

1 Non-linear mendelian randomization: detection of biases using  
2 negative controls with a focus on BMI, Vitamin D and LDL  
3 cholesterol.

4  
5

6 Fergus W Hamilton<sup>1,2</sup>, David A Hughes<sup>1</sup>, Wes Spiller<sup>1</sup>, Kate Tilling<sup>1</sup>, George Davey Smith<sup>1</sup>

7

8 1. MRC Integrative Epidemiology Unit, University of Bristol

9 2. Infection Science, North Bristol NHS Trust

10

11

12

13

14 Corresponding author:

15

16 Fergus Hamilton

17 MRC Integrative Epidemiology Unit

18 University of Bristol

19 Oakfield House

20 Oakfield Road

21 BS8 2PS

22

23 [Fergus.hamilton@bristol.ac.uk](mailto:Fergus.hamilton@bristol.ac.uk)

24 +447743165499

25

26 Abstract (299 words)

27

28 Mendelian randomisation (MR) is an established technique in epidemiological investigation,  
29 using the principle of random allocation of genetic variants at conception to estimate the causal  
30 linear effect of an exposure on an outcome. Extensions to this technique include non-linear  
31 approaches that allow for differential effects of the exposure on the outcome depending on the  
32 level of the exposure. A widely used non-linear method is the residual approach, which  
33 estimates the causal effect within different strata of the non-genetically predicted exposure (i.e.  
34 the "residual" exposure). These "local" causal estimates are then used to make inferences  
35 about non-linear effects. Recent work has identified that this method can lead to estimates that  
36 are seriously biased, and a new method - the doubly-ranked method – has been introduced as a  
37 possibly more robust approach. In this paper, we perform negative control outcome analyses in  
38 the MR context. These are analyses with outcomes onto which the exposure should have no  
39 predicted causal effect. Using both methods we find clearly biased estimates in certain  
40 situations. We additionally examined a situation for which there are robust randomised  
41 controlled trial estimates of effects - that of low density lipoprotein cholesterol (LDL-C)  
42 reduction onto myocardial infarction, where randomised trials have provided strong evidence of  
43 the shape of the relationship. The doubly-ranked method did not identify the same shape as the  
44 trial data, and for LDL-C and other lipids they generated some highly implausible findings.  
45 Therefore, we suggest that until there is extensive simulation and empirical methodological  
46 work demonstrating that these methods generally produce meaningful findings use of them is  
47 suspended. If authors feel it is imperative that they report results from them there should be  
48 strong justification for this, and a number of sanity checks (such as analysis of negative and  
49 positive control outcomes) should be provided.

50

51

52 **Keywords:** Mendelian randomisation; epidemiology; non-linear; Vitamin D; Body Mass Index;  
53 **LDL-C; instrumental variable**

54

55

56

57

58

59



61 Introduction:

62

63 Mendelian randomization (MR) is now a widely used epidemiological technique contributing to  
64 improved causal inference in settings in which unmeasured confounding may bias estimates  
65 [1–3]. MR findings have been consistent with randomised trial data in many settings [4], and  
66 have assisted causal inference as a component of triangulation of evidence [5,6] Recently,  
67 there have been developments in MR in which attempts are made to extend causal inference  
68 from a whole population (e.g. the overall effect of increasing vitamin D on all-cause mortality),  
69 to strata of that population (e.g. only those with low levels of vitamin D) [7–11]. This has the  
70 potential to answer important questions around the linearity of exposure-outcome effects. The  
71 development of the “residual” method led to MR studies that purported to show striking non-  
72 linear effects [12–18], in particular with respect to vitamin D and BMI as exposures. However,  
73 recent analyses have suggested that this method may be susceptible to serious bias, including  
74 that likely related to amplification of selection bias within strata [3,19–21]. In this manuscript,  
75 we explore a number of issues that suggest caution in using non-linear approaches at present.

76

77 *Development of the residual method and concerns about inconsistent causal estimates*

78

79 In the residual method the measured levels of the exposure (e.g. Vitamin D) are regressed onto  
80 the genetic instrument (e.g. a polygenic risk score for Vitamin D). The residuals of that  
81 regression represent the IV-free exposure (sometimes referred to as the “non-genetic” portion of  
82 Vitamin D). These residuals are then split into strata (often 10, or 100 strata) ordered by the  
83 mean IV-free exposure value. Therefore, the sample in stratum 1 have the lowest level of “IV-  
84 free” Vitamin D, and the sample in stratum 10 (or 100) have the highest value. MR is then run  
85 within each stratum, and estimates are combined by parametric statistical methods (such as  
86 fractional polynomials) to estimate a non-linear curve. The requirement for stratification based  
87 on the residuals rather than the exposure itself is driven by the desire to avoid collider bias [22]  
88 as this would be induced by stratification on the exposure itself.

89

90 There has been a recent focus on the performance of this method. In particular, one study  
91 identified a non-linear predicted causal effect of Vitamin D on a number of outcomes,  
92 suggesting that increasing Vitamin D in those who were below the median level might produce  
93 considerable benefit (as opposed to the null overall causal effect) [23]. In some analyses,  
94 despite a precisely estimated null or harmful overall effect, every strata had a protective effect,

95 with some effects substantial. Following the journal and the authors being alerted to this [19]  
96 the authors referred to it “as a logical impossibility if all estimates have a causal  
97 interpretation”[23] after the journal retracted the paper[24]. Several other analyses applying  
98 identical methods to the same exposure, with overlapping data and outcomes, have,  
99 unsurprisingly produced similar and likely equally spurious findings [15,16].

100

101 Following the documentation that the findings were erroneous the authors of the initial non-  
102 linear Vitamin D analysis [23] re-evaluated the residual method used in the study. In a recent  
103 paper [25] they showed via simulation studies that the residual method was biased under  
104 certain situations, in part relating to the “constant genetic effect” assumption – that the  
105 genotype-exposure effect is linear and thus constant across strata (i.e. the coefficient for IV-  
106 exposure remains the same across strata). For Vitamin D, the constant genetic effect  
107 assumption was shown not to be satisfied, with much larger genotype-exposure effect  
108 estimates in top strata than in the bottom strata [25]. They also noted that this heterogeneity in  
109 IV-exposure association was not identified in the residual method, and suggest that this is a  
110 flaw in the residual approach. It is not clear why this would have led to the J-shaped curve  
111 observed in the residual method, however. To address this challenge a different way of  
112 stratifying the exposure was developed, called the “doubly-ranked” method [11].

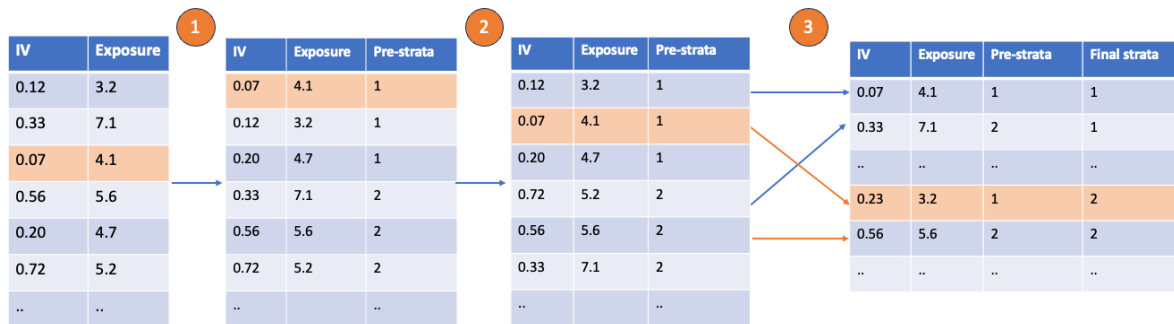
113

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

124

125 **Figure 1:** A worked example of strata generation in the doubly-ranked approach. In the first step,  
126 all participants are sorted by their level of the instrumental variable (IV) into pre-strata. In the  
127 second step, participants are sorted by their level of exposure within each pre-strata. In the third  
128 step, the first participant from each pre-strata is placed in the first final strata. The second

129 participant in each pre-strata is placed in the second final strata, and so on. In the orange, a  
 130 single participant is followed. This participant has the lowest level of IV so is placed in pre-strata  
 131 1. They have the second lowest level of exposure within this pre-stratum so are placed in the  
 132 second final stratum.  
 133



134  
 135

136 A more detailed exposition is given elsewhere [11]. The doubly-ranked method is suggested  
 137 (using simulation studies) by the authors to be insensitive to changes in the exposure to  
 138 genotype association over strata, i.e. does not rely on the constant genetic effect assumption to  
 139 generate unbiased estimates. The retracted non-linear vitamin D paper[18] – which claimed  
 140 substantial non-linearity in the causal effect of vitamin D – has been replaced by a revised paper  
 141 applying the doubly-ranked approach which suggests no meaningful non-linearity [23,24].  
 142

143 Despite the development of the doubly-ranked method solving some of the issues evident in  
 144 both empirical and simulation studies of the residual method, there remain a number of  
 145 uncertainties about the reliability of both methods to provide unbiased causal estimates. In this  
 146 paper we evaluate the residual and doubly-ranked methods using “negative control outcome  
 147 (NCO) analyses [5,6,26] in UK Biobank, a large and widely researched cohort study. NCO  
 148 analyses are analyses performed on control outcomes that should lead to a null result, and  
 149 where identification of a non-null estimate suggests bias. We additionally examined a situation  
 150 for which there are robust randomised controlled trial estimates of effects - that of the effect of  
 151 low density lipoprotein cholesterol (LDL-C) reduction on myocardial infarction, where  
 152 randomised trials have provided strong evidence of the shape of the relationship [27,28]. We  
 153 compared these estimates from trials with those generated by both non-linear MR methods. We  
 154 present our findings in the context of a recent preprint from the principle developer of the  
 155 doubly-ranked method, and draw attention to highly implausible findings in this report [29].

156 Materials and Methods:

157

158 *Data source*

159

160 This study was performed in UK Biobank [30]. UK Biobank recruited around 500,000 participants  
161 between 2006-2010 from 22 sites across the UK. Participants were invited via post and had a  
162 range of interviews on recruitment to record previous life events, demographics, and medical  
163 history. Additionally, participants had blood samples taken for biochemical testing and had a  
164 range of physical and anthropometric measures performed (e.g. body mass index, blood  
165 pressure). Subsequently, participants had record linkage to electronic health record data for  
166 secondary care and national death data (for >99% of participants).

