Rich and fat? Isolating the causal effect of obesity on income among rural Chinese residents by Mendelian randomization

Wenyan Xu, College of Economics and Management, China Agricultural University, 100083, China, xuwenyan@cau.edu.cn

Qiran Zhao, College of Economics and Management, Academy of Global Food Economics and Policy (AGFEP), China Agricultural University, 100083, China, zhaoqiran@cau.edu.cn

Wei Si, College of Economics and Management, Academy of Global Food Economics and Policy (AGFEP), China Agricultural University, 100083, China, siweicau@126.com

Chen Zhu, College of Economics and Management, Academy of Global Food Economics and Policy (AGFEP), China Agricultural University, 100083, China, zhuchen@cau.edu.cn

Selected Paper prepared for presentation at the 2024 Agricultural & Applied Economics Association Annual Meeting, New Orleans, LA; July 28-30, 2024

Rich and fat? Isolating the causal effect of obesity on income among rural Chinese residents by Mendelian randomization

Wenyan Xu[1], Qiran Zhao[1,2], Wei Si[1,2], Chen Zhu[1,2,3*]

[1] College of Economics and Management, China Agricultural University, 100083, China

[2] Academy of Global Food Economics and Policy (AGFEP), China Agricultural University, 100083, China

[3] Beijing Food Safety Policy and Strategy Research Base, China Agricultural University, 100083, China

Abstract: While the adverse health consequences of obesity are well-documented, the causal effect between obesity and economic outcomes, particularly individual income in the labor market, has yielded inconsistent findings. To shed light on this complex effect, we employed the genetic variants within the Mendelian randomization framework to investigate the causal relationship between obesity and income. Our one-sample Mendelian randomization analysis revealed that obesity reduced residents' annual income by 609.42 CNY based on a sample of 441 people collected from rural China in 2019. Results from covering over 400,000 European people using one-sample and two-sample Mendelian randomization methods further support our findings. We additionally find that childhood obesity is causally linked with reduced income and educational attainment in adulthood. These results contribute to a more nuanced understanding of the economic costs of obesity and take action to alleviate health inequities among individuals of diverse socioeconomic backgrounds.

Keywords Obesity; Income; Mendelian randomization; Genetic instruments; China

Paper type Research paper

1 Introduction

Over the past several decades, obesity has been rising rapidly in China. Specifically,

* Corresponding author: Chen Zhu, E-mail: zhuchen@cau.edu.cn

the prevalence of obesity in children and adolescents aged 6-17 years has ascended to approximately 7.9%, while among adults, the obesity rate has attained a notable 16.4% (Zeng *et al*., 2021). At the same time, the gap between urban and rural obesity rates is also narrowing, and the proportion of obese children in rural areas is significantly higher than that in urban areas. Nowadays, obesity has emerged as one of China's most pressing health challenges, posing a potentially worrying burden on the Chinese healthcare system. It is responsible for over 76400 premature deaths annually (GBD, 2019). Obesity also brings huge external economic costs to society. Without effective containment or prevention measures by governments, the Chinese financial cost of overweight or obesity is projected to reach 4180 hundred million yuan by 2030, about 22% of public medical expenditures (Wang *et al*., 2021).

In fact, obesity not only significantly impacts people's health and national economy but also influences their labor market outcomes in daily life. Yet the correlation between obesity and income is complex and contradictory primarily due to the reverse causality. Besides, the relation seems to change with economic development and income thresholds. For example, certain studies have found that the prevalence of obesity tends to increase when low-income countries become richer, suggesting a positive correlation between income and obesity (Malik *et al*., 2013). Conversely, others have documented that obesity rates are highest among the least affluent groups in high-income countries, suggesting that this relationship is negative (Giskes *et al*., 2008). As a consequence, economists have tried many methods to understand how obesity affects individuals' income. Three strategies are usually used to solve the endogeneity problem caused by reverse causality and unobservable factors affecting income and obesity simultaneously, including employing the fixed effect model, replacing peoples' present body mass index (BMI) with a lagged value of body mass index in the regression and leveraging instrument variables of obesity. However, it is noteworthy that these methodologies may not entirely resolve the endogeneity problem, and individual factors impacting obesity and labor market outcomes may still be encompassed within the error term.

Hence, in this paper, we use the one-sample and two-sample Mendelian randomization (MR) methods to explore the causal effect between obesity and income among Chinese rural residents based on a unique dataset collected between 2019 and 2021 from two hundred thirty-five villages of eight provinces. MR is a method for evaluating causal inference, which employs genetic variants as instrumental variables (IV). The genetic basis of MR relies on the random allocation of genes during meiosis in humans. This process closely mirrors the random assignment into control or treatment groups in randomized controlled trials (RCT). As genes are determined before birth and remain unchanged thereafter, they can effectively help us avoid the endogenous problems caused by reverse causation.

Our investigation demonstrates that obesity adversely affects Chinese rural residents' income using ordinary least squares, instrumental variables, and one-sample Mendelian Randomization methods. We find obesity reduces residents' annual income by 609.42 CNY by utilizing the one-sample MR method based on a sample of 441 people collected from rural China in 2019-2021. Notably, this estimate is deemed more accurate than the one derived from using a relative's obesity status as the instrumental variable. To increase the external validity and ensure the robustness of the results, we use the UK Biobank data covering over 400,000 participants to compare whether this finding is internationally generalizable. The analysis further discloses a negative influence of obesity on income in the United Kingdom, affirming our initial findings. Two-sample Mendelian randomization analysis also further supports the adverse effects of obesity on individual income. Moreover, our study extends focus to childhood obesity, revealing a causal link between childhood obesity and diminished income in adulthood. These findings contribute to understanding the long-term consequences of obesity across various life stages.

Our findings have several contributions to the previous literature. Firstly, we provide novel and robust instrumental variables and utilize the one-sample and two-sample Mendelian randomization methods to study the impact of obesity on income. Secondly, we extend the literature on estimating the causal effect of obesity on income to the world's largest developing country—China. Earlier studies demonstrated that a higher BMI lower the individual's income in some developed countries (Cawley, 2004; Katsaiti and Shamsuddin, 2016; Slade, 2017). In contrast to developed nations, fat bodies are often positively perceived in low-income countries, and obese people are more likely to be rich. Macchi (2023) found obese people earn around 80 dollars monthly more than normal-weight people in Kampala. However, little is known about how obesity affects Chinese residents' income. China has had the highest number of people who are obese or overweight around the world in recent years (NCD Risk Factor Collaboration, 2019). So, it is essential to explore how obesity affects Chinese residents' income. Thirdly, we employ data from the UK Biobank to confirm our findings further. Finally, our study identifies the long-term consequences of obesity.

