

Body mass index and all-cause mortality in HUNT and UK Biobank studies: revised non-linear Mendelian randomization analyses

Stephen Burgess^{1,2,3}, Yi-Qian Sun^{4,5,6}, Ang Zhou^{1,7}, Christopher Buck¹, Amy M Mason^{2,3}, Xiao-Mei Mai⁸

1. Medical Research Council Biostatistics Unit, University of Cambridge, Cambridge, UK.
2. Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK.
3. Victor Phillip Dahdaleh Heart and Lung Research Institute, University of Cambridge, Cambridge, UK.
4. Department of Clinical and Molecular Medicine, Faculty of Medicine and Health Science, Norwegian University of Science and Technology, Trondheim, Norway.
5. Department of Pathology, Clinic of Laboratory Medicine, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway.
6. Center for Oral Health Services and Research Mid-Norway (TkMidt), Trondheim, Norway.
7. Australian Centre for Precision Health, Unit of Clinical and Health Sciences, University of South Australia, Adelaide, Australia.
8. Department of Public Health and Nursing, Faculty of Medicine and Health Science, Norwegian University of Science and Technology, Trondheim, Norway.

Corresponding author: Stephen Burgess (sb452@medschl.cam.ac.uk)

Funding statement: This work was supported by the Wellcome Trust (grant number 225790/Z/22/Z), the United Kingdom Research and Innovation Medical Research Council (grant number MC_UU_00002/7), the British Heart Foundation (RG/18/13/33946) and the National Institute for Health and Care Research (NIHR) Cambridge Biomedical Research Centre (BRC-1215-20014; NIHR203312). YQS was supported by a researcher grant from The Liaison Committee for Education, Research and Innovation in Central Norway (project ID 2018/42794). AMM is funded by the National Institute for Health and Care Research (NIHR) Blood and Transplant Research Unit (BTRU) in Donor Health and Behaviour (NIHR203337). The views expressed are those of the authors and not necessarily those of the NIHR, NHSBT or the Department of Health and Social Care.

Competing interest statement: None declared.

Word count: 299 (abstract), 4357 (main text).

Key words: Obesity paradox, reverse causation, genetic epidemiology, causal inference, doubly-ranked stratification.

Acknowledgments

The Trøndelag Health Study (HUNT) is a collaboration between the HUNT Research Centre (Faculty of Medicine and Health Sciences, NTNU, Norwegian University of Science and Technology), Trøndelag County Council, Central Norway Regional Health Authority and the Norwegian Institute of Public Health.

This research has been conducted using the UK Biobank Resource under Application Number 7439.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

ABSTRACT

Objectives: To estimate the shape of the causal relationship between body mass index (BMI) and mortality risk in a Mendelian randomization framework.

Design: Mendelian randomization analyses of two prospective population-based cohorts.

Setting: Individuals of European ancestries living in Norway or the United Kingdom.

Participants: 56,150 participants from the Trøndelag Health Study (HUNT) in Norway and 366,385 participants from UK Biobank recruited by postal invitation.

Outcomes: All-cause mortality and cause-specific mortality (cardiovascular, cancer, non-cardiovascular non-cancer).

Results: A previously published non-linear Mendelian randomization analysis of these data using the residual stratification method suggested a J-shaped association between genetically-predicted BMI and mortality outcomes with the lowest mortality risk at a BMI of around 25 kg/m². However, the “constant genetic effect” assumption required by this method is violated. The re-analysis of these data using the more reliable doubly-ranked stratification method still indicated a J-shaped relationship, but with less precision in estimates at the lower end of the BMI distribution. Evidence for a harmful effect of reducing BMI at low BMI levels was only present in some analyses, and where present, only below 20 kg/m². A harmful effect of increasing BMI for all-cause mortality was evident above 25 kg/m², for cardiovascular mortality above 24 kg/m², for non-cardiovascular non-cancer mortality above 26 kg/m², and for cancer mortality above 30 kg/m². In UK Biobank, the association between genetically-predicted BMI and mortality at high BMI levels was stronger in women than in men.

Conclusion: This research challenges findings from previous conventional observational epidemiology and Mendelian randomization investigations that the lowest level of mortality risk is at a BMI level of around 25 kg/m². Our results provide evidence that reductions in BMI will only increase mortality risk for a small proportion of the population, and increases in BMI will increase mortality risk for those with BMI above 25 kg/m².

Strengths and limitations of the study

- Mendelian randomization design minimizes bias due to confounding and reverse causation
- Large sample sizes enable powerful analyses even in low BMI individuals
- Validity of the genetic variants as instrumental variables cannot be verified
- Bias due to selection could be non-negligible and could vary across strata
- All estimates are averaged across a stratum of the population; individual effects of raising or lowering BMI may vary between individuals

INTRODUCTION

Body mass index (BMI) is a convenient and accessible measure of obesity. The epidemiological relationship between BMI and mortality is complex, with many observational studies conducted in Western countries showing a J-shaped relationship between BMI and mortality risk in the general population, such that mortality risk is lowest for those in the upper normal weight (BMI 20.0-24.9)¹⁻³ or even the overweight (BMI 25.0-29.9)⁴⁻⁶ category rather than the lower normal weight (BMI 18.5-19.9) or underweight category (BMI <18.5). However, such findings may not reflect the causal relationship between BMI and mortality, as observational associations are influenced by confounding and reverse causation.

Mendelian randomization is an epidemiological technique for the analysis of observational data designed to avoid bias due to confounding and reverse causation⁷. Rather than assessing associations between BMI and mortality directly, we assess associations between genetic variants that predict BMI levels and mortality risk⁸. This is analogous to the analysis of a randomized controlled trial for weight loss⁹, which would assess associations between the randomized intervention and the outcome, rather than associations with measured BMI in trial participants. According to Mendel's laws of inheritance, genetic variants should be uncorrelated with traits that they do not affect, and hence should be independent of potential confounders. As genetic variants are determined at conception, genetic associations should be protected from reverse causation¹⁰. These properties mean that genetic variants are plausible instrumental variables; an instrumental variable is a variable that behaves as if it has been randomized in the population¹¹. Hence, associations between genetic predictors of BMI and mortality provide insights into the causal effect of BMI on mortality.

Non-linear Mendelian randomization is an extension of standard Mendelian randomization to investigate the shape of the causal relationship between an exposure and an outcome¹². A typical randomized trial estimates an average causal effect, representing the average impact of a population-wide shift in the distribution of an exposure; similarly, a standard Mendelian randomization analysis estimates a population-averaged causal effect¹³. A widely used method for non-linear Mendelian randomization stratifies the population based on levels of the exposure, and estimates 'localized average causal effects', representing the average impact of a population-wide shift in the distribution of the exposure for that stratum of the population¹³.

