



Holmes, M. V., Richardson, T. G., Ference, B. A., Davies, N. M., & Davey Smith, G. (2021). Integrating genomics with biomarkers and therapeutic targets to invigorate cardiovascular drug development. *Nature Reviews Cardiology*, 18(6), 435-453.  
<https://doi.org/10.1038/s41569-020-00493-1>

Peer reviewed version

Link to published version (if available):  
[10.1038/s41569-020-00493-1](https://doi.org/10.1038/s41569-020-00493-1)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via [dx.doi.org/https://doi.org/10.1038/s41569-020-00493-1](https://doi.org/10.1038/s41569-020-00493-1). Please refer to any applicable terms of use of the publisher.

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

## ***Nature Reviews* referee guidelines**

### **Review articles**

*Nature Reviews* publishes timely, authoritative articles that are of broad interest and exceptional quality. Thank you for taking the time to help us to ensure that our articles meet these high standards.

Review articles in *Nature Reviews* journals provide accessible, authoritative and balanced overviews of a field or topic. These articles are targeted towards readers from advanced undergraduate level and upwards, including researchers, academics and clinicians, and should be accessible to readers working in any discipline.

Please submit your report in narrative form and provide detailed justifications for all statements. Confidential comments to the editor are welcome, but it is helpful if the main points are stated in the comments for transmission to the authors.

Please note that all *Nature Reviews* articles will be thoroughly edited before publication and all figures will be redrawn by our in-house art editors. We therefore request that you concentrate on the scientific content of the article, rather than any minor errors in language or grammar.

### **Please consider and comment on the following points when reviewing this manuscript:**

- Is the article timely and does it provide a useful addition to the existing literature?
- Are the scope and aims of the article clear?
- Are the ideas logically presented and discussed?
- Is the article accessible to a wide audience, including readers who are not specialists in your own field?
- Does the article provide a balanced overview of the literature? Please bear in mind that it may not be possible to cover all aspects of a field within such a concise article.
- Does the article provide new insight into recent advances?
- Is the discussion fair and accurate? Although our authors are encouraged to be opinionated, they should not ignore alternative points of view.
- Do the figures, boxes and tables provide clear and accurate information? Are there any additional or alternative display items that you think that the authors should include?
- Are the references appropriate and up-to-date? Do they reflect the scope of the article?
- Are you aware of any undeclared conflicts of interest that might affect the balance, or perceived balance, of the article?

1 **Integrating genomics with biomarkers and therapeutic targets to invigorate cardiovascular**  
2 **drug development**

3 Michael V. Holmes<sup>1\*</sup>, Tom G Richardson<sup>2</sup>, Brian A. Ference<sup>3</sup>, Neil M Davies<sup>2,4</sup> and George Davey  
4 Smith<sup>2</sup>

5 <sup>1</sup>Medical Research Council Population Health Research Unit at the University of Oxford, Oxford,  
6 UK.

7 <sup>2</sup>Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, UK.

8 <sup>3</sup>Centre for Naturally Randomised Trials, University of Cambridge, Cambridge, UK.

9 <sup>4</sup>K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, NTNU,  
10 Norwegian University of Science and Technology, Norway.

11 \*e-mail: [michael.holmes@ndph.ox.ac.uk](mailto:michael.holmes@ndph.ox.ac.uk)

12

13 Abstract | Drug development in cardiovascular disease is stagnating, with lack of validated  
14 targets a major roadblock to innovation. Human genetics can provide compelling evidence of  
15 causation through approaches such as Mendelian randomization (MR). Such genetic support  
16 can increase the probability of a clinical trial succeeding. MR is being widely applied to  
17 quantitative traits (e.g. adiposity and blood lipids) with the aim of identifying risk factors for  
18 disease that are both causal and amenable to therapeutic modification. However, there are  
19 important differences between genetic investigations of a biomarker (such as HDL-cholesterol)  
20 and a drug target that seeks to modify the same biomarker of interest (such as CETP), with  
21 implications on the methodology, interpretation and applications of MR to drug development.  
22 Differences include the comparative nature of the genetic architecture, i.e. that of a biomarker

23 typically being polygenic, whereas that for a protein drug target being monogenic. and the  
24 potential for drug targets to show pleiotropic disease associations which may differ to those of  
25 the biomarker it is designed to modify. In this review, we compare and contrast the role of MR  
26 as applied to quantitative traits versus drug targets, describing differences between approaches  
27 and the function each plays. Our review will put into context findings emerging from this  
28 burgeoning field of scientific investigation, with the aim of clarifying the aetiological roles of  
29 biomarkers in disease, prioritizing candidate drug targets based on genetic-guided credibility of  
30 causation and in so doing, improve the odds of success in the drug development pipeline.

31

32

### 33 **Introduction**

34 Cardiovascular disease (CVD), comprising ischemic heart disease, stroke and peripheral vascular  
35 disease, represents the leading cause of death globally, accounting for 15.2 million deaths (or ~  
36 1 in 3 deaths) in 2016 alone <sup>1</sup>. While non-pharmacological prevention of CVD is a key  
37 component of public health approaches to preventing onset and sequelae of CVD, a  
38 cornerstone to disease prevention and treatment is the use of pharmacological therapies.

39 In order for drugs to become licensed and available to patients, they must undergo  
40 rigorous investigations, including early studies of tolerability, followed by phase II trials  
41 (typically dose-ranging in nature) and pivotal phase III outcome randomized controlled trials  
42 which serve to show clinical efficacy for the intended outcome (in this case, ideally a reduction  
43 in risk of cardiovascular disease) while showing lack of clinically-meaningful adverse events.  
44 Such phase III trials typically represent the final hurdle prior to applying for a marketing

45 authorization. If the drug is found to be efficacious for the primary outcome and safe, the  
46 regulators are likely to grant the applicant a license to market the drug. For interventions for  
47 cardiovascular disease, the process, from date of first testing in humans (clinical entry) through  
48 to marketing authorization takes a median of 10 years<sup>2</sup>, with this duration remaining largely  
49 unchanged since the 1970s<sup>2</sup>.

50 A notable trend observed over recent decades is that the number of drugs being  
51 developed for cardiovascular disease has stagnated.<sup>3</sup> While the average number of new drug  
52 approvals for any indication have fluctuated between 25/year (from 2000-2009) to 41/year  
53 (from 2010-2018)<sup>4</sup>, the corresponding values for cardiovascular indications were 2.6/year and  
54 3.2/year<sup>2</sup>. This has occurred on the background of a rapid increase in R&D expenditures in real  
55 terms<sup>3</sup> (with the median cost of bringing a drug to market between 2009-2018 being  
56 approximately \$1bn<sup>5</sup>), which together with large numbers of generic drugs being approved,  
57 collectively impedes the return on investment for new drug development. One of the major  
58 issues in developing drugs is failure due to lack of efficacy (accounting for ½ of phase II and III  
59 failures)<sup>6</sup>, with therapies targeting risk factors that are not causal representing a major  
60 roadblock to innovation<sup>3</sup>. Another major hurdle is safety (accounting for ¼ of failures<sup>6</sup>), with  
61 therapies being abandoned due to both dose-related and idiosyncratic adverse events. It is  
62 largely for these reasons that fewer than 1 in 10 drugs succeed in running the gauntlet from  
63 phase I through to marketing authorization, with the success rate being disease-specific<sup>7</sup>. Of  
64 note, the comparative probability of success (i.e. of a drug being launched) when a candidate  
65 drug enters each phase of clinical trials is lower when the therapeutic indication is  
66 cardiovascular disease as compared to other therapeutic areas<sup>8</sup>. Between 2010-2017, there was

67 a 4% probability of launch for cardiovascular drugs entering phase I RCTs versus 9% for  
68 anticancer therapies. For drugs entering phase II RCTs the comparable probabilities of launch  
69 were 10% for those with cardiovascular indications and 23% for anticancer indications and for  
70 drugs entering phase III, values were 46% for cardiovascular versus 59% for anticancer.<sup>8</sup>

71 Human genetics can provide a tool by which the causal role of an exposure (such as a  
72 complex phenotype and/or a drug target) can be characterized<sup>9</sup>. Performed well, such studies  
73 called Mendelian randomization (MR) analyses, can provide rigorous evidence of causation<sup>9,10</sup>.  
74 The presence of genetic evidence to support causation increases the probability several-fold of  
75 a therapeutic targeting the same biomarker being successful in a phase III outcome trial and  
76 obtaining a marketing authorisation<sup>11 12</sup>. With increasing availability of large-scale phenotyping  
77 (and genotyping), MR is being widely applied to quantitative traits (e.g. markers of adiposity,  
78 such as body mass index<sup>13-16</sup>; blood lipids such as HDL-cholesterol<sup>17-19</sup>; inflammation-related  
79 proteins such as interleukin-6 receptor<sup>20,21 12</sup>) with the aim of discovering modifiable risk  
80 factors for disease that are causal and amenable to therapeutic modification. Such risk factor  
81 biomarkers are typically modified pharmacologically through drug targets (the terms  
82 biomarkers and drug targets are described below). However, important differences exist  
83 between genetic investigations of a quantitative trait (such as HDL-cholesterol) as compared to  
84 a drug target (such as CETP<sup>22-25</sup>), with major implications on their interpretation, role and  
85 applications to drug development.

86 In this Review, we compare and contrast the role of MR as applied to quantitative  
87 polygenic traits versus (typically monogenic) drug targets, describing the main differences  
88 between the approaches and the function each plays in guiding the process of drug

89 development. Our Review puts into context findings emerging from this burgeoning field of  
90 scientific investigation. We aim to build on our previous review<sup>26</sup> which described challenges in  
91 interpreting estimates from MR in cardiovascular disease but which did not contrast between  
92 therapeutic targets and complex biomarkers.<sup>26</sup> We draw themes across historical<sup>27 28 29</sup> and  
93 more contemporary<sup>30 31 32</sup> articles that describe the application of human genetics to  
94 therapeutic targets. We comparatively appraise the motivation, application, methodologies and  
95 interpretation of MR as applied to drug-targets versus complex biomarkers, including an  
96 evaluation of when findings from the two aetiological approaches (i.e. MR of a drug target  
97 versus MR of a complex biomarker) may diverge. Our overarching ambition is to guide the  
98 investigator who is concerned with elucidating the aetiological roles of biomarkers in disease,  
99 prioritizing candidate drug targets based on credible genetic evidence of causation and in doing  
100 so, attempting to improve the success and thus efficiency of the drug development pipeline.

101

### 102 **Mendelian randomization to elucidate causality**

103 We can exploit the properties of the human genome (specifically, the random segregation of  
104 alleles and their non-modifiable nature) to make causal inferences about the relationships  
105 between risk factors and disease<sup>10</sup>. The process by which genetic variants are inherited in  
106 offspring from their parents occurs at random, owing to Mendel's law of independent  
107 assortment. Within families, this process ensures that genetic variants are unlikely to be  
108 associated with confounding factors that can results in bias in other study designs. For this  
109 reason, when MR was introduced by analogy with a randomized controlled trial, it was in the  
110 context of between-sibling studies, owing to the true randomization of genetic information

111 (outside those in linkage).<sup>33 9</sup> However, most MR studies to date have used datasets of  
112 unrelated individuals and make the assumption that what is true within families (i.e. the  
113 random inheritance of genetic variants from parents to offspring), holds across the population  
114 (i.e. the random distribution of variants)<sup>34</sup>. In other words, the assumption is made that when  
115 individuals are grouped in the population by their genotype, they should be similar in all  
116 regards other than one group possesses one or more genetic variant(s) and the other group  
117 another set of variants – if those genetic variants influence an exposure of interest, the  
118 comparison of disease associations across the groups permits causal inference. However, there  
119 is increasing evidence that genetic variants associated with some phenotypes are not entirely  
120 randomly distributed across the population in relation to potential confounding factors, e.g. for  
121 social or behavioural phenotypes such as educational attainment, intelligence and smoking<sup>35</sup>.  
122 This is (at least partially) due to what have been called “dynastic effects” that relate to the  
123 environment provided by parents to their offspring<sup>35 36</sup>. Of note, phenotypes that are more  
124 genetically proximal (e.g. gene expression, protein levels and many biomarkers; see **Figure 1**),  
125 are likely to be considerably less prone to these biases.

126 Many of the benefits and potential limitations of MR have been extensively described in prior  
127 articles<sup>9,26,37,3836</sup>. There have been a number of cases when MR has retrospectively found  
128 similar findings to that of randomized controlled trials (RCTs) (e.g. statins<sup>39</sup> and drug-target  
129 MRs using LDL-C lowering variants in/around *HMGCR*<sup>40</sup>; varespladib which pharmacologically  
130 targets secretory phospholipase A2-IIA<sup>41</sup> and genetic variants in *PLA2G2A*<sup>42</sup>) and there are  
131 cases where MR has predicted the outcome of trials too (such as use of genetic variants in  
132 *CETP*<sup>23</sup> predicting the outcomes of a clinical trial of anacetrapib, a cholesteryl ester transfer



133 protein (CETP) inhibitor) including findings that have yet transpire clinically (e.g. despite LDL-C  
134 lowering genetic variants in/around *PCSK9* showing associations with higher risk of T2D<sup>43,44</sup>, an  
135 excess risk of T2D has yet to reliably transpire from patients treated with PCSK9 inhibitors<sup>45</sup>).

136 Other examples of notable MR studies are described in **Box 1**. Concerns about the plausibility of  
137 the assumptions and overuse of MR have been raised.<sup>46,36</sup> For example, there are well-founded  
138 concerns over the challenges of interpreting the implications of life-long estimates derived from  
139 MR in the context of pharmaceutical interventions that are typically used later in life (e.g. the  
140 issue of ‘time zero’)<sup>47</sup>, although pragmatically, well-conducted MR studies can (and often do)  
141 provide credible causal evidence<sup>10</sup>. There are multiple cases where MR can complement RCTs  
142 and examples where MR can enhance information that would require several phase III RCTs of  
143 multiple doses of drugs— e.g. providing evidence of dose-response relationships across the  
144 physiological distribution of a trait (described in **Box 2**). In a nutshell, MR can be used  
145 throughout the lifecycle of a drug, as illustrated in **Figure 2**.

146

### 147 **Why focus on biomarkers?**

148 Biomarkers are phenotypes that can be measured objectively and provide quantifiable  
149 indications of biological processes<sup>48</sup>. The study of the roles that biomarkers play in causing  
150 disease is of interest for several reasons. First, such information provides aetiological  
151 knowledge, which enriches our understanding of biological processes that underpin disease and  
152 can stimulate new avenues of investigation. For example, whilst inflammation in general has  
153 long been postulated to influence vascular disease<sup>49,50</sup>, MR has been able to test whether  
154 particular elements in the inflammatory response cause<sup>20</sup> or are caused by<sup>51</sup> disease. This

155 evidence facilitates exploration of other inflammatory biomarkers, new therapeutic targets<sup>52</sup>  
156 and repurposing of existing drugs to new indications<sup>53</sup>. Second, an understanding of how  
157 biomarkers cause disease can help clarify the sequence of intermediaries that lie on a causal  
158 pathway from a risk factor through to disease, permitting development of therapies that act on  
159 risk factors along the causal pathway to ameliorate the consequences of harmful exposures (an  
160 example being drugs for hypertension that inhibit angiotensin converting enzyme and  
161 angiotensin II receptor antagonists, which act along the renin-angiotensinogen system  
162 pathway). This is illustrated in **Figure 3** using inflammation as an example, and serves to  
163 highlight the value of randomized evidence for prioritising where on the causal pathway to  
164 intervene (as pharmacological modification of C-reactive protein is unlikely to be beneficial to  
165 CHD <sup>51</sup>, whereas randomized evidence demonstrates the value of targeting biomarkers further  
166 upstream, e.g. interleukin-6 receptor <sup>21,54</sup> or interleukin 1-beta <sup>52</sup>). A further example is the  
167 causal role that adiposity plays in CHD: adiposity increases systolic blood pressure<sup>13,16</sup> which  
168 mediates part of the effect of BMI on risk of CAD <sup>55</sup>. Mendelian randomization can elucidate  
169 such causal intermediaries and help develop new pharmacological approaches to ameliorating  
170 the effects of harmful exposures that are challenging to modify at the population level (see **Box**  
171 **3** and **Figure 4**).

