

Association of Obesity with Health Care Costs: A Mendelian Randomization study

Christoph Kurz, Michael Laxy

Institute of Health Economics and Health Care Management, Helmholtz Zentrum München

May 7, 2019

Abstract

Causal effect estimates for the association of obesity with health care costs can be biased by reversed causation and omitted variables. In this study, we use genetic variants as instrumental variables to overcome these limitations. We estimate the effect of body mass index (BMI) and waist-to-hip ratio (WHR) on total health care costs using data from a German observational study. We find that the model using genetic instruments identifies additional annual costs of 189€ for a one unit increase in BMI, and additional 1165€ for a 0.1 unit increase in WHR. This is more than two times higher than estimates from linear regression without instrumental variables. We found little evidence of a non-linear relationship between BMI or WHR and health care costs. Our results imply that the use of genetic instruments can be a powerful tool for estimating causal effects in health economic evaluation that might be superior to other types of instruments where there is a strong association with a modifiable risk factor.

1 Introduction

Overweight and obesity are global public health concerns in terms of economic costs and effect on health ([Wang et al., 2011](#)). The projected global prevalence of obesity will reach 18% in men and surpass 21% in women in 2025 ([NCD Risk Factor Collaboration, 2016](#)). Higher body mass index (BMI), expressed in units of kg/m^2 , is a risk factor for type II diabetes mellitus, cardiovascular diseases and certain types of cancer, resulting in shortened life expectancy ([Hu et al., 2004](#)). According to the [World Health Organization \(2000\)](#), individuals are classified as obese with a BMI $\geq 30 \text{ kg}/\text{m}^2$ or a waist-to-hip ratio (WHR) ≥ 0.85 for women and ≥ 0.9 for men. While life style factors like physical activity and nutrition are important drivers of obesity, there is also a significant genetic component: at least 34% of BMI variation can be explained by genetic loci ([Yang et al., 2011](#); [Maes et al., 1997](#)).

Accurate estimates of health care costs are crucial to judge the trade offs between medical possibilities, their financial viability, as well as quality and fairness in any health care system. Because of practical and ethical issues, many health risk conditions (e.g. obesity) cannot be assessed in randomized controlled trials (RCTs) and cost data collected in RCTs often has limited generalizability ([Drummond et al., 2015](#)). On the other hand, observational studies can estimate the relation between health care costs and health conditions, but generally cannot identify causal effects ([Cawley and Meyerhoefer, 2012](#)). Many previous studies that estimated the association of obesity with health care costs arise from observational studies (e.g. [Finkelstein et al. \(2009\)](#); [Trasande et al. \(2009\)](#)) that are unable to fully account for unmeasured confounders. For example, factors like social deprivation or bias by self-reported height and weight measures are often not accounted for. In addition, the direction of the omitted variable bias is usually unclear. [Cawley and Meyerhoefer \(2012\)](#) discuss that some people became obese after suffering an injury or chronic depression, and have higher medical costs because of the injury or depression. [Carreras-Torres](#)

et al. (2018) find that higher BMI increases the risk of being a smoker, but smoking itself lowers BMI (Dare et al., 2015).

The presence of reverse causation and omitted variables motivated the use of instrumental variable (IV) analysis. The aim of the IV method is to exclude a correlation between the explanatory variables and the error term in a regression analysis. This is done by replacing the explanatory variables with other variables that are closely related to them but do not correlate with the error term or are a linear combination of other explanatory variables.

For example, in the case of health care costs of obesity, Cawley and Meyerhoefer (2012) used the BMI of a biological relative as an IV. Compared to a standard ordinary least squares (OLS) model, the authors found a fourfold increase of marginal costs of obesity, and a threefold increase in the marginal costs of one unit of BMI. Studies by Black et al. (2018) and Kinge and Morris (2018) found similar differences between OLS and IV models. They both used the BMI of a biological relative as an IV.

However, a potential concern with this approach is that unobserved characteristics may be correlated with both a person's own BMI and their relative's BMI, for example when they live in the same household (Böckerman et al., 2019).

In this paper, we use an IV approach called *Mendelian Randomization* (MR). MR is originally a concept from genetic epidemiology that uses genetic variants as instrumental variables to overcome some limitations in cases where genetic polymorphisms have a well-known effect on modifiable risk factors like BMI (Thomas and Conti, 2004). If the IV assumptions hold, genetic variants can be used to estimate the effect of BMI and WHR on health care costs. Because genotypes are assigned randomly when passed from parents to offspring, the population genotype distribution is assumed to be unrelated to confounders and can therefore serve as valid instruments. This random segregation of alleles, according to Mendel's law of independent assortment, mimics a natural experiment in which individuals are randomly assigned into groups based on their exposure (Ebrahim and Davey Smith, 2008).

Recent literature suggested that MR can be a powerful tool for economic evaluation (Dixon et al., 2016), however few studies have yet used it (Ding et al., 2009; Fletcher and Lehrer, 2011; Norton and Han, 2008; Cawley et al., 2011; Willage, 2017; von Hinke et al., 2016; Böckerman et al., 2019).

In this study we use Mendelian Randomization to estimate the effect of BMI and WHR on total annual health care costs and compare the results to a simple OLS model.

The rest of the paper is organized as follows. In Section 2 we first outline our data in Subsection 2.1. We then detail the MR approach in Subsection 2.2 and potential problems regarding the validity of this approach in Subsections 2.3 to 2.5. We present results in Section 3. Finally, we conclude in Section 4 with implications of the results.

2 Material and Methods

2.1 Data

We use data from the KORA (Cooperative Health Research in the Augsburg Region) F4 study (2006-2010, n=3080). KORA is a population-based research platform in the region of Augsburg, a city in the south-west of Germany, and two surrounding districts running population-based epidemiological studies. A more detailed description of KORA can be found in Holle et al. (2005). Participants were interviewed at the study center regarding demographic and disease-related parameters, health care utilization and medications. Weight and height measurements were per-

formed by trained staff. Education was classified in 3 groups: *basic education* (≤ 9 years of schooling), *medium education*, and *higher education* (≥ 12 years of schooling, required to enter university). Alcohol consumption was measured as average daily dose (in grams), physical activity was categorized in *high* ($>2\text{h/week}$), *moderate* (up to 2 hours/week), and *no activity*. The groups of working status were: *working*, *unemployed*, *retired*, and *other*.