However, stratifying the population such that the instrumental variable assumptions still hold in the strata of the population is tricky. Stratifying on the exposure directly would break randomization, as exposure levels are influenced by the instrumental variables. Hence an individual in the lowest stratum of the population defined by the exposure with a genetic predisposition to high values of the exposure would be more likely to have low values of the confounders. This is an example of collider bias¹⁴: the exposure is a common effect of the instrument and exposure—outcome confounders, and so stratification on the exposure leads to a conditional association between the instrument and confounders. The initial proposal for non-linear Mendelian randomization was to stratify on the residual exposure, defined as the residual from regression of the exposure on the instrumental variables¹³. The residual exposure is independent of the instrumental variables by construction, and the instrumental variables will be independent of confounders within strata of the residual exposure under a "constant genetic effect" assumption¹⁵. However, if the effect of the genetic instrumental variables on the exposure varies between individuals, then this residual stratification method can lead to severe bias in stratum-specific estimates^{16,17}. An alternative method, known as the doubly-ranked method, allows the instrumental variable effects on the exposure to vary between individuals provided that a weaker "rank-preserving assumption" holds¹⁸. The rank-preserving assumption states that the ranking of individuals according to their exposure levels would be the same at all levels of

the instrumental variable. The doubly-ranked method has been shown to be more reliable than the residual method in simulation studies^{17,18}.

In a previous paper, we presented evidence for the effect of BMI on mortality from non-linear Mendelian randomization using the residual stratification method in the Norwegian Trøndelag Health (HUNT) and UK Biobank studies. A J-shaped association between genetically-predicted BMI and mortality was observed in each study population overall, although a monotonic increasing association (that is, an always-increasing association) was seen in never-smokers¹⁹. Here, we present a re-analysis of the same datasets using the doubly-ranked stratification method. We first investigate whether the constant genetic effect assumption holds for this example, and explore the validity of the instrumental variable assumptions in strata of the population. We then present non-linear Mendelian randomization findings from the doubly-ranked stratification method.

METHODS

We present abbreviated descriptions of the datasets included in the analysis; detailed descriptions are in the original paper¹⁹.

The HUNT study

We used data from the second wave (1995-1997) of the HUNT study on 65,229 individuals living in the northern part of Trøndelag and over the age of 20 years²⁰. Participants were followed up until 15 June 2023 or their date of death. The original paper only considered mortality outcomes up to April 2015. By extending the follow-up period, we increase the number of mortality events considered in the analysis from 12,015 to 18,836. We excluded participants without data on BMI or genetic variants for BMI, leaving 56,150 individuals for analysis.

The UK Biobank study

The UK Biobank cohort comprises around 500,000 participants (94% of self-reported European ancestry) aged 40 to 69 years at baseline. They were recruited between 2006-2010 in 22 assessment centres throughout the UK, and followed up until 28 February 2021 or their date of death²¹. Again, the original paper only considered mortality outcomes up to February 2016. By extending the follow-up period, we increase the number of mortality events considered in the analysis from 10,344 to 25,021. We performed detailed quality control procedures on UK Biobank participants and on genetic variants as described previously²². In total, 366,385 unrelated European ancestry participants were included in analyses.

SNPs and genetic score used as instrumental variables

77 single nucleotide polymorphisms (SNPs) were selected as candidate instrumental variables for BMI based on European sex-combined analyses in a genome-wide association study of the GIANT consortium²³. Two of these variants (rs12016871 and rs2033732) were not available in the HUNT study, and a further two variants (rs13021737 and rs16951275) were excluded from the analyses due to association with smoking status in the HUNT study. An externally-weighted score was calculated for each individual by multiplying the number of BMI-increasing alleles the individual carries by the variant's association with BMI from the GIANT study (**Supplementary Table 1**), and summing across the remaining 73 variants. Overall, the genetic score explained 2.0% and 1.6% of the variance in BMI in the HUNT and UK Biobank studies respectively, corresponding to F-statistics of 1121 and 5964.

We assessed the instrumental variable assumptions within strata by estimating associations between the genetic score and various traits that are competing risk factors: smoking status (ever versus never), alcohol status (current versus other), education level (post-secondary or higher versus other), occupation (currently employed versus other), age at recruitment, and sex.

Study design

We performed non-linear Mendelian randomization analyses to estimate the shape of the association between genetically-predicted BMI and the outcome. Our primary analysis considered all-cause mortality as the outcome. We also conducted subgroup analyses considering men and women separately. In addition, we studied associations with cause-specific mortality events (cardiovascular, cancer, and non-cardiovascular non-cancer) in UK Biobank. We did not perform analyses stratified by smoking status (as presented in the original paper), as smoking status is influenced by BMI²⁴, and hence this stratification may introduce collider bias.

Statistical analyses

To estimate the non-linear relationship between genetically-predicted BMI and mortality risk, a fractional polynomial method was applied¹². In brief, first we divided the sample into 100 strata using the doubly-ranked method¹⁸. For comparison, we also present results from stratification using the residual method for the primary outcome of all-cause mortality. We then calculated the linear Mendelian randomization estimate, referred to as a localized average causal effect (LACE), in each stratum of the population as a ratio of coefficients: the association of the genetic score with the outcome divided by the association of the genetic score with the exposure. Associations with the exposure (BMI) were obtained from linear regression; associations with the outcome (mortality) were obtained from Cox proportional hazards regression, using age as the timescale. All associations were adjusted for age, age-squared, sex, and centre (for UK Biobank).

We performed meta-regression of the LACE estimates against the mean of the exposure in each stratum in a flexible semiparametric framework using the derivative of fractional polynomial models of degrees 1 and 2. Under the Mendelian randomization assumptions, the slope of the curve represents the average causal effect of the exposure on the outcome for the stratum with that exposure level. A positive causal effect is evident at a particular exposure level when the lower and upper confidence limits for the curve both have a positive slope; and similarly an inverse causal effect is evident if both confidence limits have a negative slope. The reference point in analyses is set to 25 kg/m²; however, this reference point is arbitrarily chosen, and the key indicator of an increasing or decreasing effect is not whether confidence intervals include or exclude the reference point, but rather whether the slope of the curve is positive or negative at a given exposure value.

We report two tests for non-linearity: a linearity test, which assesses whether a non-linear model fits the LACE estimates better than a linear model, and a trend test, which tests for a linear trend amongst the LACE estimates. Rejection of the linearity test indicates that a linear model is a suboptimal fit to the data; failure to reject means any improvement in fit for the best fitting non-linear model over the linear model is no more than would be expected due to chance alone.

As LACE estimates from the doubly-ranked method can be sensitive to specification of the analytic sample, we repeated analyses 100 times for each dataset omitting a small number of individuals in each iteration (12 individuals were removed at random in each iteration), and then combined estimates across iterations using Rubin's rules.

All statistical analyses were performed in R (version 4.3.1), and non-linear Mendelian randomization analyses were performed using the SUMnlmr package²⁵.