172 Third, when drug targets modify several traits, an exploration of the comparative role  
173 that such traits plays in disease aetiology through multivariable MR approaches can shed light  
174 on the likely causal pathway to disease<sup>56</sup>. For example, blood lipid traits (including  
175 apolipoprotein B, low-density lipoprotein cholesterol and triglycerides) each associate  
176 independently with risk of CHD. Conducting MR analyses individually on these traits shows that

177 each is related to a higher risk of CHD. However, when appraised together, the evidence  
178 supports apolipoprotein B as being necessary for lipid-mediated atherogenesis by LDL-C to  
179 occur.<sup>57,58</sup> This can help pinpoint the trait(s) scientists should focus on when developing lipid-  
180 modifying drugs to treat CHD – i.e. all things being equal, apolipoprotein B can potentially be  
181 used as a reliable surrogate marker for the expected reduction in relative risk of CHD for a novel  
182 lipid-modifying therapy. Fourth, for environmental exposures (e.g. alcohol), knowledge of  
183 causation can help frame public health policy. In this case, a trial would be challenging and/or  
184 unethical to conduct, despite previous attempts, however MR provides credible information on  
185 causation that can inform public health policies.<sup>59 60</sup>

186

### 187 **Why focus on drug targets?**

188 While valuable from an aetiological perspective (and for public health), it is generally infeasible  
189 to pharmacologically modify a non-protein biomarker in isolation. This is because therapeutic  
190 targets tend to modify phenotypes through specific mechanisms that are unique and thus  
191 dissimilar to the totality of the genetic/environmental architecture of the trait (as in the  
192 example of cholesteryl ester transfer protein (CETP) inhibitors and high-density lipoprotein  
193 cholesterol (HDL-C)<sup>22</sup>, illustrated in **Figure 5**).

194 Thus, the MR estimates of the effects of a complex biomarker may differ from the effects of a  
195 protein drug target that is under investigation on the basis that it modifies the same complex  
196 biomarker. For this reason, in order to potentially inform on the expected clinical effects of a  
197 therapy, we ideally need to conduct de novo MR studies of therapeutic targets that modify the  
198 biomarker of interest. In other words, MR of drug targets is useful because it uses genetic

199 variants that mimic as closely as possible the action of the therapeutic target of interest. This  
200 can potentially provide some of the most reliable evidence of the effects of modifying the  
201 biomarker, via a specific target, on long-term health outcomes. Beyond simply providing  
202 quantitative evidence of whether a therapy targeting a given drug target is likely to be  
203 efficacious and safe, drug target MR can provide a treasure trove of additional information that  
204 can feed into multiple facets of trial design (summarised in **Box 2**).

### 205 206 **Biomarker versus drug target**

207  
208 There is an important distinction between a complex (polygenic) biomarker and an (oligogenic)  
209 drug target, which we outline in **Figure 1**. This has a bearing on the motivation, conduct,  
210 application and interpretation of MR studies as applied to complex biomarkers versus drug  
211 targets (**Table 1** and **Box 4**). The key question and motivation of a biomarker MR is typically  
212 whether the biomarker plays a role in disease, which is of interest from a biological perspective  
213 (as it may highlight new approaches to preventing disease) and from a public health perspective  
214 (as lifestyle modification mediated alterations in the biomarker may yield large benefits in  
215 public health <sup>61</sup>). In contrast, the key question and motivation of drug-target MR is whether  
216 modifying the drug target alters risk of disease and thus whether a phase III clinical trial of a  
217 therapy that modifies the drug target is likely to demonstrate clinical benefit to patients,  
218 enabling applications for marketing authorisations and health technology appraisals. Other  
219 features and distinctions between biomarker and drug target MR are described in **Table 1** and  
220 throughout this review.

221

222 We consider a complex biomarker to be a non-protein trait (e.g. metabolites and traits such as  
223 adiposity, blood pressure and lipids; **Figure 1**) that is not encoded for by any individual gene.

224 Complex traits have polygenic architectures, with multiple genetic variants affecting the  
225 biomarker through discrete and/or overlapping pathways.

226

227 Most drug targets are proteins. Both conventional (small molecules) and newer therapies (e.g.  
228 biological monoclonal antibodies, antisense RNA inhibitors, and the consequences of gene  
229 editing) typically target proteins. Thus, if the therapeutic aim is to treat or prevent disease  
230 through the pharmacological modification of a complex biomarker, this is typically conducted  
231 by exploiting the vertically pleiotropic effects of the protein on the complex biomarker (see **Box**  
232 **5** which describes pleiotropy as applied to drug target MR). Proteins are typically encoded for  
233 by a unique (single) gene making it theoretically possible to generate a genetic instrument for  
234 the protein of interest that is specific (so-called 'cis' protein instruments that are close to, or in,  
235 the relevant gene, see **Box 6**). Thus a genetic instrument utilising cis-acting variants for a drug  
236 target is typically monogenic but can sometimes be oligogenic if more than one region of the  
237 genome (as in SNPs from more than one locus) is involved in encoding the final protein product.  
238 For example, haemoglobin is a heterotetramer that is made up of alpha and beta globin  
239 polypeptides, each encoded by discrete genes<sup>62</sup>. Thus a drug-target MR of haemoglobin  
240 selecting cis-acting variants from across discrete protein-encoding genes would be oligogenic,  
241 in contrast to a monogenic instrument for CRP<sup>51</sup>, encoded for by a single locus. Notably, if more  
242 than one SNP is used as an instrument from the same locus, this remains a monogenic  
243 instrument.

244 Cis variants typically associate with protein concentrations in the blood or gene expression in  
245 tissues (**Box 7**). Cis variants for proteins (cis-protein quantitative trait loci; cis-pQTL) are  
246 generally preferable for use as instruments in drug-target MR as compared to cis variants for  
247 mRNA expression (cis-expression quantitative loci; cis-eQTL) as the potential for pre-  
248 translational pleiotropy<sup>32</sup> (i.e. horizontal pleiotropy that arises prior to translation of mRNA into  
249 proteins) is potentially diminished when selecting variants based on their properties as cis-  
250 pQTLs versus cis-eQTLs. Another reason is that protein concentrations can act as a marker of  
251 drug efficacy, allowing a quantitative appraisal of effects of drug versus genetic instrument,  
252 which may be particularly valuable when the causal intermediaries (i.e. biomarkers  
253 downstream of the drug target which mediate the effect on disease risk) are unknown. Such  
254 cis-acting variants are likely to provide an accurate and reliable estimate of the causal effects of  
255 modifying that protein. However, there are specific caveats to the use of *cis* acting variants – for  
256 example the issue of *cis*-pQTLs altering epitope binding sites leading to potential false negative  
257 findings (as described in **Table 2**) and to protein MRs in general – e.g. the issue of targeting  
258 proteins (e.g. IL6) versus their receptors (e.g. IL6R). In this latter example, genetic variants used  
259 to mimic pharmacological inhibition of IL6R (and thus the effects of inhibiting pathways  
260 downstream of IL6) lead to a paradoxical increase in IL6 concentrations in the blood<sup>21</sup>, with the  
261 interpretation of such findings described in more detail elsewhere<sup>26</sup>. While genetic variants  
262 genomically distant from the protein-encoding gene (so-called *trans* variants) may associate  
263 with protein levels at conventional GWAS significance thresholds, they are more likely to have  
264 horizontally pleiotropic effects (e.g. by affecting the expression of multiple genes), making the  
265 interpretation of MR estimates based on trans-acting variants challenging (**Figure 6**). An

266 example here is the use of SNPs to instrument C-reactive protein from regions of the genome  
267 other than the *CRP* locus, which might lead to the erroneous conclusion that C-reactive protein  
268 causes CHD<sup>63</sup> – when restricting to SNPs within the *CRP* locus, no such effect of C-reactive  
269 protein on risk of CHD is identified<sup>51</sup>.

270

### 271 **What does this mean in practice?**

272

273 Consider pharmacological therapies that seek to lower risk of CVD through the  
274 modification of a biomarker, LDL cholesterol. Such pharmacological interventions work by  
275 targeting proteins and enzymes involved in the synthesis or metabolism of LDL cholesterol. LDL  
276 cholesterol lowering pharmacological drugs are thus unlikely to solely alter LDL cholesterol in  
277 physiological isolation, but rather, target a specific part of its synthesis and/or metabolism.  
278 Other than LDL apheresis<sup>64</sup> (an extracorporeal method for the elimination of LDL cholesterol,  
279 used in patients with familial hypercholesterolaemia) it is not possible to directly and  
280 specifically remove LDL cholesterol from the blood stream. Modifying a complex exposure (such  
281 as LDL cholesterol) through a drug target (e.g. HMGCR or PCSK9) is likely therefore to lead to a  
282 phenotypic modification of LDL cholesterol that may differ to that if LDL cholesterol were  
283 removed in isolation (e.g. by apheresis). This is evidenced by, for example, examining the  
284 metabolic effects of statins and *HMGCR* genotype on refined measures of blood lipids where  
285 widespread effects on multiple lipoproteins are seen, together with alterations to fatty acids.<sup>65</sup>  
286 I.e. statins have wide-reaching effects on the lipidome beyond that of LDL cholesterol, and  
287 some of these may lead to differences in disease risk which are discrete to those solely

288 attributable to LDL cholesterol. One example of this is ovarian cancer, where a protective effect  
289 of an instrument based on *HMGCR* is seen, in contrast to instruments for other LDL-cholesterol  
290 lowering agents<sup>66</sup>. Furthermore, discrete classes of drug targets that are designed to modify the  
291 same biomarker can display heterogeneous effects on blood lipid biomarkers, likely a reflection  
292 of the unique underlying mechanism(s) by which the drug targets exert their physiological  
293 effects. For example, *HMGCR* and *PCSK9* differ in their effects on the cholesterol concentration  
294 of apolipoprotein-B containing lipoproteins<sup>67</sup>, despite the motivation of both drugs being to  
295 prevent CHD through the lowering of LDL cholesterol. At a more extreme level, while both CETP  
296 inhibitors and *HMGCR* lower CHD through the pharmacological reductions of apolipoprotein  
297 B<sup>22</sup>, their metabolomic signatures differ wildly<sup>68</sup>. Similar differences are also seen between  
298 *ANGPT3* and *ANGPT4* (enzymes which are involved in the metabolism of triglycerides)<sup>69</sup>.  
299 Thus, a drug target may have different effects as compared to a biomarker, and different drugs  
300 that modify the same biomarker may differ in their effects on risk factors and disease  
301 outcomes.

302           A similar situation is true for adiposity. BMI is a polygenic trait, and a genetic instrument  
303 for BMI consists of many SNPs that associate with BMI through multiple mechanisms.  
304 Therefore, the relationship of the BMI polygenic score (PGS) with disease likely recapitulates  
305 the effects of modifying BMI by an average of these polygenic effects. Environmental changes  
306 in BMI (e.g. through cosmetic liposuction or weight loss surgery) might lead to effects on  
307 cardiovascular risk factors and disease that are specific to each surgical procedure. Lowering  
308 BMI through conventional dieting might lead to different effects on disease as the  
309 environmental manipulation of BMI in this setting (as opposed to surgery) might have differing



310 roles in disease. In contrast, existing pharmacological approaches to lowering BMI - e.g. drugs  
311 acting on the central nervous system or peripherally (e.g. GLP-1 receptor agonists <sup>70</sup>), while  
312 lowering BMI, do so through very different mechanisms. Therefore, the effects on disease risk  
313 of different pharmacological (and surgical <sup>71,72</sup>) approaches to modifying BMI, may potentially  
314 differ. This can be empirically tested by conducting drug-target MR using instruments for GLP-  
315 1R (to approximate the effects of GLP-1R agonists) and comparing disease associations to the  
316 equivalent BMI-lowering derived from a BMI PGS that is depleted for *GLP1R* variants. This  
317 approach was recently applied in a study investigating the vascular effects of a bone mineral  
318 density increasing therapy<sup>73</sup>. Early data from phase III clinical trials suggested an excess of  
319 vascular disease events in those randomized to receive a monoclonal antibody inhibitor of  
320 sclerostin. The pharmaceutical company developing the drug submitted unpublished data to  
321 regulators, which led to an attenuation of the CHD risks arising from sclerostin inhibition.  
322 However, genetic evidence suggested that increasing BMD via sclerostin inhibition was likely to  
323 increase risk of CHD<sup>73</sup>. This was further investigated by comparing the effects on CHD of  
324 modifying BMD implied by genetic variants in the *SOST* gene (which encodes the drug target  
325 sclerostin) to the effects implied by other BMD associated genetic variants across the genome.  
326 Very similar disease associations were found suggesting that BMD increasing therapies in  
327 general, including sclerostin inhibitors, are likely to increase risk of CHD (rather than  
328 representing a case of target-mediated pleiotropy, i.e. an effect specific to sclerostin inhibition;  
329 see **Box 5** for definition).

330

331 **When might findings from MR of a complex biomarker, MR of a drug target and findings from**  
332 **an RCT diverge?**

333 There are several scenarios where the relationships of a biomarker with disease may differ to  
334 that which the drug target (either in MR or RCT) shows. These are described in **Table 2** together  
335 with illustrations, and below. We describe pleiotropy in **Box 5**.

336

337 *Target-mediated pleiotropy (or, 'specificity' of target):* A biomarker may be non-causal for  
338 disease, but the drug target may show clinical efficacy (scenario 3 in **Table 2**). The exemplar for  
339 this is CETP inhibitors that were primarily developed on the basis that they increased HDL-C  
340 (rather than for any effect on cholesterol in apolipoprotein B containing lipoproteins) <sup>74</sup>. It is  
341 perhaps ironic that CETP inhibitors exert their relatively modest cardioprotective effects on the  
342 basis that in addition to their HDL-C raising effects, more potent CETP inhibitors also lower  
343 apolipoprotein B <sup>22,23,75</sup>. Had CETP not exerted such apolipoprotein B lowering effects, then all  
344 trials of CETP inhibitors (including that of anacetrapib <sup>75</sup>) would have likely failed owing to lack  
345 of efficacy. Indeed, this is precisely what was seen with earlier CETP inhibitors (e.g. dalcetrapib  
346 <sup>76</sup>) that increased HDL-C but had little or no effect on LDL-C or apoB. These findings are in  
347 keeping with evidence from human genetics in East Asians, where *CETP* variants in the China  
348 Kadoorie Biobank that were strongly associated with HDL-C, but not with apolipoprotein B, did  
349 not associate with CVD <sup>77</sup>.

350

351 *Off-target effects of therapies.* When a drug exerts properties that alter pathways that are  
352 discrete to the target of interest, and when these discrete pathways lead to altered risk of

353 disease, then the findings from MR and RCTs may not converge (scenario 4 in **Table 2**). An  
354 example of this is the blood pressure effects of CETP inhibitors <sup>74</sup>: while all CETP inhibitors show  
355 modest SBP increasing effects, this appears to be most pronounced for torcetrapib <sup>78 74</sup>. Of  
356 note, such SBP associations have not been identified in drug-target MRs using HDL-C raising  
357 variants in *CETP* <sup>79 77</sup>, for reasons that are unclear. This discrepancy between drug-target MR  
358 and RCT highlights the vital need for clinical trials for elucidating the effects of treatments in  
359 patients – i.e. we do not advocate drug target MR as a substitute for clinical trials (other than  
360 perhaps when findings from MR provide strong evidence against a plausible causal effect,  
361 indicating that a trial would be unlikely to succeed), but rather, we see drug-target MR to be  
362 complementary to and enhance the design and interpretation of conventional randomized  
363 trials.

364  
365 *Disease onset vs progression*: Most MR studies use either case control studies where the case  
366 has a first event and controls are disease free, or prospective cohorts of the general population  
367 with individuals followed up for incident events. In both circumstances, the SNP to disease  
368 association captures the risk of incident disease. Using these types of data, drug target MR  
369 would estimate whether modifying the biomarker or drug target affects the risk of developing a  
370 first event. However, in the setting of RCTs, often for fiscal reasons (to lower the number of trial  
371 participants, increases in statistical power, and also magnify absolute differences in effect  
372 between treatment groups in order to improve cost effectiveness analyses), and because  
373 prevention of disease progression is clinically important, trial participants often have prevalent  
374 disease at enrolment and drugs are tested to elucidate whether treatment slows disease

375 progression and/or prevent clinical manifestations of recurrent disease. A prime example of this  
376 in cardiovascular disease is that many trials are set in patients either at high risk of CVD (who  
377 likely have prevalent subclinical atherosclerosis at entry into the trial) or patients who have in  
378 the past experienced a CHD event. In this scenario, the trial typically estimates the effect of a  
379 drug on recurrent disease or disease progression.