167

168 We analysed data on the subset of participants who were minimally related and of White British  
169 ancestry (n = 385,290), defined by relationship analyses [31]. Specifically, we used the MRC  
170 Integrative Epidemiology Unit genetically quality controlled data, described in detail elsewhere,  
171 excluding highly related participants [31].

172

173 *Exposures and data source*

174

175 This study focussed on Vitamin D (measured as nmol/L) BMI (measured as kg/m<sup>2</sup>), and LDL-C  
176 and triglycerides (TG) (measured as mmol/L) as exposures, as all of these were measured in  
177 participants on recruitment to UK Biobank, and have been widely used in non-linear MR  
178 analyses [17,23,29,32]. Details on measurement are available via the UK Biobank website. Our  
179 study sample for each analysis included only participants who had these values recorded  
180 successfully on recruitment to UK Biobank and had genetic data available. Sample sizes  
181 differed slightly due to missing values for each exposure (e.g. due to assay failure, inability to  
182 give blood, etc). For triglycerides, we also extracted data from a subsequent visit, around 2  
183 years later, for around 15,000 participants.

184

185 We then identified genetic variants to use in MR for all exposures. To identify variants associated  
186 with BMI, we used a large recent meta-analysis of BMI which did not include any participants of  
187 UK Biobank [33]. We extracted all variants which were associated with BMI ( $p < 5 \times 10^{-8}$ ) and then  
188 clumped them by linkage disequilibrium to include only independent variants ( $r^2 < 0.001$ , kb =  
189 10,000). LD clumping was performed using the *TwoSampleMR* package in R and was based on

190 LD from the European ancestry participants from the 1000 Genome's project [34]. In total we  
191 included 68 variants (**Supplementary Table S1**). For Vitamin D, to match as close as possible to  
192 a previously described non-linear MR analysis, we used the same set of variants that they used  
193 [23]. These have previously identified to be associated with Vitamin D levels in meta-analysis  
194 from four well characterised genomic regions known to be involved in Vitamin D metabolism  
195 (*GC, DHCR7, CYP2R1, CYP24A1*)[35]. In that analysis, 21 variants were used but 3 were  
196 unavailable in our data so in total 18 variants were used to generate our polygenic risk score  
197 (**Supplementary Table S2**). The three missing variants were all rare (minor allele frequency  
198 <0.005) in 1000 Genomes European subset) and were therefore unlikely to substantially affect  
199 our instrument. For LDL-C and TG we used summary statistics from the most recent analysis of  
200 the Global Lipids Genetics Consortium (GLGC) excluding UK Biobank participants. Again, we  
201 took independent variants ( $p < 5 \times 10^{-8}$ ,  $r^2 < 0.001$ , kb = 10,000) from across the genome. In total  
202 we included 296 variants for LDL-C and 366 variants for TG (**Supplementary Table S3**).

203

204 To generate an individual PRS for each participant we used PLINK v2.0.4 [36], with the summary  
205 effect generated by summed weighted allele scores, with each allele weighted by its effect on  
206 the exposure.

207

### 208 *Genotyping and imputation*

209

210 This analysis was performed on quality controlled genetic data held within the MRC Integrative  
211 Epidemiology Unit. Full details of quality control are available elsewhere [31] but a summary is  
212 provided here. The full data release for this study contains the cohort of successfully genotyped  
213 samples (n=488,377). 49,979 individuals were genotyped using the UK BiLEVE array and  
214 438,398 using the UK Biobank axion array. Pre-imputation QC, phasing and imputation are  
215 described elsewhere [30]. In brief, prior to phasing, multiallelic SNPs or those with  $MAF \leq 1\%$   
216 were removed. Phasing of genotype data was performed using a modified version of the  
217 SHAPEIT2 algorithm [37]. Genotype imputation to a reference set combining the UK10K  
218 haplotype and HRC reference panels was performed using IMPUTE2 algorithms [38]. The  
219 analyses presented here were restricted to autosomal variants using a graded filtering with  
220 varying imputation quality for different allele frequency ranges. Therefore, rarer genetic variants  
221 are required to have a higher imputation INFO score (INFO>0.3 for  $MAF > 3\%$ ; INFO>0.6 for  $MAF$   
222  $1-3\%$ ; INFO>0.8 for  $MAF 0.5-1\%$ ; INFO>0.9 for  $MAF 0.1-0.5\%$ ) with  $MAF$  and Info scores having  
223 been recalculated on an in-house derived 'European' subset. Data quality control Individuals



224 with sex-mismatch (derived by comparing genetic sex and reported sex) or individuals with sex  
225 chromosome aneuploidy were excluded from the analysis (n=814).

226

227

228 *Degree of relatedness:*

229

230 Estimated kinship coefficients using the KING toolset identified 107,162 pairs of related  
231 individuals. An inhouse algorithm was then applied to this list and preferentially removed the  
232 individuals related to the greatest number of other individuals until no related pairs remain.  
233 These individuals were excluded (n=79,448). Additionally, 2 individuals were removed due to  
234 them relating to a very large number (>200) of individuals.

235

236 *Outcomes and covariates*

237

238 Our primary outcomes for this study were age and sex for BMI and Vitamin D exposures, and  
239 myocardial infarction when LDL-C was the exposure. We chose age and sex as key negative  
240 control outcomes as these cannot be caused by the exposures in question and any non-null  
241 estimates in any strata suggest the analyses are biased [39]. However, there are well  
242 established selection effects relating to BMI, age, and sex into studies like UK Biobank (see e.g.  
243 [40]). It is plausible that these selection effects are amplified by non-linear methods and could  
244 lead to bias. This still, of course, means the apparent findings are unreliable.

245

246 Myocardial infarction was extracted from the UK Biobank algorithmic definition (Data Showcase  
247 field 42000). MI outcomes were defined as prevalent (occurring before study enrolment),  
248 incident (occurring after study enrolment), or both (including both incident and prevalent  
249 cases).

250

251 We extracted covariates including UK Biobank recruitment centre, smoking status (coded as  
252 current or never/ex) and the first 5 genetic principal components from UK Biobank directly.  
253 Smoking status was included for some sensitivity analyses given the known bidirectional  
254 association between smoking status and BMI [41].

255

256 *Negative control outcome analysis – BMI and Vitamin D*

257

258 To perform our negative control outcome analysis we first performed conventional MR then  
259 used the *SUMnlmr* package in R to perform both the residual and doubly-ranked method on our  
260 individual level data [42]. Conventional MR was performed using a two stage residual inclusion  
261 approach in the *OneSampleMR* R package [43], using both the whole cohort and then in strata  
262 of age and sex. We first tested the effect of vitamin D and BMI (individually) on age and sex,  
263 firstly without covariates (unadjusted), and then age (when sex was an outcome), sex (when age  
264 was an outcome), and also including UK Biobank recruitment centre, and the first genetic five  
265 principal components. Non-linear MR was performed using the standard settings in the  
266 *SUMnlmr* package, with a Gaussian distribution for linear outcomes, and a binomial distribution  
267 for binary outcomes. We chose to use ten strata for most analyses; but performed sensitivity  
268 analyses using different numbers of strata.

269  
270 For BMI, we performed an analysis stratified by smoking status as this had been performed as a  
271 key analysis in an as yet uncorrected high-profile paper using non-linear MR [17] As BMI has a  
272 bidirectional association with smoking, this could induce collider bias, but we repeated these  
273 analyses to determine if estimates were more biased when stratifying by smoking status.

274  
275 *LDL cholesterol on myocardial infarction*

276  
277 In a further analysis, we focussed on the role of LDL-C as an exposure and myocardial infarction  
278 as an outcome. We provide the background to this positive control analysis in **Box 1**.

279  
280 We performed two analyses. First, for completeness we replicated the negative control analysis  
281 discussed above (with age and sex as outcomes) using LDL-C as an exposure. Subsequently, we  
282 performed NLMR analyses on the effect of LDL-C on a) incident, b) prevalent, and c) all  
283 myocardial infarction to identify if effects matched randomised trial data.

284  
285 As non-linear MR results could be altered by statins, which have a large effect on LDL-C levels,  
286 we performed sensitivity analyses in a cohort including only statin users, only non-statin users,  
287 and those under the age of 50 (where statin use was rare (~5%) and therefore not likely to bias  
288 estimates). Finally, we also fitted a model adjusting for statins, age, sex, and the first five genetic  
289 principal components. We recognise that the approach of analysing only those who use/do not  
290 use statins or adjusting for statins generates collider bias[22] (as statin use is almost entirely  
291 downstream of LDL-C levels) but we included this for completeness.

292 **Box 1:** Purpose of positive control analysis and background to the relationship between LDL-C  
293 and MI

The purpose of positive control analyses is to perform an analysis whereby the true result is already known, and a comparison can be made to the results of a presented analysis and this truth. In this paper, we chose the effect of LDL-C on myocardial infarction.

We chose this example because this is a well-studied exposure outcome relationship, where conventional MR estimates match in direction to randomised trial estimates, although MR estimates are around 40% larger, presumably due to the lifelong effect of lipid reduction conferred by genotype compared to that of drugs [1,44,45]. In addition there are a large number of randomised trials with differing levels of baseline LDL-C [27]. These individual trial estimates have been meta-regressed to identify if the effectiveness of LDL-C reduction depends on the baseline LDL-C [27,28], so we can calculate the shape of the LDL-C and MI relationship.

In 2010, a meta-analysis of 26 trials (largely statin vs control) identified no clear evidence of the effect of baseline LDL-C on the effectiveness of LDL-C therapy (on all major vascular events), with a summary OR of 0.78 (95% CI 0.76-0.80) per 1mmol/L reduction in LDL-C [28]. In a more recent 2018 meta-analysis of 34 trials which reported specific LDL-C effects on myocardial infarction [27], the overall OR per 1 mmol/L LDL-C reduction was similar (0.76 95% CI 0.72-0.80), but in meta-regression they identified a small decrease in rate ratio (RR) of 0.90 (95% CI 0.84-0.97) with each 1 mmol/L increase in LDL-C (i.e. the effectiveness of therapy was greater in those with higher LDL-C levels). In those with an LDL-C of more than 4mmol/L, the OR per 1mmol/L reduction was 0.64 (95% 0.53-0.78), while in those trials with a baseline LDL-C of less than 2.5mmol/L, the OR was 0.85 (95% CI 0.76-0.92). The summed data therefore suggests a mostly linear effect with a moderate increase in effectiveness in those with *higher* LDL-C.