Resource utilization regarding physician visits, in- and out-patient hospital treatments, rehabilitations and medication were assessed by extrapolating the utilization of the asked time horizon (7 days – 12 months) to one full year. Health care costs were estimated by multiplying utilization of physician visits, in- and out-patient hospital treatments, and rehabilitations with the unit costs provided by [Bock et al. \(2015\)](#). The costs of prescription medication was estimated using pharmacy retail prices, based on patient information on medication name and national drug codes. All costs were adjusted to 2011 prices in €. Table 1 presents an overview of the data set. Underweight individuals ($\text{BMI} < 18.5$, $n = 10$) were removed because this group is especially vulnerable to bias due to other pre-existing disease ([Kent et al., 2017](#)). Observations with missing or implausible information were also removed, resulting in a sample size of $n = 2779$.

Single nucleotide polymorphisms (SNPs) were available from Affymetrix Axiom genotyping for all participants. These genotypes were entered into imputation via the Michigan imputation server with HRC reference panel version r1.1 ([McCarthy et al., 2016](#)). We selected the 97 genetic loci that are highly associated ($p < 5 \times 10^{-8}$) with BMI variation, based on a recent consortium-based genome-wide association study (GWAS) by [Locke et al. \(2015\)](#). For WHR, we used the 48 SNPs found by the GWAS study by [Shungin et al. \(2015\)](#). Some SNPs are relevant for both BMI and WHR. Gene selection based on large GWAS make ideal candidates for MR studies ([Taylor et al., 2014](#)). All analyses were conducted in R ([R Core Team, 2019](#); [Kleiber and Zeileis, 2008](#)).

2.2 (One sample) Mendelian randomization and two-stage least-squares

Mendelian randomization is a type of IV analysis that uses genetic variants as instruments. The random allocation of genetic variants at conception mean that these variants are less likely to violate some of the assumptions of IV analysis than non-genetic instruments, even describing it as nature’s randomized controlled trial ([Smith, 2006](#)).

To estimate the impact of BMI and WHR on medical spending, we used a two-stage least-squares (2SLS) model. In the first step of the 2SLS approach, the endogenous regressor of interest (BMI or WHR) is regressed on all valid instruments. Since the instruments are exogenous, this approximation of the endogenous variables will not correlate with the error term.

In the second step, the outcome (total annual health care costs) is regressed on the fitted values of obesity from the first stage.

Stage 1: Regress independent variable X (BMI or WHR) on instruments Z (genetic loci):

$$X = Z\delta + \varepsilon$$

Stage 2: Regress Y (medical expenditures) on the predictions $\hat{X} = Z\hat{\delta} = Z(Z^T Z)^{-1} Z^T X$ from the first stage:

$$Y = \hat{X}\beta + \eta.$$

The 2SLS estimates are then

$$\beta_{2SLS} = (X^T Z(Z^T Z)^{-1} Z^T X)^{-1} X^T Z(Z^T Z)^{-1} Z^T Y.$$

	Overall
Annual total health care costs (mean (sd))	1,559 (3,645)
BMI (mean (sd))	27.7 (4.8)
WHR (mean (sd))	0.88 (0.9)
Age (mean (sd))	56.2 (13.2)
Sex	
Female (n (%))	1,410 (50.7)
Male (n (%))	1,369 (49.3)
Alcohol consumption g/day (mean (sd))	14.42 (19.5)
Education	
basic (n (%))	1,447 (52)
medium (n (%))	675 (24)
higher (n (%))	657 (24)
Working status	
working (n (%))	1,829 (66)
unemployed (n (%))	55 (2)
retired (n (%))	574 (21)
other (n (%))	321 (12)
Physical activity	
not active (n (%))	1,829 (66)
moderate (n (%))	55 (2)
active (n (%))	574 (21)
Observations n	2,779

Table 1: Summary statistics for the KORA F4 data set. Health care costs are expressed in 2011-€.

All 2SLS models control for age, sex, alcohol intake, education, working status, and physical activity.

Our approach, using individual level data, is called *one-sample* MR. Another popular approach is two-sample MR, using summarized data from GWAS consortia in a meta analysis. The main advantage of two-sample MR is increased statistical power, but the quality of the pooled results is dependent on that of the individual studies. That means, two-sample MR has to use the summary results presented, even when they have been adjusted for covariates that you would rather they had not been adjusted for. It is important to not view genetic susceptibility for complex conditions in isolation but along with lifestyle and environmental factors (Sugrue and Desikan, 2019), a major advantage of one-sample MR. See Lawlor (2016) for a detailed comparison of one-sample and two-sample MR.

For comparison, we compute linear OLS non-IV models that regress health care costs on BMI and WHR and again adjust for the same covariates as before. We use the Durbin-Wu-Hausman endogeneity test (Hausman, 1978) to evaluate whether there is any evidence that the instrumental variable estimate differs from the OLS estimate. In this test, a significant results indicates disagreement between OLS and IV estimates.

2.3 Pleiotropy

A common problem in MR analyses is the presence of pleiotropic effects, i.e. a genetic variant has associations with more than one risk factor on different causal pathways (Lawlor et al., 2008). This can be especially problematic for polygenic risk factors like BMI, where the influence of genetic variants is less specific (VanderWeele et al., 2014). To test violations of the IV assumptions, we used three approaches. First, we checked, in the Phenoscanner database (Staley et al., 2016), if genes related to BMI and WHR are also significantly associated with other determinants of health care costs. We found that some SNPs are related to obesity-related illnesses like type 2 diabetes and high blood pressure. As in Böckerman et al. (2019) we assume that the associations with obesity-related conditions occur because of the SNPs' association with high BMI and WHR, but we cannot definitely rule out other pathways. In a second check, we used the MR-Egger method (Bowden et al., 2015) to test for pleiotropy. The MR-Egger method performs a regression of the SNP-outcome associations on the SNP-exposure associations with the intercept left unconstrained to test for evidence of bias-generating pleiotropy. As a third test for pleiotropy we conducted Sargan's test for overidentification (Sargan, 1958). The null hypothesis in this test is that all exogenous instruments are in fact exogenous and uncorrelated with the model residuals, and all over-identifying restrictions are therefore valid.