380         If a risk factor is causal for first disease manifestation, but not for recurrent disease,  
381 findings from biomarker and/or drug target MRs for incident disease may be qualitatively  
382 different to estimates from an RCT on disease progression (scenario 7 in **Table 2**). Although in  
383 CVD, risk factors for onset (e.g. blood pressure, lipids) appear to be similar for first and  
384 recurrent disease, it is plausible that some risk factors for disease incidence and progression  
385 may differ. For example, there may be drug targets that are specific for plaque initiation and  
386 others that are specific for plaque rupture: if plaque initiation and rupture have differing roles  
387 in incident vs recurrent disease, then biomarkers (and drug targets) that modify plaque may  
388 demonstrate heterogeneity of disease associations between MR studies (of first events) and  
389 RCTs (of recurrent events), and vice versa. Another example is atrial fibrillation (AF) – for  
390 example a cause of AF may be hypertension, however once AF is established, its persistence is  
391 likely to be influenced to a greater extent by drugs involved in controlling the rate and rhythm,  
392 as opposed to drugs treating the initiating cause (e.g. lowering blood pressure).

393         Returning to the comparison of the aetiology for disease onset vs that of disease  
394 progression, how can this discrepancy be reconciled? Methods are under development for  
395 applying MR to disease progression,<sup>80</sup> which take into account issues such as index event bias  
396 <sup>81</sup>. Such endeavours are facilitated by the accumulation of large datasets (e.g. GENIUS-CHD,

397 which is a consortium of studies with established disease at study entry with follow-up for  
398 recurrent disease<sup>82</sup>).

399

400 *Pleiotropy of genetic instrument.* If genetic variants possess horizontally pleiotropic properties,  
401 then drug-target MR estimates may be invalid (scenario 5A in **Table 2**). While the susceptibility  
402 to such bias is likely to be reduced by selecting *cis*-acting variants as instruments (i.e. SNPs  
403 in/around the gene encoding the protein of interest; **Box 6**), it is plausible that local linkage  
404 disequilibrium can distort findings. Colocalization, which is a statistical approach that quantifies  
405 the extent to which two traits share the same underlying causal variant, and thus can be used  
406 to infer vertical pleiotropy<sup>83-85</sup>, can be used to test for such bias<sup>86</sup>, but can suffer from issues of  
407 lack of power and/or availability of suitable data from relevant tissues.

408

409 *Adaptation:* Individuals at high genetic risk of elevated LDL-C are more likely to be prescribed  
410 lipid lowering drugs, which is an environmental adaptation to genetic risk. Another example is  
411 individuals at higher genetic predisposition to obesity who smoke more<sup>87,88</sup>, which could  
412 represent an attempt to self-administer smoking for its nicotinic effects on appetite  
413 suppression and weight loss<sup>89</sup>. When such adaptation occurs to genetic predisposition to traits,  
414 this can distort effect estimates (typically biasing them to the null).

415

416 *Canalization:* Also called developmental compensation<sup>90</sup>, canalization is the process by which  
417 adaptation to a major perturbation owing to a genetic variant arises from a compensatory  
418 physiological or anatomical mechanism such that the phenotypic consequences are not as

419 pronounced as would be expected<sup>9</sup>. If present, canalization would represent a mechanism by  
420 which findings from a drug-target MR would diverge from those of an equivalent clinical trial.  
421 Reassuringly, in the case of drug-target MR, there is little evidence to support presence of such  
422 compensatory mechanisms – e.g. genetic variants with large phenotypic effects appear to have  
423 the expected dose-response association with disease as compared to genetic variants of smaller  
424 phenotypic effect<sup>57 40</sup>.

425

426 *Time-dependent effects:* If the exposure to disease process occurs over many years, then effect  
427 estimates from Mendelian randomization may be much greater than the comparative estimates  
428 derived from phase III clinical trials (which are typically shorter in duration), when scaled to the  
429 same difference in exposure. Thus, it may be necessary to scale the MR estimate in order to  
430 obtain the equivalent effect estimate that would be predicted to be obtained in a phase III  
431 clinical trial for a given duration. The timing of causal effects for each exposure-disease pair is  
432 likely to differ. For example, the optimal timing for lower LDL-C is likely to be different to the  
433 optimal timing for BMI lowering interventions<sup>91</sup>. Approaches such as estimating the causal  
434 effects of a biomarker at different points in life using multivariable MR may help provide  
435 evidence about timing (as has been conducted for BMI<sup>91</sup>), however this approach has not been  
436 used for drug targets, and may not be possible owing to the limited repertoire of (particularly  
437 cis-acting) genetic instruments. Fortunately, however, it is possible, that in the absence of  
438 target mediated pleiotropy of individual drug targets, once the specific biomarker-outcome  
439 relationship is characterized, including an understanding of the differences between treatment  
440 effects originating from a treatment trial of a specified duration as compared to lifelong

441 exposure instrumented through MR, such properties are likely be similar for drugs that modify  
442 the same biomarker including drugs in discrete classes. This is what is seen with multiple LDL-C  
443 lowering drug classes: MR estimates of the effects of LDL-C lowering and risk of CHD are  
444 approximately 3-fold greater than estimates from RCTs<sup>92</sup>, and as such, we can naively scale MR  
445 estimates of LDL-C lowering drug target MR studies by this same scaling factor in order to  
446 provide some indication of the possible magnitude of the effect estimate obtained in treatment  
447 trials. The process to establishing this is several-fold: first conducting MR of the complex  
448 biomarker (and ideally the drug-target) followed by a quantitative comparison to the phase III  
449 cardiovascular outcome trial (CVOT) result in order to estimate the comparable differences in  
450 magnitudes of effect owing to e.g. duration of exposure, linearity of dose-response and  
451 presence of time-specific effects; second, an appraisal of consistency of disease associations  
452 across multiple MRs of discrete drug classes scaled to the same difference in biomarker; and  
453 third, once a scaling factor for e.g. life-long estimates from MR as compared to short-term  
454 effects of RCTs is derived for an exposure-outcome pair, applying the same scaling factor to  
455 multiple drug classes to inform of likely treatment effects in prospectively-planned clinical  
456 trials<sup>92,93</sup>. A fourth check would be, for example using non-linear MR approaches<sup>94</sup>, that the  
457 dose response of the biomarker and disease risk is log-linear to facilitate scaling of effects in  
458 drug-target MR<sup>95</sup>. Such dose-response can also be approximated using individual drug-target  
459 MRs of varying effect size on the exposure<sup>96</sup>. The approach outlined here to obtaining a scaling  
460 estimate has generally held true for lipid-lowering therapies and risk of cardiovascular  
461 disease<sup>22,23,40,92</sup> but there are several potential reasons why the derivation of such a scaling for  
462 other therapeutic target-indication pairs may pose a challenge. One such reason is that the

463 scaling factor is extremely likely to differ for each therapeutic target-indication pair, but such  
464 scaling is, as a rule, unlikely to be different when the therapeutic target modifies the same final  
465 common pathway (except in the case of target-mediated pleiotropy). Thus, in the absence of  
466 target-mediated pleiotropy, it is possible that a scaling factor for broad groups of target-  
467 indication pairs may be constructed, which provides value beyond more limited 'qualitative'  
468 inferences (e.g. presence/absence of causation) that can be inferred from MR.

469

470

### 471 **Putting the pieces together**

472 How then are we to design approaches which identify therapeutic targets for disease  
473 prevention? One such framework is outlined in **Figure 7**. Complex (polygenic) biomarkers are  
474 modified pharmacologically by targeting (monogenic) proteins that, through vertical pleiotropy,  
475 modify the complex biomarker of interest. A first step, typically through complex biomarker  
476 MR, is to establish causation of the complex biomarker and clarify whether modifying the  
477 biomarker may represent a valid therapeutic goal to modifying disease risk. Since, as described  
478 above, findings from a biomarker MR cannot be assumed to be identical to those derived from  
479 drug target MR performed on the basis that a drug target modifies the same biomarker of  
480 interest, the next step is usually to conduct a drug-target MR. Data from these various sources  
481 can be used to triangulate evidence and predict the likely outcome of a therapeutic trial. In  
482 comparing a *cis*-pQTL instrument (representing a therapeutic target) to a phenotypic PGS  
483 (comprised of all phenotype-associated genetic variants minus SNPs in/around the gene  
484 encoding the drug target) it may be possible to identify potential differences between the



485 expected effect of the biomarker on a disease, versus the effect by modifying it via a specific  
486 drug target. These comparisons can identify if perturbation of the drug-target pathway of  
487 interest has a differential effect from other potential targets that affect the same biomarker of  
488 interest. If a biomarker is identified to have a particular relationship with disease (e.g. long  
489 latency or time dependency), then individual drug targets that modify the same biomarker  
490 ought to display the same attributes in their relationship with disease. Drug-target MR also  
491 heralds multiple opportunities for additional exploration – for example when the precise  
492 mediators and adverse effect profiles are not known between a drug target and disease  
493 outcome, explorations through MR-PheWAS can provide new insights<sup>66</sup>. For example,  
494 colchicine<sup>53</sup> and icosapent ethyl<sup>97</sup> have both demonstrated efficacy in reduction of CVD in  
495 recent large-scale phase III trials, however the precise mechanisms by which these drugs act,  
496 i.e. the causal intermediaries, remains elusive. A proteomic exploration of individuals treated  
497 with colchicine or icosapent ethyl might identify proteins that these drugs modify. MR studies  
498 of these proteins for CHD would characterize, for each protein, whether it is likely to be on the  
499 causal pathway from therapeutic target to disease. This might permit development of more  
500 specific drugs that could display fewer adverse events (as demonstrated in **Figure 3**) and  
501 facilitate accurate titration to dose-response, with knowledge of the causal intermediary. Thus,  
502 discrepancies between biomarker and drug-target<sup>97</sup> can be informative for elucidating the  
503 aetiology of disease and with it, shed light on new approaches to treating and preventing  
504 disease. Enhancements to drug-target MR described in **Box 8** can also address criticisms<sup>98</sup> often  
505 levied at RCTs (with such criticism subsequently rebuked<sup>99-101</sup>) including that individuals  
506 recruited into trials are usually highly-selected (thus hampering generalizability), and that

507 findings from RCTs apply to the group-level rather than to individuals, hampering progress in  
508 personalised medicines. Finally, in writing an article about the conduct of drug-target MR, it  
509 behooves us to stress that the scientific (and pharmaceutical) community has a responsibility in  
510 applying this scientific approach to promote the equitable development of therapies, ensuring  
511 that advances in drug development (including those catalysed by drug-target MR) benefit  
512 minority under-represented groups, as we describe in **Box 9**. In summary, the use of human  
513 genetics can facilitate drug development throughout the lifecycle.

514

## 515 **Conclusions**

516 In this review, we have contrasted MR of a biomarker with that of a drug target that seeks to  
517 modify the same biomarker. A biomarker is typically a complex phenotype that is optimally  
518 instrumented through MR using multiple genetic variants across the genome. In contrast, drug  
519 target MR typically focuses on a protein trait using *cis*-acting genetic variants in an individual  
520 gene that encodes the protein. The motivation for biomarker and drug target MR differs, with  
521 biomarkers informing the effects of potential public health measures and shedding light on  
522 causal risk factors for disease that motivate their pharmacological modification, while drug  
523 target MRs estimate the likelihood of a specific therapeutic (typically designed to modify a  
524 complex biomarker) showing efficacy in a clinical trial. In both scenarios, human genetics can be  
525 harnessed as an invaluable tool to elucidate causal roles in disease aetiology and through  
526 careful application, such studies can be used to guide the design and interpretation of clinical  
527 trials of medicines.

528

530 **References**

- 531 1 The top 10 causes of death (Updated January 2017) Available at:  
532 <http://www.who.int/mediacentre/factsheets/fs310/en/>. *World Health Organization*
- 533 2 Beierlein, J. M. *et al.* Landscape of Innovation for Cardiovascular Pharmaceuticals: From  
534 Basic Science to New Molecular Entities. *Clin Ther* **39**, 1409-1425 e1420,  
535 doi:10.1016/j.clinthera.2017.06.001 (2017).
- 536 3 Fordyce, C. B. *et al.* Cardiovascular drug development: is it dead or just hibernating? *J*  
537 *Am Coll Cardiol* **65**, 1567-1582, doi:10.1016/j.jacc.2015.03.016 (2015).
- 538 4 Darrow, J. J., Avorn, J. & Kesselheim, A. S. FDA Approval and Regulation of  
539 Pharmaceuticals, 1983-2018. *JAMA* **323**, 164-176, doi:10.1001/jama.2019.20288 (2020).
- 540 5 Wouters, O. J., McKee, M. & Luyten, J. Estimated Research and Development  
541 Investment Needed to Bring a New Medicine to Market, 2009-2018. *JAMA* **323**, 844-  
542 853, doi:10.1001/jama.2020.1166 (2020).
- 543 6 Harrison, R. K. Phase II and phase III failures: 2013-2015. *Nat Rev Drug Discov* **15**, 817-  
544 818, doi:10.1038/nrd.2016.184 (2016).
- 545 7 Alteri, E. & Guizzaro, L. Be open about drug failures to speed up research. *Nature* **563**,  
546 317-319, doi:10.1038/d41586-018-07352-7 (2018).
- 547 8 Dowden, H. & Munro, J. Trends in clinical success rates and therapeutic focus. *Nat Rev*  
548 *Drug Discov* **18**, 495-496, doi:10.1038/d41573-019-00074-z (2019).
- 549 9 Davey Smith, G. & Ebrahim, S. 'Mendelian randomization': can genetic epidemiology  
550 contribute to understanding environmental determinants of disease? *Int J Epidemiol* **32**,  
551 1-22 (2003).
- 552 10 Davies, N. M., Holmes, M. V. & Davey Smith, G. Reading Mendelian randomisation  
553 studies: a guide, glossary, and checklist for clinicians. *BMJ* **362**, k601,  
554 doi:10.1136/bmj.k601 (2018).
- 555 11 Nelson, M. R. *et al.* The support of human genetic evidence for approved drug  
556 indications. *Nat Genet* **47**, 856-860, doi:10.1038/ng.3314 (2015).
- 557 12 Zheng, J. *et al.* Phenome-wide Mendelian randomization mapping the influence of the  
558 plasma proteome on complex diseases. *Nature Genetics (In Press)* (2020).
- 559 13 Lyall, D. M. *et al.* Association of Body Mass Index With Cardiometabolic Disease in the  
560 UK Biobank: A Mendelian Randomization Study. *JAMA Cardiol* **1**, 882-889,  
561 doi:10.1001/jamacardio.2016.5804 (2017).
- 562 14 Dale, C. E. *et al.* Causal Associations of Adiposity and Body Fat Distribution With  
563 Coronary Heart Disease, Stroke Subtypes, and Type 2 Diabetes Mellitus: A Mendelian  
564 Randomization Analysis. *Circulation* **135**, 2373-2388,  
565 doi:10.1161/CIRCULATIONAHA.116.026560 (2017).
- 566 15 Emdin, C. A. *et al.* Genetic Association of Waist-to-Hip Ratio With Cardiometabolic  
567 Traits, Type 2 Diabetes, and Coronary Heart Disease. *JAMA* **317**, 626-634,  
568 doi:10.1001/jama.2016.21042 (2017).

