

1 **Do smoking and alcohol behaviours influence risk of type 2 diabetes? A Mendelian**
2 **randomisation study.**

3 Short title: Mendelian randomisation of smoking, alcohol and type 2 diabetes

4 Zoe E. Reed^{1,2}, Hannah M. Sallis^{2,3,4}, Rebecca C. Richmond^{2,3}, Angela S. Attwood^{1,2}, Deborah
5 A. Lawlor^{2,3}, Marcus R. Munafò^{1,2,5}

6 1 School of Psychological Science, University of Bristol, Bristol, UK

7 2 MRC Integrative Epidemiology Unit at the University of Bristol, Bristol, UK

8 3 Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

9 4 Centre for Academic Mental Health, Bristol Medical School, University of Bristol, Bristol,
10 UK

11 5 National Institute for Health Research Bristol Biomedical Research Centre, University
12 Hospitals Bristol NHS Foundation Trust and University of Bristol, Bristol, UK

13 Corresponding author: Zoe Reed, zoe.reed@bristol.ac.uk

14 **Abstract**

15 **Background:** Previous studies suggest that smoking and higher alcohol consumption are
16 both associated with greater risk of type 2 diabetes (T2D). However, studies examining
17 whether these associations reflect causal relationships are limited and do not consider
18 continuous glycaemic traits. The aim of the study was to determine whether there are
19 causal effects of smoking and alcohol consumption on T2D risk and related glycaemic traits.

20 **Methods and Findings:** We conducted both two-sample and one-sample MR to examine the
21 effects of lifetime smoking index (LSI) and alcoholic drinks per week on T2D and continuous
22 traits (fasting glucose, fasting insulin and glycated haemoglobin, HbA1c). For two-sample
23 MR we used results from genome-wide association studies (GWAS) of LSI (N=462,690),
24 alcohol consumption (N=941,280), T2D (N= 148,726 cases and 965,732 controls) and
25 continuous traits (N=149,289 to 209,605). We used inverse variance weighting (IVW) for our
26 main analyses and conducted several sensitivity analyses to explore violation of MR
27 assumptions. We compared two-sample MR to one-sample MR results for alcohol effects on
28 T2D and HbA1c in UK Biobank (N=336,984). Only these analyses were conducted to avoid
29 sample overlap and due to data availability. The main IVW two-sample MR results suggested
30 possible causal effects of higher LSI on T2D risk (OR per 1SD higher LSI=1.42, 95% CI=1.22 to
31 1.64); however, sensitivity analyses did not consistently support this finding, and there was
32 evidence of potential horizontal pleiotropy. There was no robust evidence that higher drinks
33 per week influenced risk of T2D from our main IVW two-sample MR analyses (OR per 1 SD
34 higher log-transformed drinks per week=1.04, 95% CI=0.40 to 2.65), despite evidence of
35 causal effects on higher fasting glucose (difference in mean fasting glucose in mmol/l per
36 1SD higher log-transformed drinks per week=0.34, 95% CI=0.09 to 0.59). One-sample MR

37 results suggested a possible causal effect of higher drinks per week on T2D risk (OR per 1 SD
38 higher log-transformed drinks per week=1.71, 95% CI: 1.24 to 2.36), but in contrast, lower
39 HbA1c levels (difference in mean SD of log transformed HbA1c (mol/mol) per 1 SD higher
40 log-transformed drinks per week=-0.07, 95% CI: -0.11 to -0.02). Key limitations include
41 limited generalisability of results due to analyses being conducted in European populations,
42 and potential selection bias in UK Biobank influencing results.

43 **Conclusion:** Our results suggest effective public health interventions to prevent and/or
44 reduce smoking and alcohol consumption are unlikely to reduce the prevalence of T2D.

45 **Keywords:** type 2 diabetes, alcohol, smoking, Mendelian randomisation, UK Biobank,
46 glycaemic traits.

47

48 **Introduction**

49 Type 2 diabetes (T2D) is a common chronic condition that is known to increase risk of
50 macro- and micro- vascular atherosclerotic diseases (1–4). Over the last 30 years,
51 prevalence and incidence of T2D has increased markedly (5), and the age at which it is first
52 diagnosed has decreased (6). These changes are largely thought to be driven by the global
53 obesity epidemic (7), including increased testing for T2D in people who are obese. Whilst
54 higher body mass index (BMI) is an established causal risk factor for T2D (8), other risk
55 factors have been proposed.

56 Both smoking and alcohol have been suggested as potential risk factors that may causally
57 affect T2D. Observationally, smoking is associated with higher risk, with heavier smokers
58 having the greatest risk (9–12). Additionally, former smokers seem to have a higher risk for
59 T2D compared to never smokers, with risk lowering over time since they quit (10,13). For
60 alcohol consumption, there is a long history of observational studies suggesting a J-shaped
61 association with cardiovascular diseases (14–16), with some studies finding evidence of a
62 similar pattern with T2D (17,18). This slightly higher risk among those who report no alcohol
63 consumption may be an artefact, for example due to misreporting, or because some people
64 stop drinking (or never start) for health reasons (11,19). Despite this apparently higher risk
65 at lower levels, across most of the distribution, higher alcohol consumption is associated
66 with higher risk for T2D. The associations of smoking and alcohol with T2D might be causal
67 or they might be influenced by confounding due to limited adjustment for socioeconomic
68 position and related factors or only partially accounting for these in previous studies. In this
69 study we aim to determine whether the relationships between smoking and alcohol and
70 T2D are causal.

71 Mendelian randomisation (MR) is a causal inference method, which commonly uses genetic
72 variants, typically single nucleotide polymorphisms (SNPs), as instrumental variables (IVs)
73 for the exposure of interest. MR is less prone to confounding by socioeconomic,
74 environmental and behavioural characteristics, or reverse causation than conventional
75 observational studies (20). It can be biased by violation of the core assumptions underlying
76 MR (Box 1).

Box 1. Core assumptions of Mendelian Randomisation

*1) The genetic IV is robustly associated with the exposure of interest in the relevant population (**relevance**).*

May cause biased results if there are weak instruments (i.e., a statistically weak association of the genetic instrument with the exposure, which would bias results towards the null in two-sample MR and towards the confounded association in one-sample MR. In this study, comparing results from two- and one-sample MR for some effects is useful as, if both give consistent results this gives us greater confidence that neither have weak instrument bias. Population relevance is particularly important in two-sample MR, where it is important to ensure that the underlying population is the same in both samples and consistent with the population that we want to make inference to. Therefore, in this study we restricted analyses to include results from the genetic instrument-exposure and genetic instrument-outcome associations in white European adult populations only.

*2) There is no confounding between the genetic IV and the outcome (**independence**).*

May be violated if there is population stratification, assortative mating, or confounding by dynastic factors. In this study we tried to minimise population stratification by only including results / data from participants of European ancestry and adjusting for principal components in our one-sample MR analyses.

*3) The genetic IV is only associated with the outcome via the exposure, and there are no direct effects of the genetic IV on the outcome (**exclusion restriction**).*

May be violated if there is horizontal pleiotropy (i.e. the genetic variants influence risk factors for the outcome, independently of the exposure of interest. In this study we explored the likelihood of unbalanced horizontal pleiotropy influencing our main two-sample MR IVW results through comparing those results to results from several pleiotropy robust sensitivity analyses.

77

78 Two previous two-sample MR studies have found evidence for a potential causal effect of
79 smoking initiation (21) and lifetime smoking (22) on risk of T2D. The former study used data
80 from a genome wide association study (GWAS) of diabetes with 74,124 cases and 824,006

81 controls and focused on smoking initiation, which does not capture other important
82 smoking related behaviours, like smoking duration and intensity. The latter study used data
83 from the same diabetes GWAS and was an atlas study that examined many outcomes, and
84 therefore did not focus on this relationship, nor did it consider continuous traits related to
85 diabetes.

86 We identified two previous MR studies examining alcohol and T2D. The first reported no
87 effect. It used one SNP from the alcohol dehydrogenase 1B gene which encodes an enzyme
88 involved in alcohol metabolism (i.e., it directly influences the amount of alcohol consumed
89 in those who have ever drunk alcohol), in a one-sample MR approach within 261,991 adults
90 of European ancestry (with 14,549 cases) (23). This is a useful approach as there is unlikely
91 to be bias due to horizontal pleiotropy. Therefore, further exploration, using two-sample
92 MR and exploring underlying continuous traits may strengthen the conclusion of no effect, if
93 results were consistent. The second study also used a one-sample MR approach and a
94 genetic risk score for alcohol within UK Biobank (UKBB) (N=408,540 with 33,656 cases) and
95 found a potential causal effect of higher alcohol intake on T2D risk, with the strongest
96 effects found in heavier drinkers in analyses stratified by alcohol intake (24). Our study
97 extends this work by including continuous traits related to diabetes and examining this
98 within a two-sample MR approach as well.