2.4 Invalid instruments

As the number of biomarker-associated variants is constantly increasing through GWAS, selection of the most appropriate instruments is an important issue (Swerdlow et al., 2016). Using too many genetic variants as instruments can lead to spurious estimates and increased type 1 error rates. This implies that, even if a set of multiple instruments is valid, (i.e. they are not associated with confounding factors, have no direct effect on the outcome, and are at least weakly associated with the exposure) the 2SLS estimator can still be biased towards the conventional regression estimate (Davies et al., 2015; Bound et al., 1995). A weak instrument only explains a small proportion of the variation in the risk factor. Using many weak instruments can still result in weak instrument bias (Bound et al., 1995; Staiger and Stock, 1997).

For this, [Belloni et al. \(2012\)](#) suggest selecting optimal instruments in the first stage regression by least absolute shrinkage and selection operator (LASSO) variable selection. A similar approach with explicit application to MR was proposed by [Kang et al. \(2016\)](#). Additional work by [Windmeijer et al. \(2018\)](#) recommends to use the adaptive LASSO to retain the oracle properties ([Zou, 2006](#)). In the adaptive LASSO for the first stage, the goal is to minimize

$$\min_{\delta \in \mathbb{R}^p} \left\{ \frac{1}{n} \|X - Z\delta\|_2^2 + \lambda \|w\delta\|_1 \right\},$$

where λ is a penalization parameter obtained by cross-validation, and $w = 1/\hat{\delta}$ is an initial estimate of δ (e.g., obtained by least squares). The LASSO procedure shrinks many δ parameters to effectively zero, pruning them out of the regression. According to [Keele and Morgan \(2016\)](#) we additionally perform an F-test on the first stage regression to test for weak instruments.

2.5 Non-linearity

Previous research found evidence of a non-linear relationship between BMI and health care costs ([Laxy et al., 2017](#); [Cawley and Meyerhoefer, 2012](#)), and MR is often unable to detect such non-linearities because genetic variants usually only explain a small percentage of the variance in the risk factor ([Staley and Burgess, 2017](#)). To test for non-linear effects of BMI and WHR on health care costs in our sample, we fit a generalized additive model (GAM) ([Hastie and Tibshirani, 1986](#)) where we regress the linear predictor (health care costs) linearly on unknown smooth functions of some predictor variable (BMI or WHR), adjusted for the same covariates as in the MR model.

3 Results

LASSO selects 64 valid genetic instruments for BMI and 15 for WHR. The relevant SNPs and their first stage regression estimates are shown in [Figure 1](#). In both IV analyses, we find little evidence of weak instruments, because the weak instruments test p -value is < 0.0001 in both cases and the F-statistics are above the traditional threshold of 10 ([Stock and Yogo, 2005](#)).

The Sargan test does not provide evidence of pleiotropic effects in both BMI and WHR IV models (see p -values in [Table 2](#) and [Table 3](#)). This is supported by the MR Egger intercept test (see [Figure 2](#)). The Durbin-Wu-Hausman test, however, suggests no strong evidence of differences between the OLS and IV estimates.

[Table 2](#) presents the results of the association between BMI and health care costs and [Table 3](#) the results of WHR and health care costs. In the case of BMI, the OLS model finds an effect of 75.5€ for a one unit increase of BMI. In contrast, the IV model estimates an effect of 188.5€ for a one unit increase of BMI.

For WHR, a 0.1 increase of WHR is associated with increased costs of 487€ for the OLS model and 1165€ for the IV model. In both cases, the IV model using genetic instruments estimates the effect on costs almost 2.5 times higher than the OLS model. Other significant cost drivers or savers are age and higher education for the BMI model, and, in addition, sex in the WHR model. Especially in the WHR case (see [Table 3](#)), many associations lose its significance at the 0.05 level when moving from OLS to IV analysis.

We also detect no proof of strong non-linear effects of BMI or WHR on health care costs in our sample, according to the GAM models, see [Figure 3](#).

4 Discussion

This paper is one of the first that uses a one-sample MR approach to estimate the marginal health care costs of a prevalent clinical condition. MR offers new opportunities for reliable causal inference in health economic research within the framework of observational research designs. Using MR, our findings indicate that a one unit increase of BMI increases total medical spending by 188.5€ and a 0.1 unit increase of WHR increased total medical spending by 1165€. The OLS model, that doesn't use instrumental variables, only finds additional spending of 75.5€ for BMI and 487€ for WHR. This demonstrates that a MR study with genetic instruments may detect more hidden bias than a non-IV analysis, leading to higher estimated effects. It is important to note that OLS estimates average treatment effects, while IV estimates local average treatment effects. This may partially account for the differences in the effect estimates (Imbens and Angrist, 1994). Our study confirms the results in Cawley and Meyerhoefer (2012), Black et al. (2018), and Kinge and Morris (2018) that imply much higher costs of obesity in the IV analysis compared with the non-IV analysis. These studies use the BMI of the respondent's oldest child as instrument and report that the IV-estimate for the association between BMI and health care costs is between two to four times higher than the non-IV approach. One possible explanation for the even greater difference between the IV and non-IV estimates in the US by the Cawley and Meyerhoefer (2012) study might be that subgroups with higher obesity prevalence have reduced access to health care in the US, while Germany has a social insurance system covering almost all citizens (Davis and Ballreich, 2014). However, as it is likely that, due to the shared environmental exposures, a child's BMI is correlated with relevant covariates (Dubois et al., 2012), the use of such non or quasi genetic IVs is discussed controversial and might lead to biased estimates as well.

Another advantage of our data is that height, weight, waist and hip measurements of individuals in the sample were performed by trained staff. This is more accurate than self-reported measurements where people tend to under-estimate their weight and over-estimate their height, resulting in an under-estimation in BMI (Gorber et al., 2007).

