Non-linear mendelian randomization: evaluation of biases using negative controls with a focus on BMI and Vitamin D

Fergus W Hamilton[1,2], David A Hughes[1], Wes Spiller[1], Kate Tilling[1], George Davey Smith[1]

1. MRC Integrative Epidemiology Unit, University of Bristol
2. Infection Science, North Bristol NHS Trust

Corresponding author:

Fergus Hamilton
MRC Integrative Epidemiology Unit
University of Bristol
Oakfield House
Oakfield Road
BS8 2PS

Fergus.hamilton@bristol.ac.uk
+447743165499

Abstract

Mendelian randomisation (MR) is an established technique in epidemiological investigation, using the principle of random allocation of genetic variants at conception to generate predicted causal effects of an exposure on an outcome. Extensions to this technique include non-linear approaches, that allow for differential effects of the exposure on the outcome depending on the level of the exposure. One popular method is the residual approach, which allows for stratification of the exposure via the level of the exposure and generation of causal estimates within each stratum. These "local" causal estimates are then used to make inferences about non-linear associations. Recent work has identified that under certain situations this method can lead to estimates that are seriously biased, and a new method - the doubly-ranked method – has been introduced as a possibly more robust approach. In this paper, we perform an analysis using negative control outcomes using both methods and find biased estimates in certain situations. We suggest caution in use of these methods and, if used, the performance of a number of sanity checks such as negative control outcomes in their reporting.

Introduction:

Mendelian randomization (MR) is now established as a key technique in epidemiology, to help overcome bias due to unmeasured confounding.[1,2] Estimates from MR have been consistent with randomised trial data in many settings,[3] and have assisted causal inference as a component of triangulation of evidence[4] Recently, there have been developments in MR, where an attempt is made to extend causal inference from a whole population (e.g. the effect of increasing vitamin D on all-cause mortality), to strata of that population (e.g. only those with low levels of vitamin D).[5–11] This has the potential to enhance causal inference and answer key questions around the linearity of exposure-outcome associations. In particular, the development of the "residual" method led to a number of MR studies that purported to show non-linear associations.[5,10,12–20] However, recent analyses have suggested that this method may be susceptible to bias, including e.g. that related to selection into strata. In this manuscript, we explore the development of the method and a number of issues that suggest caution in using the method at present.

Development of the residual method and concerns about inconsistent causal estimates

In the residual method the measured levels of the exposure (e.g. Vitamin D) are regressed onto the genetic instrument (e.g. a polygenic risk score for Vitamin D). The residuals of that regression represent the IV-free exposure (sometimes referred to as the "non-genetic" portion of Vitamin D). These residuals are then split into strata (often 10, or 100 strata) ordered by the mean IV-free exposure value. Therefore, the sample in stratum 1 have the lowest level of "IV-free" Vitamin D, and the sample in stratum 10 (or 100) have the highest value. MR is then run within each stratum, and estimates combined by parametric statistical methods (such as fractional polynomials) to estimate a non-linear curve. The requirement for stratification based on the residuals rather than the exposure itself is driven by the desire to avoid collider bias[21] as this would be induced by stratification on the exposure itself.

There has been a recent focus on the performance of this method. In particular, one study identified a non-linear association between Vitamin D and a number of outcomes, suggesting that increasing Vitamin D in those who were deficient might have considerable benefit (as opposed to the null overall causal effect).[11] Given the known distribution of variables , the findings from this analysis were not empirically plausible. Point estimates in each stratum suggested a protective effect, substantial in some, but there was a small adverse overall causal estimate.[22] Several other analyses

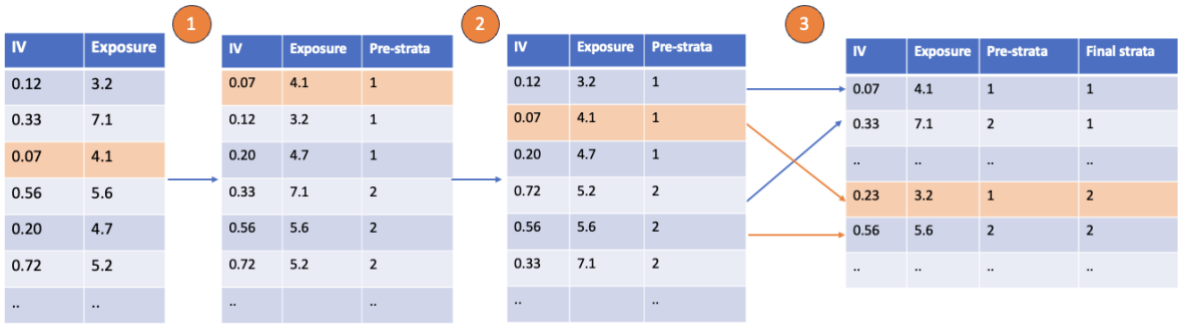
have used identical methods applied to the same exposure, and using largely overlapping data, have unsurprisingly produced similar and equally spurious findings.[15,16,23]