99 The aim of this study was to use MR to explore the effects of lifetime smoking and alcohol
100 consumption (drinks per week) on T2D risk and underlying glycaemic traits. This adds to
101 previous studies by exploring both T2D and related continuous traits, exploring effects of
102 both smoking and alcohol behaviours in the same study, undertaking more sensitivity
103 analyses to test genetic instrument validity, and the influence of unbalanced horizontal

104 pleiotropy on our results, and comparing results from our main two-sample MR with those
105 from one-sample MR, where this was possible, and data was available. In this study we used
106 lifetime smoking as one exposure as this can be applied to non-smokers (unlike smoking
107 heaviness) too and allows for a richer phenotype incorporating a range of smoking
108 behaviours. We used alcohol consumption as our other exposure (i.e., average number of
109 drinks participants reported consuming each week across all types of alcohol) to capture
110 drinking over the whole distribution.
111

112 **Methods**

113 **Exposure GWAS and selection of genetic instruments**

114 We used the largest GWAS of lifetime smoking index (LSI) (25) and alcohol consumption
115 (drinks per week) (26), avoiding sample overlap with the outcome GWAS, which can bias
116 results. Both GWAS only included participants of European ancestry and with complete
117 genotype and phenotypic data (for relevant smoking and alcohol phenotypes), resulting in
118 462,690 participants from UKBB in the LSI GWAS and 941,280 participants from GSCAN
119 (GWAS and Sequencing Consortium of Alcohol and Nicotine) in the drinks per week GWAS.
120 GWAS adjusted for principal components to further control for population substructure.
121 SNPs (and associations with the relevant exposures) were selected if they reached genome-
122 wide statistical significance ($p \leq 5 \times 10^{-08}$) and were independent (i.e., we excluded SNPs in
123 linkage disequilibrium; r^2 of 0.001; window of 10,000 kb; European 1000 genomes reference
124 panel). For any palindromic SNPs we tried to infer the positive strand based on allele
125 frequencies, but if this was not possible, these SNPs were excluded. Where a SNP was
126 available for the exposure and not the outcome, we attempted to identify proxy SNPs using
127 LDlink (27) and an LD r^2 threshold of >0.8 . After exclusions and identifying any proxies, we
128 searched for the remaining LSI SNPs in the outcome GWAS (118 SNPs for all outcomes) and
129 the remaining alcohol consumption SNPs in the outcome GWAS (70 SNPs for all outcomes).
130 Details of the exposure GWAS, including derivation of the LSI and drinks per week of alcohol
131 are provided in Supplementary Materials Section 1. To summarise, the LSI reflects a
132 combination of smoking related behaviours including smoking status, duration and
133 heaviness, where never smokers have a score of 0. Drinks per week reflects the average
134 number of drinks/glasses consumed per week by participants.

135 LSI is in standard deviation (SD) units, therefore, in our MR analyses we explore effects per 1
136 SD higher LSI. To give context, 1 SD higher LSI value is equivalent to an individual smoking 20
137 cigarettes per day for 15 years and stopping 17 years ago or smoking 60 cigarettes a day for
138 13 years and stopping 22 years ago. Natural log-transformed drinks per week were used in
139 the GWAS, therefore, in our MR analyses we explore effects per 1 SD higher log-
140 transformed drinks per week. To give context, in UKBB (the sample used in our one-sample
141 MR) 1 SD was equal to 2.14 of the log transformed drinks per week.

142 **Outcome GWAS and harmonisation of exposure SNPs**

143 We obtained associations of the exposure SNPs with outcomes from the largest GWAS of
144 T2D (28), fasting glucose (29), fasting insulin (29) and glycated haemoglobin (HbA1c) (29).
145 We only used GWAS data including participants of European ancestry, resulting in 148,726
146 cases and 965,732 controls from the Million Veteran Program, DIAMANTE and Biobank
147 Japan for T2D. The continuous traits all used data from the Meta-Analyses of Glucose and
148 Insulin-related traits Consortium (MAGIC), with 209,605 participants with data for fasting
149 glucose, 158,550 with data for fasting insulin and 149,289 with data for HbA1c. GWAS
150 summary statistics for the exposure and outcome were harmonised so that the SNP allele-
151 exposure and SNP allele-outcome associations were in the same direction. Details of these
152 GWAS can be found in Supplementary Materials Section 2. To summarise, the T2D GWAS
153 included cases with a diagnosis of T2D and controls without, fasting glucose was measured
154 in mmol/l, fasting insulin in pmol/l in serum and HbA1c as a percentage. Therefore, in our
155 MR analyses results are reported as the odds of T2D and the difference in mean fasting
156 glucose (mmol/l), fasting insulin (pmol/l) and HbA1c (NGSP percent or equivalent) per 1 SD
157 higher LSI or log-transformed drinks per week.

158 **UK Biobank data for one-sample Mendelian randomisation**

159 We used data from the UKBB, a large population-based prospective health research
160 resource of 503,317 participants (5.5% response of those invited), recruited between 2006
161 and 2010, aged between 38 and 73 years and from the UK (30). Further details are included
162 in the Supplementary Materials (Section 3 and on the study website
163 (www.ukbiobank.ac.uk)). We were only able to assess the effect of drinks per week on T2D
164 and HbA1c because the LSI genetic instruments were obtained from UKBB and the glucose
165 and insulin measures in UKBB were from non-fasting samples.

166 *Drinks per week*

167 The drinks per week phenotype was constructed from responses to questions on the
168 average weekly intake of a range of different alcoholic beverages (defined as number of
169 glasses they had). Where this information was not available, weekly consumption was
170 estimated from measures of average monthly intake (see Supplementary Materials, Section
171 4 for further details). Data were natural log-transformed due to being right skewed and
172 standardised (1 SD was equal to 2.14 of the log transformed drinks per week).

173 *Type 2 diabetes*

174 We derived possible or probable T2D using the Eastwood algorithm (31) (see Supplementary
175 Materials Section 5). In one-sample MR analyses we excluded individuals who had possible
176 or probable type 1 diabetes as per the Eastwood algorithm.

177 *HbA1c*

178 Serum HbA1c (mol/mol) was assayed using five Bio-Rad Variant II Turbo analysers, values
179 outside of the reportable range of 15 to 184 mmol/mol, or invalidated for any other reason,
180 were excluded (further information can be found at
181 https://biobank.ctsu.ox.ac.uk/crystal/ukb/docs/serum_hb1ac.pdf). These analysers used
182 high performance liquid chromatography (HPLC) to determine the relative concentration of
183 HbA1c in packed red blood cells, from blood samples (approximately 9ml) collected at
184 recruitment. Many studies examining continuous traits related to T2D exclude participants
185 with a diabetes diagnosis or above thresholds indicative of diabetes. This means that results
186 are not necessarily applicable to the whole population from which the study sample is
187 drawn and can result in selection bias (32). On the other hand, people with a diagnosis of
188 diabetes will have made changes to their lifestyles and/or be on medications that impact
189 these continuous traits and associations with them. Therefore, we conducted analyses with
190 and without excluding those with possible or probable type 1 or type 2 diabetes using the
191 Eastwood algorithm and those who had a HbA1c measure of $\geq 6.5\%$ (or 48 mmol/mol) at
192 baseline. Data were natural log-transformed due to being right skewed and standardised (1
193 SD was equal to 0.15 log mmol/mol). As UKBB did not collect fasting samples we have not
194 conducted one-sample MR on fasting glucose and insulin.

195 *Genetic data*

196 A total of 488,377 participants had genotyped samples. Pre-imputation quality control,
197 phasing and imputation are described elsewhere (33) and summarised in the
198 Supplementary Materials (Section 6).

199 **Statistical analysis**

200 We pre-registered the analysis plan for this study on the Open Science Framework in March
201 2021 (<https://osf.io/ygucn>). All analyses were conducted in R (34) (version 3.6.2).

202 *Two-sample Mendelian randomisation analyses*

203 We conducted two-sample MR analyses using the TwoSampleMR package in R (35).

204 We used the inverse-variance weighted (IVW) method as our main analysis (36). This fits a
205 linear regression model of the mean SNP-outcome value on mean SNP-exposure value
206 across all SNPs and constrains the intercept of the regression slope to be zero, with the
207 slope providing an unbiased effect estimate under the assumption that there is no
208 horizontal pleiotropy (37). Sensitivity analyses used to explore this assumption were done
209 using MR-Egger (38), weighted median (39), MR Pleiotropy RESidual Sum and Outlier (MR-
210 PRESSO) (40) and Generalised Summary-data-based MR (GSMR) (41) methods.

