Effects of apolipoprotein B on lifespan and risks of major diseases including type 2 diabetes: a mendelian randomisation analysis using outcomes in first-degree relatives

Tom G Richardson, Qin Wang, Eleanor Sanderson, Anubha Mahajan, Mark I McCarthy, Timothy M Frayling, Mika Ala-Korpela, Allan Sniderman*, George Davey Smith*, Michael V Holmes*

Summary

Background Apolipoprotein B (apoB) is emerging as the crucial lipoprotein trait for the role of lipoprotein lipids in the aetiology of coronary heart disease. In this study, we evaluated the effects of genetically predicted apoB on outcomes in first-degree relatives.

Methods Data on lipoprotein lipids and disease outcomes in first-degree relatives were obtained from the UK Biobank study. We did a univariable mendelian randomisation analysis using a weighted genetic instrument for apoB. For outcomes with which apoB was associated at a false discovery rate (FDR) of less than 5%, multivariable mendelian randomisation analyses were done, including genetic instruments for LDL cholesterol and triglycerides. Associations between apoB and self-reported outcomes in first-degree relatives were characterised for 12 diseases (including heart disease, stroke, and hypertension) and parental vital status together with age at death. Estimates were inferred causal effects per 1 SD elevated lipoprotein trait (for apoB, 1 SD=0·24 g/L). Replication of estimates for lifespan and type 2 diabetes was done using conventional two-sample mendelian randomisation with summary estimates from genome-wide association study consortia.

Findings In univariable mendelian randomisation, genetically elevated apoB in participants was identified to lead to a shorter lifespan in parents (fathers: 0·89 years of life lost per 1 SD higher apoB in offspring, 95% CI 0·63–1·16, FDR-adjusted p=4·0×10⁻¹⁰; mothers: 0·48 years of life lost per 1 SD higher apoB in offspring, 0·25–0·71, FDR-adjusted p=1·7×10⁻⁴). The effects were strengthened to around 2 years of life lost in multivariable mendelian randomisation and were replicated in conventional two-sample mendelian randomisation (odds ratio [OR] of surviving to the 90th centile of lifespan: 0·38 per 1 SD higher apoB in offspring, 95% CI 0·22–0·65). Genetically elevated apoB caused higher risks of heart disease in all first-degree relatives and a higher risk of stroke in mothers. Findings in first-degree relatives were replicated in two-sample multivariable mendelian randomisation, which identified apoB to increase (OR 2·32 per 1 SD higher apoB, 95% CI 1·49–3·61) and LDL cholesterol to decrease (0·34 per 1 SD higher LDL cholesterol, 0·21–0·54) the risk of type 2 diabetes.

Interpretation Higher apoB shortens lifespan, increases risks of heart disease and stroke, and in multivariable analyses that account for LDL cholesterol, increases risk of diabetes.

Funding British Heart Foundation, UK Medical Research Council, and UK Research and Innovation.

Introduction

Blood lipid concentrations have an important role in major vascular diseases, including coronary heart disease and stroke.[1] LDL cholesterol is the conventionally used lipid trait in observational and genetic epidemiological studies[2,3] and interventional trials.[4] However, multiple sources of evidence show that the number of atherogenic lipoprotein particles, as estimated by apolipoprotein B (apoB), is a more accurate measure of the risk that lipoprotein lipids pose than the concentration of cholesterol that they contain, as estimated by either LDL cholesterol or non-HDL cholesterol.[5–8] This has been shown in mendelian randomisation studies,[9–11] the role of

which has been amplified by the development of statistical tools, including multivariable mendelian randomisation.[12] This approach permits the evaluation of the causal effects of specific traits while holding other traits constant, avoiding the introduction of collider bias.[13]

Although the crucial role of apoB in coronary heart disease is increasingly emerging,[5] the comparative importance of apoB versus LDL cholesterol or triglycerides in other diseases remains less clear. In this study, we explore a broad repertoire of outcomes, taking advantage of events reported in first-degree relatives in UK Biobank. In studies using outcomes in first-degree relatives, the genetic instrument is constructed in

Correspondence to:
Dr Tom G Richardson, Medical
Research Council Integrative
Epidemiology Unit, University of
Bristol, Bristol BS8 2BN, UK
**tom.g.richardson@bristol.
ac.uk**

Research in context

Evidence before this study

Previous observational and mendelian randomisation studies have shown that circulating concentrations of apolipoprotein B (apoB) are of crucial importance to lipid-mediated atherogenesis, manifested as coronary heart disease. We searched PubMed on Feb 19, 2021, with terms including "apolipoprotein B" AND ("mendelian randomization" OR "mendelian randomisation") AND "longevity", which yielded no results. A broader search term of "apolipoprotein B" AND "longevity" identified 17 results. Although some studies evaluated associations of plasma concentrations of apoB or genetic variants in *APOB* with lifespan, none explored the causal role of apoB in their analyses or used an approach that quantified its causal role when considering genetically predicted LDL cholesterol or triglyceride levels.

Added value of this study

In this study, we explored the effects of genetically predicted elevations in apoB on multiple endpoints occurring in first-

degree relatives, including lifespan, and sought replication of findings using more conventional methods to exploit the statistical power from data available in large-scale genome-wide association study consortia. We identified that apoB had a deleterious effect on longevity, shortening the lifespan by a period of months to years. Furthermore, apoB caused higher risks of coronary heart disease and stroke in first-degree relatives. Finally, apoB was identified to increase the risk of type 2 diabetes using multivariable mendelian randomisation methods, in contrast to LDL cholesterol, which appeared to the lower risk of type 2 diabetes.

Implications of all the available evidence

Our findings support apoB as being the major lipoprotein entity implicated in the aetiology of coronary heart disease and stroke and identify that higher apoB decreases lifespan and increases the risk of type 2 diabetes. These findings highlight the crucial role of apoB in causing cardiometabolic disease, which collectively shortens the lifespan.

individuals who share around 50% of their DNA with those in whom outcomes occur, which should halve the effect size. Associations identified in this way can provide insights into likely underlying causal relationships, which, if consistent with more conventional mendelian randomisation approaches, bolster confidence that the associations are real.

In this study, we explored the association between an apoB polygenic instrument and self-reported outcomes in first-degree relatives in participants of the UK Biobank for 12 diseases (including heart disease, stroke, hypertension, and Alzheimer's disease) and parental vital status together with age at death. We assessed the rigour of these findings using multivariable mendelian randomisation by including genetic instruments for other major atherogenic lipoprotein lipids (LDL cholesterol and triglycerides), and comparing findings to previously published results.[10,12] Finally, we sought replication of our findings for lifespan and type 2 diabetes using findings from large-scale genome-wide association study (GWAS) consortia, given that multivariable analyses using parental data showed evidence of an effect of apoB on these outcomes in the current study.