Because participants in our study were asked to self-report their health service usage over the past 3–12 months, recall bias cannot be excluded. Health care utilization was priced with average reference value, therefore actual costs might deviate. However, this should not influence the validity of the study results (Evans and Crawford, 1999). In addition, we were unable to consider cost categories such as presenteeism, premature death or out-of-pocket payments for medication. Although the problem of recall bias and average unit costs are likely to lead to an underestimation of absolute costs, this should not strongly influence the cost increase between BMI and WHR units that were the focus of this study.

Other studies found similar results. Within the BMI range of 25 to 45 kg/m², a study in the United States by Wang et al. (2006) described an increase in medical and pharmaceutical costs of \$202.3 per BMI unit. A review of 75 international studies by Kent et al. (2017) reported a median increase in mean total annual health care costs of 36% for obese individuals compared with individuals of healthy weight. All of these studies rely on traditional non-IV estimation.

Having weak or invalid instruments can be a problem in MR analysis. Davies et al. (2015) show that 2SLS is especially vulnerable to weak instrument bias in MR and propose the limited information maximum likelihood and the continuously updating estimator as unbiased alternatives to 2SLS. However, these methods are difficult to implement and interpret. For this reason, we used the adaptive LASSO variable selection in the first state of 2SLS to rule out invalid and weak instruments. The SNPs we used have been discovered in GWAS and should be valid.

The Durbin-Wu-Hausman test p -value suggests no significant difference between OLS and IV

estimates. However, the effects being estimated by the two methods may not be the same—the MR estimate reflects the effects of lifelong perturbations in the risk factor, whereas OLS regression results may reflect more acute effects (Davies et al., 2018). According to Burgess and Thompson (2015), it would be fallacious to assume that a non-significant result means the OLS estimate is unconfounded. MR estimates are almost always less precise and have wider confidence intervals than OLS regression, so tests for difference often have low statistical power.

It must also be noted that the KORA F4 study was part of the GWAS consortium that discovered the relevant SNPs for BMI in the study by Locke et al. (2015). Because the same data was used at the GWAS discovery stage and in our analysis, it is possible that chance correlation between SNPs and confounders can lead to overestimation of the SNP-trait effect (Haycock et al., 2016). This is the so-called winner’s curse or Beavis effect (Göring et al., 2001). However, the KORA F4 study was only a small fraction of the more than 300,000 individuals in this GWAS, so it should not strongly affect the power of our analysis.

There are alternative ways to include genetic information into the IV model. For example, an allele score can be generated from the genetic variants and included as a single instrument in the association model. This allele score is calculated per individual as the weighted or unweighted sum of the number of risk or trait increasing alleles (Teumer, 2018). However, Palmer et al. (2012) showed that an unweighted score has lower power than adding multiple IVs into the 2SLS, and using an appropriately weighted allele score performs similarly to adding each valid SNP as an instrument. The effect estimate is a little less biased when using a weighted allele score but has less precision (and power) compared to the multiple IV 2SLS estimator that we used.

Burgess et al. (2014) note that instrumental variable estimates using a linear model may not reflect causal effects for large changes in the exposure. In our case we find the relationship between BMI and costs and WHR and costs approximately linear and therefore the linear model is justified.

In the 2SLS estimator, we assume normal distribution of cost data which is not present because of skewness, positivity, and heavy tails. This can also affect the validity of the standard errors and confidence intervals (Guo et al., 2018; Kang et al., 2018). Still, the use of a linear model is justified because the skewness and tail distribution are not extreme and the large sample size guarantees near-normality of sample means because of the Central Limit Theorem (Mihaylova et al., 2011). However, more sophisticated generalized linear models using different error distributions could be examined (Duan et al., 1983; Kurz, 2017).

In conclusion, we showed that MR can be a viable tool in health economic analyses. We found more than two times higher costs for BMI and WHR one unit increases in our IV model than in the OLS model. Because the association of genetic variants with BMI and WRH is still weak, results have to interpreted carefully.

Acknowledgements

We thank Konstantin Strauch, Thomas Meitinger, Harald Grallert, Christine Meisinger, and Annette Peters for providing the data.

Competing interests

All authors declare no competing interests.