211 MR Egger is identical to IVW with the exception that the intercept reflects the best fitted
212 regression model and is not constrained to zero (38). The slope provides a causal estimate
213 controlling for potential unbalanced horizontal pleiotropy. As with IVW, this approach is also
214 subject to the Instrument Strength Independent of Direct Effect (InSIDE) assumption, which
215 may be violated if any pleiotropic effects are all via a single factor that is correlated with the
216 association of the genetic instrument with the exposure (i.e., instrument strength). A non-
217 null intercept from MR Egger is indicative of unbalanced horizontal pleiotropy and we used
218 the p-value for the intercept to assess this.

219 The weighted median provides an unbiased causal estimate if no more than 50% of the
220 weight of the SNPs used in the genetic instrument are influenced by horizontal pleiotropy

221 (i.e. results may be biased if one single SNP or several SNPs cumulatively contribute 50% of
222 the weight and are horizontally pleiotropic) (39).

223 MR-PRESSO is used to detect and correct for potential horizontal pleiotropic outliers in the
224 instrument (40). It comprises of three stages to test this. An initial global test assesses
225 whether the total residual sum of squares (RSS) is similar to that expected by chance. Then
226 any outliers are identified by examining the RSS of each SNP. Finally, the extent to which
227 these outliers effect the causal estimate is evaluated by using the distortion test. From this
228 analysis we get an uncorrected and an outlier-corrected causal estimate and we use results
229 from the global and distortion tests to detect horizontal pleiotropy and test for distortion
230 between the estimate before and after correction, respectively. For MR-PRESSO the
231 precision of the p-values is determined by the number of elements to simulate, specified in
232 the model. We used a value of 2,000 for all analyses, except for lifetime smoking on fasting
233 glucose in both main and additional analyses and fasting insulin in the main analyses, where
234 we used 3,000 due to the model being unable to estimate the p-value with the lower value.

235 The GSMR approach allows estimation of a causal effect including SNPs in the instrument
236 that are correlated, by estimating the LD between SNPs from a reference sample (41). The
237 GSMR model also removes outliers which may be associated with confounding factors by
238 assessing heterogeneity across SNPs, using the heterogeneity in dependent instrument
239 (HEIDI) test, and models the SNP-exposure estimate error which other MR methods do not
240 include.

241 In addition, we explored between SNP heterogeneity, which might be an indicator of
242 horizontal pleiotropy or violation of other assumptions, using Cochran's Q, where a p-value
243 <0.05 may indicate the presence of between SNP heterogeneity. We also assessed

244 heterogeneity between SNPs, whilst adjusting for any horizontal pleiotropy for the MR-
245 Egger method, using the Rucker's Q-test, again with a p-value threshold of <0.05.

246 The IVW and MR-Egger methods assume that there is no measurement error in the SNP-
247 exposure estimates (39), known as the 'NO Measurement Error' (NOME) assumption. The
248 extent of the NOME assumption violation can be quantified using regression dilution I-
249 squared statistics, where a lower value indicates greater violation. An I-squared of less than
250 0.9, indicates that MR-Egger estimates should be interpreted with caution due to regression
251 dilution and where this is the case, we have conducted simulation extrapolation (SIMEX)
252 corrections as a sensitivity analysis. The SIMEX approach is a bias adjustment method which
253 provides an estimate for the case where NOME had been satisfied. We also estimated the
254 mean F-statistic for each analysis, which indicates instrument strength, where a value under
255 10 may indicate a weak instrument (39).

256 Overall, by using these different methods, which make different assumptions, we were able
257 to assess the robustness of evidence for causal effects against violations of the MR
258 assumptions. We were interested in whether there was evidence of causal effects.
259 However, we have previously shown that causal effect estimates when using exposures
260 related to cigarette smoking may be unreliable (42). Therefore, we considered consistency
261 of evidence (e.g., direction of effect estimate, p-value as a measure of strength of evidence
262 against the null) across analyses to guide our inference regarding whether or not a causal
263 effect may be operating, but did not attempt to directly estimate the magnitude of any such
264 effect (43).

265 Finally, the GWAS used for the outcomes of fasting glucose and fasting insulin adjust for
266 body mass index (BMI), which can bias our results. Multivariable MR (MVMR) analyses

267 including BMI can help overcome this issue and provide unbiased estimates of the exposure
268 of interest (LSI and drinks per week) on the outcome (44) (see Supplementary Materials
269 Section 7 for further details).

270 *One-sample Mendelian randomisation analyses*

271 One-sample MR analyses were conducted using the OneSampleMR and Applied
272 Econometrics with R (AER) packages, respectively, in R (45,46). We generated weighted
273 allele genetic risk scores in UKBB for alcohol consumption using the per-allele regression
274 coefficients from each independent genome-wide significant SNP for each exposure as
275 weights and then summing those weighted values (see Supplementary Materials Section 8).
276 We used two-stage least squares regression with adjustment for age, sex, the first 10
277 principal components (PCs) (derived from PC analysis of UKBB genotype data, imputed to a
278 reference set combining UK10K haplotype and Haplotype Reference Consortium [HRC]
279 reference panels), assessment centre and genotyping chip. Two genotyping chips were, the
280 UKBB axiom array (which 90% of participants were genotyped with) and the UK BiLEVE
281 array. The latter was used for those in the UK BiLEVE study (47), which was oversampled for
282 smokers, and therefore adjusting for genotyping chip may introduce collider bias. Therefore,
283 we performed analyses with and without adjustment for chip.

284 **Ethics**

285 All studies that contributed to the exposure and outcome GWAS used in MR analyses had
286 ethics approval and participant consent for their data to be used in genetic analyses. UKBB
287 (data used in one-sample MR analyses) received ethics approval from the UK National
288 Health Service Research Ethics Committee (11/NW/0382).

289 **Data availability**

290 Access details for the GWAS data used in this study are outlined in Supplementary Table S1.

291 UK Biobank data are available through a procedure described at

292 <http://www.ukbiobank.ac.uk/using-the-resource/>.

293 Analysis code is available from the University of Bristol's Research Data Repository

294 (<http://data.bris.ac.uk/data/>), at: **To be added on publication.**

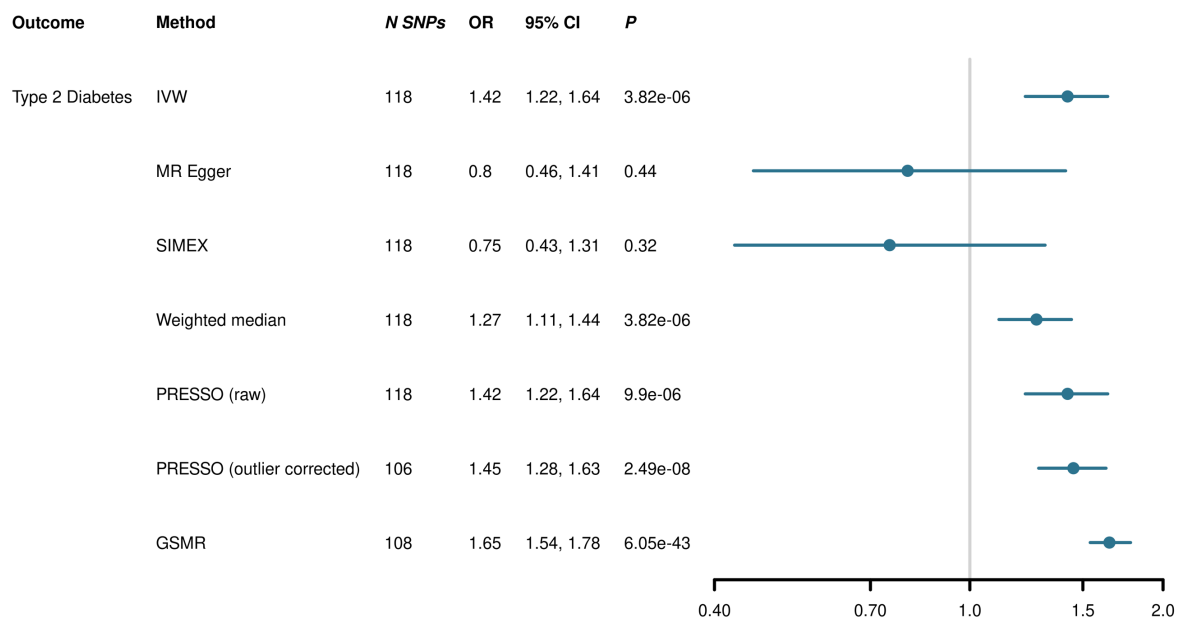
295

296 **Results**

297 **Two-sample Mendelian randomisation with lifetime smoking as the exposure**

298 Mean F-statistics for LSI were all 44.27 (Supplementary Table S2). Figure 1, Figure 2 and
299 Supplementary Table S3 provide the results from the main IVW and all sensitivity analyses
300 of the effects of LSI on outcomes. For T2D, the main IVW result suggested a causal effect of
301 higher LSI on T2D risk (OR per 1SD higher LSI=1.42, 95% CI=1.22 to 1.64). Results from
302 weighted median, PRESSO (with and without outlier correction) and the GSMR sensitivity
303 analyses were consistent with this. By contrast MR-Egger and SIMEX adjusted MR-Egger
304 results were in the opposite direction, though with wide confidence intervals (OR per 1SD
305 higher LSI=0.80, 95% CI=0.46 to 1.41 for MR Egger and OR=0.75, 95% CI=0.43 to 1.31 for
306 SIMEX adjusted MR-Egger). There was also evidence of between SNP heterogeneity
307 (Cochran's Q p-value = 5.42×10^{-66} ; Rucker's Q p-value = 1.10×10^{-62}) and potential bias due to
308 unbalanced horizontal pleiotropy based on the MR-Egger and SIMEX adjusted MR Egger
309 intercepts ($p=0.04$ and $p=0.02$, respectively) and the MR-PRESSO global test ($p<0.0003$).