Understanding that obesity lowers Chinese residents' income is essential to strive for common prosperity and mitigate developmental disparities with developed countries. In China, most rural households' incomes tend to move downwards in the income distribution (Li and Cheong, 2016). Low-income farmers are more likely to consume low-priced and energy-dense food, leading to obesity increased (Fan *et al*., 2017). Therefore, they have higher health vulnerability and are more likely to fall into the health-poverty trap (Li *et al*., 2023). Our findings support actions reducing body weight from obesity to non-obesity, such as healthy living habits, which are also effective ways to increase rural families' income and alleviate income inequities. In addition, solutions to mitigate obesity will reduce the nation's economic burden and foster sustainable economic development.

The remainder of the paper is organized as follows. In the second section, we briefly review the relevant literature. In the third section, we present the theoretical framework. In the fourth section, we describe the sample collection and data used in this work. In the fifth section, we offer the empirical strategy adopted in this study and validate the use of genetic instruments. We report the empirical results in the sixth section, and the final section draws conclusions.

2 Literature review

Several papers have studied the elusive link between obesity and income and attempted to deal with the potential endogeneity of weight when examining the causal impact of obesity on income or wages. Yet the results remain complex and contentious.

Some scholars regressed individual wages on a lagged value of past body mass index to handle the reverse causality, namely, the lower income would induce poorer health. They found that obese women have lower family incomes than those whose BMI is in the recommended range, but the results for men are weaker and mixed (Averett and Korenman, 1993). However, it does not preclude the potentially endogenous factors. Alauddin Majumder (2013) adopted a fixed-effects model and replaced contemporaneous weight variables with one-year lags of weight variables to estimate. He discovered except for white males who had an obesity wage premium, wages of all other ethnic-gender groups remained unaffected by obesity. A noticeable caveat is while this adjustment eliminates the contemporaneous effect of wages on obesity, it does not account for genetic and non-genetic factors of lagged body weight that may be related to the current wage.

To overcome this limitation, others used instrumental variables (IV) from multiple sources to address bias caused by reverse causality or unobserved confounders affecting income and obesity simultaneously. Cawley (2004) and Slade (2017) applied siblings' body mass index (BMI) as an instrumental variable (IV) and demonstrated that BMI adversely affected wages, which is equal to the wage effect of roughly three years of work experience. Katsaiti and Shamsuddin (2016) used the father's BMI as IV and also identified a negative and significant impact of obesity on wage earnings in Germany. However, a study using 808 monozygotic (MZ) twin pairs suggested the significant inverse association between adult BMI and wages is because of unmeasured earnings endowments and BMI, and the association would disappear if they control for endowments common to monozygotic (MZ) twins (Behrman and Rosenzweig, 2002). Considering that workers may self-select jobs based on their body mass, Moro et al. (2019) utilized a job selection model and revealed the effect of BMI on wages is not statistically significant in jobs requiring a lower level of personal interactions. In addition, some studies showed overweight men experience a wage premium, albeit with diminishing returns (Caliendo and Gehrsitz, 2016; Sarrias and Iturra, 2022).

These studies have not reached consistent conclusions, possibly due to data limitations that make the effectiveness of these instrumental variables, such as weak instrumental variables and the inability to satisfy the exclusivity of instrumental variables, not be carefully examined. Individual factors impacting obesity and labor market outcomes may remain in the error term (Biener *et al*., 2018; Norton and Han, 2008). Besides, some studies actually examined the effect of BMI on wages or earnings rather than the effect of obesity on income. Hence, we investigate the causal effect of obesity on earnings by leveraging genetic variants related to obesity as novel IV within the Mendelian Randomization (MR) framework.

3 Theoretical framework

In the early 1960s, many economists had already begun to consider health as a form of human capital. However, it was not until 1962 that Mushkin (1962) formally established the perspective of incorporating health into human capital. Subsequently, building upon Becker's human capital theory, Grossman (1972) regarded individual health as a capital stock that depreciates gradually with advancing age. The initial health stock's quality is influenced partly by innate factors and acquired conditions. He further conceptualized health as a capital stock that enhances both consumption levels and life satisfaction, serving as both an investment and a consumable. When health as human capital is treated as a consumable, illness results in "disutility," necessitating the direct inclusion of health in the utility function. Conversely, when viewed as an investment, the returns from health manifest in prolonged life expectancy or increased healthy working years.

Obesity may significantly influence income as a critical component of health capital (Schultz, 1961). According to the health selection theory, health status serves as one of the screening mechanisms for social mobility. Individuals with better health are more likely to attain higher socioeconomic status, while those with poorer health tend downward (West, 1991). Extensive research has demonstrated obesity may increase an individual's risk of being unhealthy (Cercato and Fonseca, 2019), so it may lead to their socioeconomic status moving downward. Furthermore, the labor market may engage in screening based on individuals' physiques according to market signaling theory (Spence, 1978), potentially resulting in discrimination against and less favorable hiring and promotion prospects for individuals who are obese (Katsaiti and Shamsuddin, 2016). This bias may stem from some employers' prejudice, erroneously viewing obesity as indicative of a lack of discipline and self-control. Such discrimination can impede the career development of individuals struggling with obesity. Besides, an individual's health status generally does not stabilize after a certain age; earlier health often influences health at different points in the subsequent life course. The prevalence of adult obesity may be shaped not solely by factors manifesting during adulthood but may trace its origins back to childhood or even earlier stages of life. Therefore, childhood obesity may impact the formation of health, cognitive, and non-cognitive abilities in adulthood, consequently affecting labor supply and efficiency in adulthood, ultimately determining labor income (Heckman, 2007).

4 Data

4.1 Sample collection

Our study utilizes a unique dataset collected by the College of Economics and Management of China Agricultural University in 2019 and 2021 from two hundred thirty-five villages of eight provinces (Heilongjiang, Henan, Hebei, Zhejiang, Yunnan, Xinjiang, Shandong, and Anhui). Before data collection, all participants were informed of the purpose of the research and voluntarily signed informed consent. This survey collected participants' demographic and socioeconomic information, such as educational attainment, birth year, and annual income. Excluding people who did not finish their questionnaire, we have 3039 samples.

It is worth noting that, given the relatively high cost associated with genetic sequencing, with each individual's sequencing expenses approaching approximately 500 Chinese Yuan, we ultimately obtained 476 observations containing their genetic information. Despite the relatively modest sample size, we employed a random stratified sampling methodology, drawing samples from eight provinces in China to ensure the representativeness of the dataset. After excluding individuals who do not meet quality control criteria, the study ultimately comprises 441 valid observations.

