

1           **Lipid Fractions and Contrasting Risks of Coronary Artery Disease and**  
2                           **Diabetes: A Mendelian Randomization Study**

3  
4                           Subtitle: Lipids, coronary artery disease and diabetes

5  
6           Jon White, PhD<sup>1\*</sup>, Daniel I Swerdlow, MD PhD<sup>2,3</sup>, David Preiss, PhD<sup>4</sup>, Zammy  
7           Fairhurst-Hunter, MSc<sup>5</sup>, Brendan J Keating, PhD<sup>6,7</sup>, Folkert W Asselbergs, MD  
8           PhD<sup>8,9,10</sup>, Naveed Sattar, MD PhD<sup>11</sup>, Steve E Humphries, PhD<sup>12</sup>, Aroon D Hingorani,  
9                           MBBS PhD<sup>2,10</sup>, Michael V Holmes, MD PhD<sup>4\*</sup>

10  
11  
12       1. UCL Genetics Institute, University College London, Gower Street, London WC1E  
13       6BT

14       2. Farr Institute at University College London, Gower Street, London WC1E 6BT

15       3. Department of Medicine, Imperial College London, London W12 0HS

16       4. Clinical Trial Service Unit & Epidemiological Studies Unit (CTSU), Nuffield  
17       Department of Population Health, University of Oxford, Richard Doll Building, Old  
18       Road Campus, Roosevelt Drive, Oxford OX3 7LF

19       5. The Wellcome Trust Centre for Human Genetics, Roosevelt Drive, University of  
20       Oxford, Oxford, OX3 7BN

21       6. Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA

22       7. Department of Surgery, University of Pennsylvania, 3400 Spruce Street,  
23       Philadelphia, PA 19104, USA

24       8. Department of Cardiology, Division Heart & Lungs, UMC Utrecht, the Netherlands

25 9. Durrer Center for Cardiogenetic Research, ICIN-Netherlands Heart Institute,  
26 Utrecht, the Netherlands

27 10. Institute of Cardiovascular Science, University College London, Gower Street,  
28 London, United Kingdom

29 11. Institute of Cardiovascular and Medical Science, University of Glasgow, Glasgow

30 12. Cardiovascular Genetics, BHF Laboratories, Institute Cardiovascular Science,  
31 University College London, Gower Street, London WC1E 6BT  
32  
33

34 \* Correspondence to:

35 **Dr Jon White**; UCL Genetics Institute, University College London, Gower Street,  
36 London WC1E 6BT, United Kingdom

37 Email: [jon.white@ucl.ac.uk](mailto:jon.white@ucl.ac.uk)

38 Tel: +44 (0)20 7679 2212

39 and

40 **Dr Michael V Holmes**; Clinical Trial Service Unit & Epidemiological Studies Unit  
41 (CTSU), Nuffield Department of Population Health, University of Oxford, Richard  
42 Doll Building, Old Road Campus, Roosevelt Drive, Oxford OX3 7LF, United  
43 Kingdom

44 Email: [michael.holmes@ndph.ox.ac.uk](mailto:michael.holmes@ndph.ox.ac.uk)

45 Tel: +44 (0)1865 74 3644  
46

47 **Date of revision: 13<sup>th</sup> May 2016**

48 **Word count (text only; limit=3100): 3097**

49    **Abstract**

50

51    **Importance** Low-density lipoprotein cholesterol (LDL-C) is causally related to  
52    coronary artery disease (CAD) but the relevance of HDL-C and triglycerides is  
53    uncertain. LDL-C lowering by statins modestly increases diabetes risk, but it is  
54    unknown if this effect is specific to statins.

55

56    **Objective** To investigate the relationships of three routinely measured lipid fractions  
57    with CAD and diabetes through Mendelian randomization (MR), using conventional  
58    MR and making use of newer approaches such as multivariate MR and MR-Egger  
59    that addresses pleiotropy of genetic instruments.

60

61    **Design** We used published data from genome wide association studies to construct  
62    genetic instruments and applied them to investigate associations between lipid  
63    fractions and risk of CAD and diabetes using MR approaches that took into account  
64    pleiotropy of genetic instruments.

65

66    **Main outcomes and measures:** coronary artery disease and diabetes

67

68    **Results** We constructed genetic instruments composed of 130 SNPs for LDL-C  
69    (explaining 7.9% of its variance), 140 SNPs for HDL-C (6.6% of variance) and 140  
70    SNPs for triglycerides (5.9% of variance). A 1-SD (genetically instrumented)  
71    elevation in LDL-C (equivalent to 38 mg/dL) and triglycerides (equivalent to 89  
72    mg/dL) were associated with higher CAD risk: odds ratios (OR) and 95% confidence  
73    intervals (CI) were 1.68 (1.51-1.87) for LDL-C and 1.28 (1.13-1.45) for triglycerides.

74 The corresponding OR for HDL-C (equivalent to a 16mg/dL increase) was 0.95  
75 (0.85-1.06). All three lipid traits were associated with a lower risk of diabetes. The  
76 OR and 95% CI limits were 0.79 (0.71-0.88) for LDL-C; 0.83 (0.76-0.90) for HDL-C  
77 and 0.83 (95%CI: 0.72, 0.95) for TG, per 1-SD elevation.

78

79 **Conclusions and Relevance** Routinely measured lipid fractions exhibit contrasting  
80 relationships with risk of CAD and diabetes. Increased LDL-C, HDL-C and TG are  
81 associated with reduced risk of diabetes. This information will be relevant to design  
82 of clinical trials of lipid modifying agents, which should carefully monitor for  
83 dysglycemia and incident of diabetes.

