for each wave of the YFS separately (2001, 2007, 2011), regressing the economic outcome for that year on BMI from that

year and other characteristics in that year (the exception is parental education, which is recorded on a single year); these results are presented in Table A7.

With the smaller samples and reduced power that comes from examining each wave separately, only two of the wave-specific estimates is statistically significant. The 32 SNP IV model indicates that a one-unit increase in BMI lowers earnings by 19.2% in 2001, compared to 7.1% over 2001-2012. The results are significant at a 10% and 5% level, respectively. The estimated effect of BMI on employment is very similar for the individual year 2001 (-2.2%) and the average of the period 2001-2012 (-1.8%), but the former is not statistically significant whereas the latter is. The second statistically significant result is that in year 2001 one-unit increase in BMI is associated with 13.9% higher social income transfers (conditional on receiving any) in the 97 SNP GRS IV model.

Although the point estimates in Table A7 are not significantly different from those in Table I, some are of the opposite sign. There are several possible explanations for the difference in point estimates. For example, it could be the result of idiosyncratic variation in labor market outcomes in a single year. Measurement error in yearly outcomes leads to attenuation bias. We strongly prefer the use of long-term measures of outcomes due to the fact that they dampen e.g. the effects of business cycle fluctuations on outcomes in a small open economy such as Finland. Conversely, contemporaneous weight (used in the wave-specific regressions) may be more relevant than one's historic weight to outcomes in that year.

B. EXTENSION 2: CONTROL FOR GENETIC PREDISPOSITION TO OBESITY-RELATED ILLNESSES

A general concern about using genetic markers as IVs is that genes may affect multiple things (pleiotropy). As an extension we control for the genetic risk score for

blood pressure, total cholesterol, and triglycerides (Teslovich *et al.*, 2010; International Consortium for Blood Pressure Genome-Wide Association Studies, 2011). If obesity affects labor market outcomes primarily because our genetic risk score (GRS) also predicts obesity-related illnesses, we would expect that directly controlling for the risk scores for those obesity-related illness to result in a reduction in the point estimate of the IV estimate for BMI. However, the genetic risk scores for other diseases are not statistically significantly correlated with genetic risk score for BMI (Appendix Table A8); for this reason it is not surprising that we find that controlling for genetic risk scores for other diseases has little impact on the point estimate of the IV coefficient on BMI. For example, the IV models based on 32 SNP GRS indicate that a one-unit increase in BMI reduces log average wages by 6.9% when we do not control for the other risk scores (Table I), and by 7.3% when we do (Table A9). A one-unit increase in BMI reduces years spent employed by 1.8% when we do not control for the other genetic risk scores (Table I), and by 1.7% when we do (Table A9). Controlling for other genetic risk scores likewise has only a small impact on the IV coefficient on BMI in the regressions for whether one received social income transfers. This suggests that the impact of BMI on labor market outcomes may not be operating through these three specific conditions.

C. EXTENSION 3: CONTROL FOR COGNITIVE PERFORMANCE

In the YFS, measures of cognitive performance are available for only a subset of the sample (1,334 out of 2,038). In the main results of this paper, we exclude cognitive performance from the set of regressors in order to avoid losing observations and thus statistical power and because cognitive performance may be affected by obesity (e.g. Sabia, 2007; Averett and Stifel, 2010). However, as an extension we estimate models in which we control for the five measures of cognitive performance:

1) overall cognitive performance 2) visual and episodic memory and visuospatial associative learning; 3) reaction time; 4) rapid visual information processing; and 5) spatial working memory. The results are presented in Appendix Table A10.

In each case, the point estimate of the coefficient on BMI is smaller in absolute magnitude and the coefficient is no longer statistically significant. However, the difference in point estimates is not statistically significant, which suggests that the lack of statistical significance in these models may be due to reduced sample size.

