

Supplementary Materials

Supplementary Checklist: STROBE-MR Reporting Guidelines.

1. TITLE and ABSTRACT

Indicate Mendelian randomization as the study's design in the title and/or the abstract.

Title and abstract

INTRODUCTION

2. Background

Explain the scientific background and rationale for the reported study. Is causality between exposure and outcome plausible? Justify why MR is a helpful method to address the study question.

Addressed in the Introduction and Methods

3. Objectives

State specific objectives clearly, including pre-specified causal hypotheses (if any).

Addressed in the Introduction.

METHODS

4. Study design and data sources

Present key elements of study design early in the paper. Consider including a table listing sources of data for all phases of the study. For each data source contributing to the analysis, describe the following:

a) Describe the study design and the underlying population from which it was drawn.

Describe also the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection, if available.

b) Give the eligibility criteria, and the sources and methods of selection of participants.

c) Explain how the analyzed sample size was arrived at.

d) Describe measurement, quality and selection of genetic variants.

e) For each exposure, outcome and other relevant variables, describe methods of assessment and, in the case of diseases, the diagnostic criteria used.

f) Provide details of ethics committee approval and participant informed consent, if relevant.

Addressed in the Methods.

5. Assumptions

Explicitly state assumptions for the main analysis (e.g. relevance, exclusion, independence, homogeneity) as well assumptions for any additional or sensitivity analysis.

Addressed in the Methods.

6. Statistical methods: main analysis

Describe statistical methods and statistics used.

a) Describe how quantitative variables were handled in the analyses (i.e., scale, units, model).

b) Describe the process for identifying genetic variants and weights to be included in the analyses (i.e., independence and model). Consider a flow diagram.

c) Describe the MR estimator, e.g. two-stage least squares, Wald ratio, and related statistics.

Detail the included covariates and, in case of two-sample MR, whether the same covariate set was used for adjustment in the two samples.

d) Explain how missing data were addressed.

e) If applicable, say how multiple testing was dealt with.

Addressed in the Methods.

7. Assessment of assumptions

Describe any methods used to assess the assumptions or justify their validity.

Addressed in the Methods.

8. Sensitivity analyses

Describe any sensitivity analyses or additional analyses performed.

Addressed in the Methods.

9. Software and pre-registration

a) Name statistical software and package(s), including version and settings used.

Addressed in the Methods.

b) State whether the study protocol and details were pre-registered (as well as when and where).

Addressed in the Methods.

RESULTS

10. Descriptive data

a) Report the numbers of individuals at each stage of included studies and reasons for exclusion. Consider use of a flow-diagram.

b) Report summary statistics for phenotypic exposure(s), outcome(s) and other relevant variables (e.g. means, standard deviations, proportions).

c) If the data sources include meta-analyses of previous studies, provide the number of studies, their reported ancestry, if available, and assessments of heterogeneity across these studies. Consider using a supplementary table for each data source.

d) For two-sample Mendelian randomization:

i. Provide information on the similarity of the genetic variant-exposure associations between the exposure and outcome samples.

ii. Provide information on extent of sample overlap between the exposure and outcome data sources.

Addressed in the Methods, Results and Supplementary Tables.

11. Main results

a) Report the associations between genetic variant and exposure, and between genetic variant and outcome, preferably on an interpretable scale (e.g. comparing 25th and 75th percentile of allele count or genetic risk score, if individual-level data available).

b) Report causal effect estimate between exposure and outcome, and the measures of uncertainty from the MR analysis. Use an intuitive scale, such as odds ratio, or relative risk, per standard deviation difference.

c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time-period.

d) Consider any plots to visualize results (e.g. forest plot, scatterplot of associations between genetic variants and outcome versus between genetic variants and exposure).

Addressed in the Results and Supplementary Tables.

12. Assessment of assumptions

a) Assess the validity of the assumptions.

b) Report any additional statistics (e.g., assessments of heterogeneity, such as I^2 , Q statistic).

Addressed in the Results, Supplementary Tables and Discussion.

13. Sensitivity and additional analyses

a) Use sensitivity analyses to assess the robustness of the main results to violations of the assumptions.

b) Report results from other sensitivity analyses (e.g., replication study with different dataset, analyses of subgroups, validation of instrument(s), simulations, etc.).

c) Report any assessment of direction of causality (e.g., bidirectional MR).

d) When relevant, report and compare with estimates from non-MR analyses.

e) Consider any additional plots to visualize results (e.g., leave-one-out analyses).