Methods

Datasets and study design

Data on lipoprotein lipids and disease outcomes in first-degree relatives were obtained from the UK Biobank study under application #15825. UK Biobank comprises approximately 500 000 participants who were recruited between 2006 and 2010 from 22 assessment centres across the UK. Details on sample handling and assays for lipoprotein lipids in UK Biobank have been described previously,[14] as have general characteristics of the cohort, such as geographical regions and recruitment processes.[15]

Details on genotyping quality control, phasing, imputation, and association testing in UK Biobank are reported elsewhere.[16] Ethical approval for this study was obtained from the UK Biobank Ethics Advisory Committee (approval number 11/NW/0382) and informed consent was collected from all participants enrolled in UK Biobank.

We sought to replicate estimates from UK Biobank using conventional two-sample mendelian randomisation. Data for lifespan were derived from a GWAS meta-analysis of cohorts with participants dichotomised according to whether they survived to an age corresponding to the 90th centile of lifespan.[17] Data for type 2 diabetes were derived from the most recent large-scale European type 2 diabetes GWAS that included data from 32 studies and excluded participants from UK Biobank.[18]

Procedures

Each lipoprotein lipid trait was normalised to have a mean of 0 and SD of 1 using inverse rank-normalisation. We did a GWAS on all UK Biobank participants with genotype and trait data after excluding individuals of non-European descent (based on K-means clustering of K=4) and standard exclusions (ie, withdrawn consent, mismatch between genetic and reported sex, and putative sex chromosome aneuploidy). BOLT-LMM version 2.3.2 (linear mixed model) software was used for the GWAS because it is robust to potential confounding caused by relatedness or population structure.[19,20] Analyses were adjusted for age, sex, fasting status, and a binary variable denoting the genotyping chip used in individuals (the UK Biobank Axiom array or the UK BiLEVE array).

Genetic instruments for each lipid-related trait were derived by linkage disequilibrium clumping of GWAS

results (appendix 1 tables S1–3). Independent genetic variants robustly associated with traits ($p<5\times10^{-8}$) were identified using a reference panel of 503 Europeans from phase 3 (version 5) of the 1000 Genomes Project.[21] Clumping parameters of $r^2<0.001$ and a distance of 1 megabase were used for instrument identification. Univariable sets of instruments for each lipoprotein trait were derived by applying linkage disequilibrium clumping on each trait's GWAS results in turn. For multivariable instruments, we combined all GWAS results before applying linkage disequilibrium clumping to ensure independent instruments were identified among these correlated traits.

Outcomes

UK Biobank participants were asked at enrolment whether their father, mother, or siblings had any of the following diseases: heart disease, stroke, high blood pressure, chronic bronchitis or emphysema, Alzheimer's disease or dementia, diabetes, Parkinson's disease, severe depression, lung cancer, bowel cancer, prostate cancer, or breast cancer. Additionally, individuals were asked whether their mother and father were still alive, and if not, at what age they had died. These questions were only asked of participants who indicated that they were not adopted as a child and regarded blood relations only. If there was any uncertainty over responses, participants were told to respond with "do not know".

Each disease outcome for father, mother, and siblings was analysed by GWAS separately based on BOLT-LMM and the quality control pipeline described previously.[22] Total numbers of cases and controls for each of these GWAS can be found in appendix 1 (table S4). Genetic estimates for all instruments were extracted from these GWAS results for two-sample mendelian randomisation analyses. In effect, we did analyses using a two-sample design (although the sample for the exposure and outcomes originated from the same dataset). Given that genetic variants for the instrument were identified from measurements in individuals separate to those that experienced disease (ie, the single nucleotide polymorphism [SNP] estimates for exposure and outcomes were in effect derived from different people, even though lipid traits were measured and data on outcomes were obtained from the same individual; appendix 2 p 2), the potential for overfitting should theoretically be reduced. However, to address issues of potential overfitting leading to false positives, we sought to replicate key findings, as described below, using data on outcomes from non-overlapping datasets.

Statistical analysis

We calculated mean F-statistics to evaluate the instrument strengths in both the univariable and multivariable mendelian randomisation using the approximation described by Bowden and colleagues[23] for the univariable analysis and the conditional F-statistic by Sanderson and colleagues[24] for the multivariable analysis.

We investigated the effect of genetically predicted higher concentrations of apoB on each of the 38 first-degree relative outcomes (described in appendix 1 table S4) in turn using univariable mendelian randomisation. We define genetically predicted apoB as the levels of circulating apoB arising from genetic variation. Genetic effects on apoB and outcomes for each instrument were harmonised and initial analyses were done using the inverse variance weight (IVW) method. Proxy SNPs were not necessary for this analysis given that both exposure and outcome datasets were from UK Biobank. All univariable mendelian randomisation analyses were repeated for each outcome using LDL cholesterol and subsequently triglycerides as the exposure. Next, we did multivariable mendelian randomisation analyses to investigate the effects of apoB, LDL cholesterol, and triglycerides simultaneously on outcomes in first-degree relatives. This permitted us to estimate the direct effect of apoB (ie, while accounting for the genetically predicted effects of LDL cholesterol and triglycerides) and allowed us to investigate the comparative causal roles of atherogenic lipoprotein traits on lifespan and disease risk, as proxied by outcome data from first-degree relatives. In this Article, we use the term direct to describe the effect of a lipid trait on disease while accounting for either mediation or confounding by another trait included in the model.

Estimates from mendelian randomisation are presented as odds ratios (ORs) per 1 SD higher lipid trait. Effect estimates for the association of lipid traits with vital status and age at death of parents are also presented in this way because, by being obtained from UK Biobank individuals at study entry, data on parental vital status are effectively cross-sectional.

For effects in the univariable mendelian randomisation analysis that survived correction for false discovery rate (FDR) of less than 5% based on estimates from the IVW method, we applied the weighted median,[25] weighted mode,[26] and MR-Egger[27] regression approaches. For multivariable analyses,[13] we additionally did multivariable MR-Egger.[28]

To replicate the findings identified for lifespan and type 2 diabetes, we did a conventional two-sample mendelian randomisation analysis in which SNPs for exposures (apoB, LDL cholesterol, and triglycerides) were derived from UK Biobank participants and SNPs for lifespan (grouping study participants based on whether they survived to the age corresponding to the 90th centile) and type 2 diabetes were derived from DIAMANTE, a previous large-scale GWAS.[17,18] The GWAS for lifespan used UK Biobank as a form of validation and the estimates we use here do not include UK Biobank. However, the latest release of data from DIAMANTE included type 2 diabetes cases from UK Biobank; since including these data might have led to overfitting of the mendelian randomisation estimate, we analysed type 2 diabetes GWAS estimates on individuals of European descent with UK Biobank data

See **Online** for appendix 1

See **Online** for appendix 2

removed. We did conventional univariable IVW two-sample mendelian randomisation. We then fitted a multivariable mendelian randomisation model simultaneously including genetic instruments for apoB, LDL cholesterol, and triglycerides to explore their direct effects on lifespan and type 2 diabetes. We did a MVMR-Egger[28] sensitivity analysis to evaluate whether the mendelian randomisation estimates were influenced by dose–response confounding of the genetic instruments.