Further development of the doubly ranked method and ongoing concerns

Consequent on this criticism the authors of the initial non-linear Vitamin D analysis[11] have re-evaluated the residual method used in the study. In a preprint[24] they showed via simulation studies that this method was biased under certain situations, in part relating to the "constant genetic effect" assumption – genotype-exposure effect is linear and resultantly constant across strata (i.e. the coefficient for IV-exposure remains the same across strata). For Vitamin D, this was shown not to be the case, with much larger genotype-exposure effect estimates in top strata than in the bottom strata.[24] They also identified that this heterogeneity in IV-exposure association was not identified in the residual method, and suggest that this is a flaw in the residual approach.[25] To deal with these challenges, they developed a similar approach to developing strata (presented fully in a separate manuscript[25]) called the "doubly-ranked" method.

In this method, the strata are developed in multiple steps (**Figure 1**). In the first step, the population is ranked by the level of the genetic instrument into pre-strata. Subsequently, within each pre-stratum, the participant with the lowest level of the exposure is taken and placed into the lowest strata. The participant with the next lowest level of the exposure in pre-strata 1 is placed into strata 2. The same process occurs for all pre-strata. The first strata therefore contains the first participant of every pre-stratum, each of which has the lowest level of exposure in the pre-stratum. The number of pre-strata (K) to be generated depends on the number of strata desired (J) and the total sample size (N) using the formula N = J x K. That is, for a sample size of 1,000, and to generate a final desired number of strata of 10, 100 pre-strata are generated.

Figure 1: A worked example of strata generation in the doubly-ranked approach. In the first step, all participants are sorted by their level of the instrumental variable (IV) into pre-strata. In the second step, participants are sorted by their level of exposure within each pre-strata. In the third step, the first participant from each pre-strata is placed in the first final strata. The second participant in each pre-strata is placed in the second final strata, and so on. In the orange, a single participant is followed. This participant has the lowest level of IV so is placed in pre-strata 1. They have the second lowest level of exposure within this pre-strata so are placed in the second final strata.



A more detailed exposition is given elsewhere.[25] This method is suggested (using simulation studies) by the authors to be insensitive to changes in the exposure to genotype association over strata, i.e. does not rely on the constant genetic effect assumption to generate unbiased estimates. Despite the development of the doubly-ranked method solving some of the issues evident in both empirical and simulation studies of the residual method, there remain a number of uncertainties about the reliability of both methods to provide unbiased causal estimates. In this paper we evaluate the residual and doubly-ranked methods using "negative control outcome" (NCO) analyses.[26] NCO analyses are analyses performed on control outcomes that should lead to a null result, and where identification of a non-null estimate suggests bias. For full details of our analytical approach, see the **Methods.**

Main:

Negative control outcomes as an assessment of the potential risk of bias

For our NCO analysis, we generated instrumental variables (IVs) for two exposures (Vitamin D and BMI) using a summed linear score across single nucleotide polymorphisms (SNPs) and performed linear and non-linear MR on participant age in the self-reported white British population in UK Biobank (details of analysis in the **Methods**). We chose these two exposures as they are both important exposures with markedly non-linear associations in observational data, and both have been studied using residual method. BMI was studied in a high profile paper in the *BMJ*,[6] and vitamin D in a paper in *Lancet Diabetes and Endocrinology*.[11] An editorial expression of concern has been posted about the latter paper.[27]

We chose age and sex as key negative control outcomes as both of these are unrelated to (autosomal) genotype so there should be a null estimate overall and for each stratum if methods are non-biased.[28] However, there are well established selection effects relating to BMI, age, and sex into studies like UK Biobank (see e.g.[29]). It is plausible that these selection effects are exacerbated in non-linear methods and could lead to bias.

In total, we included 351,005 participants in our analysis using Vitamin D as an exposure and 383,793 participants in our analysis on BMI.