Ethical approval

The HUNT study has ethical approval from The Regional Committee for Medical Research in Norway (REK). Approval for individual projects is regulated in conjunction with The Norwegian Social Science Data Services (NSD). The UK Biobank study has ethical approval from the North West Multi-centre Research Ethics Committee (MREC).

Patient involvement

No patients were involved in setting the research question or the outcome measures, nor were they involved in the design or implementation of the study. No patients were asked to advise on interpretation or writing up of results. There are no specific plans to disseminate the results of the research to study participants or the relevant patient community.

RESULTS

Study populations

Baseline characteristics of the participants are provided in **Table 1**. Genetic associations with BMI in strata of the population defined by the doubly-ranked method are displayed in **Supplementary Figure 1**, and the BMI distribution in strata of the population is given in **Supplementary Table 2**. Per 1 standard deviation increase in the genetic score, BMI was 0.358 kg/m² greater in the lowest percentile of participants in the HUNT study, and 1.126 kg/m² greater in the highest percentile. Equivalent values for UK Biobank were 0.232 kg/m² in the lowest percentile and 1.645 kg/m² in the highest percentile. It is clear that the genetic associations with BMI differ strongly at different levels of BMI, and so the assumption required for the residual stratification method is violated. We note that similar estimates of the genetic associations with the exposure from the residual stratification method are unreliable, as these estimates from the residual method are similar even if the genetic effect on the exposure varies in the population¹⁸.

Assessment of instrument validity

Associations with traits that are competing risk factors in strata of UK Biobank are displayed in **Figure 1**. As previously observed in the literature for other exposures²⁶, there are clear patterns in genetic associations with age and sex across strata. While there is no overall association of the genetic score with age or sex, there are associations in several of the strata.

As it is logically impossible that autosomal genetic variants can affect either age or sex, and as age and sex are typically measured without error, these associations must represent the effect of selection into the UK Biobank sample. While concerning, as we adjust for age and sex in our Mendelian randomization analyses, these associations will not lead to strong bias in our causal estimates (see **Supplementary Material** for a simulation study investigation into bias mitigation by adjustment for predictors of selection). For other traits, both the magnitude of associations and patterns in associations are less strong, particularly for education level and occupation, which are important measures of social class. For alcohol status and smoking status, it is impossible to distinguish between genetic associations that reflect pleiotropic or selection effects and those that reflect a causal effect of the exposure (sometimes called “vertical pleiotropy”); the latter do not violate the instrumental variable assumptions. Hence both the associations and any pattern of associations may reflect violations of the instrumental variable assumptions or selection bias, but they could also reflect downstream effects of BMI which vary across strata.

Corresponding associations in the HUNT study are shown in **Supplementary Figure 2**. While there are some patterns in the stratum associations with age and sex, these are less pronounced than for UK Biobank, potentially due to differential selection being less strong in the HUNT study.

Comparison of results from residual and doubly-ranked methods

The shape of the association between genetically-predicted BMI and all-cause mortality is displayed in **Figure 2**, and stratum-specific estimates are provided in **Supplementary Figure 3**. When stratifying the population using the residual method, there is a distinct J-shaped relationship between BMI and mortality, particularly in the UK Biobank study, with minimum risk at a BMI level of around 25 kg/m². However, as discussed above, results from the residual stratification method are unreliable here. In contrast, when stratifying using the doubly-ranked method, the relationship between BMI and mortality is far more uncertain at the lower end of the BMI distribution. In the HUNT study, the association is mildly J-shaped, with a null slope up to a BMI of around 28 kg/m², and a positive slope

above this level. For UK Biobank, there is some indication of an upturn in disease risk at very low levels of BMI (the slope of the curve is negative below 20 kg/m²), but otherwise the relationship is flat (and compatible with a null effect) up to a BMI of around 25 kg/m², and has an increasing positive slope above this level. The linearity and trend tests for the residual method indicate evidence supporting non-linearity for UK Biobank ($p_{\text{linearity}} = 5 \times 10^{-4}$, $p_{\text{trend}} = 2 \times 10^{-5}$). In contrast, these tests for the more reliable doubly-ranked method do not provide compelling evidence supporting non-linearity (**Figure 2**). Similar results were observed when inverse-normal rank transforming the exposure (**Supplementary Figure 4**).

Sex-stratified analyses

Sex-stratified analyses from the doubly-ranked method are presented in **Figure 3**, and stratum-specific estimates are provided in **Supplementary Figure 5**. In each case, there is some indication of a J-shaped relationship, but an upturn in disease risk is only evident below 20 kg/m² (below 22 kg/m² for men in the HUNT study). Otherwise, estimates are generally compatible with the null up to a BMI of 25 kg/m², and positive above this level. Evidence for a harmful effect of increases in BMI for both men and women is weaker in the HUNT study, although a harmful effect above 30 kg/m² is evident for men. The slope above 25 kg/m² is steeper for women than men in UK Biobank. The linearity and trend tests do not provide compelling evidence supporting non-linearity in any analysis (**Figure 3**).

Cause-specific mortality

Analyses investigating cause-specific mortality from the doubly-ranked method in UK Biobank are presented in **Figure 4**, and stratum-specific estimates are provided in **Supplementary Figure 6**. Again, there is some indication of a J-shaped relationship for each outcome. This is most pronounced for cardiovascular mortality, and less evident for cancer and non-cardiovascular non-cancer mortality. An upturn in disease risk at low BMI levels was only evident for cardiovascular mortality, and only below 20 kg/m². Otherwise, the relationship was compatible with the null up to around 24 kg/m² for cardiovascular mortality, 26 kg/m² for and non-cardiovascular non-cancer mortality, and 30 kg/m² for cancer mortality, and positive above this level. The slope above 25 kg/m² is steepest for cardiovascular mortality, then non-cardiovascular non-cancer, and finally for cancer mortality. The linearity and trend tests do not provide any evidence supporting non-linearity for any outcome, indicating that a non-linear model does not fit the data significantly better than a linear model for any outcome (**Figure 4**).

DISCUSSION

In this manuscript, we have investigated the shape of associations between genetically-predicted BMI and mortality outcomes to provide an indication of the causal effect of BMI on mortality risk at different levels of BMI using non-linear Mendelian randomization. While the curves generally display a J-shaped relationship, evidence for a harmful average effect of decreased BMI on mortality at low BMI levels was limited. Evidence supporting non-linearity from the doubly-ranked method was unconvincing for all mortality outcomes. In UK Biobank, the slope representing the estimated effect of BMI on mortality risk above 25 kg/m² was sharper for women than for men, and sharpest for cardiovascular mortality compared with non-cardiovascular non-cancer mortality and cancer mortality. There was evidence for a harmful effect of high BMI on all mortality subtypes.

