









# The effect of weight on labor market outcomes: An application of genetic instrumental variables

Petri Böckerman<sup>1</sup>  | John Cawley<sup>2</sup>  | Jutta Viinikainen<sup>3</sup>  | Terho Lehtimäki<sup>4</sup>  |  
Suvi Rovio<sup>5</sup>  | Ilkka Seppälä<sup>4</sup>  | Jaakko Pehkonen<sup>3</sup>  | Olli Raitakari<sup>5,6</sup> 

<sup>1</sup>Labour Institute for Economic Research and IZA, Jyväskylä University School of Business and Economics, University of Jyväskylä, Jyväskylä, Finland

<sup>2</sup>Department of Policy Analysis and Management, and Department of Economics, Cornell University, Ithaca, New York

<sup>3</sup>Jyväskylä University School of Business and Economics, University of Jyväskylä, Jyväskylä, Finland

<sup>4</sup>Department of Clinical Chemistry, Fimlab Laboratories and Finnish Cardiovascular Research Center Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland

<sup>5</sup>Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland

<sup>6</sup>Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland

## Correspondence

Petri Böckerman, University of Jyväskylä, Labour Institute for Economic Research and IZA, Helsinki, Finland.

Email: petri.bockerman@labour.fi

## Funding information

Competitive State Research Financing of the Expert Responsibility area of Kuopio; Diabetes Research Foundation of Finnish Diabetes Association; Emil Aaltonen Foundation; Finnish Foundation for Cardiovascular Research; Juho Vainio Foundation; OP Group Research Foundation; Paavo Nurmi Foundation; Palkansaaja Foundation; Robert Wood Johnson

## 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 (BMI) across many genetic loci. We estimate instrumental variable models of the effect of BMI on labor market outcomes using Finnish data that have many strengths, for example, BMI 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 nonresponse, 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

earnings, employment, genetics, obesity, overweight, social income transfers

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2018 The Authors. *Health Economics* Published by John Wiley & Sons, Ltd.

[The copyright line for this article was changed on 20 November 2018 after original online publication.]

Foundation; Signe and Ane Gyllenberg Foundation; Social Insurance Institution of Finland; Strategic Research Council funding for the project Work, Inequality and Public Policy, Grant/Award Number: 293120; Tampere and Turku University Hospitals, Grant/Award Number: X51001; EU Horizon 2020, Grant/Award Number: 755320; Yrjö Jahnsson Foundation, Grant/Award Numbers: 6664 and 6646; Tampere Tuberculosis Foundation; Finnish Cultural Foundation; Juho Vainio Foundation; Competitive State Research Financing of the Expert Responsibility area of Kuopio, Tampere and Turku University Hospitals, Grant/Award Number: X51001; Academy of Finland, Grant/Award Numbers: 117787, 121584, 124282, 126925, 129378, 134309, 286284 and 41071

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,<sup>1</sup> 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, for example, 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, Smith, & Bogin, 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, for example, Cawley (2004), Brunello and D'Hombres (2007), Kline and Tobias (2008), Lindeboom, Lundborg, and van der Klaauw (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, Farooqi, & O'Rahilly, 2000; Locke, Kahali, Berndt, et al., 2015; Pietiläinen et al., 1999). 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.

This paper contributes to the literature by using a novel instrument—genetic risk score (GRS) 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; Davey Smith, Paternoster, & Relton, 2017; Haycock et al., 2016; Tyrrell, Jones, Beaumont, et al., 2016; von Hinke, Davey Smith, Lawlor, Propper, & Windmeijer, 2016).<sup>2</sup> We utilize two different GRS for high BMI; a narrower one based on 32 single nucleotide polymorphisms (SNPs)<sup>3</sup> 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 Section 3.

In addition to the genetic IVs, the data we use have 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

<sup>1</sup>BMI is calculated as a person's weight in kilograms divided by height in meters squared.

<sup>2</sup>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, & Norton, 2011; von Hinke et al., 2016).

<sup>3</sup>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, Davey Smith, Lawlor, Propper, and Windmeijer (2012).

inefficiency and bias (see Cawley, Maclean, Hammer, & Wintfeld, 2015; Courtemanche, Pinkston, & Stewart, 2015). Second, we utilize administrative data on earnings, which avoid 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 allow 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 income transfers are also of interest because they tend to be a substantial item in government budgets, especially in countries such as 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 avoid 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, Tyrrell 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 United Kingdom.