84

85

86 Keywords: lipids, diabetes, coronary artery disease, epidemiology, Mendelian  
87 randomization

## 88    **Introduction**

89

90    Understanding the interplay between circulating lipids and risk of type 2 diabetes  
91    (T2D) and coronary artery disease (CAD) is of emerging public health importance  
92    and has implications for drug development for cardiovascular disease prevention.<sup>1,2</sup>  
93    For example, a causal influence of low-density lipoprotein cholesterol (LDL-C) on  
94    CAD is widely accepted<sup>3-5</sup> and the proposed causal role of triglycerides (TG) in CAD  
95    is gaining acceptance.<sup>6,7</sup> In contrast, the role of high-density lipoprotein cholesterol  
96    (HDL-C) in CAD remains in doubt.<sup>7-9</sup>

97

98    However, evidence has emerged that LDL-C reduction with statin therapy results in a  
99    modest increase in risk of T2D<sup>10,11</sup> (outweighed by the benefit of statins in protecting  
100    from CAD).<sup>12</sup> Whether this diabetogenic effect is a general consequence of LDL-C  
101    lowering or if it is specific to inhibition of HMG-CoA reductase remains unclear.<sup>13</sup>  
102    Moreover, the role of TG and HDL-C in the aetiology of T2D remains unclear.<sup>14</sup>

103

104    Residual confounding and reverse causality can limit causal inference from  
105    observational studies. Where a genetic instrument can be used as an instrument for an  
106    exposure, Mendelian randomization (MR) generates unbiased, unconfounded effect  
107    estimates that are sometimes interpreted as evidence of a causal role. This is because  
108    genotype is not modifiable by disease, and the random allocation of alleles at  
109    gametogenesis helps avoid bias from reverse causality and confounding, respectively.

110

111

112 A critical assumption of the MR paradigm is that the genetic instrument influences  
113 disease risk exclusively through the exposure of interest. However, genetic variant(s)  
114 used to proxy the exposure of interest can also associate with other traits, a  
115 phenomenon termed ‘pleiotropy’. When pleiotropy arises as a downstream  
116 consequence of genetic perturbation of the biomarker of interest, it is referred to as  
117 vertical pleiotropy and the MR assumption is preserved.<sup>15</sup> However, when pleiotropy  
118 arises because of the association of genetic variant(s) with additional phenotypes in  
119 alternative disease pathways (termed horizontal pleiotropy), the assumption is  
120 compromised. When MR analysis is based on multiple SNPs drawn from different  
121 regions of the genome selected systematically for their association with the biomarker  
122 of interest, additional non-systematic effects on any other biomarkers might be  
123 ‘balanced’ and the MR effect estimate could still be valid. However, if horizontal  
124 pleiotropy is unbalanced, as might occur when the set of biomarkers concerned come  
125 from closely connected pathways, MR estimates may become systematically biased  
126 (termed ‘unbalanced’ or ‘directional’ pleiotropy<sup>16</sup>), resulting in invalid effect  
127 estimates (see **Figure 1** for more details).

128

129 Recent methodological advances in MR analysis, including ‘multivariate’ MR<sup>17</sup> and  
130 ‘MR-Egger’,<sup>16</sup> provide new approaches for dealing with pleiotropic genetic  
131 instruments. In multivariate MR, adjustment is made for genetic associations with  
132 measured traits, but may not fully account for unbalanced pleiotropy<sup>18</sup>. In contrast,  
133 MR-Egger can detect and correct for unbalanced pleiotropy of the genetic instrument,  
134 even when unbalanced pleiotropy is mediated through unmeasured or unknown traits.

135

136 We used summary data from multiple major cardiometabolic genome wide  
137 association studies (GWAS) to investigate the underlying relationships between  
138 lipids, T2D and CAD using three MR approaches: (i) conventional MR that does not  
139 account for pleiotropy, (ii) multivariate MR, that adjusts for traits that may mediate  
140 unbalanced pleiotropy, and (iii) MR-Egger, that more fully accounts for unbalanced  
141 pleiotropy.

## 142   **Methods**

### 143   *Data sources*

144   We used summary-level data for lipids from the Global Lipids Genetics Consortium  
145   (GLGC),<sup>19</sup> T2D data from the DIAbetes Genetics Replication And Meta-analysis  
146   (DIAGRAM),<sup>20</sup> and CAD data from the Coronary ARtery DIsease Genome-wide  
147   Replication And Meta Analysis (CARDIoGRAM) plus The Coronary Artery Disease  
148   (C4D) Genetics, collectively known as CARDIoGRAMplusC4D consortium.<sup>21</sup>  
149   Details of the consortia and webpages for data download are provided in **Table 1**. All  
150   datasets were limited to individuals of European ancestry. Beta coefficients and  
151   standard errors were obtained for the per allele association of each SNP with all  
152   exposures and outcomes from these data sources. Where SNPs were not present in a  
153   dataset we used proxies ( $R^2 > 0.9$ ) as indicated in **Figure 2**.

