Article type: Original research article (Cardiovascular and metabolic risk)

The impact of glycated hemoglobin (HbA1c) on cardiovascular disease risk: A Mendelian randomization study using UK Biobank

Shiu Lun Au Yeung,[1*] Shan Luo,[1] C Mary Schooling[1,2]

[1]School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, People's Republic of China.

[2]City University of New York, Graduate School of Public Health and Health Policy, New York, United States of America.

*Corresponding author:

Dr Shiu Lun Au Yeung,

School of Public Health,

Li Ka Shing Faculty of Medicine,

The University of Hong Kong,

7 Sassoon Road, Pokfulam,

Hong Kong SAR, China

Email: ayslryan@hku.hk

Phone: 852-3917 6740

Fax: 852-3520 1945

Word count (Abstract): 245

Word count: 3,210

Tables: 2

Appendix: 8

Running title: HbA1c and CVD using Mendelian randomization

Key word: Glycated hemoglobin; cardiovascular disease; coronary artery disease; stroke; Epidemiology

Abstract

Objective: Glycated hemoglobin **(**HbA1c) is positively associated with cardiovascular disease (CVD) although evidence is primarily observational. Mendelian randomization studies have only examined its relation with subtypes of CVD. We examined the relation of HbA1c with CVD and its subtypes in the UK Biobank using Mendelian randomization.

Research Design and methods: We used 38 genetic variants strongly and independently related to HbA1c (n=123,665) applied to the UK Biobank (n=392,038). We used inverse variance weighting (IVW) to obtain the associations of HbA1c with CVD, coronary artery disease (CAD), stroke and its subtypes. Sensitivity analyses included MR-Egger, a weighted median and exclusion of potentially invalid SNPs. We also applied the same genetic instruments to CARDIoGRAM 1000 Genomes-based genome wide association study (GWAS) (n=184,305) as a validation for CAD.

Results: In the UK Biobank**,** HbA1c was not associated with CVD using IVW (odds ratio (OR): 1.11 per %, 95% confidence interval (CI) 0.83 to 1.48). However, HbA1c was associated with increased CAD risk (OR 1.50 per %, 95% CI 1.08 to 2.11) with directionally consistent results from MR-Egger and weighted median. The positive association with CAD was more pronounced when we excluded potentially invalid SNPs (OR 2.24 per %, 95% CI 1.55 to 3.25). The positive association was replicated in CARDIoGRAM (OR 1.52 per %, 95% CI 1.03 to 2.26). The association of HbA1c with stroke and its subtypes was less clear given low number of cases.

Conclusion: HbA1c likely causes CAD. The underlying mechanisms remain to be elucidated.

Introduction

Observational studies strongly suggest a link between type 2 diabetes and coronary artery disease (CAD) but these observations could be confounded by lack of physical activity and obesity (1-4). Randomized controlled trials such as the ACCORD trial unexpected showed intensive glycemic control did not substantially reduce the risk of cardiovascular disease (CVD) events and may even have increased overall mortality (5), which remained evident after prolonged follow up (6, 7). However, these results have not been consistently seen in all relevant trials, such as the Steno-2 Study (8). Differences in treatment regimen and sample may have contributed to these discrepancies (9). Other relevant trials such as the LEADER trial and the EMPA-REG OUTCOME trial also suggested potential beneficial effects of liraglutide and empagliflozin in reducing CVD (10, 11). Although randomized controlled trials are less vulnerable to confounding than observational studies, it is uncertain whether results from these trials, which are primarily in people with diabetes, generalize to the general population (12). Moreover, interventions in randomized controlled trials may have off-target effects. Notably some of the agents used to achieve glycemic control, such as sulphonylureas have been implicated in CVD (13).

Mendelian randomization studies, which are less prone biases particularly confounding, through use of genetic variants randomly allocated during conception, suggest a positive causal relation of dysglycaemia and diabetes with CAD (14-16). However, the number of genetic instruments used for glycated hemoglobin (HbA1c) (n=10), which represents overall blood sugar level over the previous 2-3 months, was relatively limited making assessment of potential violations of the underlying assumptions less reliable. It is also uncertain if HbA1c has an overall impact on CVD

given the heterogeneity of the phenotype and the observational nature of the evidence (17). Trials suggest glycemic traits may have different effects on different CVD subtypes, such as stroke (9), which is a major contributor to the disease burden in many regions including Asia (18), which also have high rates of diabetes (19). To date, only one Mendelian randomization study has considered the relation of glycemic traits with stroke where they found fasting glucose potentially related to large artery stroke but not to other ischemic stroke subtypes (20). However, the study did not examine the effect on overall stroke or hemorrhagic stroke.