Moreover, in our survey, all respondents who participated in the genetic sequencing needed to provide 1 mL of saliva for genotyping in a professional test tube. Then, we extracted DNA from saliva samples using the Illumina WeGene V2 array. Imputation and quality control were performed using PLINK (1.90 Beta), SHAPEIT (v2.17), and IMPUTE2 (v2.3.1). We acquired over 10 million Single Nucleotide Polymorphisms (SNPs) for each participant, which provides accurate individual genotypes of obesity genes that can be used as reliable genetic instruments.

4.2 Measures of obesity

In the survey, the researcher employed specialized equipment to measure the height and weight of the participants rather than relying on self-reported data to ensure the accurate calculation of their Body Mass Index (BMI). BMI is calculated as a weight (kg) by height squared (m^2). According to the Chinese resident's BMI reference standard, an individual is considered obese when their BMI exceeds 28. As reported in Table 1, the total interviewee's average BMI was 24.65, and 16.0% were obese. The average BMI of the respondents who included genetic information was 24.48, and 16.4% were obese. Besides, the average age of total respondents in our sample was 44.86 years old, completed 7.64 years of schooling, and earned 22700 CNY yearly.

Table 1. Summary statistics of the Chinese sample

Variable	Total sample		MR sample	
	Mean	Std. Dev.	Mean	Std. Dev.
Income (in 10,000 CNY)	2.27	2.50	1.79	2.60
BMI	24.65	3.51	24.48	3.61
Obesity	16.0%	-	16.4%	-
Male	82.9%	-	74.0%	-
Age	44.86	14.75	49.44	11.64
Years of schooling	7.64	3.32	8.25	3.47
Observations	3039		441	

5 Method

Our study seeks to demonstrate the causal effect of obesity on income. At first, we estimate regressions of the following form:

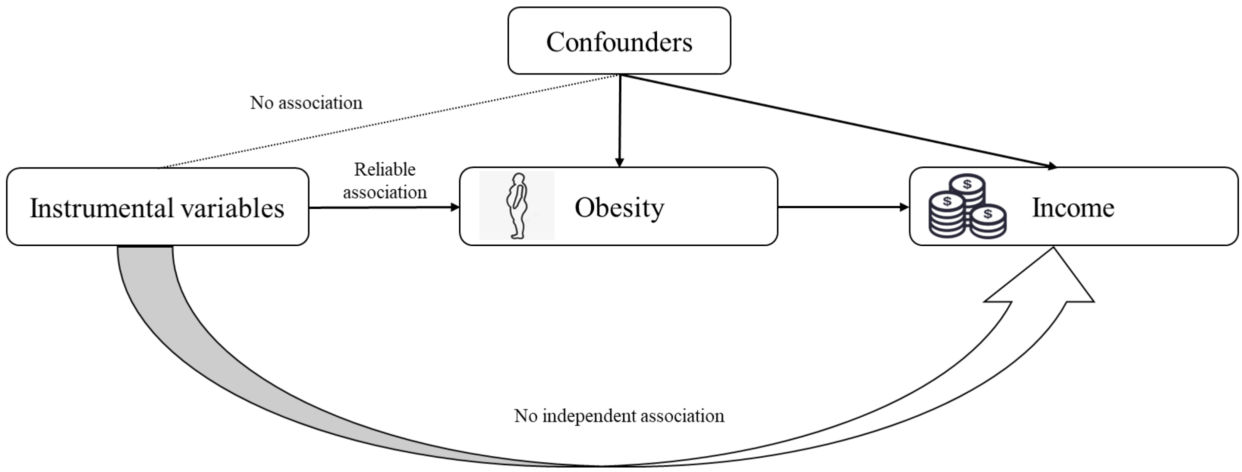
$$Y_i = \gamma Obesity_i + \alpha X_i + \mu_i \qquad (1)$$

Y_i indicates the individual's annual income, X_i is the vector of control variables and contains age, educational attainment, gender and province fixed effect. We first estimate equation (1) using ordinary least squares (OLS) to examine the association of obesity with income. But it may suffer from reverse causality and biases from unobserved confounders. For example, a negative association between obesity and income does not necessarily mean that obesity causally leads to lower income. On the one hand, income can influence obesity by changing individuals' food consumption budget constraints, labor supply, and time allocation (Ameye and Swinnen, 2019; Kwon et al., 2010). On the other hand, an individual's obesity and annual income may be jointly determined by lifestyle attitudes and culture. As a result, we use the obesity status of another individual (parents or siblings) in the family with a highly correlated gene as an instrumental variable for that individual's obesity.

5.1 Mendelian Randomization method

Obesity can be heritable, and research in behavioral genetics suggests that half of the variation in obesity is because of genetic factors (Comuzzie and Allison, 1998). In order to explore the causal effect of obesity on income, we adopt the Mendelian randomization approach by using the genetic variants of obesity as instrument variables. Mendelian randomization method is a robust causal approach that minimizes unobserved confounders' influence and reverses causality using genetic variants as instrumental variables. Since genetic variants are randomly distributed during meiosis and fertilization, similar to the random allocation of treatment groups in randomized controlled trials (RCT), they are relatively independent of self-selection behaviors and determined before birth (Lawlor *et al*., 2008). It provides strong instrumental variables and allows us to control the endogeneity of obesity and obtain consistent estimates. A conceptual diagram of our MR analysis is shown in Figure 1.

Figure 1 Mendelian randomization assumptions and associations between obesity and income



Note: Figure 1 exhibits Mendelian randomization assumptions when determining the causal relationship between a given exposure and outcome: (1) the genetic variants are closely associated with obesity, (2) the genetic variants are not associated with any potential confounders, and (3) the genetic variants do not cause the income independently from the obesity.

We adopt the following two-stage least-squares (2SLS) models within the Mendelian randomization framework to explore the causal effect between obesity and income.

First stage: $Obesity_i = \gamma Genetic\ instruments_i + \theta X_i + \epsilon_i$ (2)

Second stage: $Income_i = \beta_1 Obesity_i + \alpha X_i + \varepsilon_i$ (3)

In the regress model, $Obesity_i$ is a dummy variable. $Income_i$ represents an individual's annual income. X_i include gender, age, ancestral composition, and province fixed effects. In the first stage (Equation. (2)), the endogenous variable of obesity is regressed on the genetic instruments and other control variables. Besides, we use three criteria to select validated genetic instruments in the Chinese sample.