154

### 155   *Selection of SNPs*

156   We used 185 lipid-associated SNPs identified by Willer et al<sup>19</sup> to generate a series of  
157   genetic instruments for each of the exposures: LDL-C, HDL-C and TG. This was  
158   conducted by first restricting to a set of SNPs in low linkage disequilibrium (pairwise  
159    $R^2 < 0.2$ ). We then organized these SNPs by descending order of proportional variance  
160   ( $R^2$ , estimated from the summary statistics using the gtx() package in R) between SNP  
161   with the corresponding lipid exposure to generate a range of instruments from 5 to  
162   150 SNPs. The process used to determine the final tally of SNPs for inclusion in a  
163   genetic instrument for each lipid trait is described below.

164

### 165   *Handling of SNPs*

166 We matched SNPs across the data sources by aligning them to the same effect allele.  
167 Effect allele frequencies were checked for concordance.  
168  
169 *Mendelian randomization analyses*  
170 We used three MR approaches.  
171  
172 First, we used conventional 2-sample instrumental variable (IV) analyses, which does  
173 not make any allowance for pleiotropy. Basing our approach on the method first  
174 proposed by Johnson,<sup>22</sup> we incorporated the bootstrap suggested by Bowden,<sup>16</sup> as a  
175 way to incorporate the error in the published estimates of SNP effect on both  
176 exposure and outcome.  
177  
178 Second, we conducted multivariate MR analyses, which statistically adjusts for  
179 pleiotropy with additional phenotypes measured in the dataset.<sup>23</sup> Multivariate MR is  
180 an extension of the conventional weighted regression in which the betas for additional  
181 phenotypes are included as covariates. In this case we used all three lipid traits in the  
182 model (e.g. for the HDL-C instrument we included, thereby adjusting for, LDL-C and  
183 TG).  
184  
185 Third, we used MR-Egger,<sup>16</sup> which accounts for unbalanced pleiotropy of a genetic  
186 instrument. MR-Egger is a linear regression of estimated SNP effects (for the  
187 exposure-raising allele) on exposure against the corresponding estimates of SNP on  
188 outcome, weighted by the inverse variance of the SNP on outcome effect estimates.  
189 This differs from conventional 2-sample MR in that the regression line is not forced  
190 through the origin. Bowden et al<sup>16</sup> show that the MR-Egger estimate is unaffected by

191 net pleiotropic effects of the instrument and, indeed, the presence of unbalanced  
192 pleiotropy can be inferred if the intercept term is not zero.

193

194 For all three approaches (conventional MR, multivariate MR and MR-Egger), we  
195 conducted 10,000 bootstraps, and our effect estimate is the mean of the bootstraps  
196 with the confidence interval (CI) determined empirically and set to Bonferroni  
197 adjusted (for 6 tests) 95% (i.e. 99.2%).

198

199 *Quantifying the proportion of variance explained by the genetic instruments*

200 *i) R-Trend*

201 The proportion of variance ( $R^2$ ) of the trait explained by the genetic instrument will  
202 rise with the addition of more SNPs. However, the improvement beyond the optimum  
203 number of SNPs in the instrument will come increasingly as a result of model over-  
204 specification. We examined the ratio of  $R^2$  for the current instrument to  $R^2$  for an  
205 instrument comprising 30 more SNPs (we term this the 'R-Trend'). The trend in the  
206 ratio gives an indication of the transition from useful additional information to over-  
207 specification since it becomes asymptotic when each new SNP adds the same amount  
208 of information than the last. We judged the beginning of the asymptotic phase of the  
209 line to mark the largest useful instrument obtainable from the available data. The 30  
210 SNP window was chosen empirically because it emphasises trend; a smaller window  
211 gave a more erratic line, obscuring the trend. Calculation and use of the R-trend  
212 effectively limited the analysis to instruments comprising 155 or fewer SNPs, further  
213 restricted to 150 for presentational purposes.

214

215 *ii) Gain from adding a SNP to the instrument*

216 We estimated the benefit to  $R^2$  from adding the current SNP to the previous  
217 instrument by bootstrapping the summary statistics and calculating  $R^2$  for the  
218 instruments with  $n$  and  $(n-1)$  SNPs. Over 10,000 bootstraps we noted the number of  
219 occasions when the current instrument gave higher  $R^2$  than the previous instrument.  
220 This value was summarised as a percentage; the point at which the current instrument  
221 was no better than the previous instrument being when 50% of the runs showed a  
222 benefit.

223

#### 224 *Selection of optimal number of SNPs in genetic instrument*

225 The optimum number of SNPs was chosen by consideration of the R-ratio and the  
226 gain from adding the current SNP when presented graphically (**eFigures 1-6** in the  
227 Supplement). The optimum instrument was identified when both estimates of  $R^2$  gain  
228 were asymptotic. As we discuss later, the exact point ( $\pm 20$  SNPs) makes little  
229 difference to the conclusions. Two authors (JW and MVH) considered this  
230 independently and reached a consensus as to the number of SNPs to include for each  
231 genetic instrument for each lipid trait.

232

#### 233 *Selection of MR model to derive estimates of the underlying relationship*

234 Once we determined the optimal number of SNPs to incorporate in each instrument,  
235 we used the following decision-tree to select the MR approach to derive the estimate:

- 236 (i) if there was no evidence of unbalanced pleiotropy using the intercept  
237 derived from MR-Egger, we selected the conventional MR instrumental  
238 variable estimate as the most reliable indicator to the underlying  
239 relationship (as it retains maximal power and makes fewest assumptions)

240 (ii) if there was evidence of unbalanced pleiotropy, we used the estimate from  
 241 MR-Egger  
 242 (iii) in cases where there was discordance between conventional MR and MR-  
 243 Egger, we used multivariate MR to inform whether differences could arise  
 244 from pathways shared between the three lipids traits.