Addressed in the Results and Supplementary Tables.

DISCUSSION

14. Key results

Summarize key results with reference to study objectives.

Addressed in the Discussion.

15. Limitations

Discuss limitations of the study, taking into account the validity of the MR assumptions, other sources of potential bias, and imprecision. Discuss both direction and magnitude of any potential bias, and any efforts to address them.

Addressed in the Discussion.

16. Interpretation

a) Give a cautious overall interpretation of results considering objectives and limitations.

Compare with results from other relevant studies.

b) Discuss underlying biological mechanisms that could be modelled by using the genetic variants to assess the relationship between the exposure and the outcome.

c) Discuss whether the results have clinical or policy relevance, and whether interventions could have the same size effect.

Addressed in the Discussion.

17. Generalizability

Discuss the generalizability of the study results (a) to other populations (i.e. external validity), (b) across other exposure periods/timings, and (c) across other levels of exposure.

Addressed in the Discussion.

OTHER INFORMATION

18. Funding

Give the source of funding and the role of the funders for the present study and, if applicable, for the original study or studies on which the present article is based.

Addressed in the Funding.

19. Data and data sharing

Present data used to perform all analyses or report where and how the data can be accessed. State whether statistical code is publicly accessible and if so, where.

Addressed in the Methods.

20. Conflicts of Interest

All authors should declare all potential conflicts of interest.

Addressed in the Conflicts of interest.

Supplementary Discussion

Discussion of polygenic results. Finally, our polygenic LDL-C findings are generally in line with previous observational and genetics-based work finding that lower levels of circulating LDL-C are linked with higher T2D risk.¹ Interestingly, it has also been shown that among individuals with CVD, higher LDL-C levels are associated with reduced T2D, and families with hypercholesterolemia have demonstrated reduced T2D risk,^{2,3} supporting a growing body of literature suggesting biological relationships between lipids and T2D, and providing important insight into mechanisms underlying diabetogenesis.^{2,3} One potential explanation for the observed discrepancy in how different loci affecting LDL-C influence T2D and glycemic control is that standard measurements of LDL-C do not distinguish between its various subtypes. LDL-C is a heterogeneous particle with different subtypes, each of which might have distinct biological effects on metabolic processes, including insulin sensitivity and glucose regulation.⁴ Additionally, LDL-C undergoes changes in its composition and structure, such as altered electrophoretic mobility, increased triglyceride and ceramide content, prolonged retention of modified LDL-C in the blood, enhanced uptake by macrophages, and the formation of foam cells; the inability to

differentiate these subtypes, or more fine-grained aspects of LDL-C dynamics, in routine clinical measurements could obscure the specific pathways through which certain loci, such as *PCSK9* and *HMGCR*, influence T2D risk.^{5,6} By providing a detailed profile of lipid subtypes and their interactions with other metabolites, metabolomics can help identify the specific LDL-C subtypes that are most strongly associated with T2D risk.⁵ For instance, metabolomics data could reveal whether certain LDL-C subtypes are more prevalent in individuals with T2D or whether they interact differently with glucose and insulin metabolism compared to other subtypes. Metabolomics can capture a broader spectrum of lipid-related metabolites, offering insights into the downstream effects of LDL-C subtypes on metabolic pathways involved in glycemic control.⁷ This improved resolution could help to clarify why certain genetic variants associated with LDL-C have differing impacts on T2D.⁶ For future research, exploring the complexities of metabolomics data could be invaluable in disentangling these relationships, leading to a more refined understanding of the metabolic consequences of LDL-C variability and potentially guiding more targeted therapeutic strategies.