D. EXTENSION 4: ESTIMATE MODELS SEPARATELY BY SEX

Because of the relatively small sample size of the YFS, the main models in this paper were estimated for men and women pooled. However, previous studies of weight and labor market outcomes often found differences by sex; for example, there tends to be a greater wage penalty for excess weight among women than men in the U.S. (Cawley, 2004) and also in Finland (Sarlio-Lähteenkorva *et al.*, 2004; Johansson *et al.*, 2009). For this reason, as an extension we estimate models separately for men and women; see Appendix Tables A11 (men) and A12 (women). Interestingly, the 32 SNP genetic risk score is a more powerful instrument for men (F of 16-30) than women (F of 8-13). As found in earlier studies of weight and wages, the point estimate of the IV coefficient on BMI is larger for women than men; specifically, based on 32 SNP IV model a one-unit increase in BMI is associated with 11.0% lower wages for women (Appendix Table A12) compared to 4.6% lower wages for men (Appendix Table A11). Neither is statistically significant, however, presumably because of the smaller sample sizes (1,109 women and 929 men). Results based on 92 SNPs indicate that one-unit increase in BMI is related with 2.2 % higher wages for men and 0.5% lower wages for women but the results are not statistically significant.

The point estimates of the 32 SNP IV coefficient on BMI in the employment regressions are similar for women and men; a one-unit increase in BMI lowers the time spent working by 2.1% for women and 1.7% for men; neither is statistically significant. The respective numbers for 97 SNP IV model are 0.7% and 0.4%.

The impact of BMI on social welfare transfers is positive, large, and statistically significant for men; a one-unit increase in BMI increases the amount of social welfare transfers by 19.2%. For women the estimate is not statistically significant.

E. EXTENSION 5: ALTERNATE MEASURES OF FATNESS: OBESITY, KG

As an extension, we estimate models for additional measures of fatness: an indicator variable for obesity (BMI>=30), and weight in kilograms. Using an indicator variable for BMI>=30 allows us to focus on the effect of a BMI in the unhealthy range where the negative health effects should be more pronounced. A limitation of this approach is that there is less variation in the instrumented variable and the genetic risk score for BMI is constructed for the whole range of BMI.

Results are presented in Appendix Tables A13 (obesity) and A14 (weight in kg). Column 1 in each table presents results from OLS models. Interestingly, although the OLS coefficients on BMI were not statistically significant (in Table I), the OLS coefficients on obesity are statistically significant in earnings and employment equations. Obesity is associated with 13.3% lower wages and 3.2% fewer years employed (Appendix Table A13).

We use BMI genetic risk score as an instrument for the indicator for obesity and for weight in kg (the F-statistics range from 14 to 28 for 32 SNP GRS and from 27 to 44 for 97 SNP GRS). Appendix Table A13 presents IV results for the indicator for obesity. Results are generally consistent with the IV models for BMI: obesity

reduces earnings (although, by an implausibly large amount – 111.1%) and reduces the time spent employed by 28.6%. The IV estimates also indicate that obesity raises the probability of receiving social income transfers by 46.4 percentage points. The 97 SNP IV coefficients are statistically insignificant.

Appendix Table A14 presents results for weight in kg. Again, the results are consistent with the models for BMI. The IV models indicate that an additional 10 kg of weight reduces earnings by 27%, reduces the time spent employed by 7%, and raises the probability of receiving social income transfers by 11 percentage points. All 97 SNP IV point estimates in Table A14 are insignificant.