245

#### 246 *The inSIDE assumption*

247 Because the underlying models assume a linear dose-response, instrumental variable  
 248 (IV) effect estimates must be independent of the exposure effect in MR analysis (the  
 249 so-called ‘inSIDE rule’).<sup>16</sup> We tested the null hypothesis that the instrumental variable  
 250 effect (derived from the ratio of outcome to exposure) estimates for the SNPs in an  
 251 instrument were independent of the exposure (lipid) effect estimates for the same  
 252 SNPs for both CAD and T2D. In all scenarios, the ‘inSIDE’ assumption was satisfied  
 253 (**eTable 1** in the Supplement).

254

#### 255 *Power*

256 We followed the method of Brion et al<sup>24</sup> implemented at  
 257 <http://cnsgenomics.com/shiny/mRnd/>. Using the average number of individuals and  
 258 estimated  $R^2$  for the instrument together with the reported proportion of cases, we  
 259 adjusted the estimate of the true effect of exposure on outcome to obtain the value for  
 260 which we had 80% power at a Bonferroni adjusted alpha of 0.05/6.

261

#### 262 *Ethical Review of Study and Informed Consent of Study Participants*

263 As this report used published GWAS data available in the public domain, specific  
264 ethical review and/or consent from study participants was not sought (and had been  
265 obtained in the original studies).  
266  
267

## 268    **Results**

269

270    The pooled dataset included up to 188,577 individuals with measures of blood lipids,  
271    63,199 CAD cases and 34,840 T2D cases. The optimal number of SNPs for each lipid  
272    traits was 130 for LDL-C (explaining 7.9% of its variance), 140 for HDL-C (6.6% of  
273    HDL-C variance) and 140 for TG (5.9% of TG variance) (**eFigures 1-6** in the  
274    Supplement).

275

276

### 277    *LDL-C*

278    The genetic instrument for LDL-C showed unbalanced pleiotropy for CAD and T2D.  
279    For CAD, the estimate derived from MR-Egger was OR 1.68 (95%CI: 1.51, 1.87) per  
280    1-SD (equivalent to 38 mg/dL) genetically-instrumented higher LDL-C. This was of  
281    greater magnitude, but directionally consistent with conventional and multivariate  
282    MR estimates (**Figure 3** and **eFigure 1** in the Supplement).

283

284    For T2D, the OR was 0.79 (95%CI: 0.71, 0.88) per 1-SD higher LDL-C from MR-  
285    Egger, which, was again of greater magnitude yet directionally consistent with  
286    conventional and multivariate MR estimates (**Figure 3** and **eFigure 2** in the  
287    Supplement).

288

289

### 290    *HDL-C*

291    A 1-SD genetically instrumented elevation in HDL-C (equivalent to 16 mg/dL) did  
292    not provide conclusive evidence of a relationship between HDL-C and risk of CAD.

293 There was evidence of unbalanced pleiotropy of the HDL-C genetic instrument and  
294 the estimate for CAD from MR-Egger was OR 0.95 (95%CI: 0.85, 1.06). There was a  
295 step-wise weakening of the effect towards the null from conventional MR (OR 0.80;  
296 95%CI: 0.75, 0.86), through adjusting for LDL-C and TG in multivariate MR (OR  
297 0.86; 95%CI: 0.78, 0.96) to the MR-Egger estimate (**Figure 3** and **eFigure 3** in the  
298 Supplement).

299

300 For T2D, there was no evidence of unbalanced pleiotropy of the genetic instrument  
301 comprising 140 SNPs. The conventional MR provided an estimate of OR 0.83  
302 (95%CI: 0.76, 0.90), consistent with estimates from both multivariate MR and MR-  
303 Egger (**Figure 3** and **eFigure 4** in the Supplement).

304

305

306 *Triglycerides*

307 The TG genetic instrument showed unbalanced pleiotropy for both CAD and T2D. A  
308 1-SD genetically instrumented increase in TG (equivalent to 89 mg/dL) yielded an  
309 OR for CAD from MR-Egger of 1.28 (95%CI: 1.13, 1.45), weaker than the  
310 multivariate MR estimate and roughly half the magnitude of the conventional MR  
311 estimate (OR 1.49) (**Figure 3** and **eFigure 5** in the Supplement).

312 TG was associated with reduced risk of T2D (OR 0.83; 95%CI: 0.72, 0.95 from MR-  
313 Egger). This was dissimilar to both conventional and multivariate MR estimates  
314 (**Figure 3**). The scatter plot identified that the intercept of the MR-Egger slope was  
315 positive (**eFigure 6** in the Supplement).

316

317 *Power*

318 There was adequate powered to detect the reported estimates (**eTable 2** in the  
319 Supplement), making it unlikely that associations arose from the play of chance.

320

321 *Putting the pieces together: framework of relationships*

322 We demonstrate that elevations in LDL-C, TG and HDL-C are associated with  
323 reduced risk of T2D, with the magnitude (per 1-SD increase) being greatest for LDL-  
324 C, then TG followed by HDL-C (although the 95%CI for the effect on T2D for the  
325 three lipids overlap) (**Figure 4**). In contrast, only LDL-C and TG were associated  
326 with increased risk of CAD (with the magnitude again stronger for LDL-C than TG).