## 2 | DATA

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

The YFS is an ongoing 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.<sup>4</sup> Subjects in six age cohorts (aged 3, 6, 9, 12, 15, and 18 years) were randomly chosen from five university hospital districts of Finland using the national population register (Raitakari, Juonala, Rönnemaa, et al., 2008).<sup>5</sup> The sample is relatively small—3,596 persons participated in the study at baseline—but the richness of the data is an offsetting advantage. Eight waves of data have been collected in 3- to 9-year intervals, starting with baseline in 1980 and most recently in 2011–2012, 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 were performed for YFS subjects using the 670K Illumina platform (Sanger Institute, UK). Variation in over 670,000 known 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 GRS, which will be explained in detail in Section 3.

The second dataset that we use, the FLEED, is the source for data on employment status, salary, and other income, 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, for example, 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 LPC, is the source of information on parental education.

## 3 | METHODS

We estimate regressions of the following form:

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

<sup>4</sup>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>

<sup>5</sup>Finland is divided into 20 hospital districts, five of which are university hospital districts.

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, DeLeire, & Schwabish, 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.<sup>6</sup> 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.

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 instrumental variables (IV) estimation, that is, 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 (Averett & Stifel, 2010; Sabia, 2007).<sup>7</sup>

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 Supporting Information we also estimate models separately by gender.<sup>8</sup>

We first estimate Equation (1) using ordinary least squares (OLS) in order to estimate the conditional correlation of weight with labor market outcomes.<sup>9</sup> These correlations not only reflect any causal effect of weight on wages but also

<sup>6</sup>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.

<sup>7</sup>See Supporting Information for description of the measures for cognitive performance.

<sup>8</sup>Finland also exhibits a gender difference in the wage penalty of obesity (Johansson, Böckerman, Kiiskinen, & Heliövaara, 2009; Sarlio-Lähteenkorva, Silventoinen, & Lahelma, 2004).

<sup>9</sup>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.

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: GRS for obesity

In order to estimate the causal effect of body weight on these outcomes, we estimate models of IV in which our IV is one of the two GRS for BMI.<sup>10</sup> It is estimated that 40–70% of interindividual 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 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, Willer, Berndt, 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 GRS 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. Whereas the 32 SNP score was used by von Hinke et al. (2016), the 97 SNP score has not, to our knowledge, been used in any study examining the economic consequences of obesity.

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 zero, one, or two 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 Table S2. 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 GRS 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.<sup>11</sup>

Speliotes et al. (2010) report that the mechanisms by which these SNPs affect weight are through (a) regulators of appetite or energy balance and (b) 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 second risk score explain 2.7% of the variation in BMI (Locke et al., 2015).<sup>12</sup> (Even though it is estimated that 40–70% of interindividual 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 et al. (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 et al., 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.

<sup>10</sup>For all outcomes, the IV model is two-stage least squares.

<sup>11</sup>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.

<sup>12</sup>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 (fat mass and obesity associated gene) is associated with 20.3% greater odds of obesity. We have also estimated models in which the IV is a 32-SNP GRS 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 32-SNP GRS.



There is a possible trade-off between power and validity associated with using a broader SNP risk score (i.e., one based on more SNPs). The advantage is that a broader risk score may be more powerful (explains more variation in BMI), given that it is based on additional SNPs.<sup>13</sup> However, there is also a risk that some of those additional SNPs will also be correlated with other traits that affect labor market outcomes (pleiotropy), and thus, there may be a greater risk of bias in the IV estimates. The two GRS 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).<sup>14</sup>

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 (Kaplowitz, 2008; Wang, 2002). In summary, the other phenotypes that the obesity-related 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, McNeil, and Newhouse (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.<sup>15</sup> These comparisons will be discussed in Section 4.

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 Table S1, 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 32 coefficients on the SNPs are jointly equal to zero. In Column 2 of Table S1, 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.

<sup>13</sup>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.

<sup>14</sup>We report the results additionally controlling for height in the working paper version. The conclusions remain intact.

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

**TABLE 1** The effect of BMI on average labor market outcomes, 2001–2012

	(1)	(2)	(3)
Panel A: log of average earnings, 2001–2012	OLS	IV—32 SNP score	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	2,038	2,038	1,886
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	2,062	2,062	1,909
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	2,038	2,038	1,886
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	1,673	1,673	1,545

*Note.* 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. IV: instrumental variables; OLS: ordinary least squares; SNP: single nucleotide polymorphism. “..” means not applicable.

\*Statistically significant at the 0.10 level. \*\*Statistically significant at the 0.05 level. \*\*\*Statistically significant at the 0.01 level.