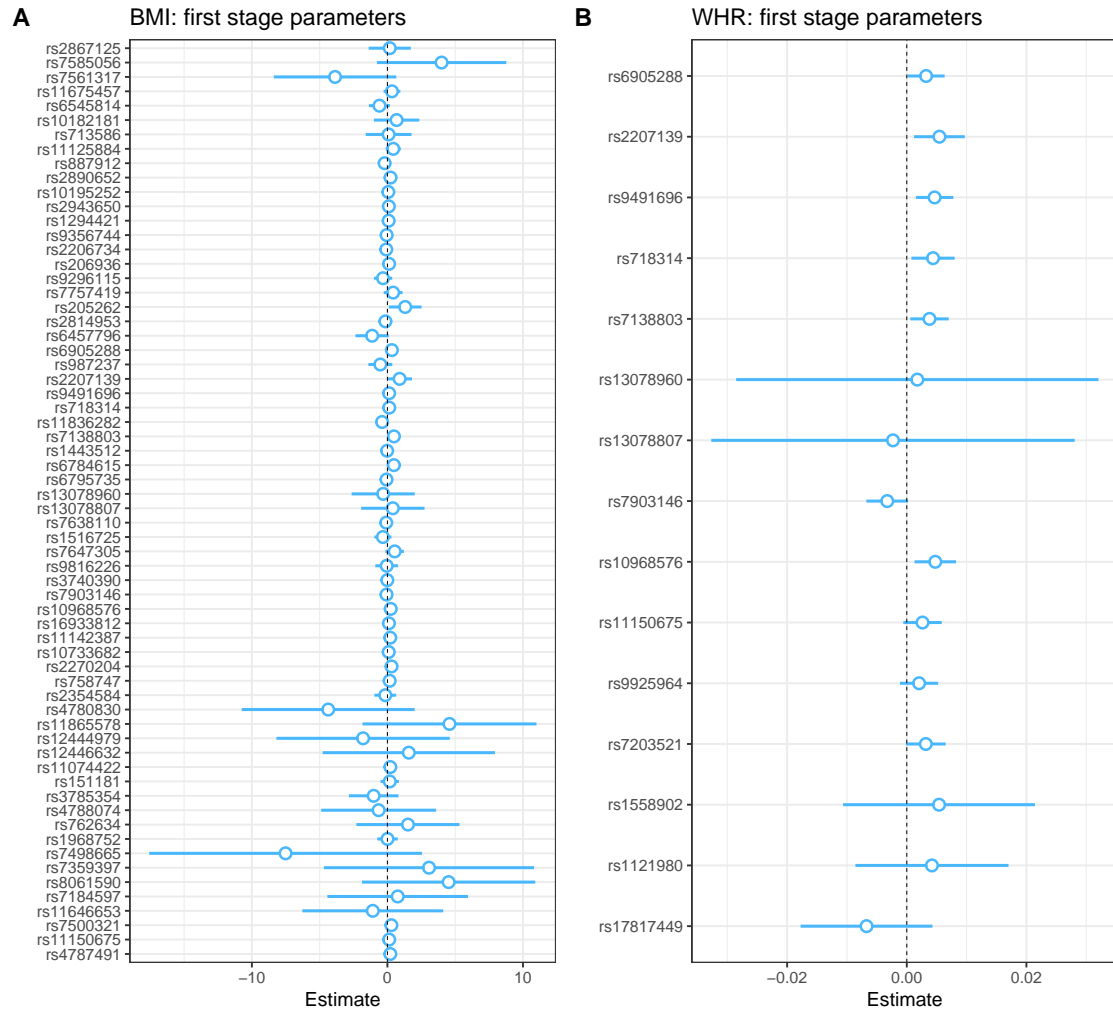


Figure 1: Forestplot for the parameter estimates and 95% confidence intervals of the first stage regression for (A) BMI and (B) WHR. Only genetic variables (SNPs) are shown.

Dependent variable:	<i>OLS</i>	<i>IV</i>
	Annual total health care costs	Annual total health care costs
Age	39.2*** (5.8)	32.6*** (7.6)
Sex = Female	188.5 (149.9)	255.5 (159.2)
BMI	75.5*** (14.9)	188.5** (84.3)
Education		
basic	Ref.	Ref.
medium	-19.2 (171.9)	108.9 (197.5)
higher	-400.7** (178.9)	-205.2 (230.8)
Alcohol consumption	-1.2 (3.7)	-1.3 (3.7)
Working status		
working	Ref.	Ref.
unemployed	368.6 (488.5)	303.2 (495.9)
retired	187.5 (182.8)	130.5 (189.4)
other	47.2 (225.4)	-4.2 (230.8)
Physical activity		
not active	Ref.	Ref.
moderate	-59.4 (160.4)	29.7 (174.8)
active	-116.6 (185.6)	61.8 (228.7)
Constant	-2,895.6*** (586.6)	-5,894.1*** (2,279.4)
Observations	2,779	2,779
Weak instruments test <i>p</i> -value		<0.0001
First stage F-statistic		12.9
Durbin-Wu-Hausman test <i>p</i> -value		0.17
Sargan test <i>p</i> -value		0.67

Note: **p*<0.1; ***p*<0.05; ****p*<0.01

Table 2: The effect of BMI on annual total health care costs using an OLS model (left column) and an IV model with genetic variants (right column). The IV results are from the second stage of the 2SLS approach. Standard errors in parenthesis.

Dependent variable:	<i>OLS</i>	<i>IV</i>
	Annual total health care costs	Annual total health care costs
Age	34.1*** (6.1)	20.9 (16.7)
Sex = Female	684.3*** (195.6)	1,436.3 (904.9)
WHR	4,872.1*** (1,133.2)	11,649.9 (8,042.3)
Education		
basic	Ref.	Ref.
medium	-48.5 (171.8)	29.8 (195.8)
higher	-431.2** (178.8)	-292.2 (243.0)
Alcohol consumption	-2.2 (3.7)	-3.6 (4.1)
Working status		
working	Ref.	Ref.
unemployed	313.5 (489.6)	176.0 (518.5)
retired	176.9 (183.2)	109.2 (200.7)
other	13.5 (226.1)	-80.9 (253.2)
Physical activity		
not active	Ref.	Ref.
moderate	-74.2 (160.5)	-12.1 (177.3)
active	-134.0 (185.8)	7.4 (250.1)
Constant	-5,517.9*** (586.6)	-11,949.9** (2,279.4)
Observations	2,779	2,779
Weak instruments test <i>p</i> -value		<0.0001
First stage F-statistic		12.0
Durbin-Wu-Hausman test <i>p</i> -value		0.39
Sargan test <i>p</i> -value		0.97

Note: **p*<0.1; ***p*<0.05; ****p*<0.01

Table 3: The effect of WHR on annual total health care costs using an OLS model (left column) and an IV model with genetic variants (right column). The IV results are from the second stage of the 2SLS approach. Standard errors in parenthesis.

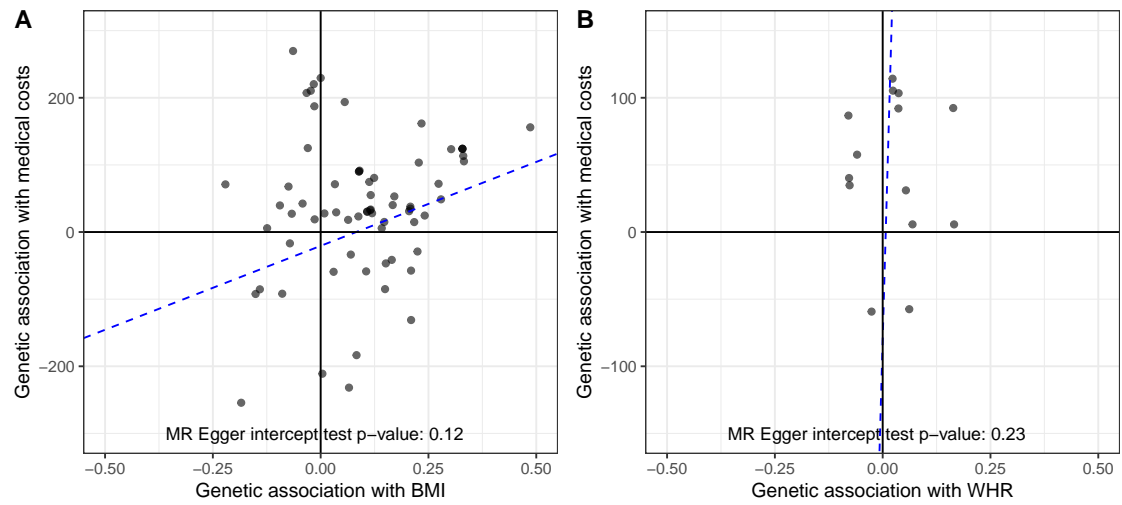


Figure 2: Plot of the gene-outcome vs gene-exposure regression coefficients for (A) BMI and (B) WHR. The dashed line represents the MR Egger regression estimate. Some outliers are not shown.