Our findings challenge “obesity paradox” results from observational epidemiology and previous Mendelian randomization analyses using the residual method, which suggest that increases in BMI might reduce risk of mortality for a substantial proportion of the general population. While the minimum of the estimated curve for all-cause mortality was at a BMI of around 23 kg/m² in UK Biobank, the curve was near flat from 22-24 kg/m², and was positive from 25 kg/m² upwards. Evidence for a harmful effect of decreased BMI at low levels of BMI (as indicated by negative slopes for both the upper and lower confidence limits of the BMI—mortality curve) was only present in some analyses, and only below 20 kg/m² when present. Amongst cause-specific mortality outcomes, this was only evident for cardiovascular mortality. In the HUNT study, there was no strong evidence for an effect of BMI on all-cause mortality at low BMI levels in overall analyses, and evidence for a harmful effect of increased BMI above around 28 kg/m². A potential reason for the discrepancy is that UK Biobank only contains participants from the age of 40 upwards, whereas the HUNT study includes younger individuals. Hence underweight strata in UK Biobank are more likely to comprise primarily of older, frailer individuals, whereas underweight strata in the HUNT study include more healthy, slim individuals.

Our investigation provides further empirical evidence that the assumptions required for the residual stratification method for non-linear Mendelian randomization are likely to be violated in practice, and that results from the residual method can be substantially different to those from the doubly-ranked method¹⁷. The ratio between the genetic association with BMI in the highest versus lowest percentile group was 3.1 in the HUNT study and 7.1 in UK Biobank, whereas the assumptions required by the residual method imply that these associations should be constant across the distribution of the exposure. Particularly in UK Biobank, the residual method provided strong evidence for non-linearity in the effect of BMI on mortality risk, giving a J-shaped relationship with the minimum risk level at around 25 kg/m². This is similar to the confounded association observed in traditional observational studies^{2,3}. In contrast, the doubly-ranked method provided little evidence for non-linearity. While there was some indication of a J-shaped curve in all analyses, estimates at the lower end of the BMI distribution were imprecise. One reason for this is that genetic associations with BMI were weaker at low levels of BMI in the doubly-ranked method, and so stratum-specific Mendelian randomization estimates are less precise in these strata.

Previous investigations into the non-linear shape of the causal relationship between BMI and mortality have used a variety of approaches. Wade *et al* used the residual stratification method for non-linear Mendelian randomization in UK Biobank²⁷, which showed a J-shaped relationship similar to the one we observe here. Jenkins *et al* investigated associations between a genetic score for BMI and mortality in UK Biobank⁵, considering analyses in the full dataset, and restricted to those with a pre-existing morbidity condition. While they did not observe evidence for non-linearity, the genetic score only explains a small proportion of variance in BMI, and so substantial non-linearity may not be

expected on this scale. The difference between average BMI levels at the lowest versus highest decile of the genetic score is less than 2 kg/m². Sulc *et al* investigated the effect of BMI on life expectancy (estimated as the mean of parental lifespan) in a Mendelian randomization framework using a polynomial approach similar to the residual method²⁸, except that instead of stratifying on the residual exposure, this approach adjusts for the residual exposure. They showed a linear effect of increased BMI on reduced life expectancy, with no indication of non-linearity. However, this approach also has been shown in simulations to be vulnerable to variation in genetic effects on the exposure¹⁸. Carslake *et al* performed an instrumental variable analysis to investigate the impact of BMI on mortality in the HUNT study²⁹, but using offspring BMI as an instrument for the BMI of study participants. This analysis did not provide evidence for a harmful effect of having low BMI. A similar analysis was performed in the 1958 British birth cohort, and a similar conclusion was reached³⁰. Blond *et al* also performed an instrumental variable analysis using offspring BMI as an instrument for the BMI of study participants in the Copenhagen School Health Records Register³¹, showing evidence of positive effects of increases in BMI on mortality at both high and low BMI levels. Finally, Kjøllesdal *et al* performed an instrumental variable analysis using early adulthood BMI as an instrument for the midlife BMI of study participants in various Norwegian health surveys³², showing mild J-shaped associations in the relationship between BMI and mortality, but far less pronounced than from observational analyses of the same dataset.

Overall, there is general consensus from different approaches that the adverse relationship of low BMI with mortality is exaggerated in observational analyses. However, variation between these results indicates the importance of modelling assumptions, both for the identification of causal effects, and for the modelling of non-linear relationships. Indeed, in our investigations, if we had used fractional polynomials of degree 1 rather than degree 2, then we would have observed monotonic increasing relationships of BMI with all outcomes. This is because the simpler degree 1 models are dominated by behaviour at the upper end of the BMI distribution, where instruments are strongest and most mortality outcomes occur. The simpler degree 1 polynomials smooth over imprecise negative estimates at the lower end of the BMI distribution, and so do not allow uncertainty in the shape of the distribution (particular at its lower end) to be accurately reflected.

Our analysis has several strengths, but also limitations. The Mendelian randomization design minimizes bias from confounding and reverse causation. The large sample sizes of the HUNT and UK Biobank studies enable powerful analyses, even in strata of the population with low BMI levels. These sample sizes also enable a fine stratification of individuals, meaning that outcome associations can be assessed in strata only consisting of low BMI (<20 kg/m²) individuals.

However, there are important limitations, which should lead to caution in the interpretation of our findings. As with all Mendelian randomization analyses, findings are dependent on the validity of the genetic variants as instrumental variables. To reduce the scope for population stratification, our analyses were limited to European ancestry participants. Our findings may therefore not be applicable to other populations. Further, recruitment into UK Biobank is dependent on age, sex, and other factors, which leads to bias in Mendelian randomization estimates. While the effect of moderate selection on Mendelian randomization estimates is often slight³³, the extent of selection bias may differ between strata, leading to differential bias across stratum-specific estimates²⁶. It is likely that age and sex are the strongest predictors of study participants, and differential genetic associations with these traits were observed in UK Biobank. For other traits, patterns in genetic associations across strata were less evident, and may reflect downstream effects of BMI rather than instrument invalidity due to pleiotropy or selection bias. As the HUNT study focused on a specific geographic area of Norway, it achieved much higher recruitment rates (around 70%, compared with

5% in UK Biobank), and so should be less affected by selection bias. This was corroborated by our analyses, as genetic associations in strata were weaker in the HUNT study. While genetic associations are protected from reverse causation, strata membership is determined by BMI levels, and hence could be subject to reverse causation. In particular, low BMI strata could contain a large proportion of individuals whose BMI levels are low because of comorbidity. However, genetic associations should still provide a reliable guide as to the extent and direction of any causal effect of BMI in these strata. Finally, all estimates represent averages across strata of the population; individual effects of raising or lowering BMI may vary between individuals. Differences in estimates for different strata may reflect the changing composition of the strata, as the characteristics of those with BMI less than 20 kg/m² are likely to be different to those with higher BMI. A non-linear curve may reflect a combination of different effects in different subgroups of the population, rather than that the relationship between BMI and mortality risk is non-linear for any individual in the population³⁴.