## 4 | EMPIRICAL RESULTS

### 4.1 | Baseline OLS estimates

OLS estimates are presented in Column 1 of Table 1.<sup>16</sup> 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.<sup>17</sup>

### 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 GRS 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 GRS, and between 52.0 and 62.7 for the 97 SNP GRS; all of these far exceed the minimum standard of  $F = 10$  suggested in Staiger and Stock (1997).<sup>18</sup>

In order to examine the validity of the genetic instruments, Table S3 presents differences in means of the observed variables for those with above- and below-average values of the BMI GRS, 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

<sup>16</sup>Table S2 reports summary statistics for our regression sample.

<sup>17</sup>We estimate several additional models to assess the robustness of the baseline results using the narrower score (Supporting Information).

<sup>18</sup>We have also estimated IV models that use both the GRS and its square as instruments. The first-stage  $F$ -statistics in these models are lower (roughly 18 for the 32-SNP score), and the results are similar, so we continue to estimate models that simply use the level of the risk score as the IV.

shows that those with a higher 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. Whereas 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 Table S3 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 (Averett & Stifel, 2010; Sabia, 2007), 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 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 Table S1.)

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, Minelli, Sheehan, & Thompson, 2015). To visually illustrate the potential heterogeneity Figure S1 plots, the genetic associations with log earnings (vertical axis) against genetic associations with the BMI (horizontal axis) for each of the 32 SNPs. (Figure S2 plots the same associations for the 97 SNPs.) Each point in Figure S1 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, Bowden, Fall, Ingelsson, & Thompson, 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 S2, 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 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 overidentification 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 1. 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<sup>19</sup> 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.<sup>20</sup> 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 based on the 32 SNP score, but

<sup>19</sup>Earnings and social income transfers have been log-transformed, 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.

<sup>20</sup>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 (Table S4).



**TABLE 2** Reduced-form IV estimates

	(1)	(2)
Panel A: log of average earnings, 2001–2012	32 SNP score	97 SNP score
Mean outcome	−0.012** (0.006)	0.046 (0.122)
<i>N</i>	9,863	9,866
	2,038	1,886
Panel B: share of years employed, 2001–2012	32 SNP score	97 SNP score
Mean outcome	−0.003** (0.002)	−0.023 (0.035)
<i>N</i>	0,857	0,859
	2,062	1,909
Panel C: indicator for social income transfers, 2001–2012 (extensive margin)	32 SNP score	97 SNP score
Mean outcome	0.005** (0.002)	0.087* (0.053)
<i>N</i>	0,821	0,819
	2,038	1,886
Panel D: log of average social income transfers, 2001–2012 (intensive margin)	32 SNP score	97 SNP score
Mean outcome	0.009 (0.010)	−0.095 (0.211)
<i>N</i>	6,836	6,831
	1,673	1,545

*Note.* 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. IV: instrumental variables; SNP: single nucleotide polymorphism.

\*Statistically significant at the 0.10 level. \*\*Statistically significant at the 0.05 level.

\*\*\*Statistically significant at the 0.01 level.

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.

### 4.3.1 | Reduced-form estimates

Table 2 presents results of reduced-form models that regress outcomes on the instrument (BMI GRS) directly, controlling for the same set of regressors as earlier. The results are consistent with those of the IV models. Raising the 32 SNP GRS 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 2).

Again, the choice of GRS makes a difference in the results. Column 2 of Table 2 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 an 8.7 percentage point (10.6%) increase in the probability of receiving any social income transfers.<sup>21</sup>

## 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: GRS for obesity based on many SNPs that are robustly associated with high BMI. Specifically, we use two such GRS and find evidence that IV model estimates are sensitive to which risk score is used.

<sup>21</sup>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 the earnings equation, there was a significant difference between the 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.

The estimates of the IV models that use the GRS 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 nonobese 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 & 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 GRS based on 32 SNPs are also consistent with the hypothesis that additional weight worsens labor market outcomes; raising the GRS 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 GRS 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: (a) the 97 SNP score but not the 32 SNP score varies by sex, age, marital status (Table S1), and father's income (Tables S1 and S3); and (b) 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 (Figures S1 and S2). 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 GRS 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, that is, 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, Brown, & Mathiowetz, 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 GRS 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.