5.2 Instrument validity

5.2.1 Relevance: the genetic IV must correlate with the endogenous variable

In our research design, the genetic instruments, including *ALDH2*, *KCNQ1*, *FGR*, *CCK*, and *MYL2,* are robustly associated with obesity among East Asian-ancestry populations according to the previous study (Wen *et al*., 2014; Akiyama *et al*., 2017). Other genetic variables have also been demonstrated to have robust associations with obesity among European-ancestry populations from a large number of existing literature in genetics. We also use the Equation. (2) to examine whether these genotypes are related to obesity in our sample. And the First-satge Cragg-Donald F-statistic exceeded the conventional cut-off of 10 for weak instruments, indicating these genetic variants are strong IVs in our MR design. So, we chose these SNPs as instrument variables.

Table 2. Information of the instrument variables

SNP	Genes	Mean	Std. Dev.
rs671	*ALDH2*	0.40	0.58
rs2237892	*KCNQ1*	0.69	0.67
rs12229654	*MYL2*	0.34	0.53
rs2076463	*FGR*	0.56	0.67
rs8192473	*CCK*	0.23	0.47
rs1229984	*ADH*	1.30	0.68
rs1051730	*CHRNA3*	0.05	0.24
rs762551	*CYP1A2*	0.40	0.49
rs1726866	*TAS2R38*	0.90	0.31
rs10246939	*TAS2R38*	0.67	0.66
rs307355	*TAS1R3*	0.37	0.54

5.2.2 Independence: the genetic IV must NOT be associated with unmeasured confounders through population stratification

Another methodological threat is population stratification. It refers to systematic differences in allele frequencies or the presence of a specific socioeconomic condition between subpopulations due to different ancestry (Zhu *et al*., 2020). Nevertheless, most existing MR research of obesity does not control population stratification through parental genetic information or individual ancestry, which may result in the genetic IV may be associated with unobserved confounders through population stratification and leading to biased results. Therefore, we control the estimated compositions of individual ancestry based on each respondent's genetic data to overcome the potential population stratification problem.

5.2.3 Exclusion: the genetic IV must have NO direct effect on income through pleiotropy

Pleiotropy is the scenario where a genetic instrumental variable (IV) can influence multiple traits (Willage, 2018). If the genetic instrumental variable of obesity is directly related to an individual's income, it will violate the exclusion condition of the instrumental variable. The pleiotropy threat may not be substantial in our study for three reasons. First, Akiyama *et al*. (2017) and Wen *et al*. (2014) have provided evidence for the credibility of these genetic instruments as IV for obesity. Second, to our knowledge, the extensive findings fail to support that the above genes are directly related to income. Third, we conduct overidentification tests, and the Sargan statistics demonstrate no violation of the exclusion restriction. As a result, these genetic variants of obesity as IV are appropriate.

6 Result

6.1 The causal effect of obesity on Chinese residents' income

Table 3, columns 1 and 2, reports estimates of the effects of obesity on income among the total sample. OLS results show that obesity is negatively associated with income, although it is insignificant. Obesity is associated with an average 214.47 CNY decrease in rural residents' annual income. Our results confirm Huang *et al*. (2016) findings. They use the fixed effect model based on CHNS data between 1991 and 2011 to analyze and discover that obesity adversely affects Chinese populations' wages. Then, we use the obesity status of the individual's relatives (parents or siblings) as IV[2]. The result shows that obesity also has an adverse effect on income. These results are consistent with Slade's (2017) research, which utilizes the sibling's BMI as IV and discovers obesity decreases whites' and Hispanics' wages. However, the result may be biased, as unobservable factors, such as a common family lifestyle, could still influence income and obesity. It may overestimate the impact of obesity on income. Therefore, we use the one-sample MR method to estimate further.

Columns 3 and 4 in Table 3[3] show that one-sample MR estimates. It demonstrates that obesity has a significantly negative impact on rural residents' income. Obesity causally decreases rural residents' annual income by 609.42 CNY. And this result is more accurate than the estimate using the relative's obesity status as IV. These results are well in accordance with previous studies demonstrating that obesity negatively affects people's wages in developed countries(Averett and Korenman, 1993; Pinkston,

2 First-stage F-statistical joint significance of the IVs exceeds the threshold of 10 weak instruments.
3 Both first-stage F-statistical joint significance of the IVs exceeds the threshold of 10 weak instruments. Both the Sargan statistics and the P-values of the overrecognition test indicate that there is no sign of violation of the exogeneity condition.

2017). It indicates obesity may not perceived as a signal of wealth in China, and there may be no weight premium.

Table 3 The causal effect of obesity on Chinese residents' income

	Total sample		MR sample	
	(1)	(2)	(3)	(4)
	OLS	Traditional IV	OLS	MR
Variables	Income	Income	Income	Income
Obesity	-214.47	-10,547.22***	-23.61	-609.42*
	(1,243.40)	(3,958.92)	(25.72)	(326.90)
Male	834.08	-7,064.17*	-31.46	-9.89
	(1,101.12)	(4,031.15)	(21.59)	(33.81)
Age	39.14	432.55***	1.01	-0.29
	(56.55)	(93.21)	(0.81)	(1.39)
Ancestry			Control	Control
Constant	16,918.08***	11,988.62***	-60.10	115.46
	(4,936.47)	(3,772.25)	(61.46)	(132.56)
Observations	3039	958	441	441
R-squared	0.12	0.05	0.06	0.08

Note: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1; in the regression model, the standard error cluster at the village level and control the year fixed effect; the ancestry include Dai, Kinh, Korean, Japanese, Han_Northern, Miao_Yao

6.2 Result generalization

6.2.1 Different sample: UK Biobank

To increase the external validity for specific people and compare whether this result differs from those in developed countries, we also use the UK Biobank data covering over 400,000 participants to analyze the causal effect of obesity on income. We chose the United Kingdom for two primary reasons. Firstly, this country exhibits notably high obesity rates, but studies on the relationship between obesity and labor market outcomes among British remain scarce. Recent evidence shows that the obesity rate in the United Kingdom has increased from 10% in 1975 to 29.5%. Secondly, the United Kingdom embodies disparate economic levels and stages of social development with China, thereby presenting an opportunity to investigate whether the influence of obesity on residents' income across diverse contextual conditions is different.

We use the UK Biobank data to examine the effect of obesity on British residents' income. The UK Biobank is a large-scale prospective cohort study. It is the world's largest established repository of human genetic cohorts and bio-samples. Since its establishment in 2006, it has gathered blood, urine, and saliva samples from 500,000 participants aged between 40 and 69 years old across various regions of the UK, along with comprehensive demographic, socioeconomic, lifestyle, genetic, and health information. After excluding people who do not prefer to answer their income and do not know their genetic information, we have 413,444 samples.