Because the apoB effects on risk of type 2 diabetes might be confounded by adiposity and fat distribution, we additionally included body-mass index (BMI) and waist–hip ratio adjusted for BMI (WHRadjBMI) in the multivariable mendelian randomisation analyses for type 2 diabetes. 641 SNPs used in the instrument for BMI were derived from 461377 UK Biobank participants using the same GWAS pipeline as described previously.[12] 398 SNPs were used in the instrument for WHRadjBMI using a previous GWAS within UK Biobank.[29] We also used the IVW, MR-Egger, and weighted median approaches in a univariable setting to investigate the potential bidirectional relationship between apoB and type 2 diabetes.

Because some outcomes (ie, prostate cancer and breast cancer) occur only or predominantly in men or women, we re-ran the GWAS stratifying by sex and subsequently derived genetic instruments and did mendelian randomisation analyses using the aforementioned methods. To quantify the extent to which coronary heart disease and type 2 diabetes mediated the effects of apoB on lifespan (selected because these were among the most reliable findings identified, and these diseases are recognised to be among the top ten causes of death globally, and therefore might be expected to mediate the relationship between apoB and lifespan), we constructed multivariable mendelian randomisation models that included apoB, type 2 diabetes, coronary heart disease, and lifespan.

Because first-degree relatives share around 50% of DNA, and because outcomes were reported multiple times among first-degree relatives, the exposure-to-outcome analyses are not truly independent and use of conventional Bonferroni approaches to address multiple testing would be overly stringent. We therefore used Benjamini-Hochberg FDR of less than 5% to guide our interpretation of results from the initial univariable mendelian randomisation analysis of apoB for all 38 outcomes. This strategy was used to highlight which outcomes to evaluate in further detail in all subsequent analyses. FDR corrections were also applied to all additional univariable and multivariable analyses using UK Biobank participant data for completeness.

BOLT-LMM software[19] was used to do GWAS analyses and identify genetic instruments. The TwoSampleMR R package was used for all mendelian randomisation analyses.[30] The ggplot2 R package was used to generate forest plots.[31] Conditional *F*-statistics were generated using the MVMR R package.

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

A median of 400304 UK Biobank participants (range 273111–454999) reported information on prevalent diseases in first-degree relatives. A median of 361816 participants (361199–364661) reported data for their siblings, 423692 (422464–426391) reported data for their mothers, and 400687 (399089–407557) reported data for their fathers (appendix 1 table S4). 180472 (39·7%) participants reported that their mothers were alive and 103919 (23·1%) reported that their fathers were alive. The mean age at death among deceased parents was 75·7 years (SD 13·3) in mothers and 70·9 years (13·1) in fathers.

Across the diseases reported as occurring in first-degree relatives (table), the prevalence was similar in fathers and mothers, with siblings having the lowest prevalence of disease. Mothers were disproportionately affected by some diseases (eg, Alzheimer's disease or dementia [36048 cases; 8·6%] and severe depression [28351; 6·7%]) compared with fathers (Alzheimer's disease or dementia: 19255 cases, 4·8%; severe depression: 15430 cases, 3·9%). The prevalence of heart disease was highest in fathers (133320 cases; 32·7%), with mothers (85620 cases; 20·1%) and siblings (37858 cases; 10·4%) being less affected. The prevalence of stroke was similar in fathers (62810 cases; 15·6%) and mothers (60880 cases; 14·3%) but lower in siblings (12031 cases; 3·3%), whereas the prevalence of type 2 diabetes was similar in all first-degree relatives (38850 cases, 9·7% in fathers; 40091 cases, 9·5% in mothers; and 31073 cases, 8·6% in siblings).

Our apoB GWAS in the UK Biobank explained 10·4% of the heritability in this trait. Linkage disequilibrium clumping identified 229 apoB-associated SNPs with an *F*-statistic of 160. For multivariable mendelian randomisation, in which genetic instruments for apoB, LDL cholesterol, and triglycerides were analysed simultaneously, an additional 197 SNPs associated with LDL cholesterol (*F*-statistic of 167) and 411 SNPs associated with triglycerides (*F*-statistic of 121) were included in the analysis. In multivariable mendelian randomisation, the conditional *F*-statistics for apoB and LDL cholesterol were similar (37 and 34), and for triglycerides the *F*-statistic was 80. Thus, in both univariable and multivariable mendelian randomisation settings, our genetic instruments would not be considered weak.

Genetically elevated apoB was associated with a lower relative odds that an individual's parents were alive (fathers: OR 0·94, 95% CI 0·92–0·97; mothers: 0·97, 0·96–0·99; figure 1A; appendix 1 table S5). Fathers were estimated to die 0·89 years (95% CI 0·63–1·16;

FDR-adjusted p=4.0×10^{-10}) earlier (corresponding to a mean of 10·7 months of life lost per 1 SD higher apoB in offspring) and mothers 0·48 years (0·25–0·71; FDR-adjusted p=1.7×10^{-4}) earlier (corresponding to a mean of 5·8 months of life lost per 1 SD higher apoB in offspring). When taking into account the effects of LDL cholesterol and triglycerides in multivariable mendelian randomisation, these estimates for apoB increased to 1·94 years of life lost (0·91–2·96; FDR-adjusted p=0·0010) in fathers and 2·02 years of life lost (1·03–3·01; FDR-adjusted p=3.9×10^{-4}) in mothers, with the inverse relative odds for being alive also becoming more pronounced (figure 1B; appendix 1 table S6).

In univariable mendelian randomisation, a higher risk of heart disease was evident in all first-degree relatives of individuals with genetically elevated apoB (fathers: OR 1·20, 95% CI 1·16–1·25, FDR-adjusted p=1.9×10^{-21}; mothers: 1·14, 1·09–1·19, FDR-adjusted p=5.4×10^{-9}; siblings: 1·27, 1·20–1·35, FDR-adjusted p=7.6×10^{-16}). Mothers of individuals with genetically elevated apoB had a higher risk of stroke (1·05, 1·02–1·08, FDR-adjusted p=4.2×10^{-4}; figure 1A). Estimates were similar, although less precise, in multivariable mendelian randomisation (figure 1B).

Parents of individuals with genetically elevated apoB had a lower risk of type 2 diabetes (fathers: OR 0·88, 95% CI 0·84–0·91, FDR-adjusted p=2.3×10^{-8}; mothers: 0·92, 0·88–0·97, FDR-adjusted p=0·0042), whereas the decrease in risk was weaker in siblings (0·95, 0·90–0·998, FDR-adjusted p=0·082; figure 1A). In multivariable mendelian randomisation, the direction of effect was reversed: higher apoB was associated with elevated risk of type 2 diabetes (mothers: 1·34, 1·06–1·70, FDR-adjusted p=0·041), although the increase in risk was weaker in fathers (1·17; 0·94–1·46, FDR-adjusted p=0·24; figure 1B).