Appendix Table A7. The effect of BMI on contemporaneous labor market outcomes

	2001			2007			2011		
	OLS	**IV (32 SNP GRS)**	**IV (97 SNP GRS)**	**OLS**	**IV (32 SNP GRS)**	**IV (97 SNP GRS)**	**OLS**	**IV (32 SNP GRS)**	**IV (97 SNP GRS)**
Panel A: Log of Earnings									
BMI	-0.015 (0.015)	-0.192* (0.104)	-0.056 (0.081)	-0.004 (0.014)	0.016 (0.108)	0.008 (0.084)	-0.006 (0.012)	-0.007 (0.086)	0.114 (0.081)
F-statistics	..	36.53	58.92	..	33.43	45.72	..	39.05	44.03
Mean outcome	8.912		8.946	9.433		9.521	9.629		9.667
N	2038		1886	1940		1492	1729		1406
Panel B: Indicator for Being Employed									
BMI	-0.002 (0.002)	-0.022 (0.015)	-0.007 (0.012)	-0.002 (0.002)	0.005 (0.014)	-0.002 (0.011)	-0.002 (0.002)	0.001 (0.011)	0.009 (0.010)
F-statistics	..	39.90	62.73	..	35.18	46.19	..	40.65	45.67
Mean outcome	0.807		0.810	0.885		0.894	0.906		0.909
N	2062		1909	1957		1502	1742		1416
Panel C: Indicator for Social Income Transfers (extensive margin)									
BMI	0.002 (0.003)	0.007 (0.019)	-0.012 (0.015)	0.002 (0.003)	-0.010 (0.019)	-0.004 (0.016)	0.000 (0.002)	0.015 (0.017)	0.011 (0.015)
F-statistics	..	36.53	58.92	..	33.49	45.72	..	39.05	44.03
Mean outcome	0.367		0.364	0.334		0.326	0.291		0.277
N	2038		1886	1940		1492	1729		1406
Panel D: Log of Social Income Transfers (intensive margin)									
BMI	0.016 (0.011)	0.155 (0.104)	0.139* (0.071)	0.013 (0.012)	0.127 (0.136)	0.039 (0.071)	-0.003 (0.015)	0.060 (0.081)	0.090 (0.075)
F-statistics	..	9.96	18.93	..	6.12	17.88	..	16.06	13.83
Mean outcome	7.726		7.721	7.853		7.795	7.893		7.909
N	748		686	647		486	503		389

Notes: The outcomes and BMI are measured in 2001, 2007 and 2011. All models include controls for the birth month and birth year effects. Gender and parental education (1980) are also controlled for in all models. The instrument used in the IV models is the BMI risk score, based on genetic markers. Angrist-Pischke multivariate F-tests of excluded instrument are reported for the IV models. Heteroscedasticity-robust standard errors are reported in parentheses: * statistically significant at the 0.10 level; ** at the 0.05 level; *** at the 0.01 level.

Appendix Table A8:
Correlations between genetic risk scores for BMI and other diseases

	Blood pressure GRS	Total cholesterol GRS	Triglycerides GRS
BMI 32 SNP GRS	0.003	-0.009	-0.024
BMI 97 SNP GRS	-0.001	-0.025	0.005

Notes: * statistically significant at the 0.10 level; ** at the 0.05 level; *** at the 0.01 level. N=2001.

Appendix Table A9: The effect of BMI on average labor market outcomes, 2001-2012 with controls for other genetic markers

Panel A: Log of Average Earnings, 2001-2012	**OLS**	**IV – 32 SNP score**	**IV – 97 SNP score**
BMI	-0.007 (0.005)	-0.073** (0.036)	0.009 (0.026)
F-statistics	…	36.69	63.24
Mean outcome	9.860	9.860	9.864
N	2001	2001	1849
Panel B: Share of Years Employed, 2001-2012	**OLS**	**IV – 32 SNP score**	**IV – 97 SNP score**
BMI	-0.002 (0.001)	-0.017* (0.009)	-0.004 (0.007)
F-statistics	..	40.09	67.46
Mean outcome	0.857	0.857	0.859
N	2024	2024	1871
Panel C: Indicator for Social Income Transfers, 2001-2012 (extensive margin)	**OLS**	**IV – 32 SNP score**	**IV – 97 SNP score**
BMI	-0.002 (0.002)	0.027* (0.015)	0.017 (0.011)
F-statistics	..	36.69	63.24
Mean outcome	0.822	0.822	0.820
N	2001	2001	1849
Panel D: Log of Average Social Income Transfers, 2001-2012 (intensive margin)	**OLS**	**IV – 32 SNP score**	**IV – 97 SNP score**
BMI	-0.005 (0.009)	0.063 (0.068)	-0.020 (0.043)
F-statistics	..	23.60	56.35
Mean outcome	6.841	6.841	6.836
N	1645	1645	1516

Notes: Earnings are measured as the log of average earnings over the period 2001-2012. Employment is measured as the average share of employment years over the period 2001-2012. Indicator for social income transfers equals one for those who have received social security transfers at least once during 2001-2012. Social income transfers are measured as the log of average transfers over the period 2001-2012, conditional on obtaining a positive amount of transfers. The mean values for the dependent variables are reported. BMI is measured in 2001. All models include controls for the birth month and birth year effects. Gender and parental education (1980) and the genetic risk scores for blood pressure, total cholesterol, and triglycerides are also controlled for in all models. The instrument used in the IV models is the BMI risk score, based on genetic markers. Angrist-Pischke multivariate F-tests of excluded instrument are reported for the IV models. Heteroscedasticity-robust standard errors are reported in parentheses: * statistically significant at the 0.10 level; ** at the 0.05 level; *** at the 0.01 level.