Table 4 reports the summary statistics of the UK Biobank sample, the average age of the sample was 56.16 years old, and they earned 56800 GBP every year. The interviewer measured the interviewee's weight and height to prevent issues arising from self-reporting errors, such as inefficiency and bias. According to the World Health Organization, a BMI that exceeds 30 is considered obese in the UK. In our sample, the interviewee's average BMI was 27.39, and 24.02% were obese. 47.65% were male, and 34.95% of respondents completed a college or university degree. The average value of the standard polygenic risk score of BMI was -0.19. If a personal polygenic score is higher than the average for the population, it means they have an increased genetic risk of obesity compared to most people.

Table 4 Summary statistics of the UK Biobank sample

Variable	Mean	Std. Dev.
Income (in 10,000 GBP)	5.68	4.48
BMI	27.39	4.76
Obesity	24.02%	-
Male	47.65%	-
Age	56.16	8.09
College or university degree	34.95%	-
Standard polygenic risk score of BMI	-0.19	0.98
No. of Observations	413,444	

In the UK Biobank sample, we also use the Equation. (2) and Equation. (3) within the MR framework to analyze. We select the standard polygenic risk score (PRS) for BMI as the instrument variable rather than the SNP used in the Chinese sample. It is generated using a Bayesian approach to meta-analyzed summary statistics GWAS data. A subsequent principal component-based ancestry-centering step is applied to center the score distributions on zero across all ancestries. The Standard PRS for BMI is calculated for all UK Biobank individuals. It is a weighted linear combination of phenotypic-related alleles on the genome (the weight is usually the estimated effect size in the GWAS). Besides, we control individual's genetic principal components to avoid the pleiotropy effect.

Table 5 reports estimates of the effects of obesity on the UK residents' income. The results exhibit that obesity significantly lowers residents' income. Non-obese people earn 4628 GBP every year more than obese people. After we use the standard polygenic risk score for BMI as IV to address the reverse causality[4], we still find obesity has a negative impact on UK residents' income. Obesity causally decreases residents' annual income by 8652 GBP. Our results confirmed weight discrimination exists in developed countries.

Table 5 One-sample MR and OLS estimates of the causal effect of obesity on the UK residents' income

	(1)	(2)
	OLS	MR
Variables	Income	Income
Obesity	-4,627.92***	-8,652.02***
	(148.99)	(704.68)
Age	-1,422.94***	-1,419.60***
	(7.94)	(7.97)
Male	6,102.15***	6,186.18***
	(126.87)	(127.79)
College	27,049.82***	26,703.23***
	(134.75)	(147.34)
Genetic principal components	Control	Control
Constant	125,211.96***	126,073.71***
	(459.59)	(483.05)
Observations	413,444	413,444
R-squared	0.18	0.18

Note: Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1; GPCA is the genetic principal components array.

6.2.2 Different method: Two-sample Mendelian Randomization method

4 Both First-stage F-statistical joint significance of the IVs far exceeds the threshold of 10 weak instruments. The first F-statistics of obesity is 1824.

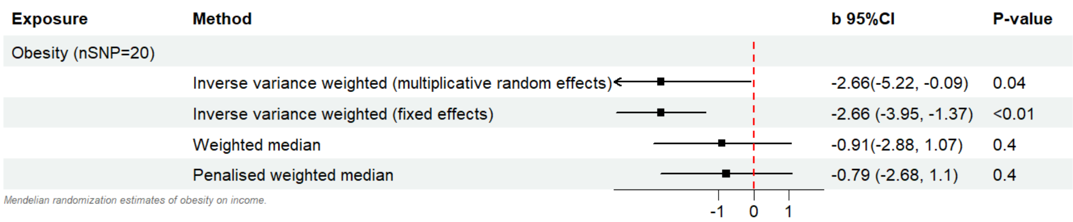
To improve the validity and robustness of this study, we have also employed the two-sample Mendelian randomization (2SMR) approach. Under the premise that all statistical assumptions are met, the orientation of parameter estimates derived from one-sample MR and two-sample MR should be consistent. In a two-sample MR analysis, genetic variants should be extracted from two different summary statistics of large-scale genome-wide association studies (GWAS). The association between genetic variation and obesity is estimated in the first GWAS, and the association between genetic variation and income is calculated in the other GWAS. In fact, its principle is similar to the two-sample instrumental variable method and the two-sample two-stage least square method proposed and developed by Angrist and Krueger (1992).

We apply four Mendelian randomization methods, including inverse variance weighting (IVW) with both multiplicative random effects and fixed effects, weighted median, and penalised weighted median, to ensure the robustness of the results. Inverse variance weighting is a primary method to infer causal effect since this method assumes that all SNPs are valid instrumental variables and can provide the highest statistical power and obtain accurate estimation results (Greco M *et al*., 2015). It is a meta-analysis of each Wald ratio used to estimate the effect of exposure on the outcome for each IV to get MR estimates (Lawlor *et al*., 2008). Furthermore, we perform a series of sensitivity analyses to avoid bias. Cochran's Q statistics using MR egger and IVW methods are used to examine the heterogeneity, with the p-value lower than 0.05 indicating the presence of heterogeneity. Applying the IVW method with the random effects model to analysis is more suitable if heterogeneity exists (Bowden *et al*., 2017). Next, we implement the leave-one-out approach to eliminate one SNP at a time and assess the individual SNP's impact on the outcome.

We adopted diagnosed obesity from the hospital to measure obesity to improve our findings' reliability in the two-sample MR analysis. Genome-wide associations for diagnosed obesity are derived from the UK Biobank covering 463,010 participants of European ancestry (GWAS ID: ukb-15541). The UK Biobank also provided summary statistics from a genome-wide association analysis of income, and the GWAS project identifier is ukb-b-7408, including 397,751 participants of European ancestry with about 9.8 million genetic variants. After statistical analysis, 48 SNPs about income attained genome-wide significance ($p<5^{*}10^{-8}$). These GWAS have adjusted for sex, age, and ten principal genotypic components.

Figure 2 reports the two-sample MR estimates. Different analysis methods show consistent evidence that obesity negatively impacts income. Specifically, in IVW with multiplicative random effects model, one standard deviation increase in obesity (β_{IVW}=-2.4,P<0.01) significantly decreases income. It further supports a causal effect that obesity lowers income, implying that reducing weight may result in additional economic benefits. We also used MR Egger and IVW to calculate Cochran's Q test results to detect heterogeneity. Heterogeneity is found in examining the effect of obesity on income (See Supplementary Table 1). It indicates that after controlling for genetic factors, the impact of obesity on income may vary across different races and genders. In addition, when we calculate the MR results of the remaining SNPs by removing one SNP at a time, we still find a negative effect of obesity on income (See Supplementary Figure 1).