Given the number of trials across different settings and the overall sample size, we can be reasonably confident of the true shape of this relationship and can compare this estimate with that generated by other methods (e.g. non-linear MR).

294

295 *Triglycerides and cancer mortality*

296

297 Finally, given a recent pre-print identifying biologically highly implausible results, we aimed to  
298 replicate an analysis of the effect of triglycerides on cancer mortality reported in the above  
299 preprint[29]. We used the same definitions of cancer mortality and performed non-linear MR  
300 adjusting for age, sex, age \* sex, age\*age, age\*age\*sex, and the first ten principal components.  
301 This choice of covariates was to match the prior analysis. We also – using the repeat sampling  
302 data available for around 15,000 participants – assessed the stability of strata for triglycerides,  
303 as these are known to be highly fluctuant. To do this we simply performed the doubly-ranked  
304 method at each time point (initial visit, repeat visit), and showed whether participants remained  
305 in the same strata using an alluvial plot.

306

307 In line with the authors of the doubly-ranked method recommendations [11] we performed  
308 multiple replicates (20) of each doubly-ranked approach and combined estimates using Rubin’s  
309 rule.

310

### 311 *Ethics*

312

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

315

### 316 *Data and code availability*

317

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

320

321

322 Results

323

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

325

326 For our negative control outcome analysis, we generated instrumental variables (IVs) for two  
327 exposures (Vitamin D and BMI) using a summed linear score across single nucleotide  
328 polymorphisms (SNPs) and performed conventional and non-linear MR on participant age in the  
329 self-reported white British population in UK Biobank.

330

331 In total, we included 351,005 participants in our analysis using Vitamin D as an exposure and  
332 383,793 participants in our analysis on BMI. Demographics of the population are in

333 **Supplementary Table S4.**

334

335 *Conventional MR estimates are close to the null*

336

337 Conventional MR estimates of the effect of Vitamin D on age (beta per nmol/L: 0.0028; 95% CI -  
338 0.004; 0.009, p = 0.39) and sex (OR per nmol/L 1.002; 95% CI 1.000 - 1.003, p = 0.06) were close  
339 to the null. MRs estimate of the effect of BMI on age (beta per kg/m<sup>2</sup>: -0.019, 95% CI -0.060 -  
340 0.023, p = 0.37) and sex (OR per kg/m<sup>2</sup> 1.01; 95% CI 1.00 - 1.02, p = 0.04) were also close to the  
341 null.

342

343 We also calculated the effect of BMI and Vitamin D on age in males and females separately, and  
344 the effect of BMI and Vitamin D on sex in deciles of age. These are reported in **Supplementary**  
345 **Table S5.** Estimates across subgroups were similar – accepting the play of chance – across  
346 these analyses.

347

348 *Non-linear MR estimates differ substantially across strata*

349

350 We then used both the residual and doubly-ranked method to generate stratum-specific  
351 estimates of the effect of each exposure on age and sex in univariable analyses. We chose to  
352 use ten strata, although results were similar using a differing number of strata (data not shown).

353

354 In contrast to the null estimate across the whole cohort, estimates across each stratum were  
355 markedly different using either method (**Figure 2**). Focussing on the residual method, for the

356 predicted effect of Vitamin D on age (**Figure 2A**), there were positive effect estimates in lower  
357 strata which decreased moving from lower to upper strata for the effect of Vitamin D on age,  
358 leading to null effects. However, the uppermost strata had a positive effect estimate. For BMI on  
359 age (**Figure 2C**), we saw a similar trend, with positive effect estimates in lower strata and  
360 negative effect estimates in higher strata.

361

362 For the effect of BMI on sex (**Figure 2D**), we saw positive estimates in lower strata and negative  
363 effects in upper strata, in line with a previous similar analysis we have performed, although that  
364 analyses used different parameters including a different instrumental variable[21]. In short the  
365 estimates would suggest that BMI increased the odds for being male in those with the lowest  
366 non-genetic BMI and increased the odds of being female in upper strata. For the effect of  
367 Vitamin D on sex (**Figure 2B**) we observed null effects in lower strata but modest positive effects  
368 in upper strata suggesting that vitamin D increases the odds of being male in those with the  
369 highest non-genetic BMI.

370

371 For the doubly-ranked method, we identified largely null estimates for the effect of Vitamin D on  
372 age (**Figure 2A**). For the effect of Vitamin D on sex (**Figure 2B**), effect estimates were similar  
373 between the residual and doubly-ranked method, with increasingly positive effect estimates  
374 when moving from lower to upper strata. For the effect of BMI on age (**Figure 2C**), we saw similar  
375 estimates to the residual method, with positive estimates in lower strata and negative estimates  
376 in upper strata. For the effect of BMI on sex (**Figure 2D**), we saw marked non-null estimates with  
377 a similar trend but, in general, estimates closer to the null for the doubly-ranked relative to the  
378 residual method. We calculated Cochran's Q to formally assess the heterogeneity of strata  
379 specific estimates. These results are reported in **Supplementary Table S6**. There was evidence  
380 – extremely strong in many cases - of strata specific estimate heterogeneity across all negative  
381 control outcome analyses except the analysis of Vitamin D on age and sex using the doubly-  
382 ranked method.

383

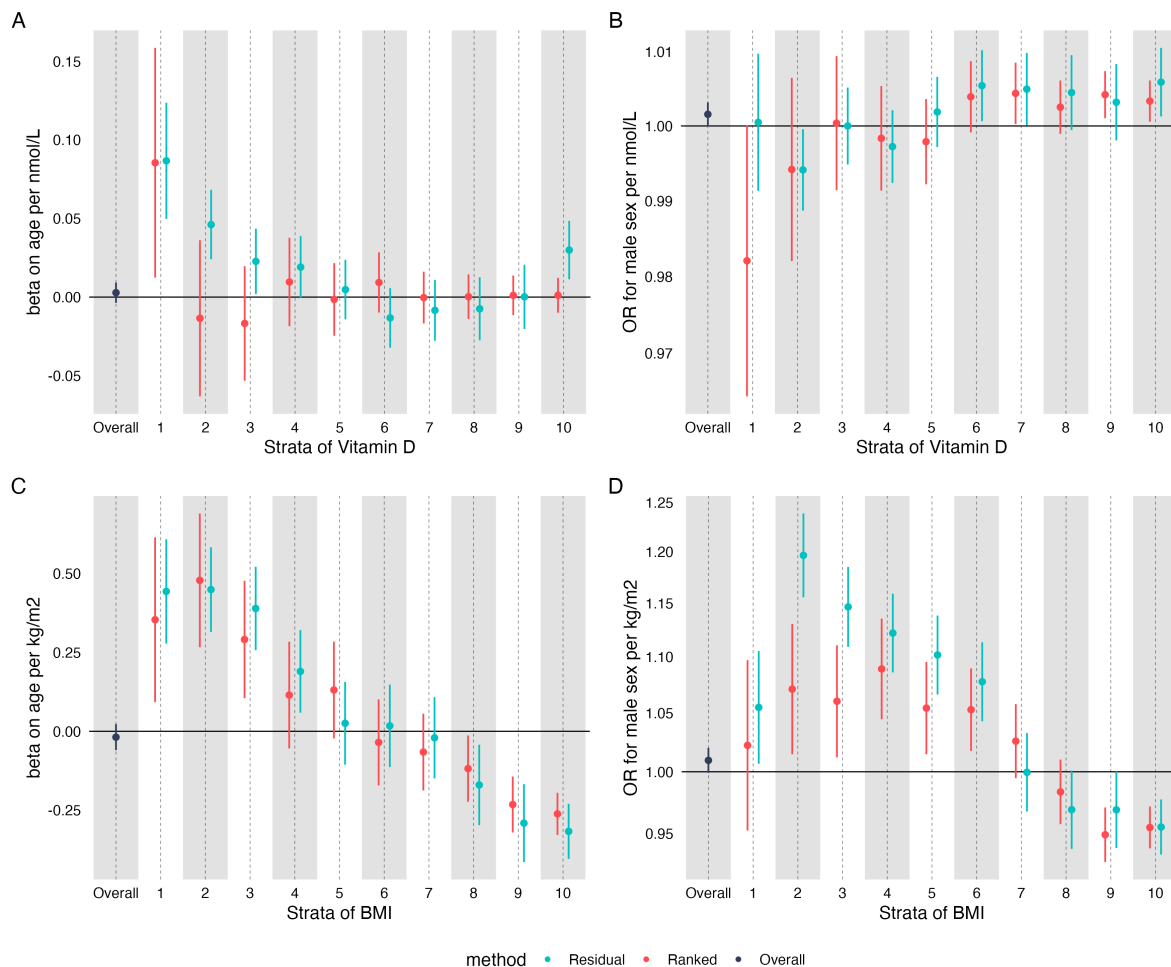
384

385

386

387 **Figure 2:** The predicted causal effect of Vitamin D on age (A), sex (B) and BMI on age (C) and sex  
388 (D). Estimates were generated in each stratum using the residual method (blue) and the doubly-

389 ranked method (red) and were unadjusted for covariates. The black estimate represents the  
 390 conventional MR estimate generated using the whole cohort.  
 391



392  
 393  
 394  
 395 Estimates were similar in analyses adjusted for covariates: age (for sex as an outcome), sex (for  
 396 age as an outcome) and the first 5 genetic principal components (**Supplementary Figure S1**).  
 397 Log-transforming Vitamin D (as recommended by the creators of the residual method) brought  
 398 estimates from the residual model closer on average to the doubly-ranked model  
 399 (**Supplementary Figure S2**), although these were still non-null across many strata.  
 400  
 401 For our BMI analysis, we then stratified by smoking status (as was performed in a previous non-  
 402 linear MR analysis in BMI, producing the headline findings from the paper[17]) and re-ran  
 403 analyses (**Supplementary Figure S3**). Smoking had been shown to have a bidirectional  
 404 relationship with BMI[41] before the non-linear *BMJ* MR paper was published[17], and therefore

405 it should have been clear that this could have produced collider bias [22]. In these analyses,  
406 stratification on smoking made considerable difference to some strata-specific estimates,  
407 although the general shape of the association remained similar.