- 569 16 Holmes, M. V. *et al.* Causal effects of body mass index on cardiometabolic traits and  
570 events: a Mendelian randomization analysis. *Am J Hum Genet* **94**, 198-208,  
571 doi:10.1016/j.ajhg.2013.12.014 (2014).
- 572 17 Voight, B. F. *et al.* Plasma HDL cholesterol and risk of myocardial infarction: a mendelian  
573 randomisation study. *The Lancet* **380**, 572-580, doi:[https://doi.org/10.1016/S0140-](https://doi.org/10.1016/S0140-6736(12)60312-2)  
574 [6736\(12\)60312-2](https://doi.org/10.1016/S0140-6736(12)60312-2).
- 575 18 White, J. *et al.* Association of Lipid Fractions With Risks for Coronary Artery Disease and  
576 Diabetes. *JAMA Cardiol* **1**, 692-699, doi:10.1001/jamacardio.2016.1884 (2016).
- 577 19 Holmes, M. V. *et al.* Mendelian randomization of blood lipids for coronary heart disease.  
578 *European Heart Journal* **36**, 539-550, doi:10.1093/eurheartj/ehv571 (2015).
- 579 20 Interleukin-6 Receptor Mendelian Randomisation Analysis, C. *et al.* The interleukin-6  
580 receptor as a target for prevention of coronary heart disease: a mendelian  
581 randomisation analysis. *Lancet* **379**, 1214-1224, doi:10.1016/S0140-6736(12)60110-X  
582 (2012).
- 583 21 IL6R Genetics Consortium Emerging Risk Factors Collaboration *et al.* Interleukin-6  
584 receptor pathways in coronary heart disease: a collaborative meta-analysis of 82  
585 studies. *Lancet* **379**, 1205-1213, doi:10.1016/S0140-6736(11)61931-4 (2012).
- 586 22 Holmes, M. V. & Davey Smith, G. Dyslipidaemia: REVEALing the effect of CETP inhibition  
587 in cardiovascular disease. *Nat Rev Cardiol* **14**, 635-636, doi:10.1038/nrcardio.2017.156  
588 (2017).
- 589 23 Ference, B. A. *et al.* Association of Genetic Variants Related to CETP Inhibitors and  
590 Statins With Lipoprotein Levels and Cardiovascular Risk. *JAMA* **318**, 947-956,  
591 doi:10.1001/jama.2017.11467 (2017).
- 592 24 Harrison, S. C. *et al.* Genetic Association of Lipids and Lipid Drug Targets With Abdominal  
593 Aortic Aneurysm: A Meta-analysis. *JAMA Cardiol* **1**, 26-33 (2018).
- 594 25 Millwood, I. Y. *et al.* Association of CETP gene variants with risk for vascular and  
595 nonvascular diseases among Chinese adults. *JAMA Cardiol* **1**, 34-43 (2018).
- 596 26 Holmes, M. V., Ala-Korpela, M. & Davey Smith, G. Mendelian randomization in  
597 cardiometabolic disease: challenges in evaluating causality. *Nat Rev Cardiol* **14**, 577-590,  
598 doi:10.1038/nrcardio.2017.78 (2017).
- 599 27 Hingorani, A. & Humphries, S. Nature's randomised trials. *Lancet* **366**, 1906-1908,  
600 doi:10.1016/S0140-6736(05)67767-7 (2005).
- 601 28 Thanassoulis, G. & O'Donnell, C. J. Mendelian randomization: nature's randomized trial  
602 in the post-genome era. *JAMA* **301**, 2386-2388, doi:10.1001/jama.2009.812 (2009).
- 603 29 Plump, A. S. & Lum, P. Y. Genomics and cardiovascular drug development. *J Am Coll*  
604 *Cardiol* **53**, 1089-1100, doi:10.1016/j.jacc.2008.11.050 (2009).
- 605 30 Khera, A. V. & Kathiresan, S. Genetics of coronary artery disease: discovery, biology and  
606 clinical translation. *Nat Rev Genet* **18**, 331-344, doi:10.1038/nrg.2016.160 (2017).
- 607 31 Hingorani, A. D. *et al.* Improving the odds of drug development success through human  
608 genomics: modelling study. *Sci Rep* **9**, 18911, doi:10.1038/s41598-019-54849-w (2019).
- 609 32 Schmidt, A. F. *et al.* Genetic drug target validation using Mendelian randomisation. *Nat*  
610 *Commun* **11**, 3255, doi:10.1038/s41467-020-16969-0 (2020).
- 611 33 KEATING, S. *et al.* The influence of HLA-matched sibling donor availability on treatment  
612 outcome for patients with AML: an analysis of the AML 8A study of the EORTC

- 613 Leukaemia Cooperative Group and GIMEMA. *British Journal of Haematology* **102**, 1344-  
614 1353, doi:10.1111/j.1365-2141.1998.896hm3674.x (1998).
- 615 34 Davies, N. M. *et al.* Within family Mendelian randomization studies. *Human Molecular*  
616 *Genetics* **28**, R170-R179, doi:10.1093/hmg/ddz204 (2019).
- 617 35 Brumpton, B. *et al.* Avoiding dynastic, assortative mating, and population stratification  
618 biases in Mendelian randomization through within-family analyses. *Nat Commun* **11**,  
619 3519, doi:10.1038/s41467-020-17117-4 (2020).
- 620 36 Davey Smith, G., Holmes, M. V., Davies, N. M. & Ebrahim, S. Mendel's laws, Mendelian  
621 randomization and causal inference in observational data: substantive and  
622 nomenclatural issues. *Eur J Epidemiol* **35**, 99-111, doi:10.1007/s10654-020-00622-7  
623 (2020).
- 624 37 Burgess, S. *et al.* Guidelines for performing Mendelian randomization investigations  
625 [version 1; peer review: awaiting peer review]. *Wellcome Open Res* **4** (2019).
- 626 38 Davey Smith, G. & Hemani, G. Mendelian randomization: genetic anchors for causal  
627 inference in epidemiological studies. *Hum Mol Genet* **23**, R89-98,  
628 doi:10.1093/hmg/ddu328 (2014).
- 629 39 Cholesterol Treatment Trialists' Collaborators *et al.* The effects of lowering LDL  
630 cholesterol with statin therapy in people at low risk of vascular disease: meta-analysis of  
631 individual data from 27 randomised trials. *Lancet* **380**, 581-590, doi:10.1016/S0140-  
632 6736(12)60367-5 (2012).
- 633 40 Ference, B. A., Yoo, W., Alesh, I. *et al.* Effect of Long-Term Exposure to Lower Low-  
634 Density Lipoprotein Cholesterol Beginning Early in Life on the Risk of Coronary Heart  
635 Disease A Mendelian Randomization Analysis. *Journal of the American College of*  
636 *Cardiology* **60**, 2631-2639, doi:10.1016/j.jacc.2012.09.017 (2012).
- 637 41 Nicholls, S. J., Kastelein, J. P., Schwartz, G. G. & *et al.* Varespladib and cardiovascular  
638 events in patients with an acute coronary syndrome: The vista-16 randomized clinical  
639 trial. *JAMA* **311**, 252-262, doi:10.1001/jama.2013.282836 (2014).
- 640 42 Holmes, M. V. *et al.* Secretory Phospholipase A2-IIA and Cardiovascular Disease. *Journal*  
641 *of the American College of Cardiology* **62**, 1966-1976,  
642 doi:<http://dx.doi.org/10.1016/j.jacc.2013.06.044> (2013).
- 643 43 Ference, B. A., Robinson, J. G., Brook, R. D. *et al.* Variation in PCSK9 and HMGCR and risk  
644 of cardiovascular disease and diabetes. *New England Journal of Medicine* **375**, 2144-  
645 2153, doi:10.1056/NEJMoa1604304 (2016).
- 646 44 Schmidt, A. F. *et al.* PCSK9 genetic variants and risk of type 2 diabetes: a mendelian  
647 randomisation study. *Lancet Diabetes Endocrinol* **5**, 97-105, doi:10.1016/S2213-  
648 8587(16)30396-5 (2017).
- 649 45 de Carvalho, L. S. F., Campos, A. M. & Sposito, A. C. Proprotein Convertase  
650 Subtilisin/Kexin Type 9 (PCSK9) Inhibitors and Incident Type 2 Diabetes: A Systematic  
651 Review and Meta-analysis With Over 96,000 Patient-Years. *Diabetes Care* **41**, 364-367,  
652 doi:10.2337/dc17-1464 (2018).
- 653 46 Mukamal, K. J., Stampfer, M. J. & Rimm, E. B. Genetic instrumental variable analysis:  
654 time to call mendelian randomization what it is. The example of alcohol and  
655 cardiovascular disease. *Eur J Epidemiol*, doi:10.1007/s10654-019-00578-3 (2019).

656 47 Swanson, S. A., et. al. Nature as a trialist? Deconstructing the analogy between  
657 Mendelian Randomization and randomized trials. *Epidemiology In press* (2017).

658 48 Strimbu, K. & Tavel, J. A. What are biomarkers? *Curr Opin HIV AIDS* **5**, 463-466,  
659 doi:10.1097/COH.0b013e32833ed177 (2010).

660 49 Libby, P. Inflammation in atherosclerosis. *Nature* **420**, 868-874,  
661 doi:10.1038/nature01323 (2002).

662 50 Clinton, S. K. & Libby, P. Cytokines and growth factors in atherogenesis. *Arch Pathol Lab*  
663 *Med* **116**, 1292-1300 (1992).

664 51 C Reactive Protein Coronary Heart Disease Genetics Collaboration *et al.* Association  
665 between C reactive protein and coronary heart disease: mendelian randomisation  
666 analysis based on individual participant data. *BMJ* **342**, d548, doi:10.1136/bmj.d548  
667 (2011).

668 52 Ridker, P. M. *et al.* Antiinflammatory Therapy with Canakinumab for Atherosclerotic  
669 Disease. *N Engl J Med* **377**, 1119-1131, doi:10.1056/NEJMoa1707914 (2017).

670 53 Tardif, J. C. *et al.* Efficacy and Safety of Low-Dose Colchicine after Myocardial Infarction.  
671 *N Engl J Med* **381**, 2497-2505, doi:10.1056/NEJMoa1912388 (2019).

672 54 Interleukin-6 Receptor Mendelian Randomisation Analysis, C. The interleukin-6 receptor  
673 as a target for prevention of coronary heart disease: a mendelian randomisation  
674 analysis. *Lancet* **379**, 1214-1224, doi:10.1016/S0140-6736(12)60110-X (2012).

675 55 Xu, L., Borges, M. C., Hemani, G. & Lawlor, D. A. The role of glycaemic and lipid risk  
676 factors in mediating the effect of BMI on coronary heart disease: a two-step, two-  
677 sample Mendelian randomisation study. *Diabetologia* **60**, 2210-2220,  
678 doi:10.1007/s00125-017-4396-y (2017).

679 56 Sanderson, E., Davey Smith, G., Windmeijer, F. & Bowden, J. An examination of  
680 multivariable Mendelian randomization in the single sample and two-sample summary  
681 data settings. *Int J Epidemiol* **48** (2019).

682 57 Ference, B. A. *et al.* Association of Triglyceride-Lowering LPL Variants and LDL-C-  
683 Lowering LDLR Variants With Risk of Coronary Heart Disease. *JAMA* **321**, 364-373,  
684 doi:10.1001/jama.2018.20045 (2019).

685 58 Richardson, T. G. *et al.* Evaluating the relationship between circulating lipoprotein lipids  
686 and apolipoproteins with risk of coronary heart disease: A multivariable Mendelian  
687 randomisation analysis. *PLoS Med* **17**, e1003062, doi:10.1371/journal.pmed.1003062  
688 (2020).

689 59 Mitchell, G., Lesch, M. & McCambridge, J. Alcohol Industry Involvement in the Moderate  
690 Alcohol and Cardiovascular Health Trial. *Am J Public Health* **110**, 485-488,  
691 doi:10.2105/AJPH.2019.305508 (2020).

692 60 O'Connor, A. Should We Be Drinking Less? *The New York Times* (2020. Available at:  
693 <https://www.nytimes.com/2020/07/10/well/eat/should-we-be-drinking-less.html>  
694 (Accessed 7th August 2020)).

695 61 ROSE, G. Sick Individuals and Sick Populations. *International Journal of Epidemiology* **14**,  
696 32-38, doi:10.1093/ije/14.1.32 (1985).

697 62 Hardison, R. C. Evolution of hemoglobin and its genes. *Cold Spring Harb Perspect Med* **2**,  
698 a011627, doi:10.1101/cshperspect.a011627 (2012).

- 699 63 Prins, B. P. *et al.* Investigating the Causal Relationship of C-Reactive Protein with 32  
700 Complex Somatic and Psychiatric Outcomes: A Large-Scale Cross-Consortium Mendelian  
701 Randomization Study. *PLoS Med* **13**, e1001976, doi:10.1371/journal.pmed.1001976  
702 (2016).
- 703 64 Bambauer, R., Bambauer, C., Lehmann, B., Latza, R. & Schiel, R. LDL-apheresis: technical  
704 and clinical aspects. *ScientificWorldJournal* **2012**, 314283, doi:10.1100/2012/314283  
705 (2012).
- 706 65 Wurtz, P. *et al.* Metabolomic Profiling of Statin Use and Genetic Inhibition of HMG-CoA  
707 Reductase. *J Am Coll Cardiol* **67**, 1200-1210, doi:10.1016/j.jacc.2015.12.060 (2016).
- 708 66 Yarmolinsky, J. *et al.* Association Between Genetically Proxied Inhibition of HMG-CoA  
709 Reductase and Epithelial Ovarian Cancer. *JAMA* **323**, 646-655,  
710 doi:10.1001/jama.2020.0150 (2020).
- 711 67 Sliz, E. *et al.* Metabolomic Consequences of Genetic Inhibition of PCSK9 Compared With  
712 Statin Treatment. *Circulation* **138**, 2499-2512,  
713 doi:10.1161/CIRCULATIONAHA.118.034942 (2018).
- 714 68 Kettunen, J. *et al.* Lipoprotein signatures of cholesteryl ester transfer protein and HMG-  
715 CoA reductase inhibition. *PLoS Biol* **17**, e3000572, doi:10.1371/journal.pbio.3000572  
716 (2019).
- 717 69 Wang, Q. *et al.* Metabolic profiling of angiopoietin-like protein 3 and 4 inhibition: a drug-  
718 target Mendelian randomization analysis. *Under Review* (2020).
- 719 70 Husain, M. *et al.* Oral Semaglutide and Cardiovascular Outcomes in Patients with Type 2  
720 Diabetes. *N Engl J Med* **381**, 841-851, doi:10.1056/NEJMoa1901118 (2019).
- 721 71 Benatti, F. *et al.* Liposuction induces a compensatory increase of visceral fat which is  
722 effectively counteracted by physical activity: a randomized trial. *J Clin Endocrinol Metab*  
723 **97**, 2388-2395, doi:10.1210/jc.2012-1012 (2012).
- 724 72 Mohammed, B. S., Cohen, S., Reeds, D., Young, V. L. & Klein, S. Long-term effects of  
725 large-volume liposuction on metabolic risk factors for coronary heart disease. *Obesity*  
726 (*Silver Spring*) **16**, 2648-2651, doi:10.1038/oby.2008.418 (2008).
- 727 73 Bovijn, J. *et al.* Evaluating the cardiovascular safety of sclerostin inhibition using  
728 evidence from meta-analysis of clinical trials and human genetics. *Science Translational*  
729 *Medicine In Press* (2020).
- 730 74 Armitage, J., Holmes, M. V. & Preiss, D. Cholesteryl Ester Transfer Protein Inhibition for  
731 Preventing Cardiovascular Events. *Journal of the American College of Cardiology* **73**, 477-  
732 487 (2019).
- 733 75 HPS3/TIMI55-REVEAL Collaborative Group *et al.* Effects of Anacetrapib in Patients with  
734 Atherosclerotic Vascular Disease. *N Engl J Med* **377**, 1217-1227,  
735 doi:10.1056/NEJMoa1706444 (2017).
- 736 76 Schwartz, G. G. *et al.* Effects of dalcetrapib in patients with a recent acute coronary  
737 syndrome. *N Engl J Med* **367**, 2089-2099, doi:10.1056/NEJMoa1206797 (2012).
- 738 77 Millwood, I. Y. *et al.* Association of CETP gene variants with risk for vascular and  
739 nonvascular diseases among Chinese adults. *JAMA Cardiol* **1**, 34-43 (2018).
- 740 78 Barter, P. J. *et al.* Effects of torcetrapib in patients at high risk for coronary events. *N*  
741 *Engl J Med* **357**, 2109-2122, doi:10.1056/NEJMoa0706628 (2007).

742 79 Sofat, R. *et al.* Separating the mechanism-based and off-target actions of cholesteryl  
743 ester transfer protein inhibitors with CETP gene polymorphisms. *Circulation* **121**, 52-62,  
744 doi:10.1161/CIRCULATIONAHA.109.865444 (2010).

745 80 Paternoster, L., Tilling, K. M. & Davey Smith, G. Genetic Epidemiology And Mendelian  
746 Randomization For Informing Disease Therapeutics: Conceptual And Methodological  
747 Challenges. *PLoS Genet In Press* (2017).

748 81 Mahmoud, O., Dudbridge, F., Smith, G. D., Munafo, M. & Tilling, K. Slope-Hunter: A  
749 robust method for index-event bias correction in genome-wide association studies of  
750 subsequent traits. *bioRxiv* (2020).

751 82 Patel, R. S. *et al.* Subsequent Event Risk in Individuals With Established Coronary Heart  
752 Disease. *Circ Genom Precis Med* **12**, e002470, doi:10.1161/CIRCGEN.119.002470 (2019).