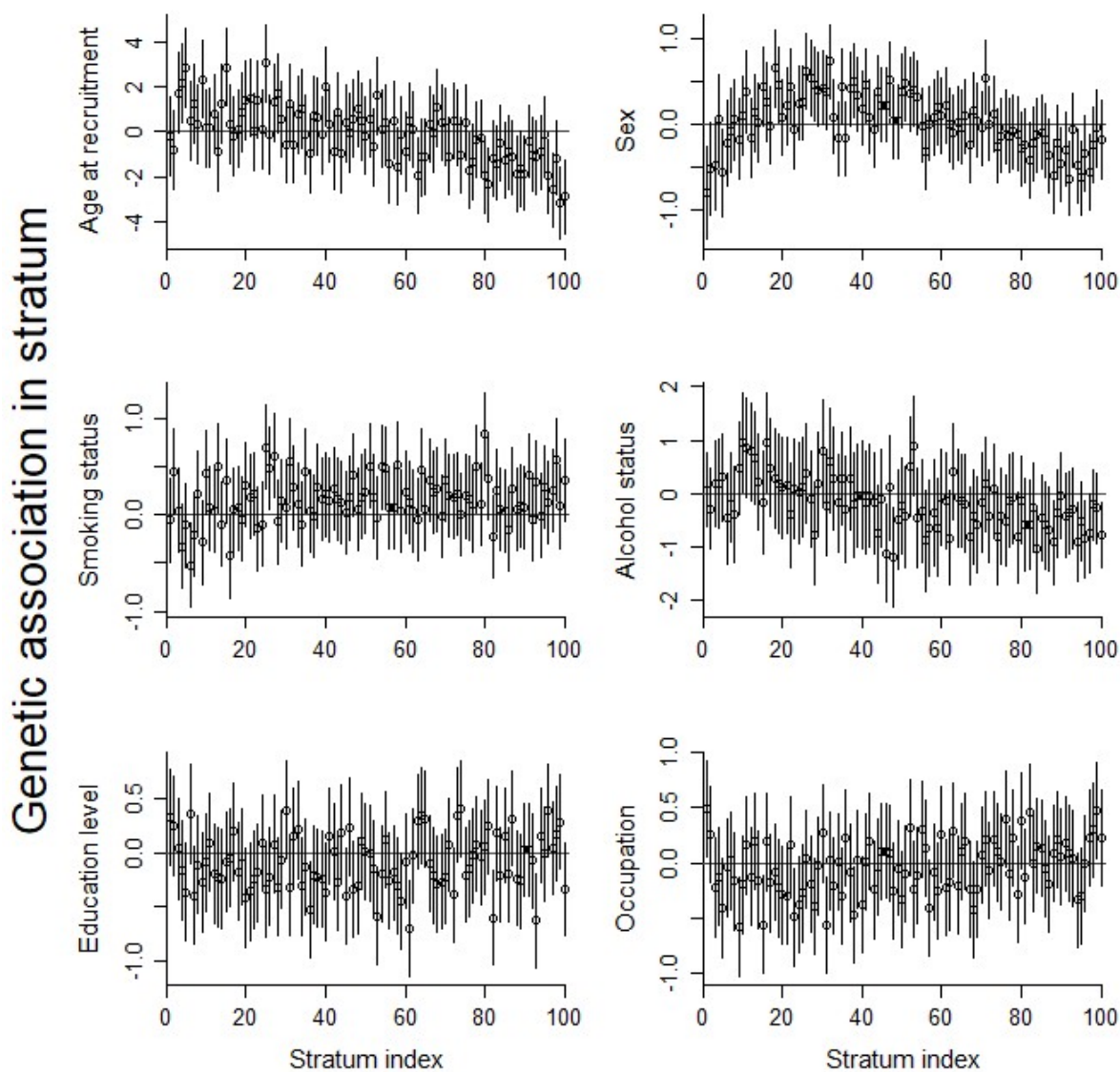
In conclusion, non-linear Mendelian randomization analyses using the doubly-ranked stratification method provide strong evidence for harmful effects of increased BMI on mortality above 25 kg/m². Evidence for a harmful effect of low BMI was only present in some analyses, and where present, only below 20 kg/m².

Table 1: Baseline characteristics of study participants in HUNT and UK Biobank studies

	HUNT	UK Biobank
Number of participants (n)	56,150	366,385
Men (%)	47.1	45.9
Age at baseline in years (SD)	49.6 (16.6)	56.7 (8.0)
Number of deaths (n)	18,836	25,021
- Cardiovascular mortality events (n)	-	5212
- Cancer mortality events (n)	-	12,880
- Other (non-cardiovascular non-cancer) events (n)	-	6692
BMI in kg/m ² (SD)	26.3 (4.1)	27.4 (4.8)
Median follow-up years	18.5	7.0
Ever-smokers (%)	55.9	46.1

BMI: body mass index; SD: standard deviation; Data are given as number of subjects (n), percentage (%), median, or mean (standard deviation). For 237 deaths in UK Biobank, sufficient data were not available to classify the event as cardiovascular, cancer, or non-cardiovascular non-cancer.

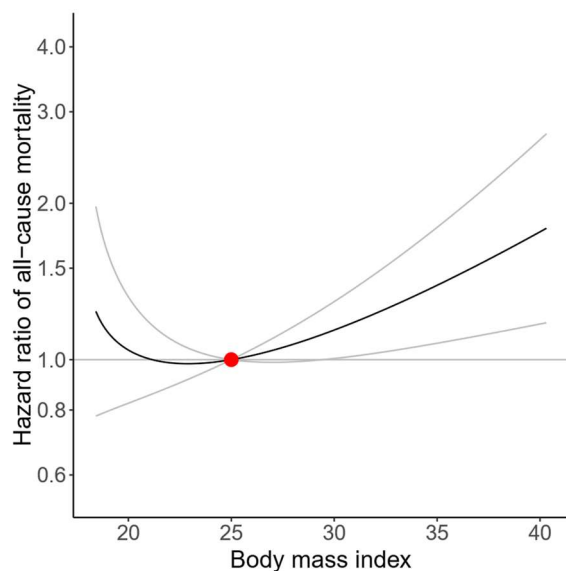
Figure 1: Genetic associations with potential competing risk factors in strata of the UK Biobank population



Points represent association estimates from linear (age at recruitment, years) or logistic regression (all other traits). Error bars represent 95% confidence intervals. Stratum 1 has lowest average BMI levels, stratum 100 has highest average BMI levels.

Figure 2: Non-linear Mendelian randomization – dose-response curve between body mass index and all-cause mortality. Using the doubly-ranked method: A) HUNT, B) UK Biobank; using the residual method (unreliable when the genetic effect on the exposure varies): C) HUNT, D) UK Biobank. Gradient at each point of the curve is the localized average causal effect. Grey lines represent 95% confidence intervals. The reference value for BMI was taken as 25 kg/m².