408

409

410

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

415

416 *LDL cholesterol and myocardial infarction*

417

418

419 To examine both methods in a scenario where we anticipate the shape of the causal  
420 relationship we examined the association between LDL-C and a key outcome: myocardial  
421 infarction. RCT data from >30 trials have demonstrated strong and broadly linear effects of LDL-  
422 C reduction on both outcomes with meta-analyses identifying slightly larger effect estimates in  
423 those with higher levels of LDL-C at baseline, while a recent MR analysis identified the opposite  
424 effect (increased effectiveness of LDL-C reduction in those with lower LDL-C[29], see **Box 1** for  
425 further background).

426

427 First, as with analyses we presented above, we performed negative control outcome analyses of  
428 the effect of increasing LDL-C on age and sex. In conventional MR, estimates were essentially  
429 null for age (beta on age per 1mmol/L increase in LDL -0.10; 95% CI -0.23, 0.03, p = 0.15), but  
430 suggested increased LDL-C was 'causal' for being more female (OR 0.95; 95% CI 0.92-0.98, p =  
431 0.004). Conventional MR estimates of the effect of LDL-C on age were similar in men and  
432 women (**Supplementary Table S6**), although there was some evidence of heterogeneity of MR  
433 estimates of LDL-C on sex across strata, p value for heterogeneity 0.03).

434

435 Compared to conventional MR estimates, divergence from the null was much greater in non-  
436 linear MR (**Figure 3**), with estimates on age as high as 5.74 (95% CI 1.79 - 9.63) years per  
437 1mmol/L LDL-C increase in one strata from the doubly-ranked method, a ~50 fold increase in

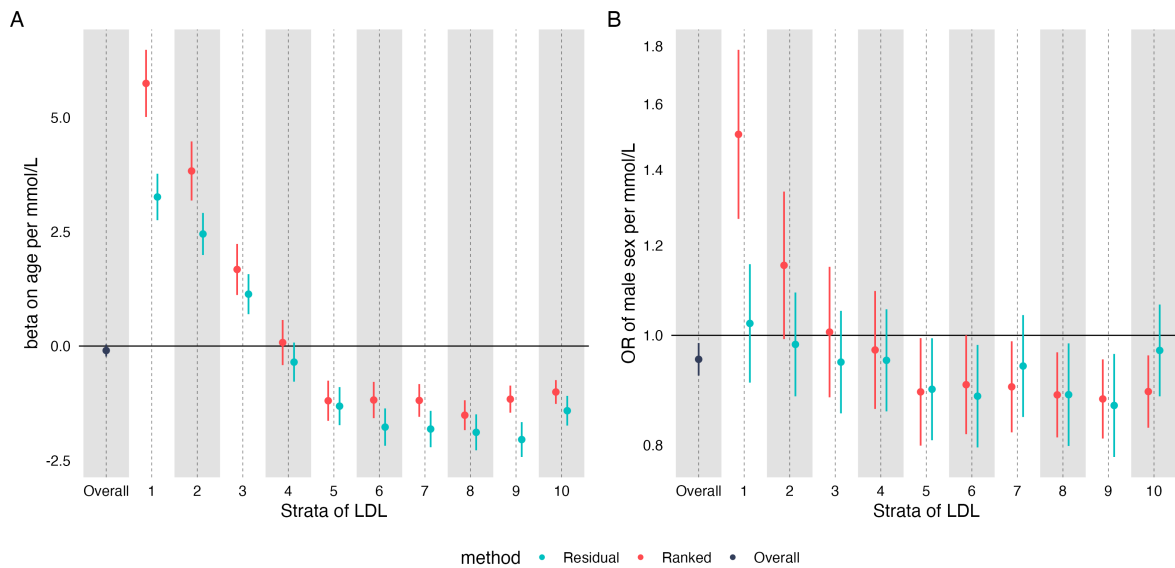


438 bias compared to the conventional MR estimate. In contrast to our NCO analyses above, the  
439 bias was more extreme for the doubly-ranked method than the residual method.

440

441 **Figure 3** The predicted causal effect of increased LDL-C on (A) age and (B) sex. Estimates were  
442 generated in each stratum using the residual method (blue) and the doubly-ranked method (red)  
443 and were unadjusted for covariates.

444



445

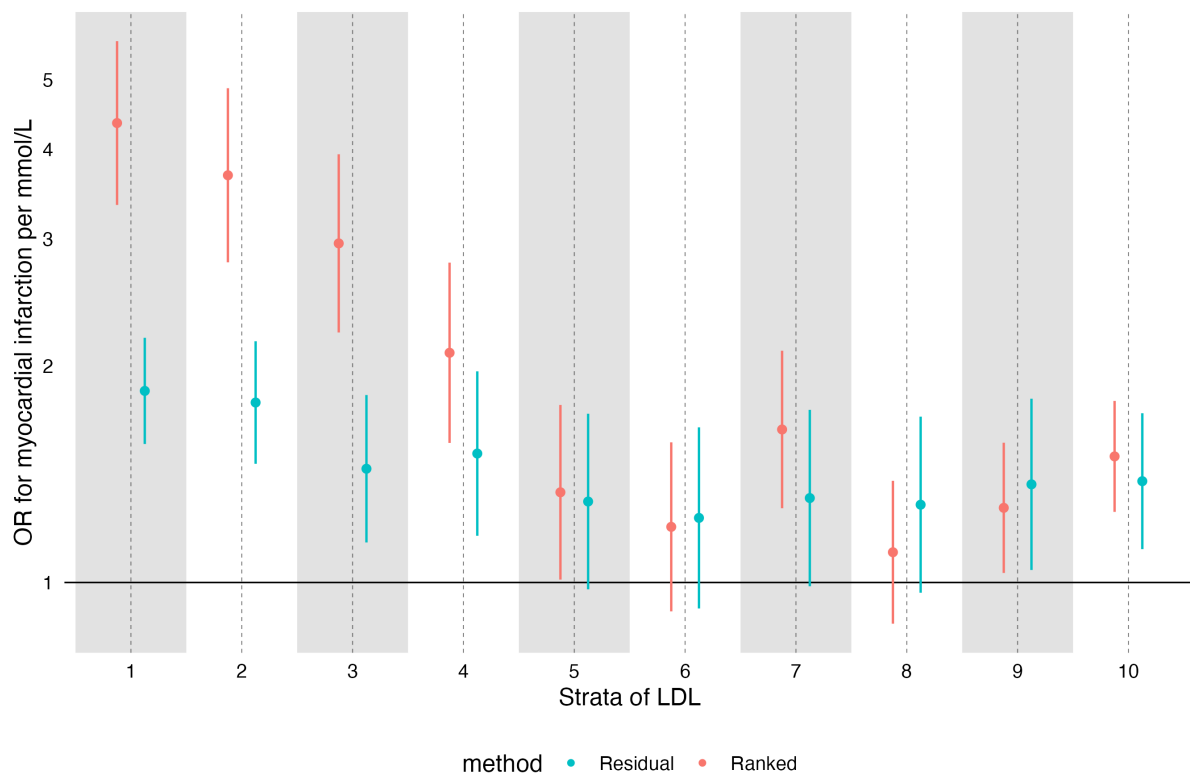
446 We then went on to perform MR of the effect of increased LDL-C on myocardial infarction. For  
447 our primary analysis we included both incident and prevalent cases of MI, the effect on incident  
448 and prevalent MI separately are shown in **Supplementary Figure 4**. In conventional MR we saw  
449 the expected effect of LDL-C (OR for MI 1.74; 95% CI 1.61-1.87). However, in NLMR we saw  
450 unexpected effects, particularly with the doubly-ranked method. For the effect on MI (**Figure 4**),  
451 we saw large differences in effects across strata, with the strongest effect in those with the  
452 lowest LDL-C, and the weakest effects in those in strata 5, 6, and 8.

453

454 **Figure 4** The estimated causal effect of increased LDL-C on MI. Estimates were generated in  
455 each stratum using the residual method (blue) and the doubly-ranked method (red) and were  
456 unadjusted for covariates.

457

458



459

460

461

462 For the residual method, the effect of LDL-C on MI was broadly stable and positive (although the  
463 trend still suggested reduction across strata,  $p = 0.02$ ). When running analyses adjusting for age,  
464 sex, and the first 5 principal components, effects were similar but less extreme for MI  
465 (**Supplementary Figure S5**), although there was still a clear negative trend ( $p = 0.001$  for using  
466 the doubly-ranked method), with effect estimates much higher in those with lower LDL-C than  
467 those with higher LDL-C.

468

469 We ran sensitivity analyses for LDL-C on MI that included a) adjusting for statin use as a  
470 covariate b) in statin users and non-statin users, and c) in under 50s, where statin use was rare -  
471 5.1% (**Supplementary Figure S6**). We recognise these estimates adjusting for statin use are  
472 highly likely to be biased due to collider bias but include these for interest. As expected,  
473 stratifying or adjusting for statin usage altered estimates dramatically. When adding statin as a  
474 covariate, estimates of LDL-C favoured protection in upper and lower strata, but were null in the  
475 middle strata (u-shaped). When analyses were performed using those only under 50, estimates  
476 seemed similar to our primary analyses but showed more variability. Similarly, analyses  
477 restricted to non-statin users looked similar to our primary analyses. Analyses in statin users  
478 had reversed estimates in most strata, with increased LDL-C associated with reduced risk of MI.