327

328

329

330

## 331 **Discussion**

332

333 We exploited data from multiple GWAS to conduct MR analyses exploring the  
334 relationships between lipids and risk of T2D and CAD. Our findings reveal a series of  
335 relationships that will help inform on potential downstream consequences of  
336 pharmacological modification of lipid levels.

337

338 While all three lipids were associated with reduced risk of T2D, it does not  
339 necessarily follow that lowering of LDL-C or TG through inhibition of specific  
340 druggable proteins (such as PCSK9) will alter risk of T2D. Large-scale genetic and  
341 clinical investigations are needed to clarify the effects of pharmacological lowering of  
342 LDL-C and TG to gauge dysglycaemic associations.<sup>25,26</sup>

343

344 Our findings are complimentary to a study by Fall et al<sup>13</sup>, that, to address pleiotropy,  
345 excluded SNPs showing strong associations with T2D, glycaemia-related traits or  
346 potential confounders such as adiposity. This manual pruning weakened the  
347 associations, yielding inconsistent conclusions. In our study, we applied novel  
348 approaches for: (i) SNP selection (to optimize the SNPs in each genetic instrument);  
349 (ii) MR (using MR-Egger, obviating the need to manually prune SNPs); that  
350 collectively allows us to make more robust conclusions about the role of lipids in  
351 T2D.

352

353 The protective effect of TG and risk of T2D that we report is novel, yet potentially  
354 counter-intuitive. Observational studies report that increases in TG are associated with  
355 an increase in risk of T2D<sup>27</sup>, however insulin resistance results in perturbations in TG

356 metabolism,<sup>28</sup> meaning that the direction of the casual relationship is not clear. While  
357 our data (suggesting TG may be protective of T2D) should be interpreted with  
358 caution, our findings are consistent with recent genetic studies in both Europeans and  
359 African-Americans.<sup>29,30</sup> Further investigations are needed to identify which TG  
360 pathways, if any, may lead to a reduction in risk of T2D.

361

362 LDL-C and TG showed robust effects on risk of CAD, however the evidence for  
363 HDL-C was far less convincing, with the estimate from MR-Egger failing to identify  
364 an effect. This is in keeping with prior MRs<sup>7,8</sup>, including the paper by Voight et al<sup>8</sup>  
365 that manually pruned pleiotropic SNPs. However, selecting only non-pleiotropic  
366 SNPs could introduce selection bias in the genetic instrument by focusing on a subset  
367 of SNPs that is not representative of any meaningful proxy of HDL-C, with the  
368 removal of potentially informative HDL-C related pathways. Our data show that  
369 adjusting for TG and LDL-C in multivariate MR does not fully account for the  
370 unbalanced pleiotropy of the HDL-C genetic instrument. MR-Egger identifies that the  
371 likely underlying relationship is that a genetically-determined higher HDL-C does not  
372 result in a reduced risk of CHD. While these findings are consistent with recent trials  
373 of therapeutics targeting HDL-C,<sup>9,31</sup> this does not preclude the possibility that a drug  
374 modifying HDL-C (or HDL particles) could reduce risk of CAD or other outcomes  
375 such as stroke.

376 The association of TG with CAD recapitulates findings from several prior MR and  
377 genetic studies.<sup>6,7</sup> Of note, the MR-Egger estimate for TG was less than half the  
378 magnitude for an equivalent increase in LDL-C (ORs 1.28 and 1.68 for TG and LDL-  
379 C, respectively, per 1-SD increment). Specific triglyceride-lowering approaches have  
380 had, at best, modest efficacy whereas statin trials have had consistently and potently

381 positive results.<sup>32,33</sup> Our data suggest that pharmacological lowering of TG should  
382 translate into CAD benefit.

383

384 This study has several advantages. First, we use the most up-to-date data available for  
385 lipids to generate the most comprehensive genetic instruments available. Second,  
386 MR-Egger enabled inclusion of GWAS-identified lipid-related SNPs in the genetic  
387 instruments, irrespective of presence of unbalanced pleiotropy. Third, using  
388 summary-level data from different sources represents an efficient study design to  
389 facilitate original investigations such as these without the cost or need for *de novo*  
390 pheno-/genotyping.

391

392 Some limitations are also worthy of note. First, estimates could be sensitive to SNPs  
393 included in the genetic instruments. However, MR estimates were stable at the point  
394 at which we selected the genetic instrument. Second, our MR analyses pertain to  
395 biomarkers rather than specific drug targets. Third, patients targeted for lipid  
396 modification may be at risk for other diseases such as heart failure or atrial fibrillation  
397 – the relevance and direction of effects on these and other endpoints could be  
398 important but were not evaluated here. Fourth, we are not able to account for statin  
399 treatment in the analyses; given that we detect the protective effect of LDL-C on risk  
400 of T2D (the scenario that statins are most likely to confound), major bias is unlikely  
401 to arise in this setting. Finally, while our data casts yet further doubt on the relevance  
402 of HDL-C in the aetiology of CAD, it remains possible that HDL lipoproteins and/or  
403 lipid compositions could play a role in the aetiology of CAD. New methods, such as  
404 <sup>1</sup>H-NMR metabolomics,<sup>34</sup> that quantify lipoprotein subclasses and lipid compositions  
405 are likely to facilitate future MR studies of HDL subclasses.