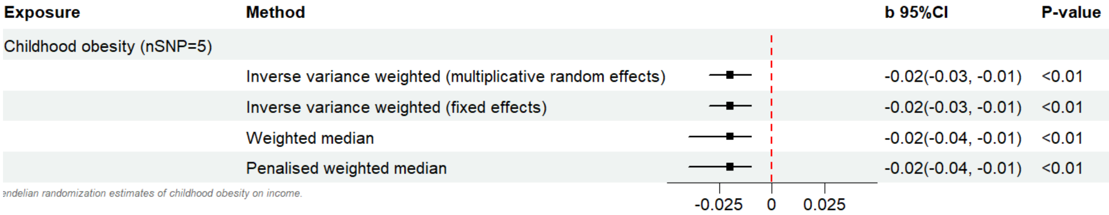
Figure 2 Two-sample MR estimates of the causal effect of obesity on income.



6.3 Further analysis: long-term effect of obesity on income

Furthermore, we want to clarify whether obesity has a long-term adverse effect on income. In order to avoid endogeneity problems caused by reverse causality, we use two-sample Mendelian randomization to analyze the impact of childhood obesity on adult income. We confirm childhood obesity (GWAS ID: ieu-a-1096) has a far-reaching negative causal effect on adulthood income. Studies have shown that educational attainment is usually positively correlated with income levels. Therefore, obesity may lead to adolescents experiencing bullying and discrimination in schools, affecting their motivation to learn and academic performance. Besides, obesity in childhood may increase the risk of disease in the future, restricting their cognitive abilities and career choices, which in turn affects future earnings. So, we should pay attention to the long-term adverse effects of residents' obesity and take some actions to prevent or mitigate childhood obesity and ultimately improve their future economic prospects.

Figure 3 Two-sample MR estimates of the causal effect of childhood obesity on income.



7 Conclusion

Our study uses genetic variants as instrument variables within the Mendelian randomization analysis framework to analyze the causal effect of obesity on residents' income. The results demonstrate a negative causal effect of obesity on income in China's rural areas and the United Kingdom. It reduces rural residents' annual income by 609.42 CNY based on a sample of 441 people collected from rural China, and the result is more accurate than the estimate using the relative's obesity status as IV. To generalize the results of our study, we also use data from the UK Biobank covering 413,444 samples and find obesity causally decreases residents' annual income by 8652 GBP. Furthermore, we use the two-sample Mendelian randomization method to analyze. They further support obesity is bad for people's income. Finally, we confirm childhood obesity has a long-term negative causal effect on adulthood income.

Our studies include several strengths. First, to our knowledge, this study is the first to evaluate the causal effect between obesity and income using the Mendelian randomization method in China. Second, we use measured height and weight to calculate BMI instead of relying on self-reported data to minimize measurement errors. Third, we use the UK Biobank sample, and the two-sample MR method acquires results that are consistent with our Chinese observations. Finally, we also confirm obesity has a long-term adverse effect on people's income.

However, some limitations in our analysis suggest promising avenues for further research. Firstly, when using the One-sample Mendelian Randomization method to assess the impact of obesity on Chinese residents' income, we observe that the MR results are only significant at the 10% significance level. This suggests a need for a cautious interpretation of our MR findings. Nevertheless, it is encouraging to note that consistent conclusions are obtained through diverse datasets and analytical approaches, affirming the results' reliability. Secondly, the available data do not support us to further explore the potential underlying mechanisms linking obesity to income. Consequently, further investigations are warranted to investigate the exact mechanisms by which obesity affects labor market outcomes.

According to previous studies, the detrimental impact of obesity on individuals' income can be attributed to several factors. Firstly, obesity is associated with an increased risk of chronic diseases like diabetes and hypertension, decreasing labor productivity (Mazhar, 2022). Additionally, obese employees may incur higher medical expenses and injuries, causing an increase in the employer's employment costs, which may result in reduced wages (Kleinman *et al*., 2014). Secondly, people may have stereotypes of obese employees; for example, obese individuals may be perceived as lacking self-control, which hinders their career development and salary. Thirdly, obesity can negatively impact employees' mental states and reaction times (Cawley, 2004). Besides, obesity may be related to the increased risk of cognitive abilities impairment, mainly learning, memory, and executive functioning (Oliveras-Cañellas *et al*., 2023). Finally, specific industries may have body type requirements or health standards that obese employees may struggle to meet, limiting their career prospects and earning potential.

This research is critical to enhance our comprehension of the economic consequences of obesity and alleviate income disparity in China. Serval policy-related suggestions may be derived from this study. First, helping the public increase their awareness of health and establishing healthy lifestyles, especially for children and adolescents, may be an effective way to alleviate the long-term economic costs of obesity. Second, limit advertising of foods high in sugar and fat and encourage consumers to choose healthy food to avoid obesity. Third, weight penalties may lead to a deviation from the efficiency and capability that should be pursued in the labor market, undermine social fairness and justice, and widen income disparity. As a result, it is essential to form a scientific and fair talent selection mechanism and avoid body discrimination, which is easily ignored in the labor market.

Reference

[1] Akiyama, M., Okada, Y., Kanai, M., Takahashi, A., Momozawa, Y., Ikeda, M., Iwata, N., Ikegawa, S., Hirata, M., Matsuda, K., Iwasaki, M., Yamaji, T., Sawada, N., Hachiya, T., Tanno, K., Shimizu, A., Hozawa, A., Minegishi, N., Tsugane, S., … Kamatani, Y. (2017). "Genome-wide association study identifies 112 new loci for body mass index in the Japanese population". *Nature Genetics*, Vol. 49 No.10, pp.1458–1467. https://doi.org/10.1038/ng.3951

[2] Alauddin Majumder, Md. (2013). "Does Obesity Matter for Wages? Evidence from the United States". *Economic Papers: A Journal of Applied Economics and Policy*, Vol.32 No.2, pp.200–217. https://doi.org/10.1111/1759-3441.12030

[3] Angrist, J. D., & Krueger, A. B. (1992). "The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples". *Journal of the American Statistical Association*, Vol 87 No.418, pp.328–336. https://doi.org/10.1080/01621459.1992.10475212