479 *Triglycerides and cancer mortality*

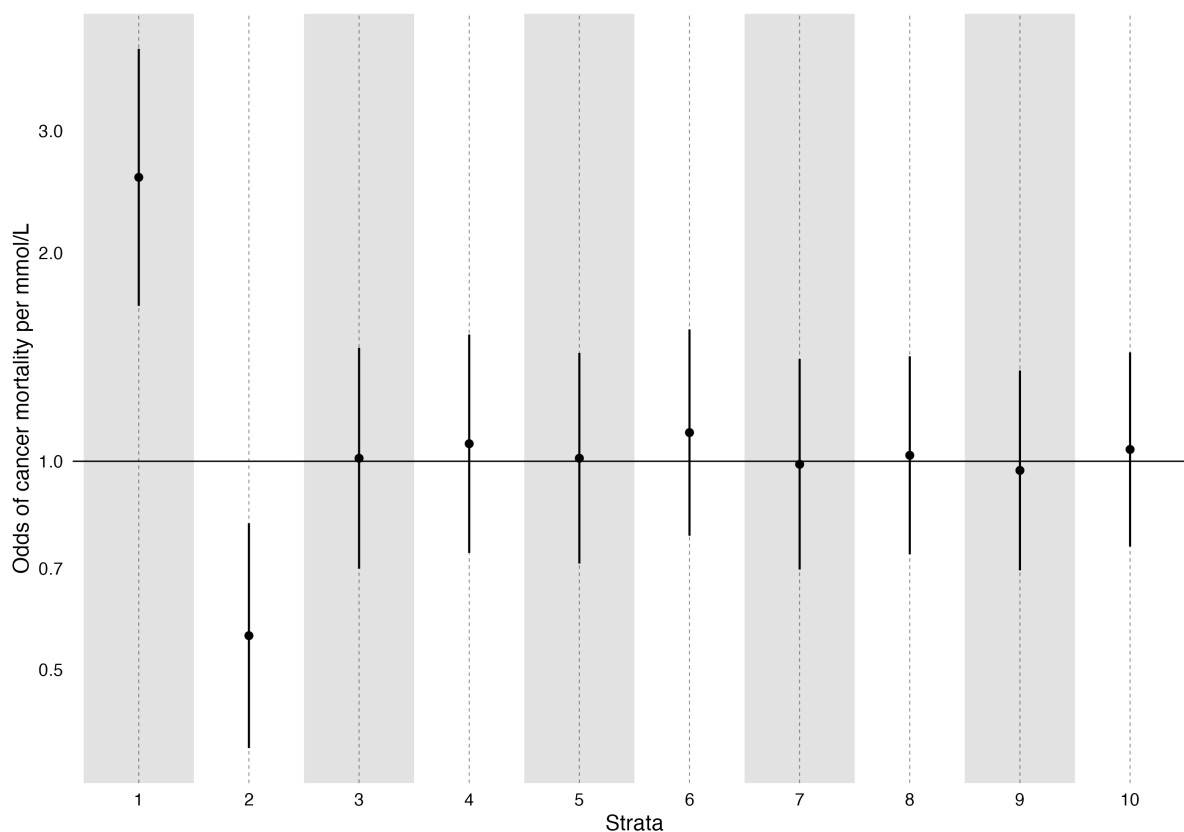
480

481 In the recent pre-print on non-linear effects of lipids[29], one analysis focussed on the effect of  
482 triglycerides on cancer mortality. The overall effect was null in both univariable and  
483 multivariable MR, but the authors report extremely implausible results: a strong positive effect  
484 in strata 1 (of ten): an OR of 2.57 (95% CI 1.67 to 3.96); but then a strong negative effect in strata  
485 2: OR 0.56 (95% CI 0.39 to 0.83). All other strata specific estimates are close to the null. We  
486 report these estimates in **Figure 5**. A Z-test comparing the lower two strata has a p value of  $1.4 \times$   
487  $10^{-7}$ , while the overall p-value for heterogeneity is  $9 \times 10^{-4}$ . These results are especially  
488 implausible given the known variability in measurements of triglycerides [46,47]. Therefore we  
489 aimed to replicate this as closely as possible, using the same dataset, exposure, outcome, and  
490 covariates. In our analysis (**Figure 6**), we were unable to replicate this finding, despite the  
491 correlation between our strata specific mean triglyceride levels being  $>0.99$ , suggesting we are  
492 analysing the same strata.

493

494 **Figure 5:** Estimates of the effect of triglycerides on cancer mortality from Yang et al [29]. Strata  
495 specific estimates generated using the doubly-ranked method.[29]

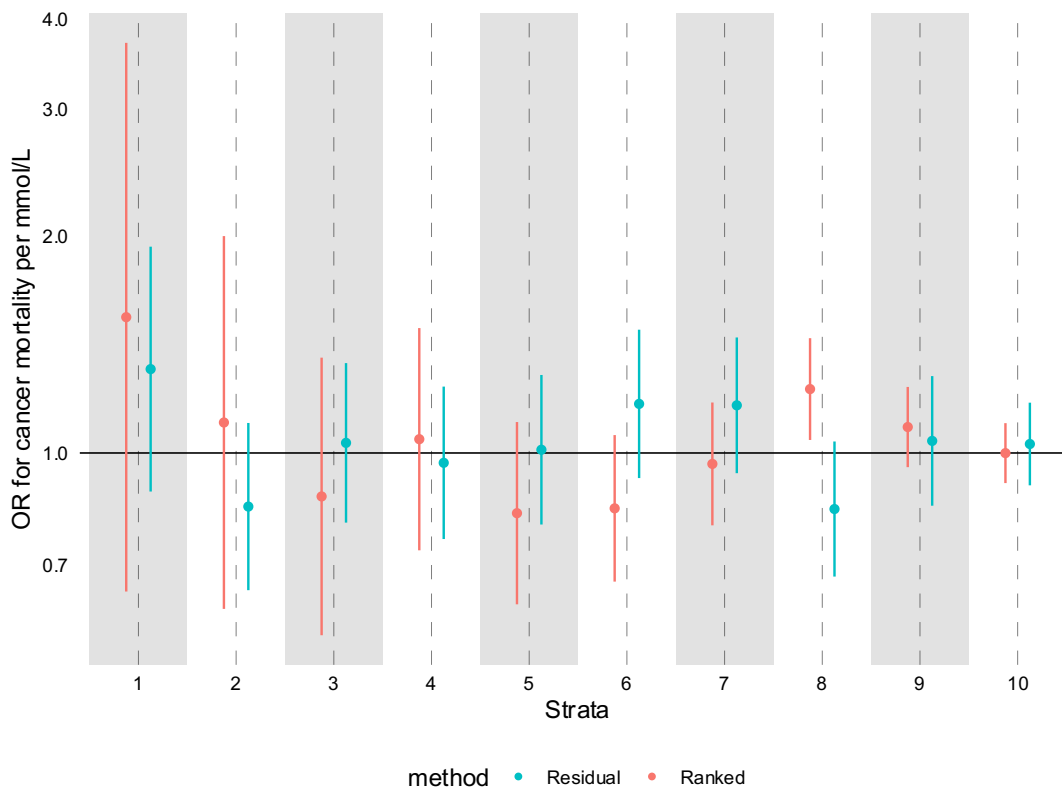
496



497

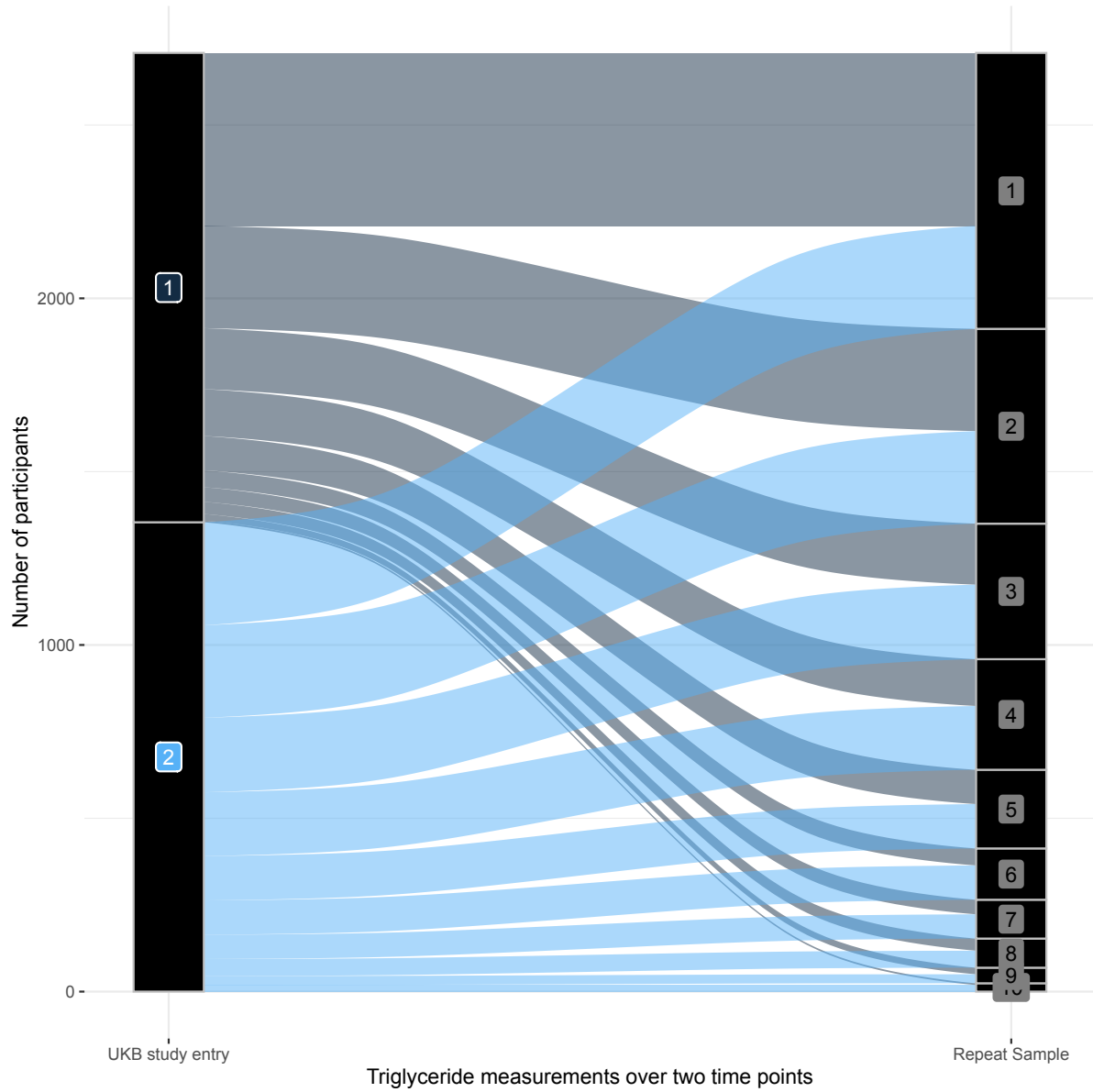
498

499 **Figure 6:** Estimates of the effect of triglycerides on cancer mortality using both the doubly-  
500 ranked (red) and residual method (blue). Estimates were generated adjusting for age, sex, age \*  
501 sex, age \* age, age \* age \* sex, and the first ten principle components.  
502



503  
504 To further assess whether the prior analysis was unreliable, we used the repeat sampling data  
505 from UK Biobank, that was taken approximately 2 years after the original visit. Data on  
506 triglyceride levels at both time points was available for 13,535 participants. Correlation  
507 between each time point was moderate (Pearson's R 0.60). As expected, when generating strata  
508 using the doubly-ranked method on the original and repeat sample, participants were often not  
509 classified in the same strata. In fact, only 37% of those participants originally in strata 1 remain  
510 in strata 1, with 22% of them in strata 2, and the rest in higher strata. For those participants  
511 originally in strata 2, only 20% remain in strata 2, with 22% now in strata 1, and the rest in higher  
512 strata. These results are visualised in the alluvial plot below (**Figure 7**).