406

407 In conclusion, our comprehensive MR investigations identify distinct relationships of  
408 major lipid subfractions and risk of CAD and T2D. LDL-C and TG increase risk of  
409 CAD. In contrast, LDL-C and HDL-C are very likely to be protective of T2D with  
410 new evidence suggesting that TG may also play a protective role. While further  
411 studies are needed to examine if specific pathways or lipid subtypes are implicated,  
412 our findings inform on potential expected downstream consequences of intervening  
413 on lipid traits and provide cautionary evidence that therapeutics that lower LDL-C  
414 and TG may have dysglycaemic effects.

415 **Acknowledgements**

416

417

418 **Funding/Support**

419 ZFH is supported by a Genomic Medicine and Statistics Wellcome Trust DPhil

420 Studentship. FWA is supported by a Dekker scholarship-Junior Staff Member

421 2014T001 – Netherlands Heart Foundation and UCL Hospitals NIHR Biomedical

422 Research Centre. SEH is a BHF Professor and is funded by PG08/008, and by the

423 National Institute for Health Research University College London Hospitals

424 Biomedical Research Centre. DIS is supported by a NIHR Academic Clinical

425 Fellowship. ADH is supported by the UCL Hospitals NIHR Biomedical Research

426 Centre and work in his laboratory is supported by BHF Programme and Special

427 Project Grants.

428

429

430 **Conflicts of Interest**

431 CTSU (University of Oxford) is the central co-ordinating centre for the REVEAL trial

432 of anacetrapib; REVEAL is funded through a grant to the University of Oxford by

433 Merck Sharp & Dohme Corp but was designed and is being conducted independently

434 of the funder. DIS is a consultant to Pfizer on work unrelated to the present analysis.

435 All other co-authors report no conflicts of interest. NS reports having received

436 honoraria for advisory boards or lectures for Amgen, Sanofi, Boehringer Ingelheim,

437 Novo Nordisk, Merck, Janssen and Astrazeneca.

438

439 **Role of the funding source**

440 The funding sources had no role in: design and conduct of the study; collection,  
441 management, analysis, and interpretation of the data; preparation, review, or approval  
442 of the manuscript; and decision to submit the manuscript for publication.

443

#### 444 **Access to data and Data Analysis**

445 JW had full access to all the data in the study and takes responsibility for the integrity  
446 of the data and the accuracy of the data analysis.

447

#### 448 **Author contributions**

449 All authors satisfy ICMJE requirements for author contribution.

450

#### 451 **Originality of Content**

452 All information and materials in the manuscript are original.

453

454

## 455     **References**

- 456     1.     Stein EA, Mellis S, Yancopoulos GD, et al. Effect of a monoclonal antibody  
457     to PCSK9 on LDL cholesterol. *The New England journal of medicine*  
458     2012;366:1108-18.
- 459     2.     Akdim F, Stroes ES, Sijbrands EJ, et al. Efficacy and safety of mipomersen,  
460     an antisense inhibitor of apolipoprotein B, in hypercholesterolemic subjects receiving  
461     stable statin therapy. *Journal of the American College of Cardiology* 2010;55:1611-8.
- 462     3.     Cannon CP, Blazing MA, Giugliano RP, et al. Ezetimibe Added to Statin  
463     Therapy after Acute Coronary Syndromes. *The New England journal of medicine*  
464     2015.
- 465     4.     Cholesterol Treatment Trialists C, Baigent C, Blackwell L, et al. Efficacy and  
466     safety of more intensive lowering of LDL cholesterol: a meta-analysis of data from  
467     170,000 participants in 26 randomised trials. *Lancet* 2010;376:1670-81.
- 468     5.     Cholesterol Treatment Trialists C, Mihaylova B, Emberson J, et al. The effects  
469     of lowering LDL cholesterol with statin therapy in people at low risk of vascular  
470     disease: meta-analysis of individual data from 27 randomised trials. *Lancet*  
471     2012;380:581-90.
- 472     6.     Do R, Willer CJ, Schmidt EM, et al. Common variants associated with plasma  
473     triglycerides and risk for coronary artery disease. *Nature genetics* 2013;45:1345-52.
- 474     7.     Holmes MV, Asselbergs FW, Palmer TM, et al. Mendelian randomization of  
475     blood lipids for coronary heart disease. *European heart journal* 2014.
- 476     8.     Voight BF, Peloso GM, Orho-Melander M, et al. Plasma HDL cholesterol and  
477     risk of myocardial infarction: a mendelian randomisation study. *Lancet*  
478     2012;380:572-80.
- 479     9.     Schwartz GG, Olsson AG, Abt M, et al. Effects of dalcetrapib in patients with  
480     a recent acute coronary syndrome. *The New England journal of medicine*  
481     2012;367:2089-99.
- 482     10.    Preiss D, Seshasai SR, Welsh P, et al. Risk of incident diabetes with intensive-  
483     dose compared with moderate-dose statin therapy: a meta-analysis. *Jama*  
484     2011;305:2556-64.
- 485     11.    Swerdlow DI, Preiss D, Kuchenbaecker KB, et al. HMG-coenzyme A  
486     reductase inhibition, type 2 diabetes, and bodyweight: evidence from genetic analysis  
487     and randomised trials. *Lancet* 2014.
- 488     12.    Ridker PM, Pradhan A, MacFadyen JG, Libby P, Glynn RJ. Cardiovascular  
489     benefits and diabetes risks of statin therapy in primary prevention: an analysis from  
490     the JUPITER trial. *Lancet* 2012;380:565-71.
- 491     13.    Fall T, Xie W, Poon W, et al. Using Genetic Variants to Assess the  
492     Relationship Between Circulating Lipids and Type 2 Diabetes. *Diabetes*  
493     2015;64:2676-84.
- 494     14.    Haase CL, Tybjaerg-Hansen A, Nordestgaard BG, Frikke-Schmidt R. High-  
495     density lipoprotein cholesterol and risk of type 2 diabetes: a Mendelian randomization  
496     study. *Diabetes* 2015.
- 497     15.    Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian  
498     randomization: using genes as instruments for making causal inferences in  
499     epidemiology. *Stat Med* 2008;27:1133-63.
- 500     16.    Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid  
501     instruments: effect estimation and bias detection through Egger regression. *Int J*  
502     *Epidemiol* 2015;44:512-25.