Appendix Table A10:
The effect of BMI on average labor market outcomes, 2001-2012 with controls for cognitive performance

Panel A: Log of Average Earnings, 2001-2012	**OLS**	**IV – 32 SNP score**	**IV – 97 SNP score**
BMI	-0.006 (0.005)	-0.021 (0.029)	0.017 (0.029)
F-statistics	…	34.63	45.62
Mean outcome	9.956	9.956	9.964
N	1334	1334	1237
Panel B: Share of Years Employed, 2001-2012	**OLS**	**IV – 32 SNP score**	**IV – 97 SNP score**
BMI	-0.000 (0.001)	-0.007 (0.008)	0.003 (0.008)
F-statistics	..	34.77	46.82
Mean outcome	0.886	0.886	0.888
N	1339	1339	1242
Panel C: Indicator for Social Income Transfers, 2001-2012 (extensive margin)	**OLS**	**IV – 32 SNP score**	**IV – 97 SNP score**
BMI	-0.003 (0.003)	0.024 (0.016)	0.010 (0.014)
F-statistics	..	34.63	45.62
Mean outcome	0.813	0.813	0.812
N	1334	1334	1237
Panel D: Log of Average Social Income Transfers, 2001-2012 (intensive margin)	**OLS**	**IV – 32 SNP score**	**IV – 97 SNP score**
BMI	0.002 (0.011)	0.058 (0.061)	-0.016 (0.050)
F-statistics	..	29.38	43.80
Mean outcome	6.747	6.747	6.740
N	1085	1085	1005

Notes: Earnings are measured as the log of average earnings over the period 2001-2012. Employment is measured as the average share of employment years over the period 2001-2012. Indicator for social income transfers equals one for those who have received social security transfers at least once during 2001-2012. Social income transfers are measured as the log of average transfers over the period 2001-2012, conditional on obtaining a positive amount of transfers. The mean values for the dependent variables are reported. BMI is measured in 2001. All models include controls for the birth month and birth year effects. Gender and parental education (1980) are also controlled for in all models. The instrument used in the IV models is the BMI risk score, based on genetic markers. Angrist-Pischke

multivariate F-tests of excluded instrument are reported for the IV models. Heteroscedasticity-robust standard errors are reported in parentheses: * statistically significant at the 0.10 level; ** at the 0.05 level; *** at the 0.01 level.

Appendix Table A11:
The effect of BMI on average labor market outcomes 2001-2012 for men

Panel A: Log of Average Earnings, 2001-2012	**OLS**	**IV – 32 SNP score**	**IV – 97 SNP score**
BMI	-0.004 (0.008)	-0.046 (0.037)	0.022 (0.037)
F-statistics	..	29.28	39.12
Mean outcome	10.082	10.082	10.075
N	929	929	859
Panel B: Share of Years Employed, 2001-2012	**OLS**	**IV – 32 SNP score**	**IV – 97 SNP score**
BMI	-0.001 (0.002)	-0.017 (0.011)	-0.004 (0.010)
F-statistics	..	29.82	40.22
Mean outcome	0.889	0.889	0.890
N	937	937	866
Panel C: Indicator for Social Income Transfers, 2001-2012 (extensive margin)	**OLS**	**IV – 32 SNP score**	**IV – 97 SNP score**
BMI	-0.004 (0.003)	0.021 (0.020)	0.019 (0.017)
F-statistics	..	29.28	39.12
Mean outcome	0.763	0.763	0.761
N	929	929	859
Panel D: Log of Average Social Income Transfers, 2001-2012 (intensive margin)	**OLS**	**IV – 32 SNP score**	**IV – 97 SNP score**
BMI	-0.007 (0.016)	0.192* (0.105)	0.000 (0.068)
F-statistics	..	15.88	34.13
Mean outcome	6.260	6.260	6.245
N	709	709	654