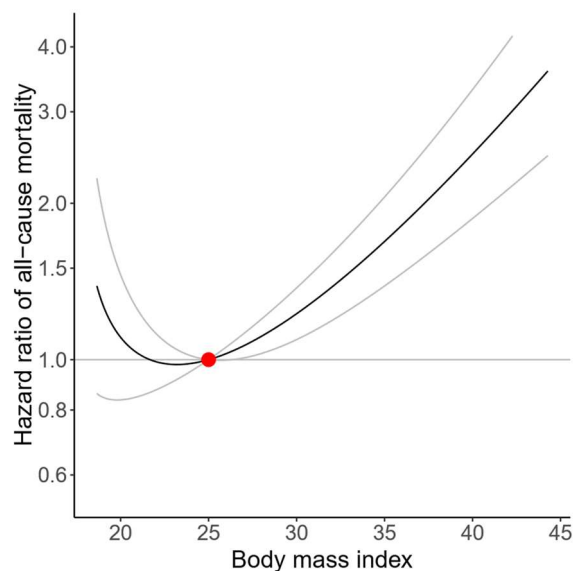
A) HUNT, doubly-ranked method:



p-value for linearity: 0.35

p-value for trend: 0.35

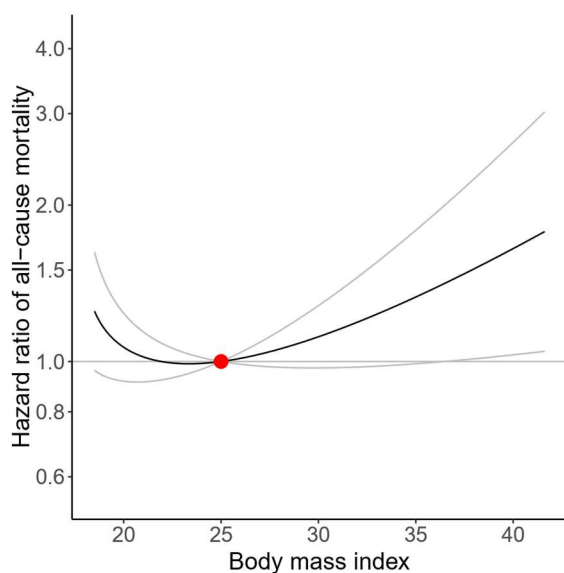
B) UK Biobank, doubly-ranked method:



p-value for linearity: 0.064

p-value for trend: 0.056

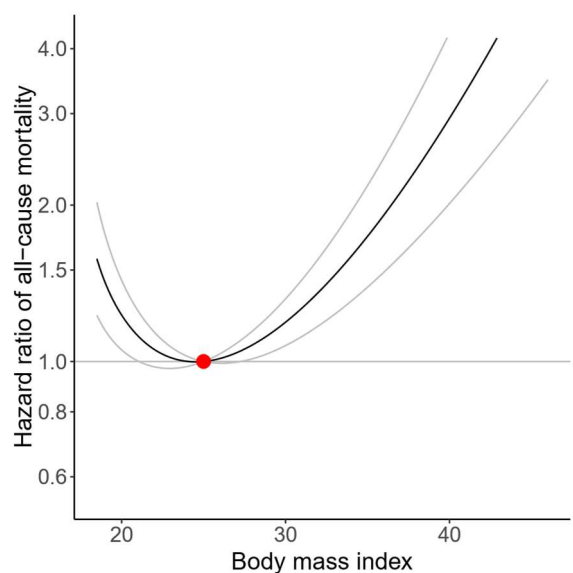
C) HUNT, residual method:



p-value for linearity: 0.12

p-value for trend: 0.12

D) UK Biobank, residual method:

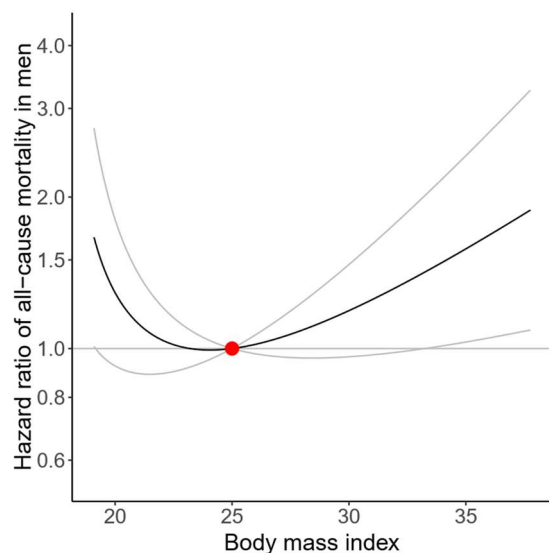


p-value for linearity: 5×10^{-4}

p-value for trend: 2×10^{-5}

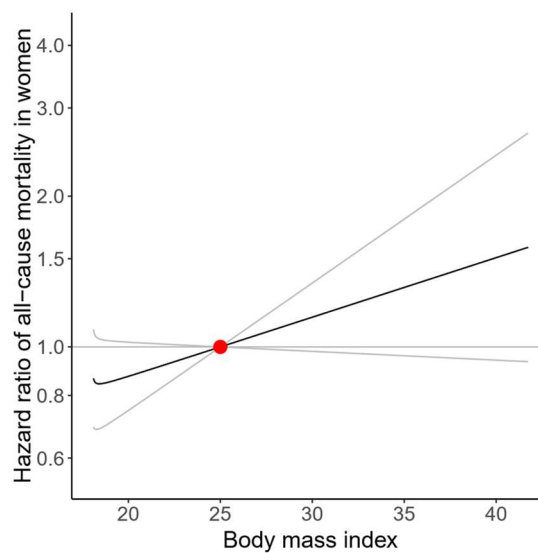
Figure 3: Non-linear Mendelian randomization – dose-response curve between body mass index and all-cause mortality in men and women respectively using doubly-ranked method. A) HUNT, men only; B) HUNT, women only; C) UK Biobank, men only; D) UK Biobank, women only. Gradient at each point of the curve is the localized average causal effect. Grey lines represent 95% confidence intervals. The reference value for BMI was taken as 25 kg/m².