Linear MR estimates are close to the null

In the whole cohort, our IV for Vitamin D did not strongly associate with either age or sex, as expected (difference in age (years) per 1 SD increase in PRS 0.01, 95% CI -0.02; 0.04, Odds ratio (OR) for being male per 1 SD increase in PRS 1.01 (95% CI 1.00;1.01), while our BMI IV was also not strongly associated with sex or age (difference in age -0.01, 95% CI -0.04 ;0.02, OR 1.00, 95% CI 1.00 ;1.01 for being male). As such, estimates from Mendelian randomization assuming a linear relationship between exposure and outcome were null.

Non-linear MR estimates differ substantially across strata

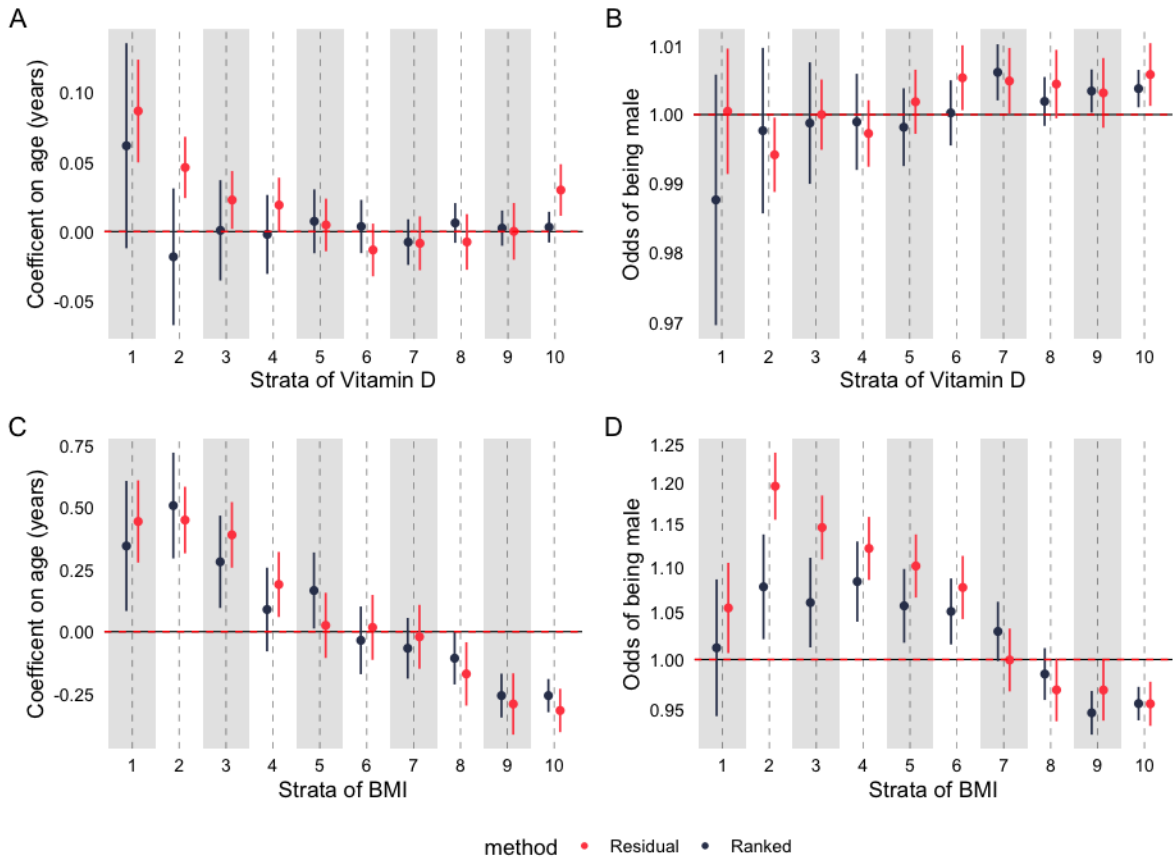
We then used both the residual and ranked method to generate stratum-specific estimates of the effect of each exposure on age and sex as univariable analysis. We chose to use ten strata, although results were similar using a differing number of strata.

In contrast to the null estimate across the whole cohort, estimates across each stratum were markedly different using either method (**Figure 2**). Focussing on the residual method, for the association with age, there were positive effect estimates in lower strata which decreased moving from lower to upper strata for the effect of Vitamin D on age, leading to a sign change. However, the uppermost strata had positive effect estimates. For BMI on age, we saw a similar effect, with positive effect estimates in lower strata and negative effect estimates in higher strata, but without the reversal to negative effect estimates in the top strata.