To address these research gaps, we implemented a two sample Mendelian randomization study to assess the relation of HbA1c with CVD and its subtypes, using genetic predictors of HbA1c from the most up-to-date GWAS of HbA1c in MAGIC (21) applied to the UK Biobank (22, 23), one of the largest population based cohorts globally with extensive phenotyping and genotyping. We also verified the association of HbA1c with CAD using the largest most extensively genotyped CAD case-control study independent of the UK Biobank, i.e., CARDIoGRAMplusC4D 1000 Genomes-based GWAS (24).

Method

This is a two sample Mendelian randomization study. We obtained genetic associations with HbA1c from MAGIC, and with CVD, and its sub-types, from the UK Biobank and CARDIoGRAMplusC4D 1000 Genomes-based GWAS.

Assumptions of Mendelian randomization

Mendelian randomization relies on 3 stringent assumptions (25). Firstly, the genetic instruments are strongly predictive of HbA1c. Secondly, the association of genetic instruments with CVD is not confounded. Lastly, the effect of the genetic instrument on CVD should be fully mediated via HbA1c (i.e. the exclusion restriction assumption).

Participants

MAGIC – genetic predictors of HbA1c

MAGIC includes a meta-analysis of GWAS of HbA1c in 159,940 adults without diabetes, including 123,665 participants of European ancestry, with imputation using the Phase 2 of the International HapMap Project reference panel (21). The mean age of the majority of the studies was over 50 years. HbA1c, National Glychohemoglobin Standardization Program (NGSP) percent, was adjusted for age, sex, study specific covariates, and genomic control. To reduce confounding by population stratification, we only selected single nucleotide polymorphisms (SNPs) reaching genome wide significance ($p < 5x10^{-8}$) in participants of European descent, which gave 43 SNPs. After removing 4 SNPs (rs11154792; rs3824065; rs10823343; and rs2408955) in linkage disequilibrium with the other SNPs ($R^2 \geq 0.05$), 39 SNPs were retained.

Genetic predictors of cardiovascular disease

UK Biobank

The UK Biobank is one of the largest Biobanks globally. It recruited 500,000+ participants (aged 40-69 years) in the United Kingdom from 2006 to 2010. Participants completed a questionnaire and physical assessment. Biochemical assays, genotyping and longitudinal follow up via record linkage to medical and mortality records are ongoing, as described in detail elsewhere (22, 23). Prevalent disease was coded using International Classification of Diseases (ICD) 9 and 10, cause of death was coded using ICD 10. Genotyping was performed using two very similar arrays, including Affymetrix UK BiLEVE Axiom array (~50,000 participants) and Affymetrix UK Biobank Axiom array (~450,000 participants). The SNPs included in this study were either directly genotyped or imputed using Haplotype Reference Consortium (HRC) panel. We restricted our analysis to people of genetically verified white British descent to reduce confounding by population stratification, as in a previous similar study (26). We also excluded participants who were extensively related (more than 10 putative third-degree relatives in the kinship table), who had poor quality genotyping (i.e. missing rate ≥1.5%), who had sex chromosome aneuploidy, or whose self-reported and genetic sex did not match. The mean age of the participants was 56.9 years.

CARDIoGRAMplusC4D 1000 Genomes-based genome wide association studies (GWAS)

CARDIoGRAMplusC4D 1000 Genomes-based GWAS is a meta-analysis of GWAS of CAD - case (n=60,801) -control (n=123,504) studies of people of mainly European descent (77%), with imputation using the 1000 Genomes phase 1 v3 reference panel (24). CAD was defined in various ways, such as diagnosis of myocardial infarction, acute coronary syndrome, chronic stable angina, or coronary stenosis >50%. Diagnoses were based on clinical diagnosis,

procedures (coronary angiography results or by-pass surgery), use of medications or symptoms that indicate angina, or self-report of a doctor diagnosis, as described elsewhere (24). CARDIoGRAMplusC4D 1000 Genomes-based GWAS adjusted for study-specific covariates (e.g. age and sex) and genomic control.