310 Figure 1. Two-sample Mendelian randomisation results of the potential causal effect of
311 lifetime smoking on type 2 diabetes



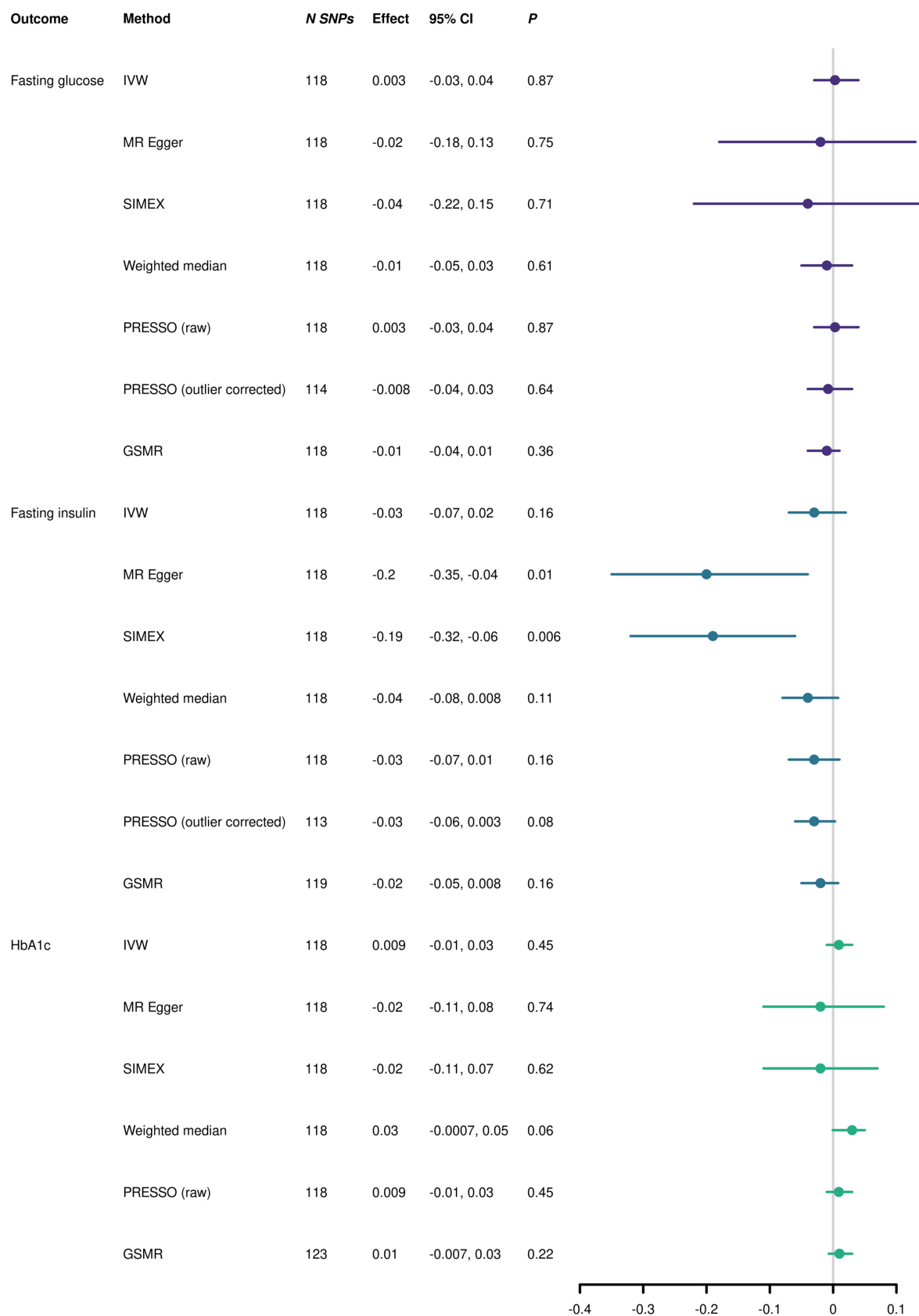
312

313 *Results are the odds ratios (OR) of type 2 diabetes per 1 SD higher lifetime smoking index score, with 95%*
 314 *confidence intervals (CI), noting that 1 SD higher LSI value is equivalent to an individual smoking 20*
 315 *cigarettes per day for 15 years and stopping 17 years ago or smoking 60 cigarettes a day for 13 years and*
 316 *stopping 22 years ago. SNP=single nucleotide polymorphism, IVW=inverse-variance weighted,*
 317 *SIMEX=simulation extrapolation, PRESSO= Pleiotropy RESidual Sum and Outlier, GSMR=Generalised*
 318 *Summary-data-based Mendelian Randomisation*

319 Our main IVW analyses also suggested that LSI was not causally related to fasting glucose
 320 (difference in mean fasting glucose in mmol/l per 1SD higher LSI=0.003, 95% CI=-0.03 to
 321 0.04), fasting insulin (difference in mean fasting insulin in pmol/l per 1SD higher LSI=-0.03,
 322 95% CI=-0.07 to 0.02) or HbA1c (difference in mean HbA1c in NGSP percent or equivalent
 323 per 1SD higher LSI=0.009, 95% CI=-0.01 to 0.03). Sensitivity analyses were mostly consistent
 324 with these results, with the exception of MR-Egger and SIMEX adjusted MR-Egger, where
 325 imprecise estimates suggested that higher LSI might reduce fasting insulin (Figure 2).
 326 Additional MVMR analyses accounting for BMI for fasting glucose and fasting insulin were in
 327 line with the main IVW results (Supplementary Table S4).

328

329 Figure 2. Two-sample Mendelian randomisation results of the potential causal effect of
 330 lifetime smoking on glycaemic traits



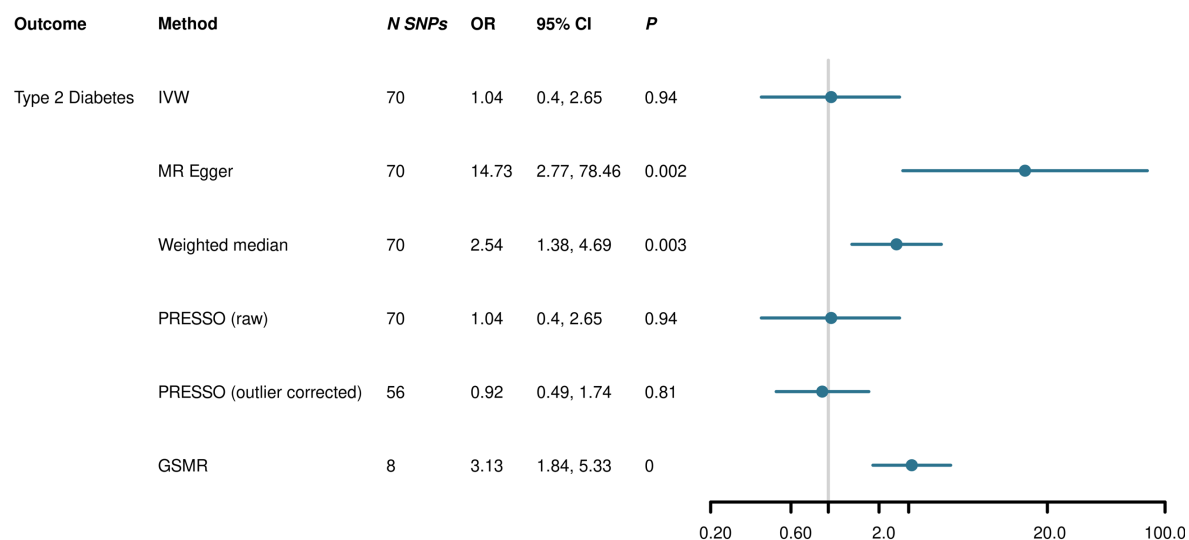
331

332 *Results are the difference in mean fasting glucose (mmol/l), fasting insulin (pmol/l) and HbA1c (NGSP*
333 *percent or equivalent) per 1 SD higher lifetime smoking index value, with 95% confidence intervals (CI),*
334 *noting that 1 SD higher LSI value is equivalent to an individual smoking 20 cigarettes per day for 15 years*
335 *and stopping 17 years ago or smoking 60 cigarettes a day for 13 years and stopping 22 years ago.*
336 *SNP=single nucleotide polymorphism, IVW=inverse-variance weighted, SIMEX=simulation extrapolation,*
337 *PRESSO= Pleiotropy RESidual Sum and Outlier, GSMR=Generalised Summary-data-based Mendelian*
338 *Randomisation*