[4] Averett, S., & Korenman, S. (1993). *The Economic Reality of the Beauty Myth* pp.4521. National Bureau of Economic Research. https://doi.org/10.3386/w4521

[5] Behrman, J. R., & Rosenzweig, M. R. (2002). The Returns to Increasing Body Weight. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.297919

[6] Biener, A., Cawley, J., & Meyerhoefer, C. (2018). "The Impact of Obesity on Medical Care Costs and Labor Market Outcomes in the US". *Clinical Chemistry*, Vol.64 No.1, pp.108–117. https://doi.org/10.1373/clinchem.2017.272450

[7] Bowden, J., Del Greco M, F., Minelli, C., Davey Smith, G., Sheehan, N., & Thompson, J. (2017). "A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization: A framework for two-sample summary data MR". *Statistics in Medicine*, Vol.36 No.11, pp.1783–1802.

https://doi.org/10.1002/sim.7221

[8] Caliendo, M., & Gehrsitz, M. (2016). "Obesity and the labor market: A fresh look at the weight penalty". *Economics & Human Biology*, Vol.23, pp.209–225. https://doi.org/10.1016/j.ehb.2016.09.004

[9] Cawley, J. (2004). "The Impact of Obesity on Wages". *Journal of Human Resources*, No.2, pp.451–474. https://doi.org/10.3368/jhr.XXXIX.2.451

[10] Cercato, C., & Fonseca, F. A. (2019). "Cardiovascular risk and obesity". *Diabetology & Metabolic Syndrome*, Vol.11 No.1, pp.74. https://doi.org/10.1186/s13098-019-0468-0

[11] Comuzzie, A. G., & Allison, D. B. (1998). "The Search for Human Obesity Genes". *Science*, Vol.280 No.5368, pp.1374–1377. https://doi.org/10.1126/science.280.5368.1374

[12] Fan, S., Cho, E. E., & Rue, C. (2017). "Food security and nutrition in an urbanizing world: A synthesis of the 2017 Global Food Policy Report". *China Agricultural Economic Review*, Vol.9 No.2, pp.162–168. https://doi.org/10.1108/CAER-02-2017-0034

[13] Giskes, K., Van Lenthe, F. J., Turrell, G., Kamphuis, C. B. M., Brug, J., & Mackenbach, J. P. (2008). "Socioeconomic Position at Different Stages of the Life Course and Its Influence on Body Weight and Weight Gain in Adulthood: A Longitudinal Study With 13-Year Follow-up". *Obesity*, Vol. 16 No.6, pp.1377–1381. https://doi.org/10.1038/oby.2008.54

[14] Greco M, F. D., Minelli, C., Sheehan, N. A., & Thompson, J. R. (2015). "Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome: Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome". *Statistics in Medicine*, Vol.34 No.21, pp.2926–2940. https://doi.org/10.1002/sim.6522

[15] Grossman, M. (1972). "On the Concept of Health Capital and the Demand for Health". *Journal of Political Economy*, Vol.80 No.2, pp.223–255. https://doi.org/10.1086/259880

[16] Heckman, J. J. (2007). "The economics, technology, and neuroscience of human

capability formation". *Proceedings of the National Academy of Sciences*, Vol.104 No. 33, pp.13250–13255. https://doi.org/10.1073/pnas.0701362104

[17] Huang, C.-C., Yabiku, S. T., Ayers, S. L., & Kronenfeld, J. J. (2016). "The obesity pay gap: Gender, body size, and wage inequalities—a longitudinal study of Chinese adults", 1991–2011. *Journal of Population Research*, Vol.33 No.3, pp.221–242. https://doi.org/10.1007/s12546-016-9170-3

[18] Katsaiti, M.-S., & Shamsuddin, M. (2016). "Weight discrimination in the German labour market". *Applied Economics*, Vol.48 No.43, pp.4167–4182. https://doi.org/10.1080/00036846.2016.1153791

[19] Lawlor, D. A., Harbord, R. M., Sterne, J. A. C., Timpson, N., & Davey Smith, G. (2008). "Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology". *Statistics in Medicine*, Vol.27 No.8, pp.1133–1163. https://doi.org/10.1002/sim.3034

[20] Li, S., & Cheong, T. S. (2016). "Convergence and mobility of rural household income in China: New evidence from a transitional dynamics approach". *China Agricultural Economic Review*, Vol.8 No.3, pp.383–398. https://doi.org/10.1108/CAER-09-2015-0126

[21] Li, W., Zhang, L., Yue, M., Ruiz-Menjivar, J., & Zhang, J. (2023). "Health vulnerability and health poverty of rice farmers: Evidence from Hubei province in China". *China Agricultural Economic Review*, Vol.15 No.2, pp.259–277. https://doi.org/10.1108/CAER-03-2021-0062

[22] Macchi, E. (2023). "Worth Your Weight: Experimental Evidence on the Benefits of Obesity in Low-Income Countries". *American Economic Review*, Vol.113 No.9, pp.2287–2322. https://doi.org/10.1257/aer.20211879

[23] Malik, V. S., Willett, W. C., & Hu, F. B. (2013). "Global obesity: Trends, risk factors and policy implications". *Nature Reviews Endocrinology*, Vol.9 No.1, pp.13–27. https://doi.org/10.1038/nrendo.2012.199

[24] Moro, A., Tello-Trillo, S., & Tempesti, T. (2019). "The Impact of Obesity on Wages: The Role of Personal Interactions and Job Selection". *LABOUR*, Vol.*33* No.2, pp.125–146. https://doi.org/10.1111/labr.12145

[25] Mushkin, S. J. (1962). "Health as an Investment". *Journal of Political Economy*, Vol.70 No.5, pp.129–157. https://doi.org/10.1086/258730

[26] NCD Risk Factor Collaboration (NCD-RisC). (2019). "Rising rural body-mass index is the main driver of the global obesity epidemic in adults". *Nature*, Vol.569 No.7755,pp. 260–264. https://doi.org/10.1038/s41586-019-1171-x

[27] Norton, E. C., & Han, E. (2008). "Genetic information, obesity, and labor market outcomes". *Health Economics*, Vol.17 No.9, pp.1089–1104. https://doi.org/10.1002/hec.1383

[28] Oliveras-Cañellas, N., Castells-Nobau, A., De La Vega-Correa, L., Latorre-Luque, J., Motger-Albertí, A., Arnoriaga-Rodriguez, M., Garre-Olmo, J., Zapata-Tona, C., Coll-Martínez, C., Ramió-Torrentà, L., Moreno-Navarrete, J. M., Puig, J., Villarroya, F., Ramos, R., Casadó-Anguera, V., Martín-García, E., Maldonado, R., Mayneris-Perxachs, J., & Fernández-Real, J. M. (2023). "Adipose tissue coregulates cognitive function". *Science Advances*, Vol.9 No.32, https://doi.org/10.1126/sciadv.adg4017

[29] Sarrias, M., & Iturra, V. (2022). "The Double Burden of Being A Woman and Obese: Evidence from the Chilean Labor Market". *Feminist Economics*, Vol.28 No.1, pp.199–231. https://doi.org/10.1080/13545701.2021.1988127

[30] Schultz, T. W. (1961). "Investment in Human Capital". *The American Economic Review*, Vol.51 No.1, pp.1–17.