A) HUNT, men only:



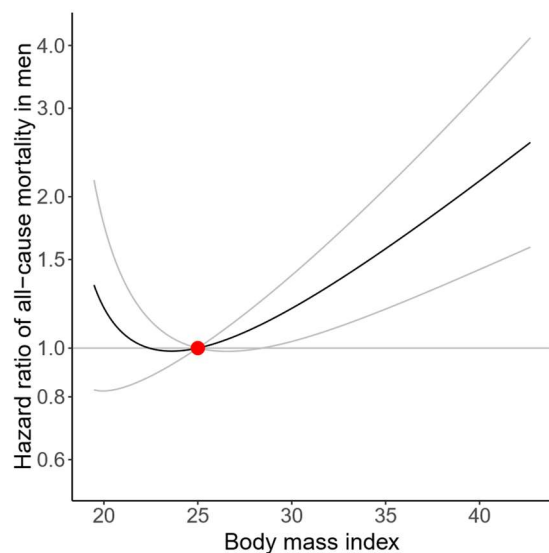
p-value for linearity: 0.050
p-value for trend: 0.22

B) HUNT, women only:



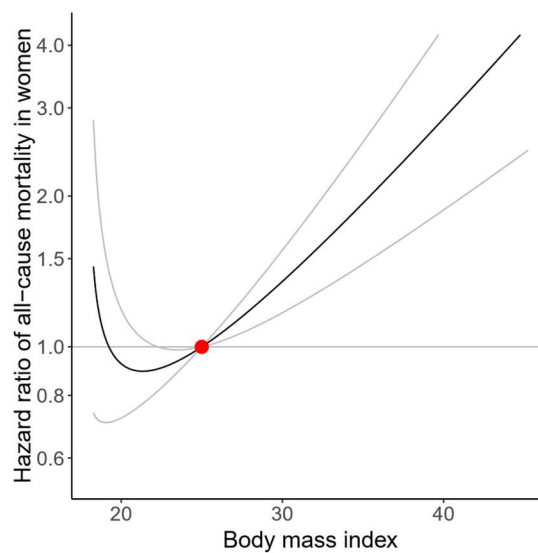
p-value for linearity: 0.99
p-value for trend: 0.79

C) UK Biobank, men only:



p-value for linearity: 0.18
p-value for trend: 0.18

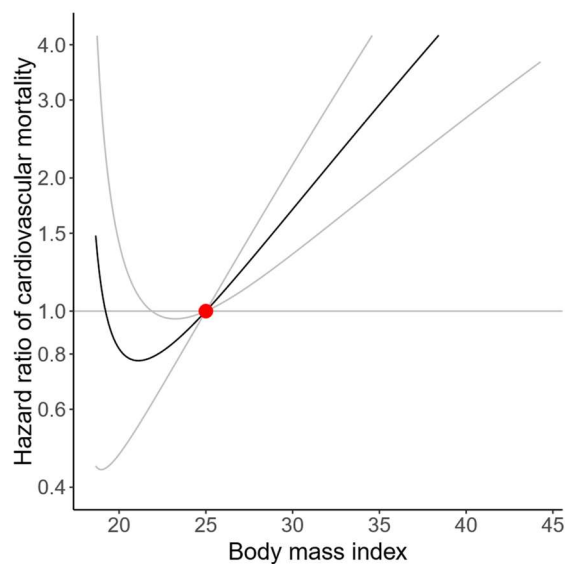
D) UK Biobank, women only:



p-value for linearity: 0.21
p-value for trend: 0.18

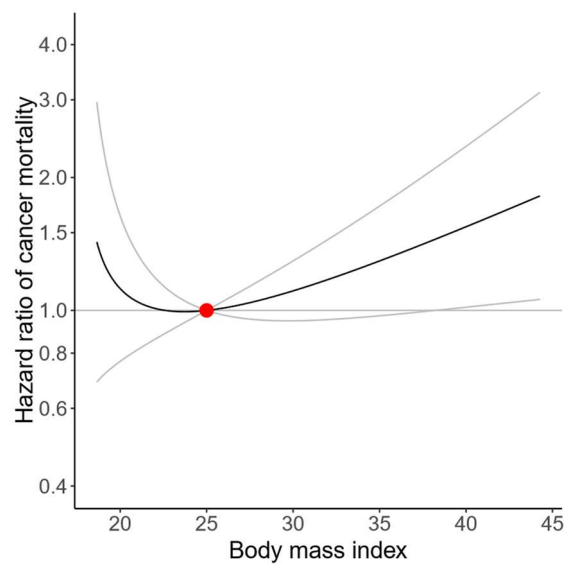
Figure 4: Non-linear Mendelian randomization in UK Biobank – dose-response curve between body mass index and cause-specific mortality using doubly-ranked method. A) Cardiovascular disease mortality; B) Cancer mortality. C) Mortality due to other causes (non-cardiovascular, non-cancer). Gradient at each point of the curve is the localized average causal effect. Grey lines represent 95% confidence intervals. The reference value for BMI was taken as 25 kg/m².

A: Cardiovascular mortality



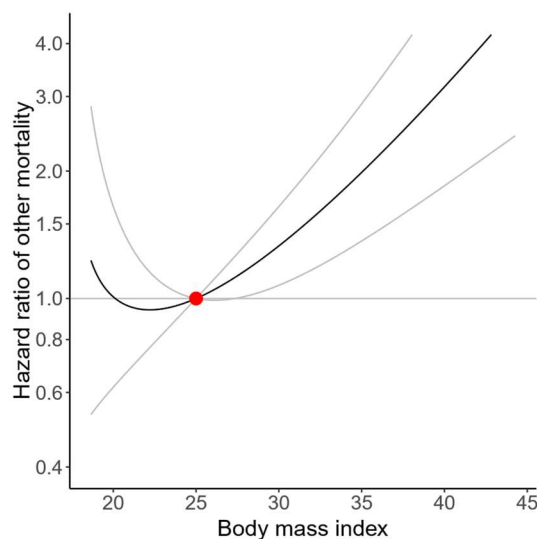
p-value for linearity: 0.99
p-value for trend: 0.92