339 **Two-sample Mendelian randomisation with drinks per week as the exposure**

340 Mean F-statistics for alcohol consumption were all 31.55 (Supplementary Table S2). Figure
341 3, Figure 4 and Supplementary Table S3 provide the results from the main IVW and all
342 sensitivity analyses of the effects of alcohol consumption on outcomes. For T2D, the main
343 IVW results suggested that alcohol consumption was not causally related to T2D (OR per 1
344 SD higher log-transformed drinks per week=1.04, 95% CI=0.40 to 2.65). Some of the
345 sensitivity analyses were consistent with these results, however, MR-Egger, weighted
346 median and GSMR estimates suggested a potential causal effect of higher drinks per week
347 on T2D risk, but confidence intervals were wide (Figure 3). There was also evidence of
348 between SNP heterogeneity (Cochran's Q p-value = 2.86×10^{-89} ; Rucker's Q p-value = 2.00×10^{-70})
349 and potential bias due to unbalanced horizontal pleiotropy based on the MR Egger
350 intercept (p=0.0005) and the MR-PRESSO global test (p< 5×10^{-04}).

351 Figure 3. Two-sample Mendelian randomisation results of the potential causal effect of
352 drinks per week on type 2 diabetes



353

354 *Results are the odds ratios (OR) of type 2 diabetes per 1 SD higher log-transformed drinks per week, with*

355 *95% confidence intervals (CI). SNP=single nucleotide polymorphism, IVW=inverse-variance weighted,*

356 *SIMEX=simulation extrapolation, PRESSO= Pleiotropy RESidual Sum and Outlier, GSMR=Generalised*

357 *Summary-data-based Mendelian Randomisation*

358 Our main IVW analyses suggested a potential casual effect of higher drinks per week on

359 higher fasting glucose (difference in mean fasting glucose in mmol/l per 1SD higher log-

360 transformed drinks per week=0.34, 95% CI=0.09 to 0.59). This was consistent across all

361 sensitivity analyses. However, there was also evidence of between SNP heterogeneity

362 (Cochran's Q p-value = 5.18×10^{-40} ; Rucker's Q p-value = 7.13×10^{-40}) and potential bias due to

363 unbalanced horizontal pleiotropy based on the MR-PRESSO global test ($p < 5 \times 10^{-04}$), but not

364 the MR-Egger intercept ($p=0.46$). Our main IVW analyses suggested that drinks per week

365 was not causally related to fasting insulin (difference in mean fasting insulin in pmol/l per

366 1SD higher LSI=0.08, 95% CI=-0.16 to 0.32) or HbA1c (difference in mean HbA1c in NGSP

367 percent or equivalent per 1SD higher LSI=-0.05, 95% CI=-0.17 to 0.07). Sensitivity analyses

368 were mostly consistent with these results, with the exception of the GSMR estimate

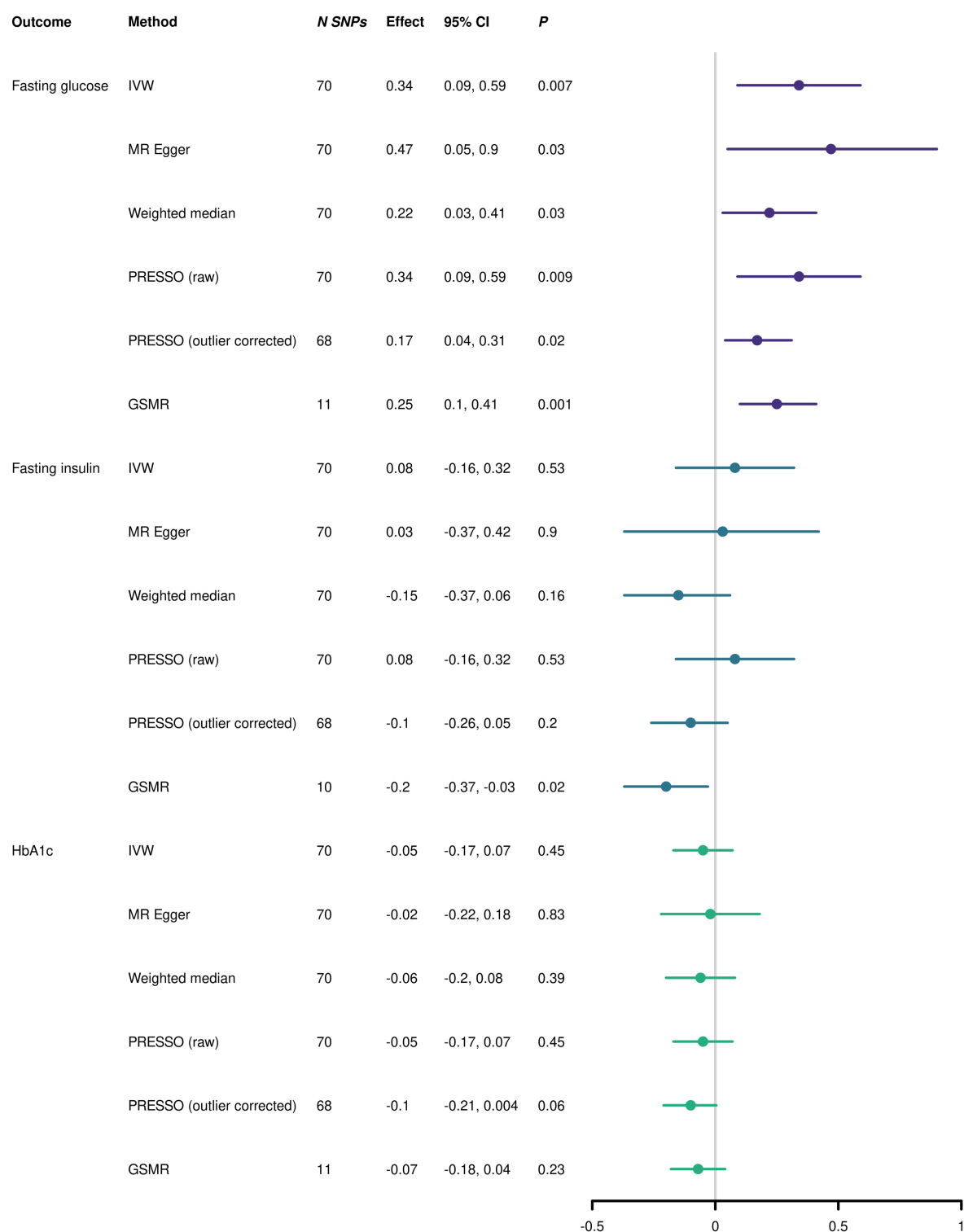
369 suggesting that higher drinks per week might reduce fasting insulin (Figure 4). The effect

370 from the additional MVMR analyses accounting for BMI for fasting glucose was attenuated

371 and did not provide evidence of a potentially causal effect. For fasting insulin, the MVMR

372 analysis results were in line with the main IVW results (Supplementary Table S4).