17. Burgess S, Dudbridge F, Thompson SG. Re: "Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects". *Am J Epidemiol* 2015;181:290-1.
18. Day FR, Loh PR, Scott RA, Ong KK, Perry JR. A Robust Example of Collider Bias in a Genetic Association Study. *American journal of human genetics* 2016;98:392-3.
19. Global Lipids Genetics C, Willer CJ, Schmidt EM, et al. Discovery and refinement of loci associated with lipid levels. *Nature genetics* 2013;45:1274-83.
20. Morris AP, Voight BF, Teslovich TM, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nature genetics* 2012;44:981-90.
21. Consortium CAD, Deloukas P, Kanoni S, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nature genetics* 2013;45:25-33.
22. Johnson T. Efficient Calculation for Multi-SNP Genetic Risk Scores. <http://cran.r-project.org/web/packages/gtx/vignettes/ashg2012.pdf> presented at the American Society of Human Genetics Annual Meeting, San Francisco, November 6–10, 2012. 2012.
23. Burgess S, Thompson SG. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *Am J Epidemiol* 2015;181:251-60.
24. Brion MJ, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. *Int J Epidemiol* 2013;42:1497-501.
25. Bonnefond A, Yengo L, Le May C, et al. The loss-of-function PCSK9 p.R46L genetic variant does not alter glucose homeostasis. *Diabetologia* 2015.
26. Tragante V, Asselbergs FW, Swerdlow DI, et al. Harnessing publicly available genetic data to prioritize lipid modifying therapeutic targets for prevention of coronary heart disease based on dysglycemic risk. *Hum Genet* 2016.
27. Tirosh A, Shai I, Bitzur R, et al. Changes in triglyceride levels over time and risk of type 2 diabetes in young men. *Diabetes care* 2008;31:2032-7.
28. Ginsberg HN, Zhang YL, Hernandez-Ono A. Regulation of plasma triglycerides in insulin resistance and diabetes. *Archives of medical research* 2005;36:232-40.
29. Klimentidis YC, Wineinger NE, Vazquez AI, Campos Gdl. Multiple Metabolic Genetic Risk Scores and Type 2 Diabetes Risk in Three Racial/Ethnic Groups. *The Journal of Clinical Endocrinology & Metabolism* 2014;99:E1814-E8.
30. Klimentidis YC, Chougule A, Arora A, Frazier-Wood AC, C-H H. Triglyceride-Increasing Alleles Associated with Protection against Type-2 Diabetes. *. PLoS genetics* 2015:e1005204.
31. Lilly to Discontinue Development of Evacetrapib for High-Risk Atherosclerotic Cardiovascular Disease <https://investor.lilly.com/releasedetail.cfm?ReleaseID=936130>. 2015.
32. Preiss D, Sattar N. Lipids, lipid modifying agents and cardiovascular risk: a review of the evidence. *Clin Endocrinol (Oxf)* 2009;70:815-28.
33. Lee M, Saver JL, Towfighi A, Chow J, Ovbiagele B. Efficacy of fibrates for cardiovascular risk reduction in persons with atherogenic dyslipidemia: A meta-analysis. *Atherosclerosis*;217:492-8.
34. Soininen P, Kangas AJ, Wurtz P, Suna T, Ala-Korpela M. Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. *Circ Cardiovasc Genet* 2015;8:192-206.

## 553 **Figure Legends**

### 554 **Figure 1. Pleiotropy and the validity of estimates derived from Mendelian** 555 **randomization.**

556 Figure legend: SNPs are used in a genetic instrument for an exposure to assess the  
557 association with risk of disease. For each exposure there is a ‘true relationship’, which  
558 we try to approximate from Mendelian randomization. For the purposes of simplicity,  
559 conventional MR is compared to MR-Egger.

560 **Vertical pleiotropy** explains where the genetic instrument associates with biomarkers  
561 (other than the exposure) that are on the causal pathway from exposure through to  
562 disease. **Horizontal pleiotropy** is where the genetic instrument associates with  
563 additional traits not on the causal pathway of the exposure of interest. When  
564 horizontal pleiotropy is **balanced**, there should be no bias in the effect derived from  
565 MR. In this scenario, the estimate obtained from conventional MR is similar to that  
566 from MR-Egger.