For the effect of BMI on sex, we saw positive estimates in lower strata (i.e. BMI odds increasing for being male in lower strata, then becoming smaller and eventually odds decreasing in upper strata), but the opposite effect with vitamin D (Vitamin D odds decreasing for being male in lower strata, then gradually becoming odds increasing in upper strata). Effect estimates were generally closer to the null for Vitamin D than for BMI.

For the ranked method, we identified largely null estimates for the effect of Vitamin D and on age. For the effect of Vitamin D on sex, effect estimates were similar between the residual and ranked method, with increasingly positive effect estimates when moving from lower to upper strata. For the effect of BMI on age, we saw similar estimates to the residual method, with positive estimates in lower strata and negative estimates in upper strata. For the effect of BMI on sex, we saw marked non-null estimates and quite different results for the doubly-ranked and residual method, with the ranked method having positive estimates in the middle strata and negative/null estimates at the lowest and upper strata.

Figure 2: The predicted causal effect of Vitamin D on male sex (A), and age (B), and BMI on age (C) and sex (D). Estimates were generated in each stratum using the residual method (red) and the ranked method (black) and were unadjusted for covariates.



Estimates were similar in analyses adjusted for covariates: age, (for sex as an outcome), sex (for age as an outcome) and the first 5 genetic principal components (**Supplementary Figure S1**). Log-transforming Vitamin D (as recommended by the creators of the residual method[24]) brought estimates from the residual model closer on average to the ranked model (**Supplementary Figure S2**), although these were still non-null across many strata.

For our BMI analysis, we then stratified by smoking status (as was performed in the previous non-linear MR analysis in BMI[6]) and re-ran analyses (**Supplementary Figure S3)**. Smoking has a known bidirectional relationship with BMI[30] and therefore has the potential to further induce collider bias. In these analyses, stratification on smoking made considerable difference to some strata-specific estimates, although the general shape of the association remained similar.

As a final comparator, we tested four more potential NCO's: birth weight, maternal smoking during pregnancy, breastfed as a baby, and birth location (latitude and longitude). All of these are very implausibly related to adult exposures, as they occur at or around the time of birth. These are

reported in the **Supplementary Results** but again show non-null estimates across some strata, with generally more non-null estimates for BMI as compared to Vitamin D, and for the residual method as compared to the ranked.

To summarise, using both the residual and doubly-ranked method we identified non-null, strata-specific associations between two exposures (Vitamin D and BMI) and a number of negative control outcomes across strata of the exposure in which the expected result is null.

228 Discussion:

229

230 Non-linear Mendelian randomization has become a commonly used approach in genetic
231 epidemiology. The original development was to consider a local average treatment effect (LATE),[9]
232 with other subsequent approaches developed.[5,8,10] One notable advancement was the development
233 of the residual method, which is now widely used.[5] Recent findings have identified that this method
234 can be biased[24] leading to estimates that are inconsistent with causal interpretation under certain
235 situations, leading to an updated 'doubly-ranked' method.[25]

236

237 In our analyses we identified that both methods showed bias in negative control outcome analyses,
238 with effect estimates on age and sex differing across generated strata, when estimates should be null
239 across all strata. It is worth considering some of the other potential reasons for the distorted
240 estimates in our negative control outcomes. One explanation is that these relate to a form of
241 selection bias into strata (e.g. certain strata are more likely to contain women, leading to subsequent
242 changes in estimates within strata).

243

244 This is particularly possible for body mass index, where selection bias (on entering studies, not on
245 entering strata) is known to alter estimates. For example, a large (~ 3.3 million person) GWAS of
246 chromosomal sex identified over 150 autosomal loci associated with male sex, including the body
247 mass increasing allele at *FTO* (OR 1.02, $P = 4.4 \times 10^{-36}$). Our findings may in part reflect the strong
248 associations between age, sex, and body mass index in the selected population of UK Biobank. This
249 would be supported by the weaker association with Vitamin D. Additionally, these results could
250 represent population stratification, which can be uncovered using negative controls.[26] Thirdly, these
251 could represent methods related issues that occur on generation of the strata unrelated to selection.
252 As we have demonstrated, stratification by both methods can lead to unusual findings and alter
253 estimates of IV-exposure associations substantially. Regardless of the reason for the bias, these
254 findings support more widespread use of negative control outcomes in Mendelian randomization
255 studies, alongside a close look at model assumptions particularly with respect to non-normal
256 exposures.