753 83 Deng, Y. & Pan, W. A powerful and versatile colocalization test. *PLoS Comput Biol* **16**,  
754 e1007778, doi:10.1371/journal.pcbi.1007778 (2020).

755 84 Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic  
756 association studies using summary statistics. *PLoS Genet* **10**, e1004383,  
757 doi:10.1371/journal.pgen.1004383 (2014).

758 85 Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts  
759 complex trait gene targets. *Nat Genet* **48**, 481-487, doi:10.1038/ng.3538 (2016).

760 86 Bovijn, J., Censin, J. C., Lindgren, C. M. & Holmes, M. V. Using human genetics to guide  
761 the repurposing of medicines. *Int J Epidemiol*, doi:10.1093/ije/dyaa015 (2020).

762 87 Censin, J. C. *et al.* Causal relationships between obesity and the leading causes of death  
763 in women and men. *PLoS Genet* **15**, e1008405, doi:10.1371/journal.pgen.1008405  
764 (2019).

765 88 Carreras-Torres, R. *et al.* Role of obesity in smoking behaviour: Mendelian  
766 randomisation study in UK Biobank. *BMJ* **361**, k1767, doi:10.1136/bmj.k1767 (2018).

767 89 Audrain-McGovern, J. & Benowitz, N. L. Cigarette smoking, nicotine, and body weight.  
768 *Clin Pharmacol Ther* **90**, 164-168, doi:10.1038/clpt.2011.105 (2011).

769 90 Zheng, J. *et al.* Recent Developments in Mendelian Randomization Studies. *Curr*  
770 *Epidemiol Rep* **4**, 330-345, doi:10.1007/s40471-017-0128-6 (2017).

771 91 Richardson, T. G., Sanderson, E., Elsworth, B., Tilling, K. & Davey Smith, G. Use of genetic  
772 variation to separate the effects of early and later life adiposity on disease risk:  
773 mendelian randomisation study. *BMJ* **369**, m1203, doi:10.1136/bmj.m1203 (2020).

774 92 Ference, B. A. How to use Mendelian randomization to anticipate the results of  
775 randomized trials. *Eur Heart J* **39**, 360-362, doi:10.1093/eurheartj/ehx462 (2018).

776 93 Burgess, S. *et al.* Association of LPA Variants With Risk of Coronary Disease and the  
777 Implications for Lipoprotein(a)-Lowering Therapies: A Mendelian Randomization  
778 Analysis. *JAMA Cardiol* **3**, 619-627, doi:10.1001/jamacardio.2018.1470 (2018).

779 94 Silverwood, R. J. *et al.* Testing for non-linear causal effects using a binary genotype in a  
780 Mendelian randomization study: application to alcohol and cardiovascular traits. *Int J*  
781 *Epidemiol* **43**, 1781-1790, doi:10.1093/ije/dyu187 (2014).

782 95 Lawlor, D. A., Harbord, R. M., Sterne, J. A., Timpson, N. & Davey Smith, G. Mendelian  
783 randomization: using genes as instruments for making causal inferences in  
784 epidemiology. *Stat Med* **27**, 1133-1163, doi:10.1002/sim.3034 (2008).



- 785 96 Ference, B. A. Using Genetic Variants in the Targets of Lipid Lowering Therapies to  
786 Inform Drug Discovery and Development: Current and Future Treatment Options. *Clin*  
787 *Pharmacol Ther* **105**, 568-581, doi:10.1002/cpt.1163 (2019).
- 788 97 Bhatt, D. L. *et al.* Cardiovascular Risk Reduction with Icosapent Ethyl for  
789 Hypertriglyceridemia. *N Engl J Med* **380**, 11-22, doi:10.1056/NEJMoa1812792 (2019).
- 790 98 Deaton, A. & Cartwright, N. Understanding and misunderstanding randomized  
791 controlled trials. *Soc Sci Med* **210**, 2-21, doi:10.1016/j.socscimed.2017.12.005 (2018).
- 792 99 Collins, R., Bowman, L., Landray, M. & Peto, R. The Magic of Randomization versus the  
793 Myth of Real-World Evidence. *N Engl J Med* **382**, 674-678, doi:10.1056/NEJMsb1901642  
794 (2020).
- 795 100 Senn, S. Individual response to treatment: is it a valid assumption? *BMJ* **329**, 966-968,  
796 doi:10.1136/bmj.329.7472.966 (2004).
- 797 101 Senn, S. Statistical pitfalls of personalized medicine. *Nature* **563**, 619-621,  
798 doi:10.1038/d41586-018-07535-2 (2018).
- 799 102 Abbate, A. *et al.* Interleukin-1 and the Inflammasome as Therapeutic Targets in  
800 Cardiovascular Disease. *Circ Res* **126**, 1260-1280, doi:10.1161/CIRCRESAHA.120.315937  
801 (2020).
- 802 103 Thompson, A. *et al.* Association of cholesteryl ester transfer protein genotypes with  
803 CETP mass and activity, lipid levels, and coronary risk. *JAMA* **299**, 2777-2788,  
804 doi:10.1001/jama.299.23.2777 (2008).
- 805 104 Holmes, M. V. & Davey Smith, G. Can Mendelian Randomization Shift into Reverse Gear?  
806 *Clin Chem* **65**, 363-366, doi:10.1373/clinchem.2018.296806 (2019).
- 807 105 Gaulton, A. *et al.* ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic*  
808 *Acids Res* **40**, D1100-1107, doi:10.1093/nar/gkr777 (2012).
- 809 106 Sliz, E. *et al.* Metabolomic Consequences of Genetic Inhibition of PCSK9 Compared with  
810 Statin Treatment. *Circulation* **138**, 2499-2512 (2018).
- 811 107 Ference, B. A. *et al.* Variation in PCSK9 and HMGCR and Risk of Cardiovascular Disease  
812 and Diabetes. *New England Journal of Medicine* **375**, 2144-2153,  
813 doi:10.1056/NEJMoa1604304 (2016).
- 814 108 Cholesterol Treatment Trialists' Collaborators *et al.* Efficacy and safety of LDL-lowering  
815 therapy among men and women: meta-analysis of individual data from 174,000  
816 participants in 27 randomised trials. *Lancet* **385**, 1397-1405, doi:10.1016/S0140-  
817 6736(14)61368-4 (2015).
- 818 109 Cholesterol Treatment Trialists' Collaborators *et al.* Efficacy and safety of more  
819 intensive lowering of LDL cholesterol: a meta-analysis of data from 170,000 participants  
820 in 26 randomised trials. *Lancet* **376**, 1670-1681, doi:10.1016/S0140-6736(10)61350-5  
821 (2010).
- 822 110 Silverman, M. G. *et al.* Association Between Lowering LDL-C and Cardiovascular Risk  
823 Reduction Among Different Therapeutic Interventions: A Systematic Review and Meta-  
824 analysis. *JAMA* **316**, 1289-1297, doi:10.1001/jama.2016.13985 (2016).
- 825 111 Rader, D. J. Apolipoprotein A-I Infusion Therapies for Coronary Disease: Two Outs in the  
826 Ninth Inning and Swinging for the Fences. *JAMA Cardiol* **3**, 799-801,  
827 doi:10.1001/jamacardio.2018.2168 (2018).

828 112 Patel, R. S. *et al.* Genetic variants at chromosome 9p21 and risk of first versus  
829 subsequent coronary heart disease events: a systematic review and meta-analysis. *J Am*  
830 *Coll Cardiol* **63**, 2234-2245, doi:10.1016/j.jacc.2014.01.065 (2014).

831 113 Mokry, L. E. *et al.* Vitamin D and Risk of Multiple Sclerosis: A Mendelian Randomization  
832 Study. *PLoS Medicine* **12**, doi:10.1371/journal.pmed.1001866 (2015).

833 114 Ridker, P. M. *et al.* Lipid-Reduction Variability and Antidrug-Antibody Formation with  
834 Bococizumab. *N Engl J Med* **376**, 1517-1526, doi:10.1056/NEJMoa1614062 (2017).

835 115 Ference, B. A., Majeed, F., Penumetcha, R., Flack, J. M. & Brook, R. D. Effect of Naturally  
836 Random Allocation to Lower Low-Density Lipoprotein Cholesterol on the Risk of  
837 Coronary Heart Disease Mediated by Polymorphisms in NPC1L1, HMGCR, or Both.  
838 *Journal of the American College of Cardiology* **65**, 1552-1561,  
839 doi:10.1016/j.jacc.2015.02.020 (2015).

840 116 Collins, R. *et al.* Interpretation of the evidence for the efficacy and safety of statin  
841 therapy. *The Lancet* **388**, 2532-2561, doi:10.1016/S0140-6736(16)31357-5 (2016).

842 117 Karjalainen, M. *et al.* Apolipoprotein A-I concentrations and risk of coronary artery  
843 disease: A Mendelian randomization study. *Atherosclerosis* **299**, 56-53 (2020).

844 118 Scott, R. A. *et al.* A genomic approach to therapeutic target validation identifies a  
845 glucose-lowering <em>GLP1R</em> variant protective for coronary heart disease.  
846 *Science Translational Medicine* **8**, 341ra376-341ra376,  
847 doi:10.1126/scitranslmed.aad3744 (2016).

848 119 Seidelmann, S. B. *et al.* Genetic Variants in SGLT1, Glucose Tolerance, and  
849 Cardiometabolic Risk. *J Am Coll Cardiol* **72**, 1763-1773, doi:10.1016/j.jacc.2018.07.061  
850 (2018).

851 120 Ordelleide, A. M. *et al.* Common variation in the sodium/glucose cotransporter 2 gene  
852 SLC5A2 does neither affect fasting nor glucose-suppressed plasma glucagon  
853 concentrations. *PLoS One* **12**, e0177148, doi:10.1371/journal.pone.0177148 (2017).

854 121 Look Ahead Research Group *et al.* Cardiovascular effects of intensive lifestyle  
855 intervention in type 2 diabetes. *N Engl J Med* **369**, 145-154,  
856 doi:10.1056/NEJMoa1212914 (2013).

857 122 Look, A. R. G. *et al.* Association of the magnitude of weight loss and changes in physical  
858 fitness with long-term cardiovascular disease outcomes in overweight or obese people  
859 with type 2 diabetes: a post-hoc analysis of the Look AHEAD randomised clinical trial.  
860 *Lancet Diabetes Endocrinol* **4**, 913-921, doi:10.1016/S2213-8587(16)30162-0 (2016).

861 123 Wanner, C. *et al.* Empagliflozin and Progression of Kidney Disease in Type 2 Diabetes. *N*  
862 *Engl J Med* **375**, 323-334, doi:10.1056/NEJMoa1515920 (2016).

863 124 Zinman, B. *et al.* Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2  
864 Diabetes. *N Engl J Med* **373**, 2117-2128, doi:10.1056/NEJMoa1504720 (2015).

865 125 Burke, J. E. & Dennis, E. A. Phospholipase A2 structure/function, mechanism, and  
866 signaling. *J Lipid Res* **50 Suppl**, S237-242, doi:10.1194/jlr.R800033-JLR200 (2009).

867 126 Millwood, I. Y. *et al.* Lipoprotein-Associated Phospholipase A2 Loss-of-Function Variant  
868 and Risk of Vascular Diseases in 90,000 Chinese Adults. *J Am Coll Cardiol* **67**, 230-231,  
869 doi:10.1016/j.jacc.2015.10.056 (2016).

- 870 127 Millwood, I. Y. *et al.* A phenome-wide association study of a lipoprotein-associated  
871 phospholipase A2 loss-of-function variant in 90 000 Chinese adults. *International Journal*  
872 *of Epidemiology* **45**, 1588-1599, doi:10.1093/ije/dyw087 (2016).
- 873 128 O'Donoghue, M. L. *et al.* Effect of darapladib on major coronary events after an acute  
874 coronary syndrome: the SOLID-TIMI 52 randomized clinical trial. *JAMA* **312**, 1006-1015,  
875 doi:10.1001/jama.2014.11061 (2014).
- 876 129 Investigators, S. *et al.* Darapladib for preventing ischemic events in stable coronary heart  
877 disease. *N Engl J Med* **370**, 1702-1711, doi:10.1056/NEJMoa1315878 (2014).
- 878 130 Gan, W. *et al.* Bone mineral density and risk of type 2 diabetes and coronary heart  
879 disease: A Mendelian randomization study. *Wellcome Open Res* **2**, 68,  
880 doi:10.12688/wellcomeopenres.12288.1 (2017).
- 881 131 Emdin, C. A. *et al.* Genome-Wide Polygenic Score and Cardiovascular Outcomes With  
882 Evacetrapib in Patients With High-Risk Vascular Disease: A Nested Case-Control Study.  
883 *Circ Genom Precis Med* **13**, e002767, doi:10.1161/CIRCGEN.119.002767 (2020).
- 884 132 Holmes, M. V., Perel, P., Shah, T., Hingorani, A. D. & Casas, J. P. CYP2C19 genotype,  
885 clopidogrel metabolism, platelet function, and cardiovascular events: a systematic  
886 review and meta-analysis. *JAMA* **306**, 2704-2714, doi:10.1001/jama.2011.1880 (2011).
- 887 133 Baigent, C. & Holmes, M. V. Variability in aspirin efficacy: all in the genes? *Eur Heart J* **40**,  
888 3393-3396, doi:10.1093/eurheartj/ehz456 (2019).
- 889 134 Doherty, A. *et al.* GWAS identifies 14 loci for device-measured physical activity and sleep  
890 duration. *Nat Commun* **9**, 5257, doi:10.1038/s41467-018-07743-4 (2018).
- 891 135 Tillmann, T. *et al.* Education and coronary heart disease: mendelian randomisation  
892 study. *BMJ* **358**, j3542, doi:10.1136/bmj.j3542 (2017).
- 893 136 Carter, A. R. *et al.* Understanding the consequences of education inequality on  
894 cardiovascular disease: mendelian randomisation study. *BMJ* **365**, l1855,  
895 doi:10.1136/bmj.l1855 (2019).
- 896 137 Kettunen, J. *et al.* Genome-wide study for circulating metabolites identifies 62 loci and  
897 reveals novel systemic effects of LPA. *Nat Commun* **7**, 11122,  
898 doi:10.1038/ncomms11122 (2016).
- 899 138 Sun, B. B. *et al.* Genomic atlas of the human plasma proteome. *Nature* **558**, 73-79,  
900 doi:10.1038/s41586-018-0175-2 (2018).
- 901 139 Haycock, P. C. *et al.* Best (but oft-forgotten) practices: the design, analysis, and  
902 interpretation of Mendelian randomization studies. *Am J Clin Nutr* **103**, 965-978,  
903 doi:10.3945/ajcn.115.118216 (2016).
- 904 140 Davey Smith, G. *et al.* STROBE-MR: Guidelines for strengthening the reporting of  
905 Mendelian randomization studies. *PeerJ Preprints* **7**:e27857v1  
906 <https://doi.org/10.7287/peerj.preprints.27857v1> (2019).
- 907 141 Burgess, S. *et al.* Using published data in Mendelian randomization: a blueprint for  
908 efficient identification of causal risk factors. *Eur J Epidemiol* **30**, 543-552,  
909 doi:10.1007/s10654-015-0011-z (2015).
- 910 142 Millwood, I. Y. *et al.* Conventional and genetic evidence on alcohol and vascular disease  
911 aetiology: prospective study of 500,000 Chinese adults  
912 . *Lancet* **393**, 1831-1842 (2019).