[31] Slade, P. (2017). "Body mass and wages: New evidence from quantile estimation". *Economics & Human Biology*, Vol.27, pp.223–240. https://doi.org/10.1016/j.ehb.2017.07.001

[32] Spence, M. (1978). Job market signaling. *Uncertainty in Economics* pp.281–306. Elsevier. https://doi.org/10.1016/B978-0-12-214850-7.50025-5

[33] Wang, Y., Zhao, L., Gao, L., Pan, A., & Xue, H. (2021). 'Health policy and public health implications of obesity in China". *The Lancet Diabetes & Endocrinology*, Vol.9 No.7, pp.446–461. https://doi.org/10.1016/S2213-8587(21)00118-2

[34] Wen, W., Zheng, W., Okada, Y., Takeuchi, F., Tabara, Y., Hwang, J.-Y., Dorajoo, R., Li, H., Tsai, F.-J., Yang, X., He, J., Wu, Y., He, M., Zhang, Y., Liang, J., Guo,

X., Sheu, W. H.-H., Delahanty, R., Guo, X., … Tanaka, T. (2014). "Meta-analysis of genome-wide association studies in East Asian-ancestry populations identifies four new loci for body mass index". *Human Molecular Genetics*, Vol.23 No.20, pp.5492–5504. https://doi.org/10.1093/hmg/ddu248

[35] West, P. (1991). "Rethinking the health selection explanation for health inequalities". *Social Science & Medicine*, Vol.32 No.4, pp.373–384. https://doi.org/10.1016/0277-9536(91)90338-D

[36] Willage, B. (2018). "The effect of weight on mental health: New evidence using genetic IVs". *Journal of Health Economics*, Vol.57, pp.113–130. https://doi.org/10.1016/j.jhealeco.2017.11.003

[37] Zeng, Q., Li, N., Pan, X.-F., Chen, L., & Pan, A. (2021). "Clinical management and treatment of obesity in China". *The Lancet Diabetes & Endocrinology*, Vol.9 No.6, pp.393–405. https://doi.org/10.1016/S2213-8587(21)00047-4

[38] Zhu, C., Chen, Q., Si, W., Li, Y., Chen, G., & Zhao, Q. (2020). "Alcohol Use and Depression: A Mendelian Randomization Study From China". *Frontiers in Genetics*, Vol.11, No.585351. https://doi.org/10.3389/fgene.2020.585351

Supplementary materials

Table 1 Pleiotropy and heterogeneity test about obesity on income.

Exposure	Pleiotropy test			Heterogeneity test					
	MR egger			MR egger			Inverse variance weighted		
	Intercept	Se	P	Q	Q_df	Q_pval	Q	Q_df	Q_pval
Obesity	-0.012	0.005	0.03	45.420	18	$3.602*10^{-4}$	59.643	19	$4.407*10^{-6}$
BMI	-0.000	0.000	0.439	1241.883	447	$5.637*10^{-76}$	1243.554	448	$5.506*10^{-76}$

Note: df: degree of freedom; MR: Mendelian randomization; Q: heterogeneity statistic Q.

Figure 1 Two-sample MR estimates of the causal effect of obesity on income.

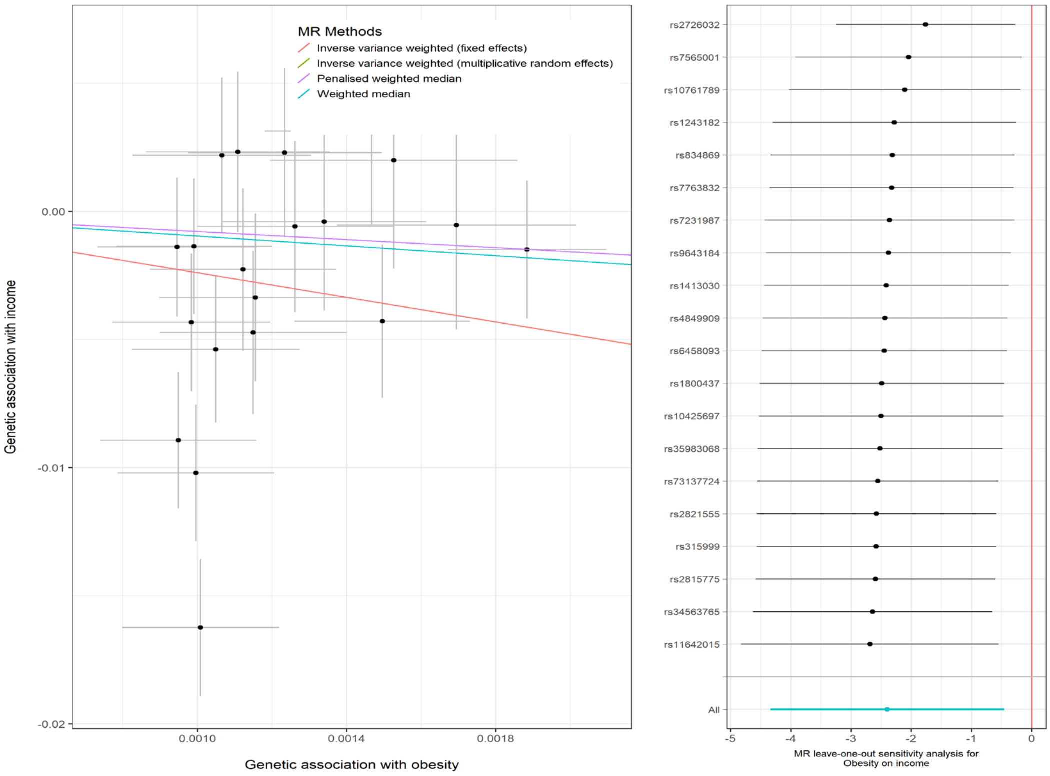


Figure 2 Two-sample MR estimates of the causal effect of childhood obesity on income.