257

258 Our findings additionally suggest that we should be cautious when applying either residual or
259 doubly-ranked MR methods, as both methods seem potentially susceptible to bias in the generation
260 of strata specific estimates. Whether this relates simply to the example of body mass index and
261 Vitamin D that we describe, or occurs for other exposures, remains to be tested.

Limitations

This paper provides an initial assessment of the ranked and residual methods for performing non-linear MR. In particular, our analysis focussed on only two exposures (BMI and Vitamin D), a small number of outcomes and included only one dataset (UK Biobank), which we limited to White British participants only. It is possible that our findings do not extend to all exposures, or only occur in certain settings. Further analyses should examine these methods on diverse datasets, exposures, and outcomes.

Conclusion:

Using a negative control outcome approach, we identified bias in non-linear mendelian randomization estimates of the effect of BMI and Vitamin D using two commonly used methods. Researchers should be cautious in applying these methods at present.

282 Methods:

283

284 *Data source*

285

286 This study was performed in UK Biobank.[31] This study recruited around 500,000 participants between
287 2006-2010 from 22 sites across the UK. Participants were invited via post and had a range of
288 interviews on recruitment to record previous life events, demographics, and medical history.
289 Additionally, participants had blood samples taken for biochemical testing and had a range of
290 physical and anthropometric measures performed (e.g. body mass index, blood pressure).
291 Subsequently, participants had record linkage to electronic health record data for secondary care and
292 national death data (for >99% of participants).

293

294 We analysed data on the subset of participants who were minimally related and of European
295 ancestry (n = 385,290), defined by relationship analyses.[32] Specifically, we used the MRC Integrative
296 Epidemiology Unit genetically quality controlled data, described in detail elsewhere, excluding highly
297 related participants.[32]

298

299 *Exposures and data source*

300

301 This study focussed on Vitamin D (measured as nmol/L) and BMI exposures (measured as kg/m^2). as
302 Both of these were measured in participants on recruitment to UK Biobank. Details on measurement
303 are available with the companion paper[31] and via the UK Biobank website. Our study population for
304 each analysis included only participants who had these values recorded successfully on recruitment
305 to UK Biobank and had genetic data available. As BMI was recorded for more participants than
306 Vitamin D (due to assay failure, inability to bleed etc), the sample sizes differed for each exposure.

307

308 To identify variants associated with BMI, we used a large recent meta-analysis of BMI which did not
309 include any participants of UK Biobank.[33] We extracted all variants which were associated with BMI
310 ($p < 5 \times 10^{-8}$) and then clumped them by linkage disequilibrium to include only independent variants
311 ($r^2 < 0.001$, kb = 10,000). LD clumping was performed using the *TwoSampleMR* package in R.[34] In total
312 we included 68 variants (**Supplementary Table S2**). For Vitamin D, to match as close as possible to a
313 previously described non-linear MR analysis, we used the same set of variants that they used.[11]
314 These have previously identified to be associated with Vitamin D levels in meta-analysis from four
315 well characterised genomic regions known to be involved in Vitamin D metabolism (*GC, DHCR7,*

316 *CYP2R1, CYP24A1*).[35] In that analysis, 21 variants were used but 3 were unavailable in our data so in
317 total 18 variants were used to generate our polygenic risk score (**Supplementary Table S3**). The three
318 missing variants were all rare (minor allele frequency <0.005) in 1000 Genomes European subset)
319 and were therefore unlikely to substantially affect our instrument.
320
321 To generate an individual PRS for each participant we used PLINK v2.0.4, with the summary effect
322 generated by weighted allele scores, with each allele weighted by its effect on the exposure.
323
324 *Genotyping and imputation*
325
326 This analysis was performed on quality controlled genetic data held within the MRC Integrative
327 Epidemiology Unit. Full details of quality control are available elsewhere[32] but a summary is provided
328 here. The full data release for this study contains the cohort of successfully genotyped samples
329 (n=488,377). 49,979 individuals were genotyped using the UK BiLEVE array and 438,398 using the UK
330 Biobank axiom array. Pre-imputation QC, phasing and imputation are described elsewhere.[31] In brief,
331 prior to phasing, multiallelic SNPs or those with MAF ≤1% were removed. Phasing of genotype data
332 was performed using a modified version of the SHAPEIT2 algorithm.[36] Genotype imputation to a
333 reference set combining the UK10K haplotype and HRC reference panels 6was performed using
334 IMPUTE2 algorithms.[37] The analyses presented here were restricted to autosomal variants using a
335 graded filtering with varying imputation quality for different allele frequency ranges. Therefore, rarer
336 genetic variants are required to have a higher imputation INFO score (Info>0.3 for MAF >3%; Info>0.6
337 for MAF 1-3%; Info>0.8 for MAF 0.5-1%; Info>0.9 for MAF 0.1-0.5%) with MAF and Info scores having
338 been recalculated on an in-house derived 'European' subset. Data quality control Individuals with
339 sex-mismatch (derived by comparing genetic sex and reported sex) or individuals with sex
340 chromosome aneuploidy were excluded from the analysis (n=814).
341
342
343 *Degree of relatedness:*
344
345 Estimated kinship coefficients using the KING toolset identified 107,162 pairs of related individuals.
346 An inhouse algorithm was then applied to this list and preferentially removed the individuals related
347 to the greatest number of other individuals until no related pairs remain. These individuals were
348 excluded (n=79,448). Additionally, 2 individuals were removed due to them relating to a very large
349 number (>200) of individuals.