373 Figure 4. Two-sample Mendelian randomisation results of the potential causal effect of
 374 drinks per week on glycaemic traits



375
 376 Results are the difference in mean fasting glucose (mmol/l), fasting insulin (pmol/l) and HbA1c (NGSP
 377 percent or equivalent) per 1 SD higher log-transformed drinks per week, with 95% confidence intervals

378 (CI). *SNP*=single nucleotide polymorphism, *IVW*=inverse-variance weighted, *SIMEX*=simulation
 379 extrapolation, *PRESSO*= Pleiotropy RESidual Sum and Outlier, *GSMR*=Generalised Summary-data-based
 380 Mendelian Randomisation
 381 One-sample Mendelian randomisation for drinks per week on type 2 diabetes and HbA1c
 382 Sample characteristics for those UKBB participants included in the one-sample MR analyses
 383 are shown in Supplementary Table S5.
 384 Table 1 provides the results from the one-sample MR analyses. Results suggested a causal
 385 effect of higher drinks per week on T2D risk (OR per 1 SD higher log-transformed drinks per
 386 week=1.71, 95% CI: 1.24 to 2.36) and on lower HbA1c levels (but only when we excluded
 387 participants with a possible or probable diabetes diagnosis or HbA1c \geq 6.5%) (difference in
 388 mean SD of log transformed HbA1c (mol/mol) per 1 SD higher log-transformed drinks per
 389 week=-0.07, 95% CI: -0.11 to -0.02). Adjusting for chip did not impact our results.
 390 Table 1. One-sample Mendelian randomisation results

| | | Main analysis | | | Adjusting for chip ¹ | |
|--------------|--|---------------|------------------------|------------------------|---------------------------------|------------------------|
| | | N | OR or effect (95% CI) | p-value | OR or effect (95% CI) | p-value |
| T2D | Excluding participants with type 1 diabetes ² | 266,005 | 1.71 (1.24 to 2.36) | 1.30x10 ⁻⁰⁹ | 1.70 (1.23 to 2.35) | 1.30x10 ⁻⁰⁹ |
| HbA1c | Including participants with diabetes ² | 253,995 | -0.02 (-0.08 to 0.03) | 0.43 | -0.02 (-0.08 to 0.03) | 0.40 |
| HbA1c | Excluding participants with diabetes ² | 242,412 | -0.07 (-0.11 to -0.02) | 0.002 | -0.07 (-0.11 to -0.02) | 0.002 |

391 Results are the odds ratio (OR) for type 2 diabetes per 1 SD higher log-transformed drinks per week or the
 392 difference in mean SD of log transformed HbA1c (mol/mol) per 1 SD higher log-transformed drinks per

393 week. ¹We ran additional sensitivity analyses adjusting for genotyping chip which is confounded with
394 being in the UK Biobank Lung Exome Variant Evaluation (BiLEVE) ²Exclusion criteria for diabetes was
395 based on having possible or probable type 1 or type 2 diabetes as determined by the Eastwood algorithm
396 or those with HbA1c $\geq 6.5\%$.

397 Discussion

398 We found evidence of a possible causal effect of higher LSI values on risk of T2D in our main
399 IVW analysis, but this was not consistent across sensitivity analyses and may be biased due
400 to unbalanced horizontal pleiotropy. Furthermore, the lack of evidence of an effect on the
401 underlying glycaemic traits suggests that we cannot confidently interpret this as a causal
402 effect on T2D risk. We found little evidence of a possible causal effect of higher drinks per
403 week on T2D risk, with no effect found for our main IVW analyses in the two-sample MR
404 and inconsistent evidence in our sensitivity analyses. However, the lack of a consistent
405 finding means that we would not interpret this as being evidence of a causal effect. We did
406 observe a possible causal effect of drinks per week on fasting glucose, which was consistent
407 across sensitivity analyses, but this was not supported by analyses with the other underlying
408 glycaemic traits and when accounting for BMI this effect was no longer observed.
409 Additionally, results from our one-sample MR analyses (only examining drinks per week on
410 T2D and HbA1c) suggest that there may be a causal effect of more drinks per week on T2D
411 risk and lower HbA1c levels. These mixed results do not provide strong evidence of causal
412 effects but do suggest that LSI and drinks per week may have some effect on some of these
413 outcomes; however, the explanations for these effects may be complicated, for example,
414 acting through pleiotropic pathways.

415 Previous studies have largely shown associations using multivariable regression, and
416 potential causal effects using MR, between smoking and T2D risk (10,12,21,22,48) and
417 between higher alcohol consumption and T2D risk (17,18,23,24). However, we note that a
418 previous MR study using a functional variant for alcohol metabolism did not find an effect
419 on T2D (23), in line with our results, suggesting it is unlikely there is an effect of alcohol on

420 T2D. In addition, our finding with decreased HbA1c in the one-sample MR analyses is in the
421 opposite direction to what we might expect given that higher levels of HbA1c are observed
422 in those with T2D. However, this may be due to the exclusion of individuals with a diagnosis
423 of diabetes, and we do not find an effect in our two-sample MR, so results should be
424 interpreted with caution. Our results suggest that future studies examining risk factors for
425 T2D should triangulate results across different analyses and consider underlying glycaemic
426 traits as outcomes in order to arrive at the correct conclusions.

427 *Limitations*

428 The key strengths of this study and how this study adds to previous literature is described in
429 the introduction and above. In terms of limitations, we were unable to explore a non-linear
430 effect as recent evidence suggests that current methods for doing so in MR are potentially
431 biased (49). Some observational studies have suggested a J-shaped association between
432 alcohol and coronary heart disease, which T2D is a risk factor for (14–16,19). However, that
433 J-shaped association, even if causal, suggests a linear association across most of the
434 distribution. Our one-sample MR analyses in UKBB may be subject to selection bias (50,51).
435 Future studies replicating these analyses in other samples would be useful to examine
436 whether selection bias in UKBB may be an issue here, however our two-sample MR analyses
437 in part overcome this for alcohol but not lifetime smoking where the GWAS was conducted
438 in UK Biobank as well. Our analyses were conducted in samples of European ancestry, due
439 to the data available, so results are not generalisable beyond this group. Finally, it is possible
440 that measurement error in the exposure or outcome could bias our results. This is more
441 likely to be the case for the exposures, which could be subjectively influenced, for example
442 misreporting of cigarettes smoked per day, time since cessation or duration of smoking and

443 number of drinks consumed per week. In particular, in UKBB drinks per week is measured
444 based on number of glasses of alcohol consumed and this is similar to the definitions used in
445 the GWAS. This doesn't account for units and glass/drink size can vary, so there is likely
446 variation in how this was reported.

447 *Conclusion*

448 In summary, we found limited evidence of a possible causal effect of higher lifetime smoking
449 index score and drinks per week score on T2D risk. We found further evidence of a possible
450 causal effect of higher drinks per week on higher fasting glucose. However, overall results
451 were not consistent across analyses and some results may be biased by horizontal
452 pleiotropy. Therefore, we do not find strong evidence of smoking and alcohol influencing
453 risk of T2D. Future research should include triangulation approaches and glycaemic traits to
454 allow for a more in depth understanding of the causal influence of risk factors on T2D.
455

456 **Acknowledgements:**

457 This research has been conducted using data from UKBB (project ID: 9142), a major
458 biomedical database (<http://www.ukbiobank.ac.uk/>). We would like to thank the research
459 participants and employees of 23andMe, inc. for making this work possible. For the purpose
460 of open access, the author(s) has applied a Creative Commons Attribution (CC BY) licence to
461 any Author Accepted Manuscript version arising from this submission. The type 2 diabetes
462 GWAS included data from the Million Veteran Program (MVP), Office of Research and
463 Development, Veterans Health Administration, and was supported by the Veterans
464 Administration (VA). The authors thank MVP staff, researchers, and volunteers, who have
465 contributed to MVP, and especially participants who previously served their country in the
466 military and now generously agreed to enroll in the study. (See
467 <https://www.research.va.gov/mvp/> for more details). The MVP GWAS data was provided
468 through dbGaP under accession number phs001672.

469 *Funding:* This work was supported in part by the UK Medical Research Council Integrative
470 Epidemiology Unit at the University of Bristol (Grant ref: MC_UU_00032/05 and
471 MC_UU_00032/07). HMS was supported by the European Research Council (Grant ref:
472 758813 MHINT). RCR was supported by Cancer Research UK (grant
473 number C18281/A29019). DAL's contribution is supported by the British Heart Foundation
474 (CH/F/20/90003 and AA/18/1/34219). MRM was supported by the National Institute for
475 Health Research Bristol Biomedical Research Centre. The views expressed in this publication
476 are those of the author(s) and not necessarily those of the NHS, the National Institute for
477 Health Research or the Department of Health and Social Care.

478 **Competing interests:** No competing interests.

479 **Author contributions:** Conceptualization: ASA, MRM, DAL; Methodology: ZER, HMS, MRM,
480 DAL; Data Curation: ZER, RCR; Formal Analysis: ZER; Investigation: ZER; Resources: ZER,
481 MRM; Writing—Original Draft: ZER; Writing—Review and Editing: ZER, HMS, RCR, ASA, DAL,
482 MRM; Supervision: DAL, MRM; Project Administration: ZER; Funding Acquisition: DAL, MRM.