- 913 143 GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature* **550**,  
914 204-213, doi:10.1038/nature24277 (2017).
- 915 144 Musunuru, K. *et al.* From noncoding variant to phenotype via SORT1 at the 1p13  
916 cholesterol locus. *Nature* **466**, 714-719, doi:10.1038/nature09266 (2010).
- 917 145 Porcu, E. *et al.* Mendelian randomization integrating GWAS and eQTL data reveals  
918 genetic determinants of complex and clinical traits. *Nat Commun* **10**, 3300,  
919 doi:10.1038/s41467-019-10936-0 (2019).
- 920 146 Kibinge, N. K., Relton, C. L., Gaunt, T. R. & Richardson, T. G. Characterizing the Causal  
921 Pathway for Genetic Variants Associated with Neurological Phenotypes Using Human  
922 Brain-Derived Proteome Data. *Am J Hum Genet* **106**, 885-892,  
923 doi:10.1016/j.ajhg.2020.04.007 (2020).
- 924 147 Suvarna, V. R. Real world evidence (RWE) - Are we (RWE) ready? *Perspect Clin Res* **9**, 61-  
925 63, doi:10.4103/picr.PICR\_36\_18 (2018).
- 926 148 Wainberg, M. *et al.* Homogeneity in the association of body mass index with type 2  
927 diabetes across the UK Biobank: A Mendelian randomization study. *PLoS Med* **16**,  
928 e1002982, doi:10.1371/journal.pmed.1002982 (2019).
- 929 149 Xu, Z. M. & Burgess, S. Polygenic modelling of treatment effect heterogeneity. *medRxiv*,  
930 doi:10.1101/2020.01.06.20016618 (2020).
- 931 150 North, T. L. *et al.* Using Genetic Instruments to Estimate Interactions in Mendelian  
932 Randomization Studies. *Epidemiology* **30**, e33-e35, doi:10.1097/EDE.0000000000001096  
933 (2019).
- 934 151 Chaturvedi, N. Ethnic differences in cardiovascular disease. *Heart* **89**, 681-686,  
935 doi:10.1136/heart.89.6.681 (2003).
- 936 152 Winkleby, M. A., Kraemer, H. C., Ahn, D. K. & Varady, A. N. Ethnic and socioeconomic  
937 differences in cardiovascular disease risk factors: findings for women from the Third  
938 National Health and Nutrition Examination Survey, 1988-1994. *JAMA* **280**, 356-362,  
939 doi:10.1001/jama.280.4.356 (1998).
- 940 153 Including Women and Minorities in Clinical Research Background. URL:  
941 [https://orwh.od.nih.gov/womens-health/clinical-research-trials/nih-inclusion-](https://orwh.od.nih.gov/womens-health/clinical-research-trials/nih-inclusion-policies/including-women-and-minorities)  
942 [policies/including-women-and-minorities](https://orwh.od.nih.gov/womens-health/clinical-research-trials/nih-inclusion-policies/including-women-and-minorities). *National Institutes of Health (Accessed 4th*  
943 *August 2020)*.
- 944 154 Tamargo, J. *et al.* Gender differences in the effects of cardiovascular drugs. *Eur Heart J*  
945 *Cardiovasc Pharmacother* **3**, 163-182, doi:10.1093/ehjcvp/pvw042 (2017).
- 946 155 Censin, J. C., Bovijn, J., Holmes, M. V. & Lindgren, C. M. Commentary: Mendelian  
947 randomization and women's health. *Int J Epidemiol* **48**, 830-833, doi:10.1093/ije/dyz141  
948 (2019).
- 949 156 Mills, M. C. & Rahal, C. A scientometric review of genome-wide association studies.  
950 *Commun Biol* **2**, 9, doi:10.1038/s42003-018-0261-x (2019).
- 951 157 Pirastu, N. *et al.* Genetic analyses identify widespread sex-differential participation bias.  
952 *bioRxiv* (2020).
- 953

## 954 Acknowledgements

955 The Medical Research Council (MRC) and the University of Bristol support the MRC Integrative  
956 Epidemiology Unit [MC\_UU\_12013/1, MC\_UU\_12013/9, MC\_UU\_00011/1]. TGR is a UKRI  
957 Innovation Research Fellow (MR/S003886/1). BAF is supported by the National Institute for  
958 Health Research (NIHR) Cambridge Biomedical Research Centre at the Cambridge University  
959 Hospitals NHS Foundation Trust. NMD is supported by a Norwegian Research Council Grant  
960 number 295989. MVH works in a unit that receives funding from the UK Medical Research  
961 Council and is supported by a British Heart Foundation Intermediate Clinical Research  
962 Fellowship (FS/18/23/33512) and the National Institute for Health Research Oxford Biomedical  
963 Research Centre.

964

#### 965 **Competing interests**

966 MVH has collaborated with Boehringer Ingelheim in research, and in adherence to the  
967 University of Oxford's Clinical Trial Service Unit & Epidemiological Studies Unit (CSTU) staff  
968 policy, did not accept personal honoraria or other payments from pharmaceutical companies.  
969 BAF reports receiving grants from Amgen, Merck & Co., Novartis, and Esperion Therapeutics;  
970 consulting or advisory board fees from Amgen, Regeneron, Sanofi, Merck & Co., Pfizer, CiVi  
971 BioPharma, and KrKA Pharmaceuticals; and grants from Merck & Co., Amgen, Novartis, Novo  
972 Nordisk, Regeneron, Sanofi, Pfizer, Eli Lilly, Mylan, Ionis, dalCOR, Silence Therapeutics, Integral  
973 Therapeutics, CiVi Pharma, KrKa Pharmaceuticals, American College of Cardiology, European  
974 Atherosclerosis Society, and European Society of Cardiology. NMD reports funding from the  
975 Global Research Awards into Nicotine Dependence (GRAND) which is an independent grant

976 awarding body funded by Pfizer. All other authors declare no conflicts of interest. GDS has  
977 received research support from GSK and Biogen.

978

979

## 980 **Key points**

981 • Mendelian randomization offers unique opportunities to explore the likely causal role of  
982 biomarkers and drug targets in disease aetiology.

983 • A biomarker is typically a complex trait that has a polygenic architecture, whereas a  
984 drug target is usually a protein that has a monogenic architecture.

985 • The motivation, application and interpretation of Mendelian randomization as applied  
986 to complex biomarkers and drug targets differs,

987 • These differences have implications for how each source of evidence contributes to  
988 aetiological insights in disease and in predicting the effects of therapies on clinical  
989 outcomes

990 • In this review, we describe these differences and offer insights into how such can be  
991 leveraged to comprehensively evaluate the causal role of a biomarker followed by an  
992 evaluation of which drug targets may be most tractable to the clinic.

993

994

995 **Figure 1. DNA, proteins, complex biomarkers and disease.** Proteins, encoded by genetic  
996 variants, are modified into complex biomarkers. Some proteins and complex biomarkers play a  
997 causal role in cardiovascular disease. If we wish to modify a complex biomarker that has a

998 putative causal role in CHD aetiology, this is typically done by therapeutic modification of an  
999 upstream protein trait. The figure illustrates the distinction between a protein trait (being  
1000 typically monogenic) and a complex biomarker (being polygenic). It is this distinction that is  
1001 critical to differentiating the motivation, conduct, interpretation and application of MR to drug  
1002 targets (i.e. proteins) as compared to complex biomarkers.

1003

1004 **Figure 2. Applications of human genetics to the lifecycle of drug development.** Mendelian  
1005 randomization can be used at each stage of the drug development pipeline. cIMT: carotid  
1006 intima medial thickness; PheWAS: phenome-wide association study

1007

1008 **Figure 3. Application of Mendelian randomization to clarify where on the causal pathway to**  
1009 **intervene.** Randomized evidence derived from treatment trials and human genetics for traits  
1010 on a causal pathway can be used to construct evidence of the likely underlying mechanisms,  
1011 and which trait to intervene on to optimally lower risk of disease. As the focus shifts upstream  
1012 (i.e. from right to left in the plot above), there is greater potential for drug targets to display  
1013 heterogeneous disease associations (including both those that may be adverse, and those that  
1014 may be beneficial) arising from target-mediated pleiotropy, described in **Box 5**. Mendelian  
1015 randomization using variants in/around the gene encoding C-reactive protein (CRP) does not  
1016 support the hypothesis that CRP causes CHD <sup>51</sup>. Exploring inflammation-related proteins more  
1017 upstream, genetic variants around the gene encoding interleukin 6 receptor (IL6R) shows an  
1018 association with risk of CHD <sup>21,54</sup>. A clinical trial of a monoclonal neutralising antibody inhibitor  
1019 to interleukin 1 beta (canakinumab) showed efficacy in a phase III cardiovascular outcome

1020 trial<sup>52</sup>. Nucleotide-binding domain leucine-rich repeat and pyrin domain containing receptor 3  
1021 (NLRP3) inflammasome (upstream of IL1b) is under exploration as a potential therapeutic target  
1022 <sup>102</sup>.

1023

1024 **Figure 4. Mendelian randomization for mediation.** MR can be used to identify intermediates  
1025 that lie on the causal pathway between polygenic (complex) biomarkers and disease. The figure  
1026 depicts an exploration of traits that mediate the effect of body mass index (used as a measure  
1027 of obesity) and risk of coronary heart disease (CHD). By identifying traits that BMI modifies and  
1028 conducting de novo MR investigations, it is possible to identify causal intermediaries (i.e. traits  
1029 that lie on the causal pathway from BMI to risk of CHD). Such may lead to the development of  
1030 new therapies that stem the tide of harmful exposures (e.g. excess adiposity) that are proving  
1031 challenging to modify at a population level. For further information, please see **Box 3**.

1032

1033 **Figure 5. Biomarker versus drug target MR.** For complex biomarker MR, a genetic instrument  
1034 typically consists of multiple SNPs selected from across the genome (a polygenic instrument)  
1035 associated with the biomarker of interest. A suite of sensitivity analyses can be used to test for  
1036 evidence of unbalanced (directional) horizontal pleiotropy<sup>37</sup>. In contrast, a drug-target MR  
1037 typically uses one or a handful of SNPs selected from the gene that encodes an individual  
1038 protein (*cis*-acting variants) in a monogenic instrument. Even when a drug target is selected on  
1039 the basis of modifying a given biomarker, the biomarker and drug target MR estimates may  
1040 differ. For example, in the example illustrated, an MR of circulating HDL-C finds little-to-no  
1041 evidence that HDL causally affects risk of CHD (once taking into account unbalanced horizontal



1042 pleiotropy<sup>18,19,58</sup>). However, using SNPs located in *CETP* associated with HDL-C at genome-wide  
1043 significant levels in a drug-target MR suggests that modifying HDL-C via CETP, is likely to reduce  
1044 risk of CHD<sup>23,103</sup>. Rather than being attributable to HDL-C, this CHD association arises due to  
1045 target-mediated pleiotropy through apolipoprotein B.

1046

1047

1048 **Figure 6. Drug-target Mendelian randomization using *cis* vs *trans*-acting SNPs.** In this example,  
1049 we have 4 *cis*-acting SNPs in the gene encoding the protein and 7 *trans*-acting SNPs from other  
1050 parts of the genome. By virtue of the mechanism by which the *trans*-acting SNPs associate with  
1051 protein levels, they are more likely to have pleiotropic effects (illustrated by the red arrow) that  
1052 may modify disease risk independently of the exposure (thus violating one of the central  
1053 assumptions of Mendelian randomization<sup>95</sup>). Care must therefore be taken in the application of  
1054 *trans*-acting variants in drug target MR as their inclusion in the genetic instrument may lead to  
1055 bias. An example relates to the use of trans-pQTL to instrument C-reactive protein, which can  
1056 distort the CHD association, leading to an erroneous conclusion.

1057

1058 **Figure 7. A framework for identifying therapeutic targets.** How might we make progress on  
1059 identifying treatment targets for disease, especially when causal risk factors are not fully  
1060 characterised? GWAS of disease can be exploited by conducting phenome-wide association  
1061 analyses (PheWAS) of disease-associated SNPs. The phenotypic signature of a genetic risk score  
1062 (GRS) derived from SNPs reliably associated with disease will likely capture multiple signals,  
1063 including associations arising due to causal pathways leading to disease, trait associations

1064 arising due to disease and other elements<sup>104</sup>. Once equipped with the phenotypic signature of a  
1065 disease GRS, de novo MRs of those traits can be conducted to establish disease causation (by  
1066 means of a conventional biomarker MR). If a biomarker is identified to cause disease, de novo  
1067 GWAS of the biomarker may identify SNPs in loci encoding therapeutic targets. An exploration  
1068 through available resources such as OpenTargets or ChEMBL<sup>105</sup> can provide additional  
1069 information to establish whether individual loci are likely to represent tractable therapeutic  
1070 targets. Drug-target MR of protein-encoding loci identified as being of relevance to risk factors  
1071 can establish the predicted effects of pharmacological modification in clinical trials and be used  
1072 to guide clinical development (as shown in **Figure 2**). As discussed in the text, care needs to be  
1073 placed in interpreting the magnitudes of effect given that MR of an exposure represents,  
1074 naively, lifetime exposure and depending on the exposure-outcome pair, this may lead to a  
1075 magnification of the effect estimates arising from MR. Of note, it may be possible to proceed  
1076 from disease GWAS through to drug-target MR without elucidating causal risk factors, but this  
1077 is likely to have less power to do so.

1078

**Table 1.** Key features of Mendelian randomization of a biomarker and a drug target

Type of MR	Biomarker MR	Drug target MR
Question	Does this biomarker cause disease?	Does modifying this drug target alter risk of disease?
Motivation	Public health (e.g. implications of obesity on disease) Biological understanding to discover biomarkers to intervene on	New therapies for the prevention of incident and recurrent disease, and disease-related sequelae
SNP selection	GWAS-derived preferable (unless protein exposure)	<i>Cis</i> -acting SNPs where possible
MR methods	Heavily reliant on sensitivity analyses to show lack of unbalanced horizontal pleiotropy	Might be less vulnerable to pleiotropy but should ideally check consistency of independent SNPs in same locus
Vertical pleiotropy	Can point to pharmacological approaches to intervention (e.g. proteomic mediators of adiposity and BMI/CAD)	Can shed light on differences in treatments (e.g. PCSK9 and HMGCR) on intermediaries (e.g. blood metabolic markers <sup>106</sup> ), which can then be used to explore whether such differences are clinically meaningful <sup>92</sup>
Horizontal pleiotropy	Need to formally evaluate, e.g. through robust MR approaches <sup>37</sup>	Confounding by linkage disequilibrium can be assessed through colocalization <sup>84</sup>
Causal inference	Evidence on whether the biomarker causes incident disease and progression of disease	Evidence on whether modifying a drug target is likely to alter risk of (i) disease onset (when the study is set in general population); or (ii) disease progression (when the study is set in individuals with established disease)
Role of PheWAS	Unknown effects of biomarker on other biomarkers and diseases	Unknown (both unintended: ADRs, and beneficial: repurposing) target-mediated effects
Limitations	Public health interventions may not reciprocate effects (as challenges in purely 'modifying' biomarker)	Cannot predict <i>off-target</i> effects of individual therapies

1081 **Table 2. Scenarios where findings from MR of phenotype, drug target and treatment trials**

1082 **may diverge**

Scenario	MR of phenotype	MR of drug target	RCT	Potential Explanation(s)* and examples
1	Causal	Causal	Causal	Biomarker causes disease; and modifying biomarker through the drug target has same effect on disease. E.g. LDL-C and apolipoprotein B (phenotype) <sup>40</sup> and HMGCR <sup>107</sup> , PCSK9, NPC1L1 (drug targets); statins <sup>39,108,109</sup> <sup>110</sup> PCSK9 mAbs and ezetimibe (drugs)
2	Non-causal	Non-causal	Non-causal	Biomarker doesn't cause disease, nor does the drug target. E.g. HDL-C <sup>58</sup> , <i>CETP</i> variants that don't associate with LDL cholesterol <sup>77</sup> , <i>CETP</i> therapeutic inhibitors that don't lower LDL-cholesterol (or apolipoprotein B) <sup>76</sup> , and apo-A1 infusions <sup>111</sup>
3	Non-causal	Causal	Causal	Drug target shows target-mediated pleiotropy that leads to beneficial effect on disease. E.g. <i>CETP</i> variants that lower apolipoprotein B <sup>23</sup> , and potent

				pharmacological CETP inhibitors <sup>75</sup>
	<p>target mediated pleiotropy</p>			
4	Non-causal	Non-causal	Causal	<p>Drug has off-target effects; E.g. CETP inhibitors and systolic blood pressure<sup>74,79</sup></p> <p>Also: drug target tested in RCT causal for recurrent events but not incident events (in which MRs conducted) (e.g. biomarkers involved in plaque stability<sup>112</sup>).</p>
	<p>(not recapitulated in MR of drug target or complex biomarker)</p>			
5	Causal	Non-causal	Causal	<p>(a) SNPs used for drug target MR confounded by correlation with other SNPs (through linkage disequilibrium) that have opposing effect on disease risk with the net effect to bias disease associations to the null; (b) Genetic variant alters an epitope binding site for assay, which has no qualitative or quantitative effect of the protein</p>

	<p>A) confounding by linkage disequilibrium</p> <p>Drug target → modifies → complex biomarker → causal → disease</p> <p>confounded genetic estimate from drug-target MR biases estimate to null</p> <p>B) epitope alteration leads to false positive for protein assay and false negative on MR of disease</p> <p>Drug target → modifies → complex biomarker → causal → disease</p> <p>false -ve (as SNP to protein estimate erroneous)</p>			
6	Causal	Non-causal	Non-causal	Biomarker causes disease, but drug target has target-mediated pleiotropy that off-sets this effect to yield net null effect
	<p>Drug target → modifies → complex biomarker → causal → disease</p> <p>target mediated pleiotropy → additional trait → causal (but opposing effect)</p> <p>no net effect → disease</p>			
7	Causal	Causal	Non-causal	Biomarker important for disease onset but not for disease progression with biomarker and drug target MRs conducted in general population but RCT conducted in patients with disease (and vice versa). Biomarker (and drug target modifying biomarker) only causal for disease during a specific time (e.g. adolescence) and clinical trial conducted during adulthood – e.g. vitamin D and multiple sclerosis <sup>113</sup> .