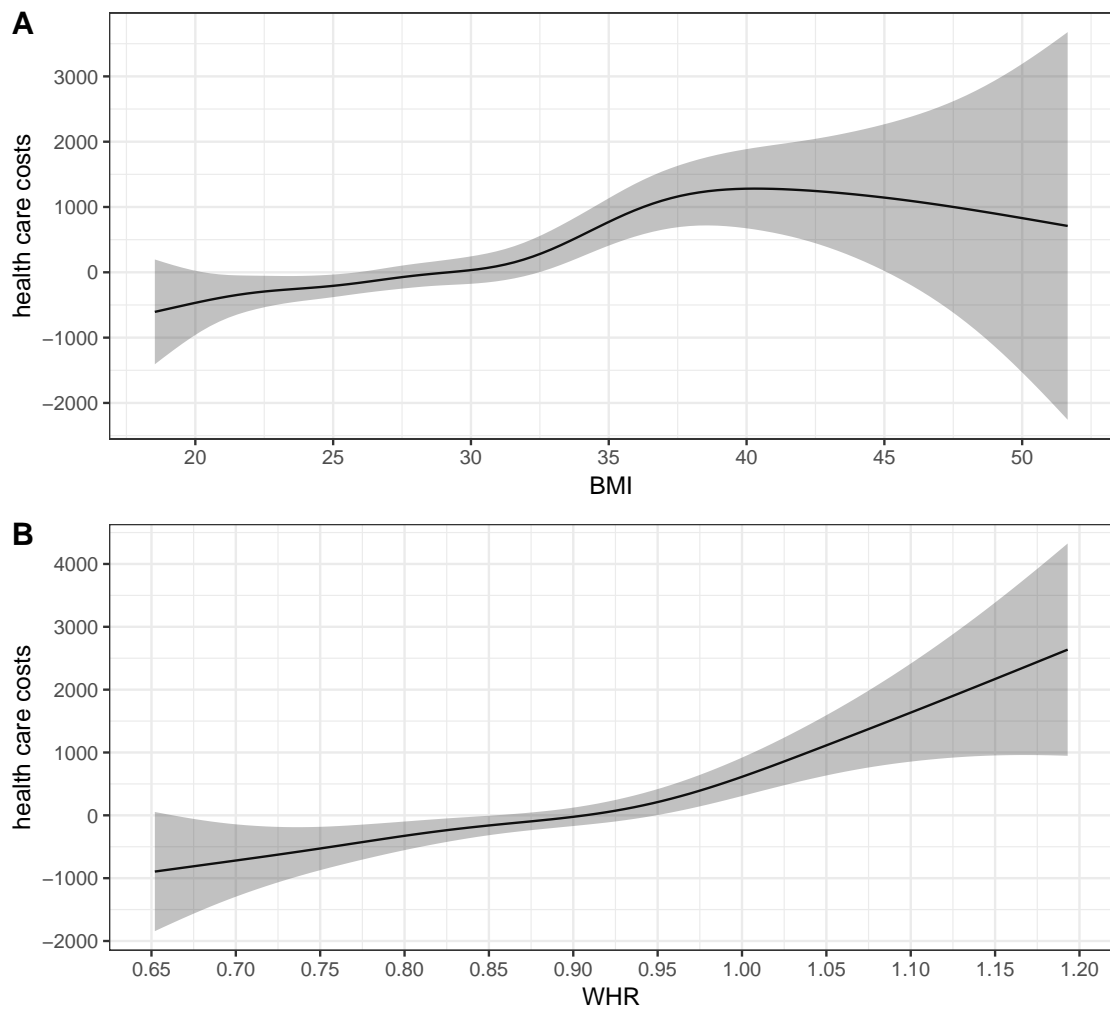


Figure 3: Plots of the smooth function from the GAM for (A) BMI and (B) WHR. The shaded area represents the 95% confidence band.

References

- Belloni, A., Chen, D., Chernozhukov, V., and Hansen, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, 80(6):2369–2429.
- Black, N., Hughes, R., and Jones, A. M. (2018). The health care costs of childhood obesity in australia: An instrumental variables approach. *Economics & Human Biology*, 31:1–13.
- Bock, J., Brettschneider, C., Seidl, H., Bowles, D., Holle, R., Greiner, W., and König, H. (2015). Calculation of standardised unit costs from a societal perspective for health economic evaluation. *Das Gesundheitswesen*, 77(1):53–61.
- Böckerman, P., Cawley, J., Viinikainen, J., Lehtimäki, T., Rovio, S., Seppälä, I., Pehkonen, J., and Raitakari, O. (2019). The effect of weight on labor market outcomes: An application of genetic instrumental variables. *Health Economics*, 28(1):65–77.
- Bound, J., Jaeger, D. A., and Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association*, 90(430):443–450.
- Bowden, J., Davey Smith, G., and Burgess, S. (2015). Mendelian randomization with invalid instruments: effect estimation and bias detection through egger regression. *International Journal of Epidemiology*, 44(2):512–525.
- Burgess, S., Davies, N. M., and Thompson, S. G. (2014). Instrumental variable analysis with a nonlinear exposure–outcome relationship. *Epidemiology*, 25(6):877.
- Burgess, S. and Thompson, S. G. (2015). *Mendelian randomization: Methods for using genetic variants in causal estimation*. Chapman and Hall/CRC.
- Carreras-Torres, R., Johansson, M., Haycock, P. C., Relton, C. L., Smith, G. D., Brennan, P., and Martin, R. M. (2018). Role of obesity in smoking behaviour: Mendelian randomisation study in uk biobank. *British Medical Journal*, 361:k1767.
- Cawley, J., Han, E., and Norton, E. C. (2011). The validity of genes related to neurotransmitters as instrumental variables. *Health Economics*, 20(8):884–888.
- Cawley, J. and Meyerhoefer, C. (2012). The medical care costs of obesity: an instrumental variables approach. *Journal of Health Economics*, 31(1):219–230.
- Dare, S., Mackay, D. F., and Pell, J. P. (2015). Relationship between smoking and obesity: a cross-sectional study of 499,504 middle-aged adults in the uk general population. *PloS One*, 10(4):e0123579.
- Davies, N. M., Hinke Kessler Scholder, S., Farbmacher, H., Burgess, S., Windmeijer, F., and Smith, G. D. (2015). The many weak instruments problem and mendelian randomization. *Statistics in Medicine*, 34(3):454–468.
- Davies, N. M., Holmes, M. V., and Smith, G. D. (2018). Reading mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *British Medical Journal*, 362:k601.
- Davis, K. and Ballreich, J. (2014). Equitable access to care—how the united states ranks internationally. *New England Journal of Medicine*, 371(17):1567–1570.