In univariable mendelian randomisation, the associations of LDL cholesterol with diseases in first-degree relatives were similar to those of apoB for all outcomes (figure 2A; appendix 1 table S7). However, effect estimates differed in multivariable mendelian randomisation. When taking into account apoB and triglycerides, higher LDL cholesterol was associated with an older age at death in both parents (fathers: increase of 1·31 years per 1 SD increase in LDL cholesterol, 95% CI 0·22–2·41, FDR-adjusted p=0·045; mothers: increase of 1·90 years, 0·84–2·96, FDR-adjusted p=0·0017) and a higher probability that mothers were alive (OR 1·16, 1·07–1·25, FDR-adjusted p=7.6×10^{-4}). For fathers, the increase in the odds of being alive was weaker (1·04, 0·94–1·16, FDR-adjusted p=0·53; figure 2B; appendix 1 table S6).

By contrast, the univariable and multivariable estimates of the association of LDL cholesterol with risk of type 2 diabetes were directionally constant: higher LDL cholesterol was associated with a lower risk of type 2 diabetes, irrespective of whether genetic associations

	Father (median N=400 687*)	Mother (median N=423 692*)	Siblings (median N=361 816*)
Age at death, years, mean (SD)†	70·9 (13·1)	75·7 (13·3)	NA
Still alive	103 919/450 333 (23·1%)	180 472/454 999 (39·7%)	NA
Alzheimer's disease or dementia	19 255/399 793 (4·8%)	36 548/423 738 (8·6%)	2094/361 264 (0·6%)
Bowel cancer	23 883/399 920 (6·0%)	22 028/423 135 (5·2%)	8920/361 508 (2·5%)
Breast cancer	NA	35 102/423 458 (8·3%)	16 586/361 809 (4·6%)
Chronic bronchitis or emphysema	46 263/402 389 (11·5%)	25 314/423 692 (6·0%)	10 325/361 823 (2·9%)
Diabetes	38 850/400 687 (9·7%)	40 091/423 892 (9·5%)	31 073/362 826 (8·6%)
Heart disease	133 320/407 557 (32·7%)	85 620/426 240 (20·1%)	37 858/363 542 (10·4%)
High blood pressure	91 242/402 899 (22·6%)	130 948/426 391 (30·7%)	77 059/364 661 (21·1%)
Lung cancer	37 443/401 624 (9·3%)	17 566/423 258 (4·2%)	8199/361 586 (2·3%)
Parkinson's disease	10 106/399 089 (2·5%)	6998/422 464 (1·7%)	2005/361 199 (0·6%)
Prostate cancer	30 945/399 670 (7·7%)	NA	5952/361 394 (1·6%)
Severe depression	15 430/399 499 (3·9%)	28 351/423 217 (6·7%)	26 368/362 315 (7·3%)
Stroke	62 810/402 616 (15·6%)	60 880/424 977 (14·3%)	12 031/361 925 (3·3%)

Data are % cases unless otherwise stated (numbers of cases and controls for each outcome are presented in appendix 1 table S4). NA=not applicable. *Value corresponds to the median number of UK Biobank participants reporting any outcome in fathers, mothers, or siblings (details are provided in appendix 1 table S4). †Age at death for fathers of 341 118 UK Biobank participants and mothers of 273 111 UK Biobank participants.

Table: Vital status, age at death, and prevalence of outcomes in first-degree relatives, as reported by UK Biobank participants

with apoB and triglycerides were included (fathers: OR 0·84, 95% CI 0·80–0·88, FDR-adjusted p=3.3×10^{-13} in univariable mendelian randomisation and 0·70, 0·55–0·89, FDR-adjusted p=3.3×10^{-3} in multivariable mendelian randomisation; mothers: 0·86, 0·82–0·90, FDR-adjusted p=2.7×10^{-9} in univariable mendelian randomisation and 0·63, 0·49–0·82, FDR-adjusted p=4.2×10^{-4} in multivariable mendelian randomisation; figure 2).

A positive relationship of LDL cholesterol with heart disease in parents and siblings (and with stroke in mothers) in univariable mendelian randomisation became weaker when accounting for apoB and triglycerides in the multivariable analysis, as did the inverse relationships with lung cancer, Parkinson's disease, and prostate cancer (figure 2).

Estimates for triglycerides were generally robust to multivariable mendelian randomisation analyses (figure 2; appendix 1 table S8). Higher triglycerides were associated with an earlier age at death in both parents (fathers: 0·68 years of life lost, 95% CI 0·47–0·88, FDR-adjusted p=0·0072; mothers: 0·58 years of life lost, 0·38–0·79, FDR-adjusted p=0·015), but the magnitude of association was smaller than for apoB in multivariable mendelian randomisation. Similarly, positive associations between triglycerides and heart disease in parents and siblings in univariable mendelian randomisation were robust to multivariable mendelian randomisation, but the
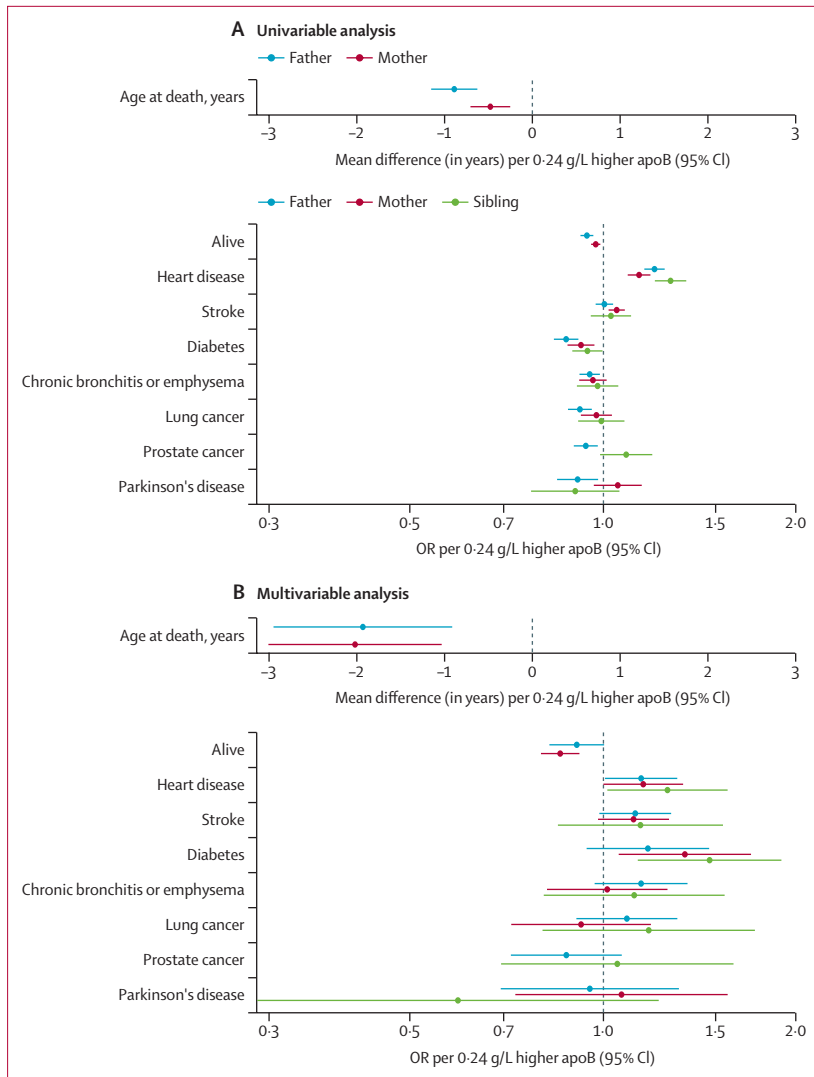
Figure 1: Univariable (A) and multivariable (B) mendelian randomisation estimates of genetically elevated apoB and risk of outcomes in first-degree relatives, including vital status and age at death
Multivariable estimates represent the direct effects of apoB, adjusted for LDL cholesterol and triglycerides. ApoB=apolipoprotein B. OR=odds ratio.