Outcomes and covariates

Our primary outcomes for this study were age and sex (negative control outcomes), and all-cause mortality. Secondary outcomes included latitude of birth, estimated birth weight, whether breastfed as a child and maternal smoking. All of these except mortality were extracted directly from UK Biobank and were recorded on recruitment. Mortality was extracted directly from death records provided by UK Biobank, which are in turn provided by national linked mortality data.

For maternal smoking and breastfeeding, participants who answered “Yes” and “No” were included, while participants with all other answers or no answer were deemed missing.

Covariates

We extracted covariates including UK Biobank recruitment centre, smoking status, and the first 5 genetic principal components from UK Biobank directly from the Researcher Access Platform. As results were similar with and without covariates, we perform all analyses unadjusted except where stated.

Negative control outcome analysis

To perform our negative control outcome analysis we first performed linear MR then used the *SUMnlmr* package in R to perform both the residual and ranked method on our individual level data.[7] Linear MR was performed using a two stage least squares (TSLS) approach using in built R functions. We tested the effect of vitamin D and BMI (individually) on age and sex, firstly without covariates (unadjusted), and then including either age (for the outcome of sex), sex (for the outcome of age), UK Biobank recruitment centre, and the first five principal components. Non-linear MR was performed using the standard settings in the *SUMnlmr* package, with a gaussian distribution for linear outcomes, and a binomial distribution for binary outcomes. We chose to use ten strata for most analyses; but performed sensitivity analyses using different numbers of strata – these did not fundamentally alter results.

In line with the authors of the doubly-ranked method recommendations[25], we performed multiple replicates (20) of each doubly-ranked approach and combined estimates using Rubin’s rule.

Ethics

This study was performed under the UK Biobank application number 81499. UK Biobank was ethically approved by the North West Multi-centre Research Ethics Committee (MREC).

Data and code availability

This analysis was performed in the UK Biobank. Data can be accessed via applicaton to the UK Biobank. We provide code used to perform our analysis in a supplement.

Funding

FH's time was funded by the GW4-CAT Wellcome Trust Doctoral Fellowship Scheme (222894/Z/21/Z). UK Biobank was funded by the Wellcome Trust, the Medical Research Council, the NIHR, and a variety of other charities (https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/about-us/our-funding).

References:

1. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?*. Int J Epidemiol [Internet] 2003 [cited 2023 May 22];32(1):1–22. Available from: https://academic.oup.com/ije/article/32/1/1/642797

2. Sanderson E, Glymour MM, Holmes MV, et al. Mendelian randomization. Nature Reviews Methods Primers [Internet] 2022 [cited 2023 May 22];2(1):1–21. Available from: https://www.nature.com/articles/s43586-021-00092-5

3. Sobczyk MK, Zheng J, Smith GD, Gaunt TR. Systematic comparison of Mendelian randomization studies and randomized controlled trials using electronic databases [Internet]. bioRxiv. 2022;Available from: https://www.medrxiv.org/content/10.1101/2022.04.11.22273633v2.abstract

4. Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. Int J Epidemiol [Internet] 2016;45(6):1866–86. Available from: http://dx.doi.org/10.1093/ije/dyw314

5. Staley JR, Burgess S. Semiparametric methods for estimation of a nonlinear exposure-outcome relationship using instrumental variables with application to Mendelian randomization. Genet Epidemiol [Internet] 2017;41(4):341–52. Available from: http://dx.doi.org/10.1002/gepi.22041