B: Cancer mortality



p-value for linearity: 0.39
p-value for trend: 0.38

C: Other (non-cardiovascular, non-cancer) mortality



p-value for linearity: 0.20
p-value for trend: 0.19

REFERENCES

1. Aune D, Sen A, Prasad M, et al. BMI and all cause mortality: systematic review and non-linear dose-response meta-analysis of 230 cohort studies with 3.74 million deaths among 30.3 million participants. *BMJ*. 2016;353:i2156. doi:10.1136/bmj.i2156
2. Bhaskaran K, dos-Santos-Silva I, Leon DA, Douglas IJ, Smeeth L. Association of BMI with overall and cause-specific mortality: a population-based cohort study of 3.6 million adults in the UK. *The Lancet Diabetes & Endocrinology*. 2018;6(12):944-953. doi:10.1016/S2213-8587(18)30288-2
3. Global BMI Mortality Collaboration, Di Angelantonio E, Bhupathiraju Sh N, et al. Body-mass index and all-cause mortality: individual-participant-data meta-analysis of 239 prospective studies in four continents. *Lancet*. Aug 20 2016;388(10046):776-86. doi:10.1016/s0140-6736(16)30175-1
4. Flegal KM, Kit BK, Orpana H, Graubard BI. Association of All-Cause Mortality With Overweight and Obesity Using Standard Body Mass Index Categories: A Systematic Review and Meta-analysis. *JAMA*. 2013;309(1):71-82. doi:10.1001/jama.2012.113905
5. Jenkins DA, Wade KH, Carlslake D, et al. Estimating the causal effect of BMI on mortality risk in people with heart disease, diabetes and cancer using Mendelian randomization. *International journal of cardiology*. 2021;330:214-220. doi:10.1016/j.ijcard.2021.02.027
6. Visaria A, Setoguchi S. Body mass index and all-cause mortality in a 21st century U.S. population: A National Health Interview Survey analysis. *PLOS ONE*. 2023;18(7):e0287218. doi:10.1371/journal.pone.0287218
7. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology*. 2003;32(1):1-22.
8. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Statistics in Medicine*. 2008;27(8):1133-63. doi:10.1002/sim.3034
9. Thanassoulis G, O'Donnell CJ. Mendelian randomization: nature's randomized trial in the post-genome era. *JAMA*. 2009;301(22):2386-2388.
10. Burgess S, Swanson SA, Labrecque JA. Are Mendelian randomization investigations immune from bias due to reverse causation? *European Journal of Epidemiology*. 2021;36(3):253-257.
11. Martens EP, Pestman WR, de Boer A, Belitser SV, Klungel OH. Instrumental variables: application and limitations. *Epidemiology*. 2006;260-267.
12. Staley JR, Burgess S. Semiparametric methods for estimation of a nonlinear exposure-outcome relationship using instrumental variables with application to Mendelian randomization. *Genet Epidemiol*. May 2017;41(4):341-352. doi:10.1002/gepi.22041
13. Burgess S, Davies NM, Thompson SG. Instrumental variable analysis with a nonlinear exposure-outcome relationship. *Epidemiology*. Nov 2014;25(6):877-85.
14. Cole SR, Platt RW, Schisterman EF, Chu H, Westreich D, Richardson D, Poole C. Illustrating bias due to conditioning on a collider. *International Journal of Epidemiology*. 2010;39(2):417-420.
15. Small DS. Commentary: Interpretation and Sensitivity Analysis for the Localized Average Causal Effect Curve. *Epidemiology*. 2014;25(6)
16. Davey Smith G. Mendelian randomisation and vitamin D: the importance of model assumptions. *The Lancet Diabetes & Endocrinology*. 2023;11(1):14. doi:[https://doi.org/10.1016/S2213-8587\(22\)00345-X](https://doi.org/10.1016/S2213-8587(22)00345-X)
17. Burgess S. Violation of the constant genetic effect assumption can result in biased estimates for non-linear Mendelian randomization. *Human Heredity*. 2022;(In Press)doi:10.1159/000531659
18. Tian H, Mason AM, Liu C, Burgess S. Relaxing parametric assumptions for non-linear Mendelian randomization using a doubly-ranked stratification method. *PLOS Genetics*. 2023;19(6):e1010823. doi:10.1101/2022.06.28.497930
19. Sun Y-Q, Burgess S, Staley JR, et al. Body mass index and all cause mortality in HUNT and UK Biobank studies: linear and non-linear mendelian randomisation analyses. *British Medical Journal*. 2019;364:l1042. doi:10.1136/bmj.l1042

20. Krokstad S, Langhammer A, Hveem K, et al. Cohort Profile: The HUNT Study, Norway. *International Journal of Epidemiology*. 2013;42(4):968-977. doi:10.1093/ije/dys095
21. Sudlow C, Gallacher J, Allen N, et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLOS Medicine*. 2015;12(3):e1001779. doi:10.1371/journal.pmed.1001779
22. Astle WJ, Elding H, Jiang T, et al. The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell*. 2016;167(5):1415-1429. doi:10.1016/j.cell.2016.10.042
23. Locke AE, Kahali B, Berndt SI, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015;518(7538):197-206. doi:10.1038/nature14177
24. Carreras-Torres R, Johansson M, Haycock PC, Relton CL, Davey Smith G, Brennan P, Martin RM. Role of obesity in smoking behaviour: Mendelian randomisation study in UK Biobank. *BMJ*. 2018;361:k1767.
25. Mason AM, Burgess S. Software Application Profile: SUMnlmr, an R package that facilitates flexible and reproducible non-linear Mendelian randomization analyses. *International Journal of Epidemiology*. 2022;51(6):2014-2019. doi:10.1093/ije/dyac150
26. Hamilton FW, Hughes DA, Spiller W, Tilling K, Davey Smith G. Non-linear mendelian randomization: evaluation of biases using negative controls with a focus on BMI and Vitamin D. *medRxiv*. 2023:2023.08.21.23293658. doi:10.1101/2023.08.21.23293658
27. Wade KH, Carslake D, Sattar N, Davey Smith G, Timpson NJ. BMI and Mortality in UK Biobank: Revised Estimates Using Mendelian Randomization. *Obesity*. 2018;26(11):1796-1806. doi:doi:10.1002/oby.22313
28. Sulc J, Sjaarda J, Kutalik Z. Polynomial Mendelian randomization reveals non-linear causal effects for obesity-related traits. *HGG Advances*. 2022;3(3):100124. doi:10.1016/j.xhgg.2022.100124
29. Carslake D, Davey Smith G, Gunnell D, Davies N, Nilsen TIL, Romundstad P. Confounding by ill health in the observed association between BMI and mortality: evidence from the HUNT Study using offspring BMI as an instrument. *International Journal of Epidemiology*. 2018;47(3):760-770.
30. Hyppönen E, Carslake D, Berry DJ, Power C, Davey Smith G. Estimating the influence of body mass index (BMI) on mortality using offspring BMI as an instrumental variable. *International Journal of Obesity*. 2022;46(1):77-84.
31. Blond K, Carslake D, Gjærde LK, Vistisen D, Sørensen TIA, Smith GD, Baker JL. Instrumental variable analysis using offspring BMI in childhood as an indicator of parental BMI in relation to mortality. *Scientific Reports*. 2021;11(1):22408.
32. Kjøllestad MKR, Davey Smith G, Ariansen I, Kinge JM, Degerud E, Næss Ø. The association between BMI and mortality using early adulthood BMI as an instrumental variable for midlife BMI. *Scientific Reports*. 2018/07/31 2018;8(1):11499.
33. Gkatzionis A, Burgess S. Contextualizing selection bias in Mendelian randomization: how bad is it likely to be? *International Journal of Epidemiology*. 2018;48(3):691-701. doi:10.1093/ije/dyy202
34. Small DS. Commentary: Interpretation and sensitivity analysis for the localized average causal effect curve. *Epidemiology*. 2014;25(6):886-888.

Author contributions: Conceptualization: Burgess, Sun, Mai; Formal analysis: Burgess, Sun; Methodology: all authors; Writing – Original Draft: Burgess; Writing – Review & Editing: all authors.

STROBE statement: Included in submission.

Data sharing statement: Data from the HUNT and UK Biobank studies are available on application to any bona fide scientific researcher. Summary statistics for genetic associations with the exposure and outcome in strata for UK Biobank are available at [to be done after initial review].