individuals to have a younger age of death than the age corresponding to the 90th survival centile) and this association was stronger when taking into account the effects of LDL cholesterol and triglycerides in multivariable mendelian randomisation. An increase in apoB by 1 SD in multivariable mendelian randomisation was associated with lower relative odds of surviving to the 90th centile of lifespan (OR 0·38, 95% CI 0·22–0·65). By contrast, the initial harmful effect of LDL cholesterol on lifespan in univariable mendelian randomisation was directionally reversed in multivariable mendelian randomisation. Likewise, triglycerides had a detrimental effect on lifespan in univariable mendelian randomisation, but the relationship became weaker in multivariable mendelian randomisation.

No clear effect of genetically predicted apoB on type 2 diabetes was identified in univariable mendelian randomisation. In contrast, higher LDL cholesterol caused a lower risk of type 2 diabetes (OR 0·84 per 1 SD higher LDL cholesterol, 95% CI 0·76–0·92) and higher triglycerides caused a higher type 2 diabetes risk (1·40 per 1 SD higher triglycerides, 1·26–1·55) in univariable mendelian randomisation. In multivariable mendelian randomisation, the estimate for triglycerides was largely unchanged (1·39 per 1 SD higher triglycerides, 1·25–1·54). However, the protective effect of LDL cholesterol against type 2 diabetes became more pronounced (0·34 per 1 SD higher LDL cholesterol, 0·21–0·54) and a strong positive association between apoB and type 2 diabetes risk emerged (2·32 per 1 SD higher apoB, 1·49–3·61; figure 3, appendix 1 table S10).

Repeating analyses using robust mendelian randomisation approaches for both univariable and multivariable mendelian randomisation led to generally consistent associations between apoB and risk of disease in first-degree relatives (appendix 1 tables S11–15, appendix 2 pp 3–6), with risk of type 2 diabetes in DIAMANTE (appendix 1 table S16, appendix 2 p 7), and with lifespan (appendix 1 table S9). The estimate of triglycerides and risk of type 2 diabetes was notable for being directionally opposite between IVW and MR-Egger in both univariable analyses for type 2 diabetes in parents (appendix 2 p 4) and in conventional mendelian randomisation in DIAMANTE (appendix 1 table S16, appendix 2 p 7). Of note, using multivariable MR-Egger and orientating SNPs so that they all associated with higher triglycerides led to an attenuation of the association between triglycerides and type 2 diabetes, suggesting that it might be explained by unbalanced horizontal pleiotropy (appendix 1 table S16). Excluding variants from in and around the *APOE* locus led to a marked attenuation of the relationship between genetic instruments for apoB and LDL cholesterol and risk of Alzheimer's disease in mothers (appendix 1 tables S17, S18). Inclusion of BMI in the multivariable mendelian randomisation for type 2 diabetes in DIAMANTE had no notable effect on the direct causal effects of apoB, LDL cholesterol, or triglycerides (appendix 1 table S19,

magnitudes of effect were attenuated. For type 2 diabetes, positive associations with triglycerides were evident in both univariable and multivariable mendelian randomisation analyses.

Given the findings for vital status and risk of type 2 diabetes in first-degree relatives, we sought to further elucidate these relationships using a conventional two-sample mendelian randomisation framework with data from two GWAS consortia: a GWAS of lifespan with 11 262 individuals surviving to the age corresponding to the 90th survival centile and DIAMANTE, including 55 927 cases of type 2 diabetes. The analyses of lifespan replicated our findings from age of death and vital status of parents (figure 3, appendix 1 table S9). ApoB was detrimental to survival (ie, higher apoB caused

appendix 2 p 8), which was also the case when accounting for WHRadjBMI (appendix 1 table S20, appendix 2 p 8). We also applied univariable mendelian randomisation to evaluate evidence of a bidirectional association between apoB and type 2 diabetes (ie, whether genetic liability to type 2 diabetes affects apoB levels). We identified weak evidence of an effect of type 2 diabetes liability on apoB in this analysis using genetic instruments derived from DIAMANTE based on the IVW and MR-Egger methods, whereas the weighted median estimate was directionally opposite to the multivariable mendelian randomisation estimate between apoB and type 2 diabetes (appendix 1 table S21). We did sex-specific analyses for diseases only occurring in women and men and found that the results were largely consistent with the sex-agnostic genetic instrument (appendix 1 table S22). Multivariable mendelian randomisation estimated that coronary heart disease and type 2 diabetes explained between 25% and 40% of the relationship between apoB and lifespan (appendix 1 table S23).

Discussion

Our findings implicate apoB in several major diseases, including heart disease, stroke, and diabetes. Importantly, a higher apoB was associated with a shorter lifespan in parents, as shown by both a lower odds that an individual's parents were alive and a younger age at their death. Inclusion of LDL cholesterol and triglycerides in multivariable mendelian randomisation strengthened these effects, suggesting that apoB was the predominant driving influence on lifespan among these three lipid traits. All of these findings were replicated using an independent dataset, verifying their robustness. Taken together, our findings show that reductions in apoB should be the primary goal of lipid lowering, because not only does this lead to lower risk of common diseases such as heart disease and stroke, but also a reduction in apoB prolongs life by a period of months to years. Importantly, the effect estimates we report in this study are likely to be diluted due to the nature of the exposure and outcome. The real magnitudes of effect, in terms of duration of life lost due to elevated apoB, are likely to be greater, as indicated in the effect estimates from the GWAS of 90th centile of survival. Our findings further strengthen the suggestion that it is the number of circulating apoB particles, rather than their lipid content, that is the critical element for atherogenesis, manifested as coronary heart disease and ischaemic stroke.[5,6,8] The cholesterol within LDL particles does play a causal role in atherogenesis but it is within the physiological framework of the trapping of apoB particles within the arterial wall.