6. 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. BMJ [Internet] 2019;364:l1042. Available from: http://dx.doi.org/10.1136/bmj.l1042

7. Mason AM, Burgess S. Software Application Profile: SUMnlmr, an R package that facilitates flexible and reproducible non-linear Mendelian randomization analyses. Int J Epidemiol [Internet] 2022 [cited 2023 May 22];51(6):2014–9. Available from: https://academic.oup.com/ije/article/51/6/2014/6659062

8. Sulc J, Sjaarda J, Kutalik Z. Polynomial Mendelian randomization reveals non-linear causal effects for obesity-related traits. HGG Adv [Internet] 2022;3(3):100124. Available from: http://dx.doi.org/10.1016/j.xhgg.2022.100124

9. Silverwood RJ, Holmes MV, Dale CE, et al. Testing for non-linear causal effects using a binary genotype in a Mendelian randomization study: application to alcohol and cardiovascular traits. Int J Epidemiol [Internet] 2014;43(6):1781–90. Available from: http://dx.doi.org/10.1093/ije/dyu187

10. Burgess S, Davies NM, Thompson SG, EPIC-InterAct Consortium. Instrumental variable analysis with a nonlinear exposure-outcome relationship. Epidemiology [Internet] 2014;25(6):877–85. Available from: http://dx.doi.org/10.1097/EDE.0000000000000161

11. Emerging Risk Factors Collaboration/EPIC-CVD/Vitamin D Studies Collaboration. Estimating dose-response relationships for vitamin D with coronary heart disease, stroke, and all-cause mortality: observational and Mendelian randomisation analyses. Lancet Diabetes Endocrinol [Internet] 2021;9(12):837–46. Available from: http://dx.doi.org/10.1016/S2213-8587(21)00263-1

12. Rogne T, Solligård E, Burgess S, et al. Body mass index and risk of dying from a bloodstream infection: A Mendelian randomization study. PLoS Med [Internet] 2020;17(11):e1003413. Available from: http://dx.doi.org/10.1371/journal.pmed.1003413

13. Pinto Pereira SM, Garfield V, Norris T, et al. Linear and Non-linear associations between vitamin D and grip strength: a Mendelian Randomisation study in UK Biobank. J Gerontol A Biol Sci Med Sci [Internet] 2022;Available from: http://dx.doi.org/10.1093/gerona/glac255

14. Zhou A, Hyppönen E. Vitamin D deficiency and C-reactive protein: a bidirectional Mendelian randomization study. Int J Epidemiol [Internet] 2023;52(1):260–71. Available from: http://dx.doi.org/10.1093/ije/dyac087

15. Sutherland JP, Zhou A, Hyppönen E. Vitamin D Deficiency Increases Mortality Risk in the UK Biobank : A Nonlinear Mendelian Randomization Study. Ann Intern Med [Internet] 2022;175(11):1552–9. Available from: http://dx.doi.org/10.7326/M21-3324

16. Zhou A, Selvanayagam JB, Hyppönen E. Non-linear Mendelian randomization analyses support a role for vitamin D deficiency in cardiovascular disease risk. Eur Heart J [Internet] 2022;43(18):1731–9. Available from: http://dx.doi.org/10.1093/eurheartj/ehab809

17. Topiwala A, Taschler B, Ebmeier KP, et al. Alcohol consumption and telomere length: Mendelian randomization clarifies alcohol's effects. Mol Psychiatry [Internet] 2022;27(10):4001–8. Available from: http://dx.doi.org/10.1038/s41380-022-01690-9

18. Gaziano L, Sun L, Arnold M, et al. Mild-to-Moderate Kidney Dysfunction and Cardiovascular Disease: Observational and Mendelian Randomization Analyses. Circulation [Internet]

2022;146(20):1507–17. Available from: http://dx.doi.org/10.1161/CIRCULATIONAHA.122.060700

19. Moksnes MR, Graham SE, Wu K-H, et al. Genome-wide meta-analysis of iron status biomarkers and the effect of iron on all-cause mortality in HUNT. Commun Biol [Internet] 2022;5(1):591. Available from: http://dx.doi.org/10.1038/s42003-022-03529-z

20. Biddinger KJ, Emdin CA, Haas ME, et al. Association of Habitual Alcohol Intake With Risk of Cardiovascular Disease. JAMA Netw Open [Internet] 2022;5(3):e223849. Available from: http://dx.doi.org/10.1001/jamanetworkopen.2022.3849