513  
514 **Figure 7:** Classification of 13,535 participants who have repeat measurements into doubly-  
515 ranked strata based on triglyceride levels. The left hand Y axis represents the original strata, the  
516 right hand Y-axis the strata at repeat sampling.



517  
518

519 This variability is not consistent with the reported estimates from Yang et al, which would  
520 suggest those in strata 1 are at greatly increased risk of cancer mortality, and those in strata 2 at  
521 greatly reduced risk, with no effect elsewhere. If this were the case, more than half of the  
522 participants in the lower two strata dramatically change their risk of cancer mortality over 2  
523 years, with some estimates flipping from greatly increased risk to protective. To present  
524 analyses which implicitly assumes this is possible is simply not credible..

525

526 Discussion:

527

528 Mendelian randomization has become a commonly used approach in biomedical research[3],  
529 with the original intention being to strengthen causal inference regarding the effect of

530 modifiable exposures on health outcomes [1]. Quantified estimates of various average benefits  
531 that would be seen from treatment are often presented in MR papers [2], with the more recent  
532 development of methods that allowed for estimation of the effect of such treatments in groups  
533 at differing levels of the exposure[7–10,48] . The residual method [10] is currently the most  
534 widely used approach. However recent findings have identified that this method can be  
535 seriously biased [19] leading to estimates that are inconsistent with causal interpretation. In  
536 particular, the central estimate for some outcomes was precisely estimated and close to the  
537 null, while all strata specific estimates favoured protection. The authors of this paper and the  
538 residual method accept this is not possible and referred to it “as a logical impossibility if all  
539 estimates have a causal interpretation”[23] after the journal retracted the paper[24] following  
540 being alerted to it being obvious seriously problematic[19]. However, whether the recently  
541 developed doubly-ranked method also leads to biased results remains unknown, leading us to  
542 perform our negative control outcome analyses.

543

544 In our negative control outcome analyses we identified that both methods showed bias, with  
545 effect estimates on age and sex differing across generated strata, when estimates should be  
546 null across all strata. It is worth considering some of the potential reasons for the distorted  
547 estimates in our negative control outcomes. One explanation is that these relate to a form of  
548 selection bias into strata which lead to associations between the genetic variants used to proxy  
549 for the exposure in the MR analyses – the so-called “instruments” – and other factors, including  
550 age and sex. This essentially reintroduces bias in MR analyses that were advanced with the  
551 intention of producing unbiased evidence on effects of exposures[1,2] .

552

553 For body mass index overall selection bias (on entering studies, not on entering strata) is  
554 known to distort estimates, and thus the non-linear analysis of this exposure may be  
555 particularly liable to bias. For example, a large (~ 3.3 million person) GWAS of chromosomal sex  
556 identified over 150 autosomal loci associated with male sex, including the body mass  
557 increasing allele at *FTO* (OR 1.02,  $P = 4.4 \times 10^{-36}$ ) [40]. Our findings may in part reflect the strong  
558 associations between age, sex, and body mass index in the selected population of UK Biobank.  
559 This would be supported by the weaker association with Vitamin D. However, this explanation is  
560 potentially less appealing for the effect of LDL-C – with the most extreme non-null estimates.  
561 Although there is likely selection onto LDL-C, there is little evidence that adjusting for  
562 participant bias in UK Biobank affects MR estimates in LDL-C [49].

563

564 Additionally, these NCO results could represent population stratification differing across strata,  
565 which can be uncovered using negative controls [26]. Thirdly, these could represent methods  
566 related issues that occur on generation of the strata unrelated to selection. As we have  
567 demonstrated, stratification by both methods can lead to substantial alteration of estimates of  
568 IV-exposure associations. Regardless of the reason for the bias, these findings support more  
569 widespread use of negative control outcomes in Mendelian randomization studies, alongside a  
570 closer look at model assumptions.

571

572 Our analysis on the effect of LDL-C on myocardial infarction raises further concerns. The  
573 evidence base supporting the effectiveness of LDL-C reduction for MI is among the strongest in  
574 modern medicine, with >30 randomised trials to date showing benefit [27]. As such, meta-  
575 regressions of these trials using their baseline LDL-C have confidently suggested increased  
576 benefit of LDL-C reduction in those with higher LDL-C. This is also consistent with decades of  
577 non-RCT evidence, and our understanding of the pathogenesis of atherosclerosis (see **Box 1** for  
578 further background). In non-linear MR (especially with the doubly-ranked method), we identified  
579 the *opposite* non-linear effect, suggesting a *decreased* effectiveness of LDL-C reduction in  
580 those with higher LDL-C. Importantly, this bias was not ameliorated by the addition of age, sex,  
581 and genetic principal components into the model. Estimates were also similar whether using  
582 incident, prevalent, or other MI, so these results are not likely to be generated by survivorship  
583 bias.

584

585 In fact, these results do not depend on our particular analytical approach and/or outcome  
586 definition. A recent preprint from the authors of the doubly-ranked method performed a similar  
587 analysis on the same dataset [29] (using coronary artery disease as an outcome, not MI), and  
588 found similar findings, and with a strong negative trend with increasing LDL-C reduction (**Figure**  
589 **8** compares both analyses). This consistent disagreement between trial data and MR data  
590 suggests a fundamental issue with the method, or with the application of this method to this  
591 data. This may reflect selection, but importantly, estimates were broadly similar in our primary  
592 analysis (unadjusted for age, sex and genetic principal components) and their analysis (which  
593 adjusts for the above), and suggests that if selection bias is occurring, it is not fixed by adjusting  
594 for age and sex.

595

596 It is therefore plausible these findings relate to other issues such as statin use (which has a  
597 large effect on LDL-C). However, adjusting for statins did not recover sensible estimates, nor

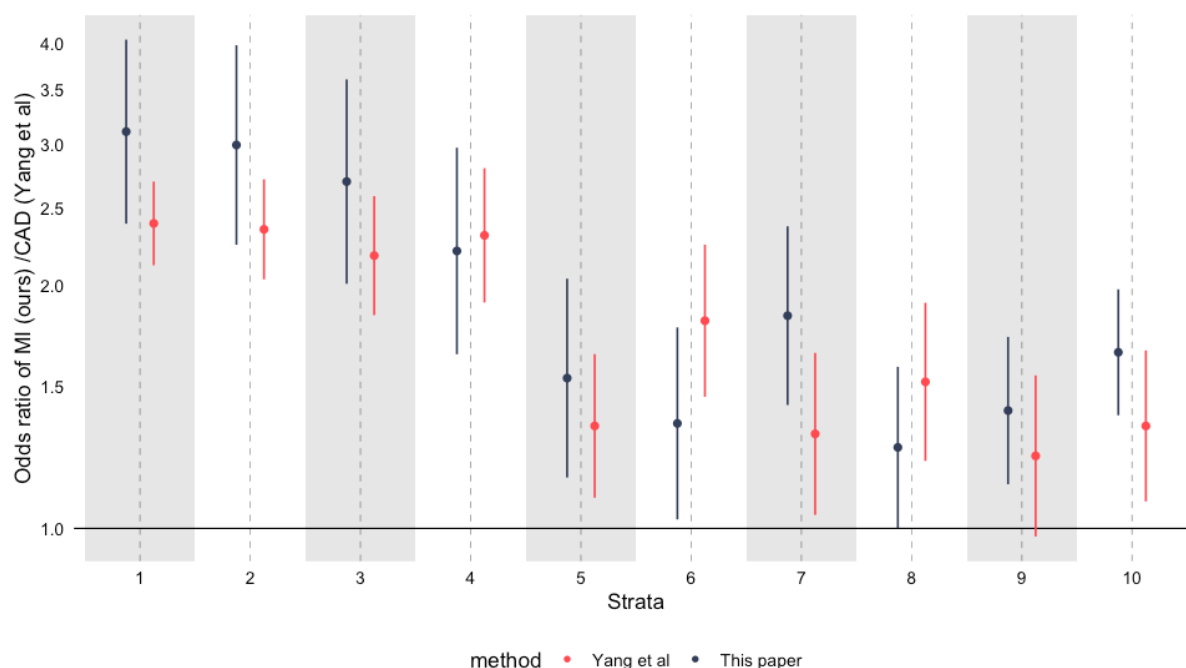
598 would this be advised as this would lead to collider bias. Even if sensible estimates were  
599 recovered, this would suggest that non-linear MR estimates require understanding and  
600 modelling of the covariate structure of the exposure and the outcome, which would bring in all  
601 the issues of traditional observational analyses that MR is designed to remove[1,2]. This  
602 particular example of LDL-C on CAD is particularly challenging for non-linear MR, as the  
603 strength of evidence supporting the effectiveness of conventional MR for the effect of LDL-C on  
604 CAD is very high, with multiple drugs having strong support from MR analyses [50].

605

606

607 **Figure 8:** The effect of increasing LDL-C on MI (our data), and coronary artery disease (CAD,  
608 Yang et al[29]). Both estimates adjusted for age, sex, and genetic principal components and  
609 using UK Biobank participants. Red estimates from Yang et al. Both estimates using the doubly-  
610 ranked method.

611



612

613

614 Further implausible analyses have been generated in the same preprint with an analysis on TG  
615 and cancer mortality. It is not possible to think of a plausible biological model where those in  
616 the bottom strata have very robustly estimated *increased* cancer mortality with higher  
617 triglycerides, whilst shifting into the 2<sup>nd</sup> lowest lowest strata have robustly estimated *decreased*  
618 cancer mortality, with there being no causal effect for all other participants. In addition given the  
619 considerable fluctuations demonstrated in triglyceride levels individuals would shift their



620 classification in one into the other strata depending on exactly when the measurement for the  
621 particular individuals were made. Importantly, we were unable to replicate the findings in this  
622 preprint[29] despite being able to generate strata specific mean estimates of TG that exactly  
623 matched theirs, and using extremely similar definitions and analytical approaches, and despite  
624 being able to replicate similar findings for the effect of LDL-C on MI as they report on CAD.