Notes: Earnings are measured as the log of average earnings over the period 2001-2012. Employment is measured as the average share of employment years over the period 2001-2012. Indicator for social income transfers equals one for those who have received social security transfers at least once during 2001-2012. Social income transfers are measured as the log of average transfers over the period 2001-2012, conditional on obtaining a positive amount of transfers. The mean values for the dependent variables are reported. BMI is measured in 2001. All models include controls for the birth month and birth year effects. Parental education (1980) is also controlled for in all models. The instrument used in the IV models is the BMI risk score, based on genetic markers. Angrist-Pischke multivariate F-tests of excluded instrument are reported for the IV models. Heteroscedasticity-robust standard errors are reported in parentheses: * statistically significant at the 0.10 level; ** at the 0.05 level; *** at the 0.01 level.

Appendix Table A12:
The effect of BMI on average labor market outcomes 2001-2012 for women

Panel A: Log of Average Earnings, 2001-2012	**OLS**	**IV – 32 SNP score**	**IV – 97 SNP score**
BMI	-0.009 (0.007)	-0.110 (0.073)	-0.005 (0.039)
F-statistics	..	10.04	21.99
Mean outcome	9.679	9.679	9.692
N	1109	1109	1027
Panel B: Share of Years Employed, 2001-2012	**OLS**	**IV – 32 SNP score**	**IV – 97 SNP score**
BMI	-0.002 (0.002)	-0.021 (0.016)	-0.007 (0.011)
F-statistics	..	12.54	24.92
Mean outcome	0.831	0.831	0.834
N	1125	1125	1043
Panel C: Indicator for Social Income Transfers, 2001-2012 (extensive margin)	**OLS**	**IV – 32 SNP score**	**IV – 97 SNP score**
BMI	0.001 (0.002)	0.051* (0.028)	0.021 (0.016)
F-statistics	..	10.04	21.99
Mean outcome	0.869	0.869	0.868
N	1109	1109	1027
Panel D: Log of Average Social Income Transfers, 2001-2012 (intensive margin)	**OLS**	**IV – 32 SNP score**	**IV – 97 SNP score**
BMI	-0.004 (0.009)	-0.079 (0.099)	-0.034 (0.062)
F-statistics	..	8.17	20.26
Mean outcome	7.260	7.260	7.262
N	964	964	891

Notes: Earnings are measured as the log of average earnings over the period 2001-2012. Employment is measured as the average share of employment years over the period 2001-2012. Indicator for social income transfers equals one for those who have received social security transfers at least once during 2001-2012. Social income transfers are measured as the log of average transfers over the period 2001-2012, conditional on obtaining a positive amount of transfers. The mean values for the dependent variables are reported. BMI is measured in 2001. All models include controls for the birth month and birth year effects. Parental education (1980) is also controlled for in all models. The instrument used in the IV models is the BMI risk score, based on genetic markers. Angrist-Pischke multivariate F-tests of excluded instrument are reported for the IV models. Heteroscedasticity-robust standard errors are reported in parentheses: * statistically significant at the 0.10 level; ** at the 0.05 level; *** at the 0.01 level.

Appendix Table A13:
The effect of obesity (BMI ≥ 30) on average labor market outcomes 2001-2012