The findings for type 2 diabetes are intriguing. In naive univariable mendelian randomisation analysis, a higher LDL cholesterol was associated with a lower risk of type 2 diabetes, a finding that was strengthened with multivariable mendelian randomisation analysis. By contrast, a higher apoB was associated with a lower risk
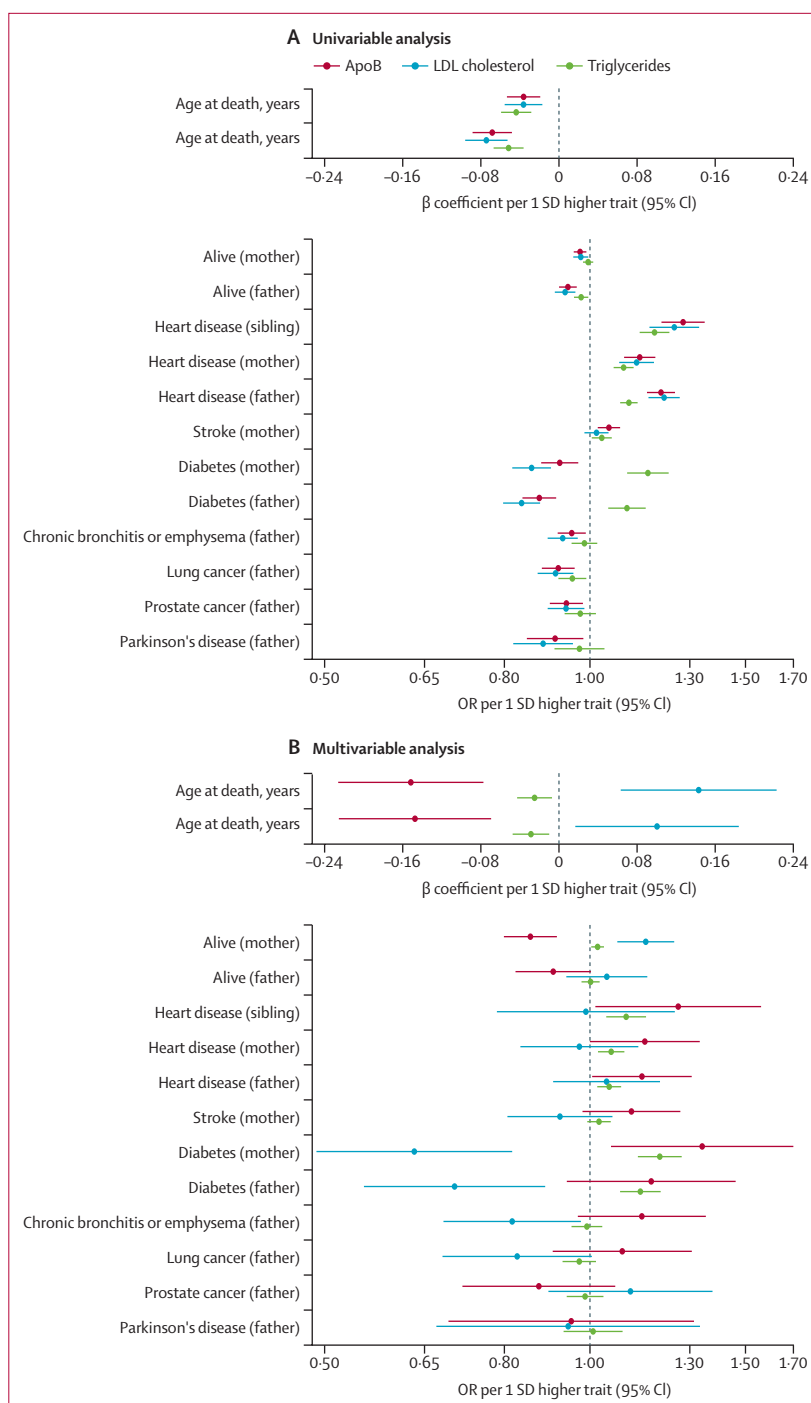

Figure 2: Univariable (A) and multivariable (B) mendelian randomisation estimates of association between genetically elevated apoB, LDL cholesterol, and triglycerides and risk of outcomes in first-degree relatives, including vital status and age at death
Multivariable estimates represent the direct effects of each lipoprotein entity, adjusted for the other two traits. ApoB=apolipoprotein B. OR=odds ratio.

of type 2 diabetes in univariable mendelian randomisation, but in multivariable mendelian randomisation, a strong, positive association between apoB and risk of
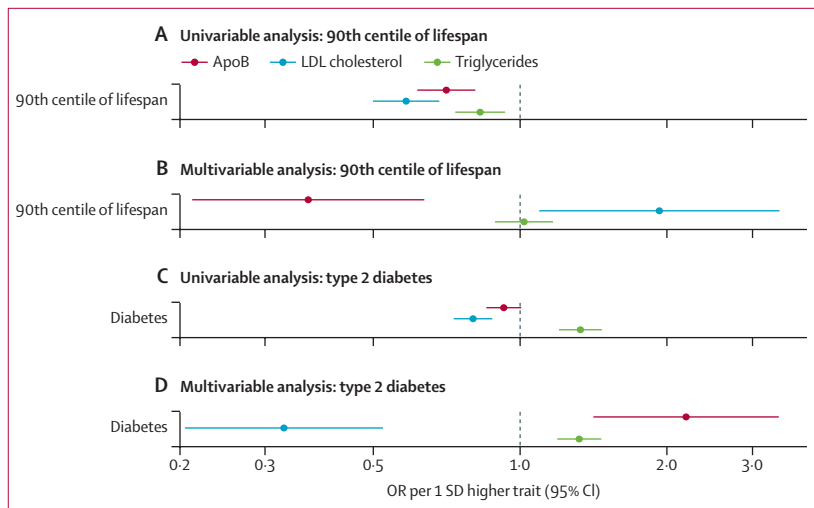
Figure 3: **Univariable and multivariable estimates of genetically elevated apoB, LDL cholesterol and triglycerides with lifespan (A, B) and risk of type 2 diabetes (C, D) using two-sample mendelian randomisation**
Multivariable mendelian randomisation estimates represent the direct effects of each lipoprotein entity, adjusted for the other two traits. ApoB=apolipoprotein B. OR=odds ratio.

type 2 diabetes emerged. This finding was replicated using data from DIAMANTE, providing further evidence of a potentially direct causal relation between apoB and type 2 diabetes.

The majority of apoB particles in plasma are LDL particles, which are cleared most efficiently by the LDL pathway. Although the liver is the major site of LDL clearance, peripheral tissues, such as the pancreas, also have LDL receptors and clear LDL particles (appendix 2 p 9).[32] Moreover, experimentally increasing cholesterol uptake or reducing cholesterol efflux produces islet cell dysfunction with reduced insulin secretion and cell proliferation.[33–38] All of these findings are consistent with the hypothesis that increased uptake of LDL particles by the LDL pathway could contribute to the pathogenesis of type 2 diabetes.