				Development of neutralizing antibodies to biological therapies e.g. development of antibodies to PCSK9 mAbs <sup>114</sup>
8	Non-causal	Causal	Non-causal	Drug-target MR using cis-pQTL (protein quantitative loci) confounded by local LD

1083

1084

1085 Footnote: \* potential explanations assume that studies are adequately powered.

1086

1087 **Box 1:** Notable examples of Mendelian randomization studies of biomarkers, drug targets and  
 1088 randomized controlled trials

MR of biomarker	MR of drug target	RCT	Comments
LDL cholesterol – shows evidence of causation with CHD <sup>19</sup>	LDL-C lowering variants in <i>HMGCR</i> , <i>PCSK9</i> , <i>NPC1L1</i> <sup>107</sup> <sup>115</sup> associate with CHD	Multiple trials of drugs targeting LDL-C lowering show cardiovascular benefit <sup>39,116</sup>	The exemplar, where MR of a biomarker, drug targets and RCTs align
HDL cholesterol <sup>17,19</sup> MRs do not provide evidence of causation, once unbalanced horizontal pleiotropy is taken into account	<i>CETP</i> <sup>23</sup> variants associate with risk of CHD; <i>APOA1</i> <sup>117</sup> variants do not associate with risk of CHD	REVEAL trial of anacetrapib <sup>75</sup> demonstrated evidence of vascular benefit; RCTs of apoA-I infusion <sup>111</sup> have thus far not shown benefit	HDL-C and apoA-I are likely to be non-causal in CVD. CETP inhibitors are likely to lower risk of CHD if they lower apoB.
BMI increases risk of T2D and CHD in MR <sup>14</sup> with one potential mechanism being increased blood glucose	<i>GLP1R</i> variants that associate with lower blood glucose show evidence of a potential causal role of glucagon-like peptide-1 receptor (GLP1R) in T2D and CHD <sup>118</sup>  <i>SGLT1</i> variants are associated with altered blood glucose and risk of vascular disease <sup>119</sup> but it has proven challenging to identify variants in	The Look-AHEAD trial showed directionally consistent evidence of lower vascular risk <sup>121</sup> but was potentially underpowered, supported by post-hoc analyses <sup>122</sup>  Monoclonal antibody agonists to GLP-1R have shown efficacy for CHD <sup>70</sup>  SGLT2 monoclonal antibody inhibitors have shown efficacy	An example were modifying a complex biomarker (BMI) with recognised harmful effects on multiple diseases at the population level is challenging, requiring pharmacological modification of a protein upstream. Notably such protein drug targets may display target-mediated pleiotropic effects that are



	<i>SLC5A2</i> (which encodes SGLT2) to instrument therapeutic inhibition of SGLT2 <sup>120</sup>	for multiple endpoints related to metabolic and vascular disease <sup>123,124</sup> .	independent of BMI (see <b>Box 5</b> )
Phospholipase A2 (PLA <sub>2</sub> ) is a family of enzymes that hydrolyse fatty acid from membrane phospholipids <sup>125</sup> . Secretory phospholipase A2-IIA (sPLA <sub>2</sub> -IIA) and lipoprotein-associated PLA2 (Lp-PLA <sub>2</sub> ) were both thought to play potential roles in atherogenesis. sPLa-IIA <sup>42</sup> and Lp-PLA <sup>126,127</sup> drug-target MRs showed lack of evidence of causal roles in CVD		Varespladib (a drug targeting sPLA <sub>2</sub> -IIA) failed in VISTA16 <sup>41</sup> ; Darapladib (which targeted Lp-PLA <sub>2</sub> ) failed in SOLID-TIMI 52 <sup>128</sup> and STABILITY <sup>129</sup> trials	Billions of US dollars were spent on developing therapies that were destined to fail for lack of target-mediated disease causation.
MR of CRP <sup>51</sup> failed to demonstrate a causal role in CHD	The interleukin 6 receptor (IL6R) is a drug target that alters CRP (see <b>Figure 3</b> ), and for which genetic evidence supports a causal role of IL6R in CHD <sup>20 54</sup>	Selective IL-1 beta inhibitor (canakinumab) <sup>52</sup> , which is upstream of IL6R demonstrates benefit in phase III cardiovascular outcome trials	An example of applying randomized evidence from biomarkers on a causal pathway to elucidate where to intervene, illustrated in <b>Figure 3</b> .
MR of bone mineral density (BMD) found that higher BMD increased risk of CHD and T2D <sup>130</sup>	BMD-associated variants in <i>SOST</i> (used as a genetic instrument for sclerostin inhibition) show higher risk of CVD <sup>73</sup>	Early evidence from RCTs of drugs that modify sclerostin suggest potential excess risk of CVD <sup>73</sup>	An example of human genetics providing evidence to support a potential target-mediated adverse effect of a therapy

1090 **Box 2: How can genetic evidence can enhance the evidence base of causation to facilitate the**  
 1091 **conduct of randomized controlled trials?**

1092 Drug target Mendelian randomization studies can provide a source of valuable information that  
 1093 is of value to understanding whether the modification of a drug target is likely to lead to clinical  
 1094 benefit.  
 1095

Feature	Application of Mendelian randomization
Trial design	When a trial might be unethical (e.g. insufficient clinical equipoise, potentially harmful intervention or when there are potential ethical issues of withholding treatment or placebo comparator unethical)
	When a previously-conducted large-scale trial was underpowered (e.g. adiposity and vascular disease in Look-AHEAD <sup>121</sup> )
	To inform on whether a trial should be conducted in particular phenotypically-defined and/or genetically-defined subgroups
	To predict the outcome of a factorial trial design, e.g. by predicting drug by drug interactions (e.g. HMGCR and CETP <sup>23</sup> )
Characteristics of the intervention	When no therapies exist (e.g. drugs not yet developed – for example prior to development of PCSK9 inhibitors)
	To deprioritise development of non-causal therapies (e.g. those targeting CRP <sup>51</sup> , sPLA <sub>2</sub> -IIA <sup>42</sup> or Lp-PLA <sub>2</sub> <sup>126</sup> )
	To explore the nature of the causal relationship across the physiological distribution of the exposure - notably individual trials of a single dose of drug are unable to speak to causal effects across the physiological distribution of exposure on which they intervene
Mechanistic insights	To explain mechanisms of effect (e.g. <b>Figures 3-4, and Box 3</b> )
	To clarify pleiotropic effects <ul style="list-style-type: none"> <li>(a) Off-target effects of a <i>drug class</i>: e.g. whether the drug has effects on an outcome beyond that which is considered to be the primary ‘exposure’ (e.g. HMGCR/statins and CHD – effect estimates are proportionate to LDL-C reductions<sup>23</sup> seen with other lipid-associated genetic variants and lipid lowering therapies, arguing against pleiotropic effects that lead to CVD reduction through other mechanisms, such as a reduction in inflammation)</li> <li>(b) Off-target effects of an <i>individual drug</i>: e.g. torcetrapib blood pressure effects vs other CETP inhibitors<sup>74</sup></li> <li>(c) To explore the full repertoire of target-mediated effects (including those that are beneficial and potentially harmful) in</li> </ul>

	order to inform clinical trial design to enable accurate synthesis of risk benefit analyses, and potential repurposing to new indications
Adaptive trial design	<p>MR estimates with disease can be used to guide the selection of diseases used in a composite primary endpoint (e.g. removing elements from a composite outcome if MR shows that one element is not causally modified by the drug target). If this information becomes available during the conduct of a trial, the constituents of the composite primary endpoint might be modified with such knowledge through an adaptive design.</p> <p>Re-estimate power based on MR point estimate<sup>92</sup> (with appropriate scaling of effect estimate, if possible), prolong recruitment to increase number of participants and/or prolong duration of follow-up</p> <p>Use of MR to inform on subgroups to alter trial eligibility criteria to enrol participants that will experience:</p> <ul style="list-style-type: none"> <li>(a) greater benefit (may allow trial to recruit fewer numbers of participants and/or have a shorter duration to lower the cost of drug development)</li> </ul> <p>E.g. a recent re-appraisal of the CETP inhibitor evacetrapib that failed to demonstrate an overall effect on CAD was examined to explore whether an interaction existed on stratifying individuals by a CAD PGS. Perhaps unsurprisingly, this analysis showed no CAD PGS by evacetrapib interaction<sup>131</sup>. However, repeating the analysis by conducting a GWAS of apoB lowering by evacetrapib and constructing a <i>de novo</i> PGS based on SNPs demonstrated to modify the drug-induced apoB lowering might well display a qualitative interaction (between evacetrapib and the PGS) on risk of CAD</p> <ul style="list-style-type: none"> <li>(b) fewer adverse effects: although it may not be possible to separate out the mechanism-based adverse effects of a therapeutic (e.g. <i>CYP2C19</i>/clopidogrel<sup>132</sup>; <i>GUCY1A3</i>/aspirin<sup>133</sup>) if an adverse event arises due to a mechanistic pathway that is separate to the mechanistic pathway leading to benefit, it may be possible to identify genetic variants that modify the propensity of the pharmacological agent to modify the pathway leading to adverse events, and by doing so, disentangle benefit from harm.</li> </ul>



1098 **Box 3: Using Mendelian randomization to explore potential mediators of the relationships of**  
1099 **complex exposures and disease risk.**

1100 One of the prevailing challenges for population health is that while we understand the causal  
1101 relationships of multiple risk factors for disease, we are unable to modify them at a population-  
1102 level in order to gain maximal benefit <sup>61</sup> (e.g. BMI and CVD). Such population-wide interventions  
1103 are likely to be identified in the future but may take decades to develop. What can be done in  
1104 the mean time? One as-of-yet untapped, but potentially lucrative, application of MR is in its  
1105 ability to identify mediators on the causal pathway from complex exposures (such as  
1106 adiposity<sup>55</sup>, physical activity<sup>134</sup>, education<sup>135,136</sup>) and disease risk (e.g. CHD, T2D). With the  
1107 repertoire of metabolites<sup>137</sup> and proteins<sup>138</sup> that are being measured in sufficient numbers to  
1108 characterize their genetic architecture through GWAS and construct reliable genetic  
1109 instruments, MR is uniquely placed to elucidate what lies on the causal pathway from  
1110 population-wide 'lifestyle' exposures (e.g. BMI) and disease. This can be conducted by  
1111 performing MR of known causal risk factors (e.g. BMI) as the exposure on proteins and  
1112 metabolites as outcomes, and, for those proteins and metabolites identified to be altered by  
1113 genetically predicted differences in BMI, conduct *de novo* MRs of these proteins and  
1114 metabolites onto disease. Through the intersection of this with drug development, it may be  
1115 possible to identify druggable biomarkers that mediate (i.e. lie on the causal pathway(s))  
1116 between complex exposures and disease . Through the pharmacological modification of such  
1117 intermediates, it may be possible to ameliorate the harmful effects of upstream complex  
1118 exposures that are challenging to modify at the population level.

1119

1120 **Box 4: How to conduct the optimal MR for a biomarker vs a drug target?**

1121 Several articles have described approaches to the conduct of MR analyses<sup>37,139-141</sup>, obviating the  
1122 need to provide an exhaustive prescription here.

1123 Rather, we consider the following scenarios:

1124 (1) *A polygenic biomarker*, for example, body mass index (BMI) or high-density lipoprotein  
1125 cholesterol (HDL-C), the optimal approach is to use a polygenic instrument with as many SNPs  
1126 as possible that reliably associate with the exposure of interest, while avoiding weak instrument  
1127 bias and the introduction of horizontal pleiotropy. This is typically conducted through the use of  
1128 SNPs reaching GWAS significance. The reason for the general preference to use as many SNPs  
1129 that reliably associate with the exposure as possible is to capture the full repertoire of the  
1130 genetic architecture of the trait. The trade-off is that this will invariably mean that many of the  
1131 SNPs in the instrument will associate with the exposure through unknown mechanisms.  
1132 However, such an approach to constructing genetic instruments permits sensitivity analyses to  
1133 be conducted allowing a thorough appraisal of the potential for violation of the instrumental  
1134 variable assumptions through e.g. horizontal pleiotropy.

1135

1136 (2) *A biomarker where the biological synthesis or metabolism is clearly defined* – e.g. the  
1137 metabolism of alcohol by aldehyde dehydrogenase (encoded for by variants in *ALDH2*). In this  
1138 case, while it may be feasible and informative to use a polygenic instrument for alcohol, it may  
1139 be preferable to use one or more SNPs (a monogenic instrument) of known function that  
1140 influence either the synthesis or metabolism of the biomarker<sup>142</sup>. The principal advantage here  
1141 is specificity of the instrument. This comes at a strategic shortcoming: that the instrument

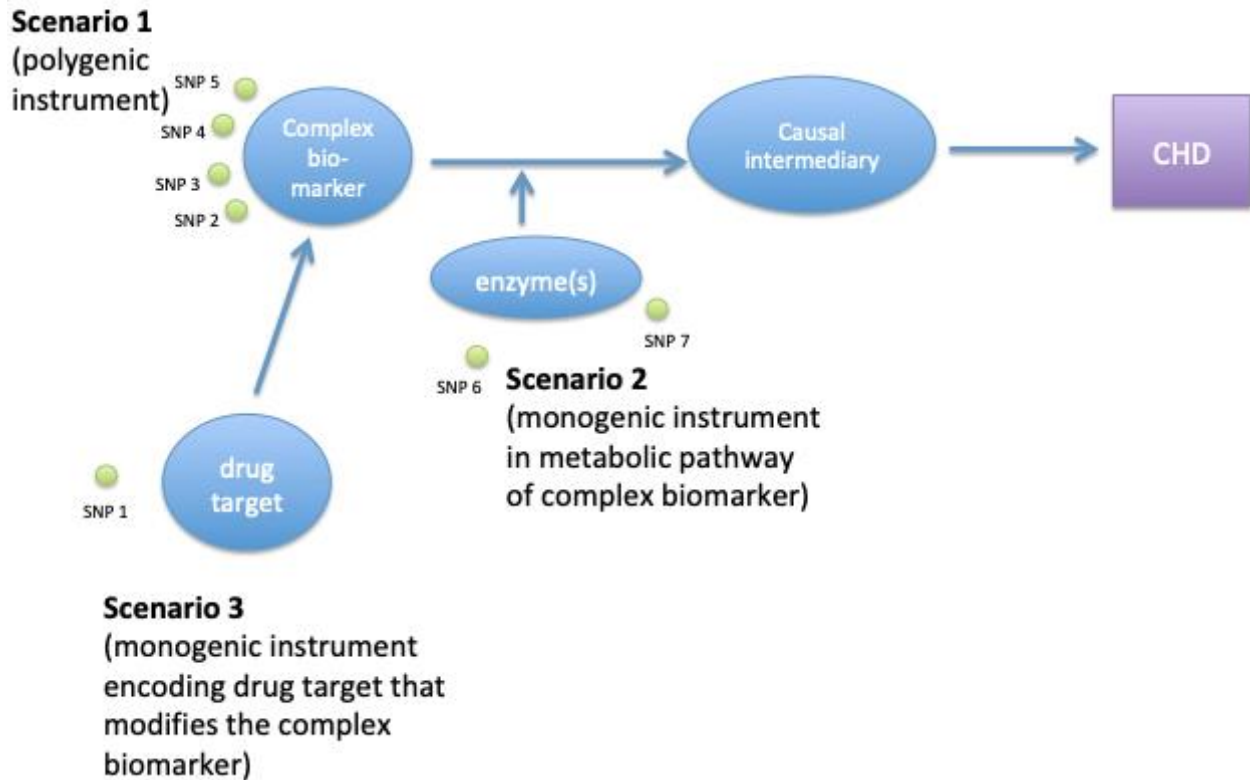
1142 specificity relies on the biological plausibility of such specificity. A phenotypic scan of  
1143 monogenic vs polygenic instrument and both instruments vs phenotype can help unravel such,  
1144 but it may be challenging to interpret, owing to the potential issue of differential confounding  
1145 (i.e. confounding differentially affecting the polygenic- vs monogenic -MR and observational  
1146 analyses).