Study strengths. This study has several strengths. The primary strength of the study is our use of non-EUR data to perform parallel population-specific MR analyses in the most comprehensive genetics-based investigation of the risk for T2D associated with lipid-lowering drug-targets to date. A recent meta-analysis of 20,692 RCTs in the United States listed in ClinicalTrials.gov from the years 2000-2020 found that while there have been positive trends in RCT enrollment, there are still substantial racial/ethnic differences in minority recruitment, reporting, and representation.⁸ Among the 20,692 RCTs (generating data from more than 4.76 million participants), Turner et al. found that European populations comprised almost 80% of all participants while only 10% of participants were AFR, 6% Hispanics/Latino, and 1% Asian (EAS and SAS combined).⁸ This study and others like it may be an important step to help address the current imbalances in representation for minority race/ethnicity participants in RCTs,⁸ and resulting health disparities^{9,10} (at least with regards to lipid-lowering therapies, including statins, which are the most prescribed for noncommunicable diseases drug worldwide^{11,12}). Another strength is our use of both proteomic and transcriptomic data to construct additional PCSK9 instruments in therapeutically and T2D-relevant tissues. For example, in mice, tissue-specific investigation of PCSK9 inhibition found that PCSK9 expression in the pancreas may be related to impaired β cell dysfunction while liver-derived PCSK9 was not related β cell dysfunction or insulin secretion.¹³ However, our PCSK9 instruments derived from pancreatic tissue (in addition to the liver tissue PCSK9 instruments) failed to find evidence of pancreatic or hepatic PCSK9 expression on T2D or glycemic markers using human data, which should reassure clinicians and patients regarding the PCSK9-T2D relationship. Other strengths include leveraging complementary MR methods incorporating genetic matrices of the underlying LD structure between drug target instrument variants to improve instrument precision, which is crucial for causal inference in MR.¹⁴ Employing complementary MR methods, heterogeneity tests, and alternate instruments – and observing consistent MR estimates across them – further strengthens causal inference.¹⁴⁻¹⁶

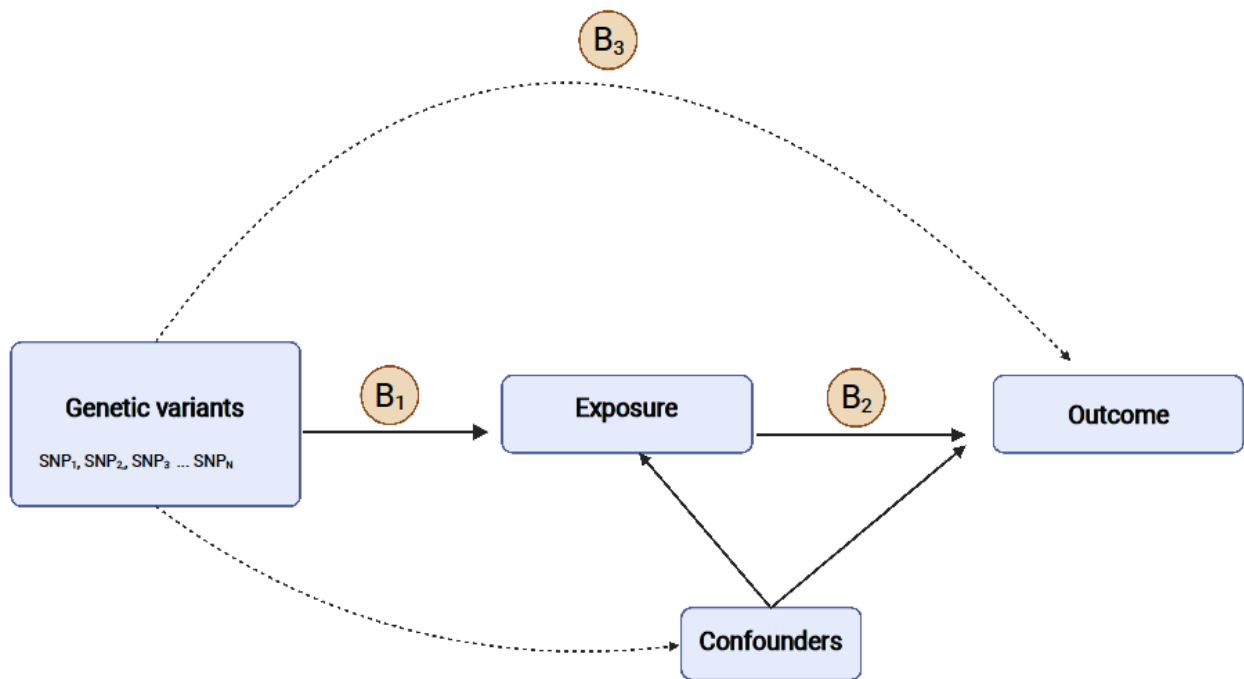


Figure S1. Mendelian Randomization Model and Assumptions B_2 is the genetic association of interest, estimated by $B_2 = B_1 / B_3$. B_1 and B_3 are the associations of the genetic variants with the exposure and the outcome. MR assumes that the genetic variants comprising the instrument for the exposure only impact the outcome of interest via the exposure and not directly, or via confounders (dotted lines).¹⁴