LDL particles tend to be depleted in cholesterol when LDL particles are overproduced.[6,39] This is the scenario approximated by our multivariable mendelian randomisation analysis when apoB is increased and LDL cholesterol kept constant. In this situation, clearance of LDL particles by the LDL pathway is increased, with the result that delivery of LDL particles to pancreatic islet cells could be increased (appendix 2 p 10). By contrast, decreased LDL clearance via the LDL pathway is characterised by LDL particles that are cholesterol-enriched.[6,39] This is the scenario approximated by our multivariable mendelian randomisation analysis when LDL cholesterol is increased and apoB kept constant, and the risk of type 2 diabetes is reduced. In support of our findings is the reduced incidence of diabetes in familial hypercholesterolaemia,[40] a disorder characterised by markedly reduced uptake of LDL particles and the cholesterol they contain via the LDL pathway.

In-vitro and mendelian randomisation analyses have shown that increased activity of the LDL pathway related to proprotein convertase subtilisin/kexin type 9 (PCSK9) deficiency is associated with an increased incidence of diabetes.[36–38,40–43] The finding that the incidence of dysglycaemia was higher in individuals with cholesterol-depleted apoB particles and lower in individuals with cholesterol-enriched apoB particles is also consistent with our hypothesis.

Alternatively, increased uptake of LDL particles by adipocytes has been posited by Faraj[44] to induce adipose tissue dysfunction with an increased inflammatory response, resulting in reduced insulin sensitivity. Faraj[44] has also suggested that cytotoxic injury to adipocytes due to increased uptake of LDL might be a mechanism by which statins and PCSK9 inhibitors, agents which increase activity of the LDL pathway, could be diabetogenic. Accordingly, we hypothesise that increased uptake of LDL particles might injure adipocytes or islet cells or both, which might explain the strong positive association between apoB and the risk of diabetes that was observed in this study.

The resilience of our findings to inclusion of adiposity measured by BMI and WHRadjBMI (indexing total and visceral adiposity, respectively) in the multivariable mendelian randomisation suggests that the associations we identify with risk of type 2 diabetes are not confounded by upstream adiposity or fat distribution measures, which are recognised to causally affect blood lipid traits.[45] Our findings suggest that therapies that do not increase pancreatic uptake of LDL particles while increasing hepatic uptake of LDL particles or precursor VLDL particles might lead to a reduction in risk of both coronary heart disease and type 2 diabetes.

This study has the usual caveats and limitations that apply to mendelian randomisation studies,[46] and also has additional issues related to the particularities of its design. First, we cannot claim that apoB directly causes outcomes in the first-degree relatives of the UK Biobank participants because the genetically elevated apoB is only an approximation to unconfounded estimates within first-degree relatives. However, the replication of our analyses for lifespan and type 2 diabetes corro~-borates our findings and therefore strengthens these hypotheses. Second, because first-degree relatives share 50% of their DNA, the effect estimates generated by our genetic instruments have been attenuated through regression dilution. Thus, the magnitudes of effect estimates for coronary heart disease, type 2 diabetes, and lifespan associations using conventional two-sample mendelian randomisation approaches are several times greater than the equivalent estimates obtained using outcomes reported as occurring in first-degree relatives. Third, because outcomes are self-reported, measurement error is likely. For example, an individual who has heart disease might be more likely to report a first-degree relative as also

having heart disease. Study participants might also have accidentally answered questions regarding non-blood relatives (eg, adopted individuals, step-parents, or step-siblings) or half-siblings, which would have weakened effect estimates. Fourth, outcomes related to siblings might represent multiple individuals collapsed into a single trait because individuals were asked about siblings only cumulatively rather than individually. This would dilute effects if more than one sibling has an outcome (eg, if two or more first-degree siblings had coronary heart disease).

The cumulative effect of these limitations probably explains why the effect estimates we report for heart disease (ORs of 1·14 to 1·27) are of considerably smaller magnitude than the approximation to within-individual effect estimates we previously reported using data from UK Biobank and CARDIoGRAMplusC4D (OR 1·7) scaled to the same difference in exposure (1 SD higher apoB in the UK Biobank participants).[12] A similar difference in the magnitudes of effect is evident between estimates of apoB with risk of type 2 diabetes in first-degree relatives (ORs of 1·17 to 1·33) versus data from DIAMANTE (OR >2). Of note, both estimates for heart disease (using CARDIoGRAMplusC4) and type 2 diabetes (using DIAMANTE) might be inflated by spectrum-bias effects arising from case-control studies that sample cases at the extremes of the phenotypic distribution. Future studies that apply alternative approaches when analysing highly correlated lipoprotein traits might therefore be valuable in corroborating our findings, such as the mendelian randomisation Bayesian model averaging approach.[11]

In conclusion, our evaluation of apoB using outcomes in first-degree relatives identified that higher apoB is detrimental to lifespan and increases the risk of coronary heart disease and type 2 diabetes. Lifestyle and pharmacological approaches to lowering apoB should have widespread beneficial effects, including preventing common diseases and prolonging life.

References
1 Borén J, Chapman MJ, Krauss RM, et al. Low-density lipoproteins cause atherosclerotic cardiovascular disease: pathophysiological, genetic, and therapeutic insights: a consensus statement from the European Atherosclerosis Society Consensus Panel. *Eur Heart J* 2020; **41:** 2313–30.

2 Ference BA, Majeed F, Penumetcha R, Flack JM, Brook RD. Effect of naturally random allocation to lower low-density lipoprotein cholesterol on the risk of coronary heart disease mediated by polymorphisms in NPC1L1, HMGCR, or both: a 2×2 factorial Mendelian randomization study. *J Am Coll Cardiol* 2015; **65:** 1552–61.

3 Holmes MV, Asselbergs FW, Palmer TM, et al. Mendelian randomization of blood lipids for coronary heart disease. *Eur Heart J* 2015; **36:** 539–50.

4 Silverman MG, Ference BA, Im K, et al. Association between lowering LDL-C and cardiovascular risk reduction among different therapeutic interventions: a systematic review and meta-analysis. *JAMA* 2016; **316:** 1289–97.

5 Sniderman AD, Pencina M, Thanassoulis G. ApoB. *Circ Res* 2019; **124:** 1425–27.

6 Sniderman AD, Thanassoulis G, Glavinovic T, et al. Apolipoprotein B particles and cardiovascular disease: a narrative review. *JAMA Cardiol* 2019; **4:** 1287–95.

7 Sniderman AD, Williams K, Contois JH, et al. A meta-analysis of low-density lipoprotein cholesterol, non-high-density lipoprotein cholesterol, and apolipoprotein B as markers of cardiovascular risk. *Circ Cardiovasc Qual Outcomes* 2011; **4:** 337–45.