1147

1148 (3) *Drug target*– in this case there are three separate approaches that can be undertaken. (i)  
1149 use only *cis* instruments; (ii) use only *trans* instruments (see Box 5 for definition); (iii) using both  
1150 *cis* and *trans* instruments. The advantages of using *cis* SNPs include specificity, and the  
1151 possibility of assessing such specificity through co-localisation. The disadvantages of using *cis*-  
1152 only includes potential lack of power owing to an insufficient proportion of variance of the  
1153 exposure explained, and potential inability to conduct sensitivity analyses that require multiple  
1154 variants (e.g. MR-Egger and weighted median approaches, *inter alia*). Use of *trans* SNPs by  
1155 definition includes SNPs that associate through either an upstream mechanism or gene  
1156 regulatory functions, that may lead to biased estimates (see **Figure 6**), but which can potentially  
1157 be (a) compared to the *cis* instrument for comparability and (b) analysed for pleiotropy using  
1158 well-established MR approaches.

1159

Figure to accompany Box 4



1160  
1161

1162

1163

1164

1165 **Box 5. Pleiotropy as applied to drug target Mendelian randomization**

1166 In this article we describe various forms of pleiotropy – namely that which is **target-mediated,**

1167 **horizontal,** and **vertical. Target mediated pleiotropy** is the situation where a drug target

1168 demonstrates effects that are discrete to those of the complex biomarker – this usually arises

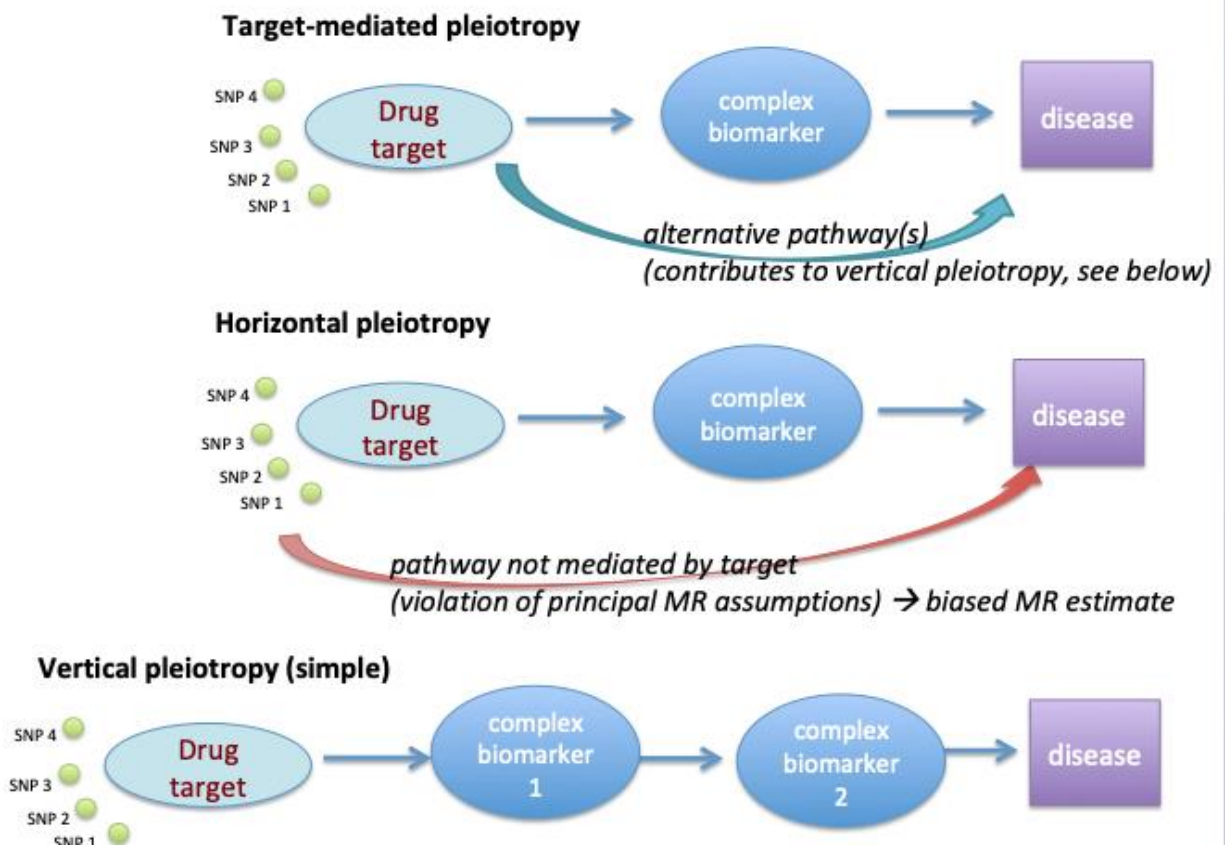
1169 from pathways branching off upstream of the biomarker (see **Figure**). **Horizontal pleiotropy** is

1170 when the genetic instrument associates with the outcome through pathways other than those



1171 which are mediated through the primary exposure (in this case a protein), and can violate the  
1172 exclusion restriction criterion of instrumental variable analysis. **Vertical pleiotropy** is the  
1173 principle on which Mendelian randomization rests – i.e. the association of the genetic  
1174 instrument with traits (and disease) owing to them being downstream consequences of the  
1175 protein drug target.

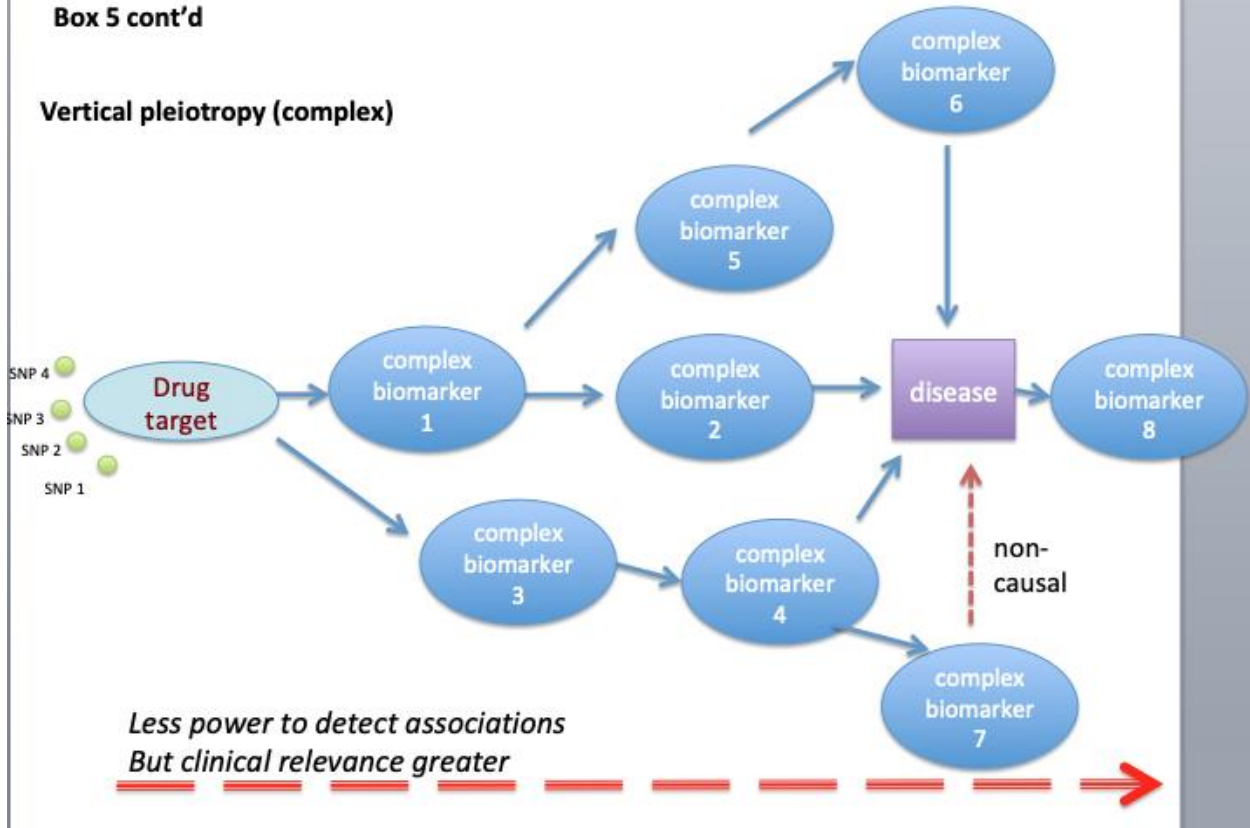
1176



1177

Box 5 cont'd

Vertical pleiotropy (complex)



1178

1179 **Box 6: *Cis* and *trans* genetic variants**

1180 Genetic variants that are within or close to gene that encodes for the protein of interest can  
1181 affect proteins either qualitatively or quantitatively, with the latter leading to e.g. lower levels  
1182 of a protein. These proximal genetic variants are known as *cis*-acting variants. By altering the  
1183 exposure of interest (i.e. the protein) through the gene that directly encodes it, such *cis*-acting  
1184 variants are likely to have effects specific to the protein of interest, and are less likely to be  
1185 pleiotropic. In contrast, genetic variants that affect a protein that are not close to the gene that  
1186 encodes the protein of interest are known as *trans*-acting variants. By virtue of being distal to  
1187 the gene-encoding locus, *trans*-acting variants are more likely to have pleiotropic effects that  
1188 affect multiple pathways (as illustrated in **Figure 6**). As a result, *trans*-acting variants can be  
1189 more challenging to interpret in protein drug-target MR studies, as their effects can be

1190 mediated by multiple pathways, some of which might be pleiotropic and lead to confounded  
1191 interpretations from MR.

1192

### 1193 **Box 7. Use of tissue-specific gene expression data for drug-target MR**

1194 Proteins are synthesized following the transcription of DNA to messenger RNA (mRNA), with  
1195 mRNA being subsequently translated, through transfer RNA, into amino acids, which are  
1196 polymerised into proteins. Characterisation of mRNA provides a means to quantify gene  
1197 expression in tissues, and evaluate pre-translational effects of genes, especially those encoding  
1198 drug targets, across different tissue types <sup>143</sup>. Increasing availability of genetic data linked to  
1199 tissue-specific mRNA expression in human tissues (so-called expression quantitative trait loci, or  
1200 eQTL) makes it possible to leverage this information to help guide instrument selection for  
1201 drug-target MR. Data from gene expression in human tissues (so-called expression quantitative  
1202 trait loci, or eQTL) are increasingly available making it possible to leverage this information to  
1203 help guide instrument selection for drug-target MR. For example, one is the earliest LDL  
1204 cholesterol signals detected by GWAS was at the locus on chromosome 1p13, likely caused by  
1205 the *SORT1* gene <sup>144</sup>. Applying the principals of MR to evaluate the genetically predicted effects  
1206 of *SORT1* expression on LDL cholesterol highlights the importance of tissue-specificity when  
1207 selecting eQTL as instruments<sup>145</sup>. The genotype-tissue expression (GTEx) project has provided  
1208 evidence that expression of *SORT1* in liver tissues affects LDL cholesterol levels (Beta=0.067,  
1209 SE=0.005, P=7.03x10<sup>-38</sup>) <sup>143</sup>. As LDL cholesterol is synthesised by the liver, using gene expression  
1210 data derived from this organ is likely the most pertinent in capturing the functional mechanism  
1211 responsible for this GWAS signal. In contrast, using eQTL derived from other tissues types which

1212 are not relevant to LDL cholesterol synthesis incorrectly implicates *PSRC1* as the putative causal  
1213 gene at this locus (e.g. skin tissue Beta=0.063, SE=0.011, P=8.38x10<sup>-09</sup>).

1214 A major limitation currently in using these tissue-specific molecular datasets is that the sample  
1215 sizes are typically modest compared to GWAS standards (e.g. the latest version of GTEx ranges  
1216 between n=21 and n=755 across tissues). This often results in having only a single independent  
1217 cis-regulatory variant for MR analyses, meaning that findings need to be subsidised by evidence  
1218 from genetic colocalization analyses (see refs<sup>83-85</sup>). Another challenge encountered is  
1219 disentangling the effects of nearby genes which are often co-expressed and thus correlated,  
1220 which may suggest that a non-causal gene is responsible for a GWAS signal (e.g. as described  
1221 above for *PSRC1* and LDL cholesterol). Furthermore, publicly available data on proteins are  
1222 predominantly derived using whole blood rather than disease-relevant human tissues, which  
1223 will be preferential for drug target validation purposes once they are accessible at the same  
1224 scale and coverage as eQTL datasets. Once this is the case, they are likely to be a valuable  
1225 source of instruments which may be capable of detecting effects potentially overlooked by  
1226 analysing data from the plasma proteome<sup>146</sup>.

1227

1228

### 1229 **Box 8. Heterogeneity of treatment effects**

1230 An increasingly seen criticism of randomized controlled trials is that they are carried out in  
1231 highly selected populations, and therefore their findings may not be generalizable to other  
1232 patient groups<sup>99,147</sup>. MR can contribute here by using genetic proxies for treatments to  
1233 examine whether the same predicted effects of either biomarker change or drug treatment are

1234 seen in different subgroups of the population. For example, an MR study of body mass index  
1235 (BMI) effects on type II diabetes yielded the same predicted relative risk reduction for a given  
1236 difference in BMI among groups at different level of risk <sup>148</sup>. Similar analyses could be done  
1237 using drug targets across risk groups. Such groups can include those taking other drugs, either  
1238 though applying MR in those taking such agents, or by proxying both drugs using MR – as, for  
1239 example, in a factorial study of MR predicted effects of PCSK9 inhibition and HMGCR inhibition  
1240 on CHD and type II diabetes <sup>107</sup>. It is also possible to test for interactions of effects of a genetic  
1241 proxy for drug treatment with genome wide genetic variation, to examine whether there is any  
1242 evidence of overall heterogeneity of effect, with a study of *HMGCR* finding no strong predicted  
1243 effect modification <sup>149,150</sup>. These analyses can be useful for prospectively designing clinical trials,  
1244 in order to maximize effect estimates, and also as an approach to stratified medicines.

1245

1246 **Box 9. Promoting equitable development of therapeutic targets through Mendelian**  
1247 **randomization**

1248 As the conduct of Mendelian randomization studies becomes increasingly feasible, facilitated  
1249 by availability of data and analytical platforms permitting rapid analyses at scale, an important  
1250 question is whether our focus is on asking the right questions? For cardiovascular disease drug  
1251 development, this should ensure that drugs are developed equitably, ensuring that effective  
1252 treatments are developed and made available for minority under-represented groups.

1253 Cardiovascular disease and risk factors for CVD disproportionately affects individuals of lower  
1254 socioeconomic status, is unevenly balanced between women and men, and incidence is higher  
1255 in individuals of certain ethnic and minority groups (e.g. Asians <sup>151,152</sup>). Women and minorities

1256 have historically been under investigated and under-represented in clinical research<sup>153</sup>. It is  
1257 plausible that the relationship between risk factors and disease may differ meaningfully  
1258 between women and men, with potential consequences on providing appropriate treatment  
1259 and preventative strategies <sup>154</sup>. In clinical trials, a difference in treatment effect is typically  
1260 investigated by a priori defined subgroup analyses where the effect of an intervention on an  
1261 outcome is stratified by sex, and analysis permits quantification of whether there is  
1262 heterogeneity in treatment effects between women and men. In Mendelian randomization, a  
1263 similar approach can be used<sup>155</sup>, but requires additional features including representation of  
1264 minority groups in GWAS. As of 2017, almost 90% of participants in GWAS were European  
1265 ancestry <sup>156</sup>. If the genetic architecture of the exposure (be it a drug target or complex  
1266 biomarker) and outcome differs between the strata under investigation (e.g. defined by sex),  
1267 then ideally this should be reflected in MR analysis by using e.g. sex-stratified GWAS in order to  
1268 select representative genetic instruments. This can pose challenges if genetic instruments are  
1269 differentially available across strata and/or if biases exist (e.g. selection biases into the study  
1270 that are differential by exposure and stratum <sup>157</sup>).

1271 In a similar vein, GWAS in non-European ethnic groups can facilitate explorations of whether  
1272 there are important differences of the causes of disease by ethnicity.

1273

1274

1275

1276