Panel A: Log of Average Earnings, 2001-2012	**OLS**	**IV – 32 SNP score**	**IV – 97 SNP score**
Obesity (BMI ≥ 30)	-0.133* (0.068)	-1.111* (0.568)	0.176 (0.465)
F-statistics	..	24.70	31.34
Mean outcome	9.863	9.863	9.866
N	2038	2038	1886
Panel B: Share of Years Employed, 2001-2012	**OLS**	**IV – 32 SNP score**	**IV – 97 SNP score**
Obesity (BMI ≥ 30)	-0.032* (0.019)	-0.286* (0.146)	-0.085 (0.126)
F-statistics	..	27.52	33.58
Mean outcome	0.857	0.857	0.859
N	2062	2062	1909
Panel C: Indicator for Social Income Transfers, 2001-2012 (extensive margin)	**OLS**	**IV – 32 SNP score**	**IV – 97 SNP score**
Obesity (BMI ≥ 30)	-0.024 (0.027)	0.464* (0.251)	0.330 (0.207)
F-statistics	..	24.70	31.34
Mean outcome	0.821	0.821	0.819
N	2038	2038	1886
Panel D: Log of Average Social Income Transfers, 2001-2012 (intensive margin)	**OLS**	**IV – 32 SNP score**	**IV – 97 SNP score**
Obesity (BMI ≥ 30)	-0.003 (0.115)	1.009 (1.149)	-0.360 (0.803)
F-statistics	..	14.72	26.97
Mean outcome	6.836	6.836	6.831
N	1673	1673	1545

Notes: Obesity indicator equals one for whose BMI was at least 30 in 2001. Earnings are measured as the log of average earnings over the period 2001-2012. Employment is measured as the average share of employment years over the period 2001-2012. Indicator for social income transfers equals one for those who have received social security transfers at least once during 2001-2012. Social income transfers are measured as the log of average transfers over the period 2001-2012, conditional on obtaining a positive amount of transfers. The mean values for the dependent variables are reported. All models include controls for the birth month and birth year effects. Gender and parental education (1980) are also controlled for in all models. The instrument used in the IV models is the BMI risk score, based on genetic markers. Angrist-Pischke multivariate F-tests of excluded instrument are reported for the IV models. Heteroscedasticity-robust standard errors are reported in parentheses: * statistically significant at the 0.10 level; ** at the 0.05 level; *** at the 0.01 level.

Appendix Table A14:
The effect of weight (kg) on average labor market outcomes 2001-2012

Panel A: Log of Average Earnings, 2001-2012	**OLS**	**IV – 32 SNP score**	**IV – 97 SNP score**
Weight (kg)	-0.000 (0.002)	-0.027* (0.014)	0.004 (0.010)
F-statistics	..	23.40	41.05
Mean outcome	9.863	9.863	9.866
N	2038	2038	1886
Panel B: Share of Years Employed, 2001-2012	**OLS**	**IV – 32 SNP score**	**IV – 97 SNP score**
Weight (kg)	-0.000 (0.000)	-0.007* (0.004)	-0.002 (0.003)
F-statistics	..	26.21	43.88
Mean outcome	0.857	0.857	0.859
N	2062	2062	1909
Panel C: Indicator for Social Income Transfers, 2001-2012 (extensive margin)	**OLS**	**IV – 32 SNP score**	**IV – 97 SNP score**
Weight (kg)	-0.001 (0.001)	0.011* (0.006)	0.007 (0.004)
F-statistics	..	23.40	41.05
Mean outcome	0.821	0.821	0.819
N	2038	2038	1886
Panel D: Log of Average Social Income Transfers, 2001-2012 (intensive margin)	**OLS**	**IV – 32 SNP score**	**IV – 97 SNP score**
Weight (kg)	-0.002 (0.003)	0.024 (0.027)	-0.007 (0.016)
F-statistics	..	14.02	35.62
Mean outcome	6.836	6.836	6.831
N	1673	1673	1545

Notes: Weight (kg) is measured in 2001. Earnings are measured as the log of average earnings over the period 2001-2012. Employment is measured as the average share of employment years over the period 2001-2012. Indicator for social income transfers equals one for those who have received social security transfers at least once during 2001-2012. Social income transfers are measured as the log of average transfers over the period 2001-2012, conditional on obtaining a positive amount of transfers. The mean values for the dependent variables are reported. All models include controls for the birth month and birth year effects. Gender and parental education (1980) are also controlled for in all models. The instrument used in the IV models is the BMI risk score, based on genetic markers. Angrist-Pischke multivariate F-tests of excluded instrument are reported for the IV models. Heteroscedasticity-robust standard errors are reported in parentheses: * statistically significant at the 0.10 level; ** at the 0.05 level; *** at the 0.01 level.