## ACKNOWLEDGMENTS

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 the Yrjö Jahnsson Foundation (Grants 6646 and 6664). Viinikainen also thanks OP Group Research Foundation for financial support. 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, IIPF 2018, 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 SF. To obtain access to the data, please contact SF, FI-00022, Helsinki, Finland. The specific instructions to obtain access to the data are available at [https://tilastokeskus.fi/tup/mikroaineistot/hakumenettely\\_en.html](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 SF, 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.

## ORCID

Petri Böckerman  <http://orcid.org/0000-0002-5372-2985>

John Cawley  <http://orcid.org/0000-0002-4805-9883>

Jutta Viinikainen  <http://orcid.org/0000-0002-4252-3147>  
 Terho Lehtimäki  <http://orcid.org/0000-0002-2555-4427>  
 Suvi Rovio  <http://orcid.org/0000-0003-0911-7478>  
 Ilkka Seppälä  <http://orcid.org/0000-0002-5595-2573>  
 Jaakko Pehkonen  <http://orcid.org/0000-0002-9684-7139>  
 Olli Raitakari  <http://orcid.org/0000-0001-9365-3702>

## REFERENCES

- Averett, S. L. (2011). Labor market consequences: Employment, wages, disability, and absenteeism. In J. Cawley (Ed.), *The Oxford handbook of the social science of obesity* (pp. 531–552). New York: Oxford University Press.
- Averett, S. L., & Stifel, D. C. (2010). Race and gender differences in the cognitive effects of childhood overweight. *Applied Economics Letters*, 17, 1673–1679.
- Barsh, G. S., Farooqi, I. S., & 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 E. Leamer, & J. Heckman (Eds.), *Handbook of econometrics* (ed., Vol. 5) (pp. 3705–3843). Amsterdam: North Holland.
- 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, S. G. (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, J. C., 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, J. C., & 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, J. A. (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.
- 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, F. D., Minelli, C., Sheehan, N. A., & Thompson, J. R. (2015). Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. *Statistics in Medicine*, 34, 2926–2940.
- Haycock, P. C., Burgess, S., Wade, K. H., 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*. New York: Oxford University Press.
- 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, P. B. (2008). Link between body fat and the timing of puberty. *Pediatrics*, 121(Supplement 3), S208–S217.
- Kline, B., & Tobias, J. L. (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, P. K., & 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, A. E., Kahali, B., Berndt, S. I., et al. (2015). Genetic studies of body mass index yield new insights for obesity biology. *Nature*, 518, 197–206.
- McClellan, M., McNeil, B. J., & Newhouse, J. P. (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, J. A. (1974). *Schooling, experience, and earnings*. New York: National Bureau of Economic Research.
- Ng, M., Fleming, T., Robinson, M., Thomson, B., Graetz, N., Margono, C., ... Gakidou, E. (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, E. C., & Han, E. (2008). Genetic information, obesity and labor market outcomes. *Health Economics*, 17, 1089–1104.
- Palmer, T. M., Lawlor, D. A., Harbord, R. M., Sheehan, N. A., Tobias, J. H., Timpson, N. J., ... Sterne, J. A. C. (2012). Using multiple genetic variants as instrumental variables for modifiable risk factors. *Statistical Methods in Medical Research*, 21, 223–242.
- Pietiläinen, K. H., Kaprio, J., Rissanen, A., Winter, T., Rimpelä, A., Viken, R. J., & Rose, R. J. (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, R. M. (2011). Bias, stigma, and discrimination. In J. Cawley (Ed.), *The Oxford handbook of the social science of obesity* (pp. 553–571). New York: Oxford University Press.
- Raitakari, O. T., Juonala, M., Rönnemaa, T., et al. (2008). Cohort profile: The cardiovascular risk in Young Finns Study. *International Journal of Epidemiology*, 37, 1220–1226.
- Rooth, D.-O. (2009). Obesity, attractiveness, and differential treatment in hiring: A field experiment. *Journal of Human Resources*, 44, 710–735.
- Sabia, J. J. (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, M. D. (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, E. K., Willer, C. J., Berndt, S. I., 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, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, 65, 447–486.
- Tyrrell, J., Jones, S. E., Beaumont, R., et al. (2016). Height, body mass index, and socioeconomic status: Mendelian randomisation study in UK Biobank. *British Medical Journal*, i582, 352.
- von Hinke Kessler Scholder, S., Davey Smith, G., Lawlor, D. A., 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, D. A., 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.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Böckerman P, Cawley J, Viinikainen J, et al. The effect of weight on labor market outcomes: An application of genetic instrumental variables. *Health Economics*. 2019;28:65–77. <https://doi.org/10.1002/hec.3828>