REFERENCES

1. Klimentidis YC, Arora A, Newell M, Zhou J, Ordovas JM, Renquist BJ, Wood AC. Phenotypic and Genetic Characterization of Lower LDL Cholesterol and Increased Type 2 Diabetes Risk in the UK Biobank. *Diabetes*. Oct 2020;69(10):2194-2205. doi:10.2337/db19-1134
2. Sacks FM, Tonkin AM, Craven T, et al. Coronary heart disease in patients with low LDL-cholesterol: benefit of pravastatin in diabetics and enhanced role for HDL-cholesterol and triglycerides as risk factors. *Circulation*. Mar 26 2002;105(12):1424-8. doi:10.1161/01.cir.0000012918.84068.43
3. Besseling J, Kastelein JJ, Defesche JC, Hutten BA, Hovingh GK. Association between familial hypercholesterolemia and prevalence of type 2 diabetes mellitus. *Jama*. Mar 10 2015;313(10):1029-36. doi:10.1001/jama.2015.1206
4. Wolska A, Remaley AT. Measuring LDL-cholesterol: what is the best way to do it? *Curr Opin Cardiol*. Jul 2020;35(4):405-411. doi:10.1097/hco.0000000000000740
5. Bonilha I, Hajdich E, Luchiari B, Nadruz W, Le Goff W, Sposito AC. The Reciprocal Relationship between LDL Metabolism and Type 2 Diabetes Mellitus. *Metabolites*. Nov 28 2021;11(12)doi:10.3390/metabo11120807
6. Morze J, Wittenbecher C, Schwingshackl L, Danielewicz A, Rynkiewicz A, Hu FB, Guasch-Ferré M. Metabolomics and Type 2 Diabetes Risk: An Updated Systematic Review and Meta-analysis of Prospective Cohort Studies. *Diabetes Care*. Apr 1 2022;45(4):1013-1024. doi:10.2337/dc21-1705
7. Djekic D, Nicoll R, Novo M, Henein M. Metabolomics in atherosclerosis. *IJC Metabolic & Endocrine*. 2015/09/01/ 2015;8:26-30. doi:<https://doi.org/10.1016/j.ijcme.2014.11.004>
8. Turner BE, Steinberg JR, Weeks BT, Rodriguez F, Cullen MR. Race/ethnicity reporting and representation in US clinical trials: A cohort study. *The Lancet Regional Health – Americas*. 2022;11doi:10.1016/j.lana.2022.100252
9. Fatumo S, Chikowore T, Choudhury A, Ayub M, Martin AR, Kuchenbaecker K. A roadmap to increase diversity in genomic studies. *Nature Medicine*. 2022/02/01 2022;28(2):243-250. doi:10.1038/s41591-021-01672-4
10. Patin E, Lopez M, Grollemund R, et al. Dispersals and genetic adaptation of Bantu-speaking populations in Africa and North America. *Science*. 2017/05/05 2017;356(6337):543-546. doi:10.1126/science.aal1988
11. Lippi G, Mattiuzzi C, Cervellin G. Statins popularity: A global picture. *Br J Clin Pharmacol*. Jul 2019;85(7):1614-1615. doi:10.1111/bcp.13944
12. Lin S-y, Baumann K, Zhou C, Zhou W, Cuellar AE, Xue H. Trends in Use and Expenditures for Brand-name Statins After Introduction of Generic Statins in the US, 2002-2018. *JAMA Network Open*. 2021;4(11):e2135371-e2135371. doi:10.1001/jamanetworkopen.2021.35371
13. 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*. Jan 21 2019;40(4):357-368. doi:10.1093/eurheartj/ehy357
14. Sanderson E, Glymour MM, Holmes MV, et al. Mendelian randomization. *Nature Reviews Methods Primers*. 2022/02/10 2022;2(1):6. doi:10.1038/s43586-021-00092-5
15. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet*. Sep 15 2014;23(R1):R89-98. doi:10.1093/hmg/ddu328
16. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife*. May 30 2018;7doi:10.7554/eLife.34408