21. Munafò MR, Tilling K, Taylor AE, Evans DM, Davey Smith G. Collider scope: when selection bias can substantially influence observed associations. Int J Epidemiol [Internet] 2018;47(1):226–35. Available from: http://dx.doi.org/10.1093/ije/dyx206

22. Davey Smith G. Mendelian randomisation and vitamin D: the importance of model assumptions [Internet]. Lancet Diabetes Endocrinol. 2023;11(1):14. Available from: http://dx.doi.org/10.1016/S2213-8587(22)00345-X

23. Navale SS, Mulugeta A, Zhou A, Llewellyn DJ, Hyppönen E. Vitamin D and brain health: an observational and Mendelian randomization study. Am J Clin Nutr [Internet] 2022;116(2):531–40. Available from: http://dx.doi.org/10.1093/ajcn/nqac107

24. Burgess S. Violation of the constant genetic effect assumption can result in biased estimates for non-linear Mendelian randomization [Internet]. bioRxiv. 2022 [cited 2022 Nov 1];2022.10.26.22280570. Available from: https://www.medrxiv.org/content/10.1101/2022.10.26.22280570v1.abstract

25. Tian H, Mason AM, Liu C, Burgess S. Relaxing parametric assumptions for non-linear Mendelian randomization using a doubly-ranked stratification method. PLoS Genet [Internet] 2023 [cited 2023 May 22];19(6):e1010823. Available from: http://dx.doi.org/10.1371/journal.pgen.1010823

26. Sanderson E, Richardson TG, Hemani G, Davey Smith G. The use of negative control outcomes in Mendelian randomization to detect potential population stratification. Int J Epidemiol [Internet] 2021;50(4):1350–61. Available from: http://dx.doi.org/10.1093/ije/dyaa288

27. Editors of The Lancet Diabetes & Endocrinology. Expression of Concern-Estimating dose-response relationships for vitamin D with coronary heart disease, stroke, and all-cause mortality: observational and Mendelian randomisation analyses. Lancet Diabetes Endocrinol [Internet] 2023;Available from: http://dx.doi.org/10.1016/S2213-8587(23)00198-5

28. Sanderson E, Macdonald-Wallis C, Davey Smith G. Negative control exposure studies in the presence of measurement error: implications for attempted effect estimate calibration. Int J Epidemiol [Internet] 2018;47(2):587–96. Available from: http://dx.doi.org/10.1093/ije/dyx213

29. Pirastu N, Cordioli M, Nandakumar P, et al. Genetic analyses identify widespread sex-differential participation bias. Nat Genet [Internet] 2021;53(5):663–71. Available from: http://dx.doi.org/10.1038/s41588-021-00846-7

30. Carreras-Torres R, Johansson M, Haycock PC, et al. Role of obesity in smoking behaviour: Mendelian randomisation study in UK Biobank. BMJ [Internet] 2018 [cited 2023 Apr 26];361:k1767. Available from: http://dx.doi.org/10.1136/bmj.k1767

31. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature [Internet] 2018;562(7726):203–9. Available from: http://dx.doi.org/10.1038/s41586-018-0579-z

32. Mitchell R, Hemani G, Dudding T, Corbin L, Harrison S, Paternoster L. UK Biobank genetic data: MRC-IEU quality control, version 2 [Internet]. 2019 [cited 2021 Dec 21];Available from: https://data.bris.ac.uk/data/dataset/1ovaau5sxunp2cv8rcy88688v

33. Locke AE, Kahali B, Berndt SI, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature [Internet] 2015 [cited 2022 Apr 28];518(7538):197–206. Available from: http://dx.doi.org/10.1038/nature14177

34. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. Elife [Internet] 2018;7. Available from: http://dx.doi.org/10.7554/eLife.34408

35. Wang TJ, Zhang F, Richards JB, et al. Common genetic determinants of vitamin D insufficiency: a genome-wide association study [Internet]. Lancet. 2010;376(9736):180–8. Available from: http://dx.doi.org/10.1016/S0140-6736(10)60588-0

36. Choi Y, Chan AP, Kirkness E, Telenti A, Schork NJ. Comparison of phasing strategies for whole human genomes. PLoS Genet [Internet] 2018;14(4):e1007308. Available from: http://dx.doi.org/10.1371/journal.pgen.1007308

37. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. G3 [Internet] 2011;1(6):457–70. Available from: http://dx.doi.org/10.1534/g3.111.001198