- Ding, W., Lehrer, S. F., Rosenquist, J. N., and Audrain-McGovern, J. (2009). The impact of poor health on academic performance: New evidence using genetic markers. *Journal of Health Economics*, 28(3):578–597.
- Dixon, P., Smith, G. D., von Hinke, S., Davies, N. M., and Hollingworth, W. (2016). Estimating marginal healthcare costs using genetic variants as instrumental variables: Mendelian randomization in economic evaluation. *Pharmacoeconomics*, 34(11):1075–1086.
- Drummond, M. F., Sculpher, M. J., Claxton, K., Stoddart, G. L., and Torrance, G. W. (2015). *Methods for the economic evaluation of health care programmes*. Oxford university press.
- Duan, N., Manning, W. G., Morris, C. N., and Newhouse, J. P. (1983). A comparison of alternative models for the demand for medical care. *Journal of Business & Economic Statistics*, 1(2):115–126.
- Dubois, L., Kyvik, K. O., Girard, M., Tatone-Tokuda, F., Pérusse, D., Hjelmberg, J., Skytthe, A., Rasmussen, F., Wright, M. J., and Lichtenstein, P. (2012). Genetic and environmental contributions to weight, height, and bmi from birth to 19 years of age: an international study of over 12,000 twin pairs. *PLOS One*, 7(2):e30153.
- Ebrahim, S. and Davey Smith, G. (2008). Mendelian randomization: can genetic epidemiology help redress the failures of observational epidemiology? *Human Genetics*, 123(1):15–33.
- Evans, C. and Crawford, B. (1999). Patient self-reports in pharmaco-economic studies. *Pharmacoeconomics*, 15(3):241–256.
- Finkelstein, E. A., Trogon, J. G., Cohen, J. W., and Dietz, W. (2009). Annual medical spending attributable to obesity: payer-and service-specific estimates. *Health Affairs*, 28(5):w822–w831.
- Fletcher, J. M. and Lehrer, S. F. (2011). Genetic lotteries within families. *Journal of Health Economics*, 30(4):647–659.
- Gorber, S. C., Tremblay, M., Moher, D., and Gorber, B. (2007). A comparison of direct vs. self-report measures for assessing height, weight and body mass index: a systematic review. *Obesity Reviews*, 8(4):307–326.
- Göring, H. H., Terwilliger, J. D., and Blangero, J. (2001). Large upward bias in estimation of locus-specific effects from genomewide scans. *The American Journal of Human Genetics*, 69(6):1357–1369.
- Guo, Z., Kang, H., Tony Cai, T., and Small, D. S. (2018). Confidence intervals for causal effects with invalid instruments by using two-stage hard thresholding with voting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.
- Hastie, T. and Tibshirani, R. (1986). Generalized additive models. *Statistical Science*.
- Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, pages 1251–1271.
- Haycock, P. C., Burgess, S., Wade, K. H., Bowden, J., Relton, C., and 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(4):965–978.
- Holle, R., Happich, M., Löwel, H., and Wichmann, H.-E. f. t. M. S. G. (2005). Kora—a research platform for population based health research. *Das Gesundheitswesen*, 67(S 01):19–25.

- Hu, F. B., Willett, W. C., Li, T., Stampfer, M. J., Colditz, G. A., and Manson, J. E. (2004). Adiposity as compared with physical activity in predicting mortality among women. *New England Journal of Medicine*, 351(26):2694–2703.
- Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475.
- Kang, H., Peck, L., and Keele, L. (2018). Inference for instrumental variables: a randomization inference approach. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*.
- Kang, H., Zhang, A., Cai, T. T., and Small, D. S. (2016). Instrumental variables estimation with some invalid instruments and its application to mendelian randomization. *Journal of the American Statistical Association*, 111(513):132–144.
- Keele, L. and Morgan, J. W. (2016). How strong is strong enough? strengthening instruments through matching and weak instrument tests. *The Annals of Applied Statistics*, 10(2):1086–1106.
- Kent, S., Fusco, F., Gray, A., Jebb, S. A., Cairns, B. J., and Mihaylova, B. (2017). Body mass index and healthcare costs: a systematic literature review of individual participant data studies. *Obesity Reviews*, 18(8):869–879.
- Kinge, J. M. and Morris, S. (2018). The impact of childhood obesity on health and health service use. *Health Services Research*, 53(3):1621–1643.
- Kleibler, C. and Zeileis, A. (2008). *Applied Econometrics with R*. Springer-Verlag, New York. ISBN 978-0-387-77316-2.
- Kurz, C. F. (2017). Tweedie distributions for fitting semicontinuous health care utilization cost data. *BMC Medical Research Methodology*, 17(1):171.
- Lawlor, D. A. (2016). Commentary: Two-sample mendelian randomization: opportunities and challenges. *International Journal of Epidemiology*, 45(3):908.
- Lawlor, D. A., Harbord, R. M., Sterne, J. A., Timpson, N., and Davey Smith, G. (2008). Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Statistics in Medicine*, 27(8):1133–1163.
- Laxy, M., Stark, R., Peters, A., Hauner, H., Holle, R., and Teuner, C. (2017). The non-linear relationship between bmi and health care costs and the resulting cost fraction attributable to obesity. *International Journal of Environmental Research and Public Health*, 14(9):984.
- Locke, A. E., Kahali, B., Berndt, S. I., Justice, A. E., Pers, T. H., Day, F. R., Powell, C., Vedantam, S., Buchkovich, M. L., and Yang, J. (2015). Genetic studies of body mass index yield new insights for obesity biology. *Nature*, 518(7538):197.
- Maes, H. H., Neale, M. C., and Eaves, L. J. (1997). Genetic and environmental factors in relative body weight and human adiposity. *Behavior Genetics*, 27(4):325–351.
- McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A. R., Teumer, A., Kang, H. M., Fuchsberger, C., Danecek, P., and Sharp, K. (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics*, 48(10):1279.
- Mihaylova, B., Briggs, A., O’hagan, A., and Thompson, S. G. (2011). Review of statistical methods for analysing healthcare resources and costs. *Health Economics*, 20(8):897–916.