625  
626  
627

628 Therefore, our findings – and the results of other, independently performed analyses - suggest  
629 that we should be cautious when applying either the residual or doubly-ranked MR methods, as  
630 both methods seem potentially more susceptible to bias in the generation of strata specific  
631 estimates than conventional MR estimates are. Whether this entirely relates to selection bias or  
632 to other methodological issues remains unknown. Corrections should be issued for papers  
633 which have produced likely misleading estimates of non-linear effects [20,21,24], especially  
634 when their misleading findings gained widespread publicity due to the efforts of authors [51].  
635 That these biases relate only to the examples of body mass index, Vitamin D, and lipid traits is  
636 unlikely, but they remain to be demonstrated for other exposures.

637  
638  
639

### 639 *Limitations*

640  
641  
642  
643  
644  
645  
646  
647  
648

This paper provides an initial assessment of the doubly doubly-ranked and residual methods for  
performing non-linear MR. In particular, our analysis focussed on three exposures (BMI, Vitamin  
D and lipids) a small number of outcomes and included only one dataset (UK Biobank), which  
we limited to White British participants. It is possible that our findings do not extend to all  
exposures, or only occur in certain settings. However, nearly all published non-linear analyses  
use UK Biobank, so these limitations likely apply to much of the published literature. Further  
analyses should examine these methods on diverse datasets, exposures, and outcomes, with  
proof of principle examples.

649  
650

### 650 *Conclusion:*

651  
652  
653

Using a negative control outcome approach, we identified bias in non-linear mendelian  
randomization estimates of the effect of BMI, Vitamin D and LDL-C using two commonly used

654 methods. Estimates of the effect of LDL-C on cardiovascular outcomes were inconsistent with  
655 RCTs and biological understanding. Until reliable evidence is presented that the methods are  
656 generating sensible findings there should be a moratorium on the further publication of non-  
657 linear MR findings using the two methods we examine in this paper.

658

659

660

661

662

663 References:

664

665 1. Davey Smith G, Ebrahim S. ‘Mendelian randomization’: can genetic epidemiology contribute  
666 to understanding environmental determinants of disease?\*. *Int J Epidemiol* [Internet]. 2003  
667 [cited 2023 May 22];32:1–22. Available from:

668 <https://academic.oup.com/ije/article/32/1/1/642797>

669 2. Sanderson E, Glymour MM, Holmes MV, Kang H, Morrison J, Munafò MR, et al. Mendelian  
670 randomization. *Nature Reviews Methods Primers* [Internet]. 2022 [cited 2023 May 22];2:1–21.

671 Available from: <https://www.nature.com/articles/s43586-021-00092-5>

672 3. Smith GD, Ebrahim S. Mendelian randomisation at 20 years: how can it avoid hubris, while  
673 achieving more? *Lancet Diabetes Endocrinol* [Internet]. 2024;12:14–7. Available from:

674 [http://dx.doi.org/10.1016/S2213-8587\(23\)00348-0](http://dx.doi.org/10.1016/S2213-8587(23)00348-0)

675 4. Sobczyk MK, Zheng J, Davey Smith G, Gaunt TR. Systematic comparison of Mendelian  
676 randomisation studies and randomised controlled trials using electronic databases. *BMJ Open*  
677 [Internet]. 2023;13:e072087. Available from: <http://dx.doi.org/10.1136/bmjopen-2023-072087>

678 5. Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. *Int J Epidemiol*  
679 [Internet]. 2016;45:1866–86. Available from: <http://dx.doi.org/10.1093/ije/dyw314>

680 6. Munafò MR, Higgins JPT, Davey Smith G. Triangulating Evidence through the Inclusion of  
681 Genetically Informed Designs. *Cold Spring Harb Perspect Med* [Internet]. 2021;11. Available  
682 from: <http://dx.doi.org/10.1101/cshperspect.a040659>

683 7. Staley JR, Burgess S. Semiparametric methods for estimation of a nonlinear exposure-  
684 outcome relationship using instrumental variables with application to Mendelian  
685 randomization. *Genet Epidemiol* [Internet]. 2017;41:341–52. Available from:

686 <http://dx.doi.org/10.1002/gepi.22041>

687 8. Sulc J, Sjaarda J, Kutalik Z. Polynomial Mendelian randomization reveals non-linear causal  
688 effects for obesity-related traits. *HGG Adv* [Internet]. 2022;3:100124. Available from:

689 <http://dx.doi.org/10.1016/j.xhgg.2022.100124>

- 690 9. Silverwood RJ, Holmes MV, Dale CE, Lawlor DA, Whittaker JC, Davey Smith G, et al. Testing for  
691 non-linear causal effects using a binary genotype in a Mendelian randomization study:  
692 application to alcohol and cardiovascular traits. *Int J Epidemiol* [Internet]. 2014;43:1781–90.  
693 Available from: <http://dx.doi.org/10.1093/ije/dyu187>
- 694 10. Burgess S, Davies NM, Thompson SG, EPIC-InterAct Consortium. Instrumental variable  
695 analysis with a nonlinear exposure-outcome relationship. *Epidemiology* [Internet].  
696 2014;25:877–85. Available from: <http://dx.doi.org/10.1097/EDE.0000000000000161>
- 697 11. Tian H, Mason AM, Liu C, Burgess S. Relaxing parametric assumptions for non-linear  
698 Mendelian randomization using a doubly-ranked stratification method. *PLoS Genet* [Internet].  
699 2023 [cited 2023 May 22];19:e1010823. Available from:  
700 <http://dx.doi.org/10.1371/journal.pgen.1010823>
- 701 12. Rogne T, Solligård E, Burgess S, Brumpton BM, Paulsen J, Prescott HC, et al. Body mass  
702 index and risk of dying from a bloodstream infection: A Mendelian randomization study. *PLoS*  
703 *Med* [Internet]. 2020;17:e1003413. Available from:  
704 <http://dx.doi.org/10.1371/journal.pmed.1003413>
- 705 13. Pinto Pereira SM, Garfield V, Norris T, Burgess S, Williams DM, Dodds R, et al. Linear and  
706 Non-linear associations between vitamin D and grip strength: a Mendelian Randomisation study  
707 in UK Biobank. *J Gerontol A Biol Sci Med Sci* [Internet]. 2022; Available from:  
708 <http://dx.doi.org/10.1093/gerona/glac255>
- 709 14. Zhou A, Hyppönen E. Vitamin D deficiency and C-reactive protein: a bidirectional Mendelian  
710 randomization study. *Int J Epidemiol* [Internet]. 2023;52:260–71. Available from:  
711 <http://dx.doi.org/10.1093/ije/dyac087>
- 712 15. Sutherland JP, Zhou A, Hyppönen E. Vitamin D Deficiency Increases Mortality Risk in the UK  
713 Biobank : A Nonlinear Mendelian Randomization Study. *Ann Intern Med* [Internet].  
714 2022;175:1552–9. Available from: <http://dx.doi.org/10.7326/M21-3324>
- 715 16. Zhou A, Selvanayagam JB, Hyppönen E. Non-linear Mendelian randomization analyses  
716 support a role for vitamin D deficiency in cardiovascular disease risk. *Eur Heart J* [Internet].  
717 2022;43:1731–9. Available from: <http://dx.doi.org/10.1093/eurheartj/ehab809>
- 718 17. Sun Y-Q, Burgess S, Staley JR, Wood AM, Bell S, Kaptoge SK, et al. Body mass index and all  
719 cause mortality in HUNT and UK Biobank studies: linear and non-linear mendelian

- 720 randomisation analyses. *BMJ* [Internet]. 2019;364:l1042. Available from:  
721 <http://dx.doi.org/10.1136/bmj.l1042>
- 722 18. Sofianopoulou E, Kaptoge SK, Afzal S, Jiang T, Gill D, Gundersen TE, et al. RETRACTED:  
723 Estimating dose-response relationships for vitamin D with coronary heart disease, stroke, and  
724 all-cause mortality: observational and Mendelian randomisation analyses. *The Lancet Diabetes*  
725 *& Endocrinology* [Internet]. 2021;9:837–46. Available from:  
726 <https://www.sciencedirect.com/science/article/pii/S2213858721002631>
- 727 19. Davey Smith G. Mendelian randomisation and vitamin D: the importance of model  
728 assumptions [Internet]. *Lancet Diabetes Endocrinol*. Elsevier BV; 2023. p. 14. Available from:  
729 [http://dx.doi.org/10.1016/S2213-8587\(22\)00345-X](http://dx.doi.org/10.1016/S2213-8587(22)00345-X)
- 730 20. Burgess S, Butterworth AS. Dose-response relationships for vitamin D and all-cause  
731 mortality - Authors' reply [Internet]. *Lancet Diabetes Endocrinol*. 2022. p. 158–9. Available from:  
732 [http://dx.doi.org/10.1016/S2213-8587\(22\)00015-8](http://dx.doi.org/10.1016/S2213-8587(22)00015-8)
- 733 21. Wade KH, Hamilton FW, Carslake D, Sattar N, Davey Smith G, Timpson NJ. Challenges in  
734 undertaking non-linear Mendelian randomization. *Obesity* [Internet]. 2023;2887–90. Available  
735 from: <http://dx.doi.org/10.1002/oby.23927>
- 736 22. Munafò MR, Tilling K, Taylor AE, Evans DM, Davey Smith G. Collider scope: when selection  
737 bias can substantially influence observed associations. *Int J Epidemiol* [Internet]. 2018;47:226–  
738 35. Available from: <http://dx.doi.org/10.1093/ije/dyx206>
- 739 23. Emerging Risk Factors Collaboration/EPIC-CVD/Vitamin D Studies Collaboration. Estimating  
740 dose-response relationships for vitamin D with coronary heart disease, stroke, and all-cause  
741 mortality: observational and Mendelian randomisation analyses. *Lancet Diabetes Endocrinol*.  
742 2024;9:837–46.
- 743 24. The Editors Of The *Lancet Diabetes Endocrinology*. Retraction and republication-Estimating  
744 dose-response relationships for vitamin D with coronary heart disease, stroke, and all-cause  
745 mortality: observational and Mendelian randomisation analyses. *Lancet Diabetes Endocrinol*  
746 [Internet]. 2023; Available from: [http://dx.doi.org/10.1016/S2213-8587\(23\)00364-9](http://dx.doi.org/10.1016/S2213-8587(23)00364-9)
- 747 25. Burgess S. Violation of the constant genetic effect assumption can result in biased  
748 estimates for non-linear Mendelian randomization [Internet]. *Hum. Hered*. 2023 [cited 2022 Nov  
749 1]. p. 2022.10.26.22280570. Available from: <http://dx.doi.org/10.1159/000531659>