567 When horizontal pleiotropy is **unbalanced** (also termed ‘directional pleiotropy’), the  
568 pleiotropy systematically biases the estimate (which can be exaggerated or  
569 diminished) in a naïve analysis using conventional MR. In the example in Figure 1,  
570 the unbalanced pleiotropy exaggerates the magnitude of the association. Conventional  
571 MR will derive a biased estimate, whereas MR-Egger, correcting for unbalanced  
572 pleiotropy, should yield a valid estimate. An example of unbalanced horizontal  
573 pleiotropy is the relationship of HDL-C and risk of CAD; the association derived  
574 from conventional MR is different to that of MR-Egger with the latter indicating that,  
575 once unbalanced pleiotropy is accounted for, there is no effect of HDL-C on risk of  
576 CAD (see **Figure 3**).  
577

### 578 **Figure 2. Pipeline for derivation of the dataset used for Mendelian**

### 579 **randomization analyses of lipid subtypes with risk of coronary artery disease** 580 **and diabetes.**

581

### 582 **Figure 3. Associations of routinely measured lipids with risk of coronary artery** 583 **disease (CAD) and type 2 diabetes (T2D) from Mendelian randomization** 584 **analyses.**

585 Figure legend: See Methods for description of the three Mendelian randomization  
586 (MR) models. Estimates for conventional MR are derived from two-sample MR that  
587 forces the slope through the origin, thereby not accounting for pleiotropy.  
588 Multivariate MR (MVMR) statistically adjusts for other lipid traits, and MR-Egger  
589 adjusts for unbalanced pleiotropy.  $R^2$  refers to proportion of variance of lipid trait  
590 explained by the genetic instrument. 95% confidence intervals (CI) are Bonferroni-

591 adjusted. To convert HDL-C and LDL-C to mmol/L, multiply by 0.0259; to convert  
592 triglycerides to mmol/L, multiply by 0.0113

593

594

595

596 **Figure 4. Cross-hair plot of a one standard deviation increase in lipids and risk**

597 **of CAD and T2D.**

598 Figure legend: All estimates derived from MR-Egger. Error bars represent 95%  
599 confidence intervals (CI) that are Bonferroni-adjusted.

600

601

602

603 **Tables.**

604 **Table 1. Details of the consortia**

605

Consortium name	Trait/ Disease	Numbers	Data Source; file
GLGC <sup>19</sup>	LDL-C, HDL-C, TG	188,577	<a href="http://www.sph.umich.edu/csg/abecasis/public/lipids2013">http://www.sph.umich.edu/csg/abecasis/public/lipids2013</a>
CARDIoGRAMplusC4D <sup>21</sup>	CAD	63,746 CAD cases, 130,681 controls	<a href="http://www.cardiogramplusc4d.org">http://www.cardiogramplusc4d.org</a>
DIAGRAM <sup>20</sup>	T2D	34,840 T2D cases and 114,981 controls	<a href="http://diagram-consortium.org">http://diagram-consortium.org</a> (v3 dataset)

606

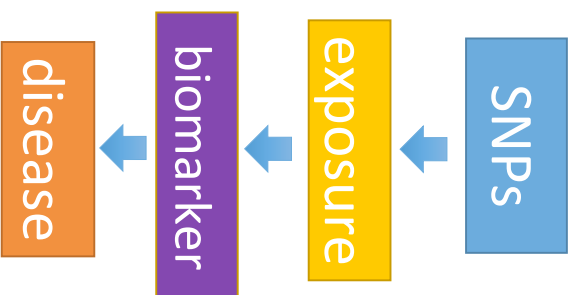
607 Abbreviations: CAD: coronary artery disease; CARDIoGRAMplusC4D: Coronary ARtery Disease Genome-wide

608 Replication And Meta Analysis (CARDIoGRAM) plus The Coronary Artery Disease (C4D) Genetics; DIAGRAM:

609 DIAbetes Genetics Replication And Meta-analysis; GLGC: Global Lipids Genetic Consortium; T2D: type 2 diabetes

610

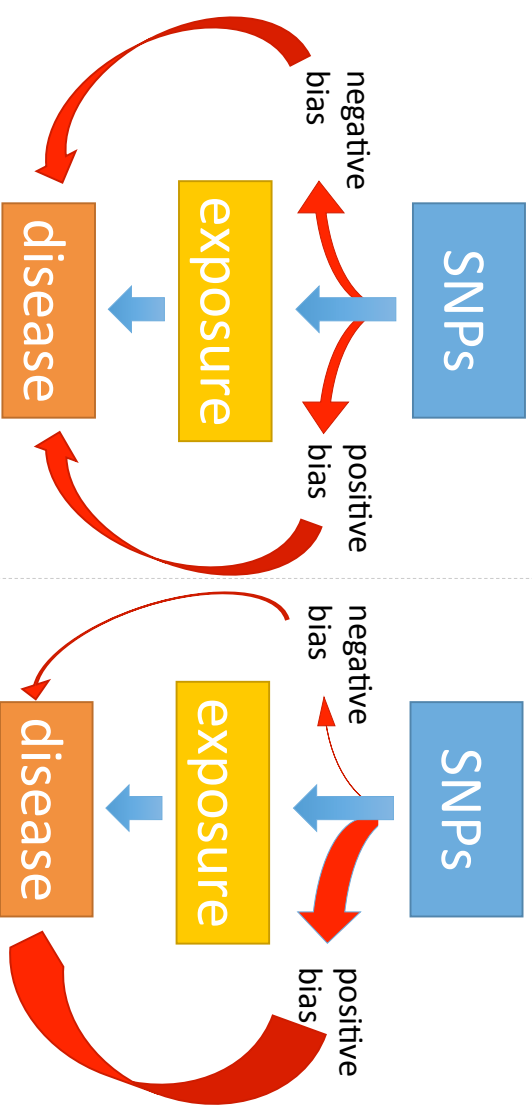
## VERTICAL PLEIOTROPY



## HORIZONTAL PLEIOTROPY

Balanced

Unbalanced



True relationship

Conventional MR

MR-Egger