483 **References**

- 484 1. Zoungas S, Arima H, Gerstein HC, Holman RR, Woodward M, Reaven P, et al. Effects
485 of intensive glucose control on microvascular outcomes in patients with type 2
486 diabetes: a meta-analysis of individual participant data from randomised controlled
487 trials. *Lancet Diabetes Endocrinol* [Internet]. 2017 Jun 1 [cited 2021 Sep 28];5(6):431–
488 7. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28365411>
- 489 2. Ahmad OS, Morris JA, Mujammami M, Forgetta V, Leong A, Li R, et al. A Mendelian
490 randomization study of the effect of type-2 diabetes on coronary heart disease. *Nat*
491 *Commun* [Internet]. 2015 Nov 28 [cited 2021 Sep 28];6(1):7060. Available from:
492 <http://www.nature.com/articles/ncomms8060>
- 493 3. Ray KK, Seshasai SRK, Wijesuriya S, Sivakumaran R, Nethercott S, Preiss D, et al. Effect
494 of intensive control of glucose on cardiovascular outcomes and death in patients with
495 diabetes mellitus: a meta-analysis of randomised controlled trials. *The Lancet*
496 [Internet]. 2009 May 23 [cited 2021 Sep 28];373(9677):1765–72. Available from:
497 <http://www.ncbi.nlm.nih.gov/pubmed/19465231>
- 498 4. Gan W, Bragg F, Walters RG, Millwood IY, Lin K, Chen Y, et al. Genetic predisposition
499 to type 2 diabetes and risk of subclinical atherosclerosis and cardiovascular diseases
500 among 160,000 Chinese adults. *Diabetes* [Internet]. 2019 [cited 2021 Sep
501 28];68(11):2155–64. Available from:
502 <http://www.ncbi.nlm.nih.gov/pubmed/31399431>

- 503 5. Lin X, Xu Y, Pan X, Xu J, Ding Y, Sun X, et al. Global, regional, and national burden and
504 trend of diabetes in 195 countries and territories: an analysis from 1990 to 2025. *Sci*
505 *Rep* [Internet]. 2020 Dec 8 [cited 2021 Sep 27];10(1):14790. Available from:
506 <https://www.nature.com/articles/s41598-020-71908-9>
- 507 6. Dabelea D, Mayer-Davis EJ, Saydah S, Imperatore G, Linder B, Divers J, et al.
508 Prevalence of type 1 and type 2 diabetes among children and adolescents from 2001
509 to 2009. *JAMA - Journal of the American Medical Association* [Internet]. 2014 May 7
510 [cited 2021 Sep 27];311(17):1778–86. Available from:
511 <http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2014.3201>
- 512 7. Apovian CM, Riffenburt KM. Perspectives on the global obesity epidemic [Internet].
513 Vol. 24, *Current Opinion in Endocrinology, Diabetes and Obesity*. 2017 [cited 2021
514 Sep 27]. p. 307–9. Available from: [http://journals.lww.com/01266029-201710000-](http://journals.lww.com/01266029-201710000-00002)
515 [00002](http://journals.lww.com/01266029-201710000-00002)
- 516 8. Xu L, Borges MC, Hemani G, Lawlor DA. The role of glycaemic and lipid risk factors in
517 mediating the effect of BMI on coronary heart disease: a two-step, two-sample
518 Mendelian randomisation study. *Diabetologia* [Internet]. 2017 Sep 9 [cited 2021 Oct
519 6];60(11):2210–20. Available from: [https://link.springer.com/article/10.1007/s00125-](https://link.springer.com/article/10.1007/s00125-017-4396-y)
520 [017-4396-y](https://link.springer.com/article/10.1007/s00125-017-4396-y)
- 521 9. Kolb H, Martin S. Environmental/lifestyle factors in the pathogenesis and prevention
522 of type 2 diabetes. *BMC Med* [Internet]. 2017 Dec 19 [cited 2020 Nov 26];15(1):131.

523 Available from: <http://bmcmmedicine.biomedcentral.com/articles/10.1186/s12916->
524 017-0901-x

525 10. Maddatu J, Anderson-Baucum E, Evans-Molina C. Smoking and the risk of type 2
526 diabetes [Internet]. Vol. 184, Translational Research. NIH Public Access; 2017 [cited
527 2020 Nov 26]. p. 101–7. Available from:
528 <http://www.ncbi.nlm.nih.gov/pubmed/28336465>

529 11. Zheng Y, Ley SH, Hu FB. Global aetiology and epidemiology of type 2 diabetes mellitus
530 and its complications [Internet]. Vol. 14, Nature Reviews Endocrinology. Nature
531 Publishing Group; 2018 [cited 2020 Nov 26]. p. 88–98. Available from:
532 <http://www.nature.com/articles/nrendo.2017.151>

533 12. Bellou V, Belbasis L, Tzoulaki I, Evangelou E. Risk factors for type 2 diabetes mellitus:
534 An exposure-wide umbrella review of meta-analyses. Nerurkar P V., editor. PLoS One
535 [Internet]. 2018 Mar 20 [cited 2020 Nov 26];13(3):e0194127. Available from:
536 <https://dx.plos.org/10.1371/journal.pone.0194127>

537 13. Pan A, Wang Y, Talaei M, Hu FB, Wu T. Relation of active, passive, and quitting
538 smoking with incident diabetes: a meta-analysis and systematic review. Lancet
539 Diabetes and Endocrinology. 2015;3(12):958–67.

540 14. Griffith C, Bogart D. Alcohol consumption: can we safely toast to our health?
541 [Internet]. Vol. 109, Missouri medicine. Missouri State Medical Association; 2012
542 [cited 2021 Jul 21]. p. 459–65. Available from:
543 <http://www.ncbi.nlm.nih.gov/pubmed/23362649>

- 544 15. O’Keefe JH, Bybee KA, Lavie CJ. Alcohol and Cardiovascular Health. The Razor-Sharp
545 Double-Edged Sword [Internet]. Vol. 50, Journal of the American College of
546 Cardiology. Elsevier; 2007 [cited 2021 Jul 21]. p. 1009–14. Available from:
547 <https://www.sciencedirect.com/science/article/pii/S0735109707020074#bib38>
- 548 16. Chokshi DA, El-Sayed AM, Stine NW. J-shaped curves and public health [Internet]. Vol.
549 314, JAMA - Journal of the American Medical Association. American Medical
550 Association; 2015 [cited 2021 Jul 21]. p. 1339–40. Available from:
551 <http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2015.9566>
- 552 17. Koppes LLJ, Dekker JM, Hendriks HFJ, Bouter LM, Heine RJ. Moderate alcohol
553 consumption lowers the risk of type 2 diabetes: A meta-analysis of prospective
554 observational studies [Internet]. Vol. 28, Diabetes Care. American Diabetes
555 Association; 2005 [cited 2021 Jul 21]. p. 719–25. Available from:
556 <http://www.ncbi.nlm.nih.gov/pubmed/15735217>
- 557 18. Baliunas DO, Taylor BJ, Irving H, Roerecke M, Patra J, Mohapatra S, et al. Alcohol as a
558 risk factor for type 2 diabetes: A systematic review and meta-analysis. Diabetes Care
559 [Internet]. 2009 Nov [cited 2021 Jul 21];32(11):2123–32. Available from:
560 <http://www.ncbi.nlm.nih.gov/pubmed/19875607>
- 561 19. Wannamethee SG, Shaper AG, Perry IJ, Alberti KGMM. Alcohol consumption and the
562 incidence of type II diabetes. In: Journal of Epidemiology and Community Health
563 [Internet]. BMJ Publishing Group Ltd; 2002 [cited 2020 Nov 26]. p. 542–8. Available
564 from: <http://www.ncbi.nlm.nih.gov/pubmed/12080164>

- 565 20. Davey Smith G, Lawlor DA, Harbord R, Timpson N, Day I, Ebrahim S. Clustered
566 environments and randomized genes: a fundamental distinction between
567 conventional and genetic epidemiology. Cardon L, editor. PLoS Med [Internet]. 2007
568 Dec 11 [cited 2021 Sep 29];4(12):1985–92. Available from:
569 <https://dx.plos.org/10.1371/journal.pmed.0040352>
- 570 21. Yuan S, Larsson SC. A causal relationship between cigarette smoking and type 2
571 diabetes mellitus: A Mendelian randomization study. Sci Rep [Internet]. 2019;9(1):1–
572 4. Available from: <http://dx.doi.org/10.1038/s41598-019-56014-9>
- 573 22. Yuan S, Larsson SC. An atlas on risk factors for type 2 diabetes: a wide-angled
574 Mendelian randomisation study. Diabetologia [Internet]. 2020 Nov 8 [cited 2020 Nov
575 26];63(11):2359–71. Available from: [http://link.springer.com/10.1007/s00125-020-](http://link.springer.com/10.1007/s00125-020-05253-x)
576 [05253-x](http://link.springer.com/10.1007/s00125-020-05253-x)
- 577 23. Holmes M V., Dale CE, Zuccolo L, Silverwood RJ, Guo Y, Ye Z, et al. Association
578 between alcohol and cardiovascular disease: Mendelian randomisation analysis based
579 on individual participant data. BMJ (Online) [Internet]. 2014;349(July):1–16. Available
580 from: <http://dx.doi.org/doi:10.1136/bmj.g4164>
- 581 24. Lu T, Nakanishi T, Yoshiji S, Butler-Laporte G, Greenwood CMT, Richards JB. Dose-
582 dependent Association of Alcohol Consumption With Obesity and Type 2 Diabetes:
583 Mendelian Randomization Analyses. J Clin Endocrinol Metab. 2023 Nov
584 17;108(12):3320–9.