- NCD Risk Factor Collaboration (2016). Trends in adult body-mass index in 200 countries from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 19·2 million participants. *The Lancet*, 387(10026):1377–1396.
- Norton, E. C. and Han, E. (2008). Genetic information, obesity, and labor market outcomes. *Health Economics*, 17(9):1089–1104.
- Palmer, T. M., Lawlor, D. A., Harbord, R. M., Sheehan, N. A., Tobias, J. H., Timpson, N. J., Smith, G. D., and Sterne, J. A. (2012). Using multiple genetic variants as instrumental variables for modifiable risk factors. *Statistical Methods in Medical Research*, 21(3):223–242.
- R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica*, pages 393–415.
- Shungin, D., Winkler, T. W., Croteau-Chonka, D. C., Ferreira, T., Locke, A. E., Mägi, R., Strawbridge, R. J., Pers, T. H., Fischer, K., Justice, A. E., et al. (2015). New genetic loci link adipose and insulin biology to body fat distribution. *Nature*, 518(7538):187.
- Smith, G. D. (2006). Randomised by (your) god: robust inference from an observational study design. *Journal of Epidemiology & Community Health*, 60(5):382–388.
- Staiger, D. and Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, 65(3):557–586.
- Staley, J. R., Blackshaw, J., Kamat, M. A., Ellis, S., Surendran, P., Sun, B. B., Paul, D. S., Freitag, D., Burgess, S., Danesh, J., et al. (2016). Phenoscanner: a database of human genotype–phenotype associations. *Bioinformatics*, 32(20):3207–3209.
- Staley, J. R. and Burgess, S. (2017). Semiparametric methods for estimation of a nonlinear exposure-outcome relationship using instrumental variables with application to mendelian randomization. *Genetic Epidemiology*, 41(4):341–352.
- Stock, J. H. and Yogo, M. (2005). Testing for weak instruments in linear iv regression. *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, page 80.
- Sugrue, L. P. and Desikan, R. S. (2019). What Are Polygenic Scores and Why Are They Important? *Journal of the American Medical Association*.
- Swerdlow, D. I., Kuchenbaecker, K. B., Shah, S., Sofat, R., Holmes, M. V., White, J., Mindell, J. S., Kivimäki, M., Brunner, E. J., and Whittaker, J. C. (2016). Selecting instruments for mendelian randomization in the wake of genome-wide association studies. *International Journal of Epidemiology*, 45(5):1600–1616.
- Taylor, A. E., Davies, N. M., Ware, J. J., VanderWeele, T., Smith, G. D., and Munafò, M. R. (2014). Mendelian randomization in health research: using appropriate genetic variants and avoiding biased estimates. *Economics & Human Biology*, 13:99–106.
- Teumer, A. (2018). Common methods for performing mendelian randomization. *Frontiers in Cardiovascular Medicine*, 5.
- Thomas, D. C. and Conti, D. V. (2004). Commentary: the concept of ‘mendelian randomization’. *International Journal of Epidemiology*, 33(1):21–25.

- Trasande, L., Liu, Y., Fryer, G., and Weitzman, M. (2009). Effects of childhood obesity on hospital care and costs, 1999–2005. *Health Affairs*, 28(4):w751–w760.
- VanderWeele, T. J., Tchetgen, E. J. T., Cornelis, M., and Kraft, P. (2014). Methodological challenges in mendelian randomization. *Epidemiology*, 25(3):427.
- von Hinke, S., Smith, G. D., Lawlor, D. A., Propper, C., and Windmeijer, F. (2016). Genetic markers as instrumental variables. *Journal of Health Economics*, 45:131–148.
- Wang, F., McDonald, T., Bender, J., Reffitt, B., Miller, A., and Edington, D. W. (2006). Association of healthcare costs with per unit body mass index increase. *Journal of Occupational and Environmental Medicine*, 48(7):668–674.
- Wang, Y. C., McPherson, K., Marsh, T., Gortmaker, S. L., and Brown, M. (2011). Health and economic burden of the projected obesity trends in the usa and the uk. *The Lancet*, 378(9793):815–825.
- Willage, B. (2017). The effect of weight on mental health: New evidence using genetic ivs. *Journal of Health Economics*.
- Windmeijer, F., Farbmacher, H., Davies, N., and Davey Smith, G. (2018). On the use of the lasso for instrumental variables estimation with some invalid instruments. *Journal of the American Statistical Association*, pages 1–12.
- World Health Organization (2000). *Obesity: preventing and managing the global epidemic*. Number 894. World Health Organization.
- Yang, J., Manolio, T. A., Pasquale, L. R., Boerwinkle, E., Caporaso, N., Cunningham, J. M., De Andrade, M., Feenstra, B., Feingold, E., Hayes, M. G., et al. (2011). Genome partitioning of genetic variation for complex traits using common snps. *Nature Genetics*, 43(6):519.
- Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476):1418–1429.