8 Borén J, Williams KJ. The central role of arterial retention of cholesterol-rich apolipoprotein-B-containing lipoproteins in the pathogenesis of atherosclerosis: a triumph of simplicity. *Curr Opin Lipidol* 2016; **27:** 473–83.

9 Ference BA, Kastelein JJP, Ginsberg HN, et al. Association of genetic variants related to CETP inhibitors and statins with lipoprotein levels and cardiovascular risk. *JAMA* 2017; **318:** 947–56.

10 Ference BA, Kastelein JJP, Ray KK, et al. Association of triglyceride-lowering LPL variants and LDL-C-lowering LDLR variants with risk of coronary heart disease. *JAMA* 2019; **321:** 364–73.

11   Zuber V, Gill D, Ala-Korpela M, et al. High-throughput multivariable Mendelian randomization analysis prioritizes apolipoprotein B as key lipid risk factor for coronary artery disease. *Int J Epidemiol* 2020; published online Nov 1. https://doi.org/10.1093/ije/dyaa216.

12   Richardson TG, Sanderson E, Palmer TM, et al. Evaluating the relationship between circulating lipoprotein lipids and apolipoproteins with risk of coronary heart disease: a multivariable Mendelian randomisation analysis. *PLoS Med* 2020; 17: e1003062.

13   Sanderson E, Davey Smith G, Windmeijer F, Bowden J. An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. *Int J Epidemiol* 2019; 48: 713–27.

14   Elliott P, Peakman TC, Biobank UK. The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. *Int J Epidemiol* 2008; 37: 234–44.

15   Sudlow C, Gallacher J, Allen N, et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015; 12: e1001779.

16   Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* 2018; 562: 203–09.

17   Deelen J, Evans DS, Arking DE, et al. A meta-analysis of genome-wide association studies identifies multiple longevity genes. *Nat Commun* 2019; 10: 3669.

18   Mahajan A, Taliun D, Thurner M, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet* 2018; 50: 1505–13.

19   Loh PR, Tucker G, Bulik-Sullivan BK, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* 2015; 47: 284–90.

20   Loh PR, Kichaev G, Gazal S, Schoech AP, Price AL. Mixed-model association for biobank-scale datasets. *Nat Genet* 2018; 50: 906–08.

21   Abecasis GR, Auton A, Brooks LD, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012; 491: 56–65.

22   Mitchell R, Elsworth BL, Raistrick CA, Paternoster L, Hemani G, Gaunt TR. MRC IEU UK Biobank GWAS pipeline version 2. Feb 20, 2019. https://data.bris.ac.uk/data/dataset/pnoat8cxo0u52p6ynfaekeigi (accessed June 1, 2020).

23   Bowden J, Del Greco M F, Minelli C, Davey Smith G, Sheehan NA, Thompson JR. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I2 statistic. *Int J Epidemiol* 2016; 45: 1961–74.

24   Sanderson E, Spiller W, Bowden J. Testing and correcting for weak and pleiotropic instruments in two-sample multivariable mendelian randomisation. *bioRxiv* 2020; published online April 3. https://doi.org/10.1101/2020.04.02.021980 (preprint).

25   Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol* 2016; 40: 304–14.

26   Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol* 2017; 46: 1985–98.

27   Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* 2015; 44: 512–25.

28   Rees JMB, Wood AM, Burgess S. Extending the MR-Egger method for multivariable Mendelian randomization to correct for both measured and unmeasured pleiotropy. *Stat Med* 2017; 36: 4705–18.

29   O'Connor LJ, Price AL. Distinguishing genetic correlation from causation across 52 diseases and complex traits. *Nat Genet* 2018; 50: 1728–34.

30   Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. *eLife* 2018; 7: e34408.

31   Wickham H. ggplot2—elegant graphics for data analysis, 2nd edn. New York, NY: Springer Verlag, 2016.

32   Dietschy JM, Spady DK, Stange EF. Quantitative importance of different organs for cholesterol synthesis and low-density-lipoprotein degradation. *Biochem Soc Trans* 1983; 11: 639–41.

33   Hao M, Head WS, Gunawardana SC, Hasty AH, Piston DW. Direct effect of cholesterol on insulin secretion: a novel mechanism for pancreatic beta-cell dysfunction. *Diabetes* 2007; 56: 2328–38.

34   Ishikawa M, Iwasaki Y, Yatoh S, et al. Cholesterol accumulation and diabetes in pancreatic beta-cell-specific SREBP-2 transgenic mice: a new model for lipotoxicity. *J Lipid Res* 2008; 49: 2524–34.

35   Cnop M, Hannaert JC, Grupping AY, Pipeleers DG. Low density lipoprotein can cause death of islet beta-cells by its cellular uptake and oxidative modification. *Endocrinology* 2002; 143: 3449–53.

36   Da Dalt L, Ruscica M, Bonacina F, et al. PCSK9 deficiency reduces insulin secretion and promotes glucose intolerance: the role of the low-density lipoprotein receptor. *Eur Heart J* 2019; 40: 357–68.

37   Kruit JK, Kremer PH, Dai L, et al. Cholesterol efflux via ATP-binding cassette transporter A1 (ABCA1) and cholesterol uptake via the LDL receptor influences cholesterol-induced impairment of beta cell function in mice. *Diabetologia* 2010; 53: 1110–19.

38   Roehrich ME, Mooser V, Lenain V, et al. Insulin-secreting beta-cell dysfunction induced by human lipoproteins. *J Biol Chem* 2003; 278: 18368–75.

39   Sniderman AD, Couture P, Martin SS, et al. Hypertriglyceridemia and cardiovascular risk: a cautionary note about metabolic confounding. *J Lipid Res* 2018; 59: 1266–75.

40   Besseling J, Kastelein JJ, Defesche JC, Hutten BA, Hovingh GK. Association between familial hypercholesterolemia and prevalence of type 2 diabetes mellitus. *JAMA* 2015; 313: 1029–36.

41   Bello-Chavolla OY, Kuri-García A, Ríos-Ríos M, et al. Familial combined hyperlipidemia: current knowledge, perspectives, and controversies. *Rev Invest Clin* 2018; 70: 224–36.

42   Schmidt AF, Swerdlow DI, Holmes MV, et al. PCSK9 genetic variants and risk of type 2 diabetes: a mendelian randomisation study. *Lancet Diabetes Endocrinol* 2017; 5: 97–105.

43   Ference BA, Robinson JG, Brook RD, et al. Variation in PCSK9 and HMGCR and risk of cardiovascular disease and diabetes. *N Engl J Med* 2016; 375: 2144–53.

44   Faraj M. LDL, LDL receptors, and PCSK9 as modulators of the risk for type 2 diabetes: a focus on white adipose tissue. *J Biomed Res* 2020; 34: 251–59.

45   Dale CE, Fatemifar G, Palmer TM, et al. Causal associations of adiposity and body fat distribution with coronary heart disease, stroke subtypes, and type 2 diabetes mellitus: a mendelian randomization analysis. *Circulation* 2017; 135: 2373–88.

46   Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003; 32: 1–22.