- 750 26. Sanderson E, Richardson TG, Hemani G, Davey Smith G. The use of negative control  
751 outcomes in Mendelian randomization to detect potential population stratification. *Int J*  
752 *Epidemiol* [Internet]. 2021;50:1350–61. Available from: <http://dx.doi.org/10.1093/ije/dyaa288>
- 753 27. Navarese EP, Robinson JG, Kowalewski M, Kolodziejczak M, Andreotti F, Bliden K, et al.  
754 Association Between Baseline LDL-C Level and Total and Cardiovascular Mortality After LDL-C  
755 Lowering: A Systematic Review and Meta-analysis. *JAMA* [Internet]. 2018;319:1566–79.  
756 Available from: <http://dx.doi.org/10.1001/jama.2018.2525>
- 757 28. Cholesterol Treatment Trialists' (CTT) Collaboration. Efficacy and safety of more intensive  
758 lowering of LDL cholesterol: a meta-analysis of data from 170 000 participants in 26  
759 randomised trials. *Lancet* [Internet]. 2010;376:1670–81. Available from:  
760 [https://doi.org/10.1016/S0140-6736\(10\)61350-5](https://doi.org/10.1016/S0140-6736(10)61350-5)
- 761 29. Yang G, Mason AM, Wood AM, Schooling CM, Burgess S. Assessing dose-response relations  
762 of lipid traits with coronary artery disease, all-cause mortality, and cause-specific mortality: a  
763 linear and non-linear Mendelian randomization study [Internet]. *bioRxiv*. 2023. Available from:  
764 <https://www.medrxiv.org/content/10.1101/2023.09.27.23296203v1.abstract>
- 765 30. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource  
766 with deep phenotyping and genomic data. *Nature* [Internet]. 2018;562:203–9. Available from:  
767 <http://dx.doi.org/10.1038/s41586-018-0579-z>
- 768 31. Mitchell R, Hemani G, Dudding T, Corbin L, Harrison S, Paternoster L. UK Biobank genetic  
769 data: MRC-IEU quality control, version 2 [Internet]. University of Bristol; 2019 [cited 2021 Dec  
770 21]. Available from: <https://data.bris.ac.uk/data/dataset/1ovaau5sxunp2cv8rcy88688v>
- 771 32. Liu H, Li J, Liu F, Huang K, Cao J, Chen S, et al. Efficacy and safety of low levels of low-density  
772 lipoprotein cholesterol: trans-ancestry linear and non-linear Mendelian randomization  
773 analyses. *Eur J Prev Cardiol* [Internet]. 2023;30:1207–15. Available from:  
774 <http://dx.doi.org/10.1093/eurjpc/zwad111>
- 775 33. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body  
776 mass index yield new insights for obesity biology. *Nature* [Internet]. 2015 [cited 2022 Apr  
777 28];518:197–206. Available from: <http://dx.doi.org/10.1038/nature14177>

- 778 34. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform  
779 supports systematic causal inference across the human phenome. *Elife* [Internet]. 2018;7.  
780 Available from: <http://dx.doi.org/10.7554/eLife.34408>
- 781 35. Wang TJ, Zhang F, Richards JB, Kestenbaum B, van Meurs JB, Berry D, et al. Common genetic  
782 determinants of vitamin D insufficiency: a genome-wide association study [Internet]. *Lancet*.  
783 2010. p. 180–8. Available from: [http://dx.doi.org/10.1016/S0140-6736\(10\)60588-0](http://dx.doi.org/10.1016/S0140-6736(10)60588-0)
- 784 36. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set  
785 for whole-genome association and population-based linkage analyses. *Am J Hum Genet*  
786 [Internet]. 2007;81:559–75. Available from: <http://dx.doi.org/10.1086/519795>
- 787 37. Choi Y, Chan AP, Kirkness E, Telenti A, Schork NJ. Comparison of phasing strategies for  
788 whole human genomes. *PLoS Genet* [Internet]. 2018;14:e1007308. Available from:  
789 <http://dx.doi.org/10.1371/journal.pgen.1007308>
- 790 38. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. *G3*  
791 [Internet]. 2011;1:457–70. Available from: <http://dx.doi.org/10.1534/g3.111.001198>
- 792 39. Sanderson E, Macdonald-Wallis C, Davey Smith G. Negative control exposure studies in the  
793 presence of measurement error: implications for attempted effect estimate calibration. *Int J*  
794 *Epidemiol* [Internet]. 2018;47:587–96. Available from: <http://dx.doi.org/10.1093/ije/dyx213>
- 795 40. Pirastu N, Cordioli M, Nandakumar P, Mignogna G, Abdellaoui A, Hollis B, et al. Genetic  
796 analyses identify widespread sex-differential participation bias. *Nat Genet* [Internet].  
797 2021;53:663–71. Available from: <http://dx.doi.org/10.1038/s41588-021-00846-7>
- 798 41. Carreras-Torres R, Johansson M, Haycock PC, Relton CL, Davey Smith G, Brennan P, et al.  
799 Role of obesity in smoking behaviour: Mendelian randomisation study in UK Biobank. *BMJ*  
800 [Internet]. 2018 [cited 2023 Apr 26];361:k1767. Available from:  
801 <http://dx.doi.org/10.1136/bmj.k1767>
- 802 42. Mason AM, Burgess S. Software Application Profile: SUMnlmr, an R package that facilitates  
803 flexible and reproducible non-linear Mendelian randomization analyses. *Int J Epidemiol*  
804 [Internet]. 2022 [cited 2023 May 22];51:2014–9. Available from:  
805 <https://academic.oup.com/ije/article/51/6/2014/6659062>
- 806 43. Palmer T, Spiller W, Sanderson E. OneSampleMR: One Sample Mendelian Randomization  
807 and Instrumental Variable Analyses. 2023.

- 808 44. Holmes MV, Davey Smith G. Dyslipidaemia: Revealing the effect of CETP inhibition in  
809 cardiovascular disease. *Nat Rev Cardiol* [Internet]. 2017;14:635–6. Available from:  
810 <http://dx.doi.org/10.1038/nrcardio.2017.156>
- 811 45. Smith GD, Ebrahim S. Mendelian randomization: prospects, potentials, and limitations. *Int J*  
812 *Epidemiol* [Internet]. 2004 [cited 2024 Jan 18];33:30–42. Available from:  
813 <https://academic.oup.com/ije/article-pdf/33/1/30/1939982/dyh132.pdf>
- 814 46. Phillips AN, Smith GD. How independent are “independent” effects? Relative risk estimation  
815 when correlated exposures are measured imprecisely. *J Clin Epidemiol* [Internet].  
816 1991;44:1223–31. Available from: [http://dx.doi.org/10.1016/0895-4356\(91\)90155-3](http://dx.doi.org/10.1016/0895-4356(91)90155-3)
- 817 47. Johansen MØ, Moreno-Vedia J, Balling M, Davey Smith G, Nordestgaard BG. Triglyceride  
818 content increases while cholesterol content decreases in HDL and LDL+IDL fractions following  
819 normal meals: The Copenhagen General Population Study of 25,656 individuals.  
820 *Atherosclerosis* [Internet]. 2023;383:117316. Available from:  
821 <http://dx.doi.org/10.1016/j.atherosclerosis.2023.117316>
- 822 48. Millwood IY, Walters RG, Mei XW, Guo Y, Yang L, Bian Z, et al. Conventional and genetic  
823 evidence on alcohol and vascular disease aetiology: a prospective study of 500 000 men and  
824 women in China. *Lancet* [Internet]. 2019;393:1831–42. Available from:  
825 [http://dx.doi.org/10.1016/S0140-6736\(18\)31772-0](http://dx.doi.org/10.1016/S0140-6736(18)31772-0)
- 826 49. Schoeler T, Speed D, Porcu E, Pirastu N, Pingault J-B, Kutalik Z. Participation bias in the UK  
827 Biobank distorts genetic associations and downstream analyses. *Nat Hum Behav* [Internet].  
828 2023; Available from: <http://dx.doi.org/10.1038/s41562-023-01579-9>
- 829 50. Allara E, Morani G, Carter P, Gkatzionis A, Zuber V, Foley CN, et al. Genetic Determinants of  
830 Lipids and Cardiovascular Disease Outcomes: A Wide-Angled Mendelian Randomization  
831 Investigation. *Circ Genom Precis Med* [Internet]. 2019;12:e002711. Available from:  
832 <http://dx.doi.org/10.1161/CIRCGEN.119.002711>
- 833 51. The sunshine vitamin that ‘D’elivers on cardio health [Internet]. Home. [cited 2023 Sep 27].  
834 Available from: [https://www.unisa.edu.au/media-centre/Releases/2021/the-sunshine-vitamin-](https://www.unisa.edu.au/media-centre/Releases/2021/the-sunshine-vitamin-that-delivers-on-cardio-health/)  
835 [that-delivers-on-cardio-health/](https://www.unisa.edu.au/media-centre/Releases/2021/the-sunshine-vitamin-that-delivers-on-cardio-health/)
- 836
- 837





839 Statements and Declarations

840

841

842 *Funding*

843

844 FH's time was funded by the GW4-CAT Wellcome Trust Doctoral Fellowship Scheme  
845 (222894/Z/21/Z). UK Biobank was funded by the Wellcome Trust, the Medical Research Council,  
846 the NIHR, and a variety of other charities ([https://www.ukbiobank.ac.uk/learn-more-about-uk-](https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/about-us/our-funding)  
847 [biobank/about-us/our-funding](https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/about-us/our-funding)). DAH was supported by the Dr. Nicholas J Timpson's Wellcome  
848 Investigator Award (202802/Z/16/Z). All authors work within the MRC Integrative Epidemiology  
849 Unit at the University of Bristol, which is supported by the Medical Research Council  
850 (MC\_UU\_00011/1). This research was funded in whole, or in part, by the Wellcome Trust  
851 [222894/Z/21/Z]. For the purpose of Open Access, the author has applied a CC BY public  
852 copyright licence to any Author Accepted Manuscript version arising from this submission.

853

854 *Author contributions*

855

856 GDS conceived the idea, while FH performed analyses and wrote the initial manuscript. DH,  
857 WS, and KT provided commentary and feedback.

858

859 *Competing interests*

860

861 GDS reports Scientific Advisory Board Membership for Relation Therapeutics and Insitro.

862

863