- 585 25. Wootton RE, Richmond RC, Stuijzand BG, Lawn RB, Sallis HM, Taylor GMJ, et al.
586 Evidence for causal effects of lifetime smoking on risk for depression and
587 schizophrenia: a Mendelian randomisation study. *Psychol Med* [Internet]. 2019 Nov 6
588 [cited 2020 Aug 18];1–9. Available from:
589 [https://www.cambridge.org/core/product/identifier/S0033291719002678/type/jour](https://www.cambridge.org/core/product/identifier/S0033291719002678/type/journal_article)
590 [nal_article](https://www.cambridge.org/core/product/identifier/S0033291719002678/type/journal_article)
- 591 26. Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, et al. Association studies of up to
592 1.2 million individuals yield new insights into the genetic etiology of tobacco and
593 alcohol use. *Nat Genet* [Internet]. 2019 Feb 14 [cited 2020 Jul 1];51(2):237–44.
594 Available from: <http://www.nature.com/articles/s41588-018-0307-5>
- 595 27. Machiela MJ, Chanock SJ. LDlink: A web-based application for exploring population-
596 specific haplotype structure and linking correlated alleles of possible functional
597 variants. *Bioinformatics* [Internet]. 2015 Nov 1 [cited 2020 Nov 25];31(21):3555–7.
598 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26139635>
- 599 28. Vujkovic M, Keaton JM, Lynch JA, Miller DR, Zhou J, Tcheandjieu C, et al. Discovery of
600 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million
601 participants in a multi-ancestry meta-analysis. *Nature Genetics* 2020 52:7 [Internet].
602 2020 Jun 15 [cited 2022 Dec 16];52(7):680–91. Available from:
603 <https://www.nature.com/articles/s41588-020-0637-y>
- 604 29. Chen J, Spracklen CN, Marenne G, Varshney A, Corbin LJ, Luan J, et al. The trans-
605 ancestral genomic architecture of glycemic traits. *Nature Genetics* 2021 53:6

- 606 [Internet]. 2021 May 31 [cited 2022 Dec 16];53(6):840–60. Available from:
607 <https://www.nature.com/articles/s41588-021-00852-9>
- 608 30. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Biobank: An Open
609 Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of
610 Middle and Old Age. *PLoS Med*. 2015 Mar 31;12(3):e1001779.
- 611 31. Eastwood S v., Mathur R, Atkinson M, Brophy S, Sudlow C, Flaig R, et al. Algorithms
612 for the Capture and Adjudication of Prevalent and Incident Diabetes in UK Biobank.
613 *PLoS One* [Internet]. 2016 Sep 1 [cited 2022 Sep 20];11(9):e0162388. Available from:
614 <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0162388>
- 615 32. Delgado-Rodríguez M, Llorca J. Bias. *J Epidemiol Community Health* (1978).
616 2004;58(8):635–41.
- 617 33. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank
618 resource with deep phenotyping and genomic data. *Nature*. 2018 Oct
619 10;562(7726):203–9.
- 620 34. R Core Team. R: A language and environment for statistical computing. R Foundation
621 for Statistical Computing, Vienna, Austria; 2016.
- 622 35. Hemani G, Zheng J, Wade KH, Laurin C, Elsworth B, Burgess S, et al. MR-Base: a
623 platform for systematic causal inference across the phenome using billions of genetic
624 associations. *bioRxiv* [Internet]. 2016 [cited 2017 Mar 26];078972. Available from:
625 <http://biorxiv.org/content/early/2016/12/16/078972>

- 626 36. Burgess S, Butterworth A, Thompson SG. Mendelian Randomization Analysis With
627 Multiple Genetic Variants Using Summarized Data. *Genet Epidemiol* [Internet]. 2013
628 Nov [cited 2024 Apr 29];37(7):658. Available from: </pmc/articles/PMC4377079/>
- 629 37. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with
630 multiple genetic variants using summarized data. *Genet Epidemiol* [Internet]. 2013
631 Nov [cited 2020 Jul 15];37(7):658–65. Available from:
632 <http://www.ncbi.nlm.nih.gov/pubmed/24114802>
- 633 38. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid
634 instruments: Effect estimation and bias detection through Egger regression. *Int J*
635 *Epidemiol*. 2015;44(2):512–25.
- 636 39. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian
637 Randomization with Some Invalid Instruments Using a Weighted Median Estimator.
638 *Genet Epidemiol* [Internet]. 2016 May;40(4):304–14. Available from:
639 <http://www.ncbi.nlm.nih.gov/pubmed/27061298>
- 640 40. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy
641 in causal relationships inferred from Mendelian randomization between complex
642 traits and diseases. *Nat Genet*. 2018 May 23;50(5):693–8.
- 643 41. Zhu Z, Zheng Z, Zhang F, Wu Y, Trzaskowski M, Maier R, et al. Causal associations
644 between risk factors and common diseases inferred from GWAS summary data. *Nat*
645 *Commun*. 2018 Dec 15;9(1):224.

- 646 42. Taylor AE, Davies NM, Ware JJ, Vanderweele T, Smith GD, Munafò MR. Mendelian
647 randomization in health research: Using appropriate genetic variants and avoiding
648 biased estimates. *Econ Hum Biol.* 2014 Mar 1;13(1):99–106.
- 649 43. Sterne JA, Davey Smith G. Sifting the evidence-what’s wrong with significance tests?
650 *BMJ [Internet].* 2001 Jan 27 [cited 2021 Jul 28];322(7280):226–31. Available from:
651 <http://www.ncbi.nlm.nih.gov/pubmed/11159626>
- 652 44. Gilbody J, Borges MC, Smith GD, Sanderson E. Multivariable MR can mitigate bias in
653 two-sample MR using covariable-adjusted summary associations. *medRxiv [Internet].*
654 2022 Jul 19 [cited 2024 Jul 9];2022.07.19.22277803. Available from:
655 <https://www.medrxiv.org/content/10.1101/2022.07.19.22277803v1>
- 656 45. Palmer T, Spiller W, Sanderson E. <https://github.com/remlapmot/OneSampleMR>,
657 <https://remlapmot.github.io/OneSampleMR/>. . 2023. *OneSampleMR: One Sample*
658 *Mendelian Randomization and Instrumental Variable Analyses.*
- 659 46. Kleibler C, Zeileis A. *Applied Econometrics with R.* New York.: Springer-Verlag; 2008.
- 660 47. Wain L V, Shrine N, Miller S, Jackson VE, Ntalla I, Artigas MS, et al. Novel insights into
661 the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary
662 disease (UK BiLEVE): a genetic association study in UK Biobank. *Lancet Respir Med.*
663 2015 Oct 1;3(10):769–81.
- 664 48. Larsson SC, Burgess S. Appraising the causal role of smoking in multiple diseases: A
665 systematic review and meta-analysis of Mendelian randomization studies.

- 666 EBioMedicine [Internet]. 2022 Aug 1 [cited 2022 Dec 20];82. Available from:
667 <http://www.thelancet.com/article/S2352396422003358/fulltext>
- 668 49. Wade KH, Hamilton FW, Carslake D, Sattar N, Davey Smith G, Timpson NJ. Challenges
669 in undertaking nonlinear Mendelian randomization. Obesity [Internet]. 2023 Dec 1
670 [cited 2024 Jul 16];31(12):2887–90. Available from:
671 <https://onlinelibrary.wiley.com/doi/full/10.1002/oby.23927>
- 672 50. Tyrrell J, Zheng J, Beaumont R, Hinton K, Richardson TG, Wood AR, et al. Genetic
673 predictors of participation in optional components of UK Biobank. Nat Commun
674 [Internet]. 2021 Dec 9 [cited 2021 Apr 9];12(1):886. Available from:
675 <http://www.nature.com/articles/s41467-021-21073-y>
- 676 51. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of
677 Sociodemographic and Health-Related Characteristics of UK Biobank Participants with
678 Those of the General Population. Am J Epidemiol. 2017 Nov 1;186(9):1026–34.
679