Exposure

The exposure was genetically predicted HbA1c (%).

Outcomes

The primary outcomes were prevalent CVD (defined as ICD9 401-459.9, ICD10 I10-I99) and its subtypes, including coronary artery disease (CAD: ICD9 410-414.9 and ICD10 I20-I25.9, ischemic stroke: ICD9 434 and 436 and ICD10 I63-I64, and hemorrhagic stroke: ICD9 430-431 and ICD10 I60-I61) based on self-reports and hospital episodes, and death from CVD (ICD10 I10-I99) or its subtypes (CAD, ICD10 I20-I25.9, ischemic stroke ICD10 I63-I64, and hemorrhagic stroke ICD10 I60-I61) from death records, following the recommended definitions of the UK Biobank Stroke Outcomes Group (27). For completeness, we also considered other CVD, i.e., all CVD excluding stroke and CAD which was mainly hypertensive diseases, and other stroke, i.e., all stroke excluding ischemic and hemorrhagic stroke. CVD, CAD and stroke mortality, based on primary cause of death, were also considered separately as secondary outcomes.

Statistical analysis

We assessed departure from Hardy Weinberg equilibrium for each SNP using chi-square tests with Bonferroni correction to correct for multiple comparison (0.05/39 = 0.00128). We used analysis of variance (continuous) and chi square test (categorical) to examine whether the genetic variants were associated with factors potentially confounding the association of HbA1c with CVD, including Townsend deprivation index, education, age, body mass index, smoking, and alcohol drinking, in the UK Biobank, with Bonferroni correction to correct for multiple comparisons (0.0002, based on 0.05/234 derived from 39 SNPs x 6 traits). We obtained the association of each SNP with CVD and its subtypes using multivariable logistic regression in the UK Biobank, adjusted for age, sex, genotyping array, and 10 principal components.

We conducted our main analysis using inverse variance weighting (IVW) with multiplicative random effects, which is a weighted regression of gene-outcome associations on gene-exposure associations for UK Biobank and CARDIoGRAMplusC4D 1000 Genomes based GWAS. Given IVW assumes no horizontal pleiotropy, which cannot be empirically assessed, we used the I^2 of the Wald estimates (SNP-outcome association divided by SNP-exposure association) to indicate the presence of invalid instruments. In the presence of invalid SNPs (i.e., SNPs that have effect on the outcome not via HbA1c), IVW will be invalid. As such, we also conducted several sensitivity analyses to assess the robustness of our results to potential violations of the Mendelian randomization assumptions since these analyses have different assumptions for validity, as described below. Although these approaches may have different statistical power (e.g. wider confidence intervals for MR-Egger), the rationale is that if these approaches give similar

conclusion regarding the association of HbA1c with the outcomes, then we are more confident in inferring that the positive findings are unlikely driven by violation of the MR assumptions (28).

Instrument strength

To assess instrument strength, we computed the F statistic for the association of genetic instruments with HbA1c, assuming the genetic variants included explained at least 4.2% of the HbA1c variance (the lower bound of variance explained based on the previous GWAS) (21). A higher F statistics indicate a stronger instrument.

Sensitivity analysis

1) MR-Egger

We conducted MR-Egger regression, which produces valid estimates even if all the genetic instruments are invalid, as long as the InSIDE (Instrument Strength Independent of Direct Effect) assumption holds (29). We also presented the intercept p-value from the MR-Egger regression, because a significant intercept indicates the IVW estimate may be invalid due to horizontal pleiotropy.

2) Weighted median

We also used a weighted median which produces valid estimates as long as more than 50% of the information is derived from valid SNPs (30).

3) Exclusion of potentially invalid SNPs

We identified potentially invalid genetic instruments (SNPs) in two ways. Firstly, we excluded SNPs related to potential causes of CAD based on the GWAS Catalog/ PhennoScanner and SNPs associated with potential confounders in the UK Biobank: (Set 1). Secondly, we additionally excluded SNPs defined as "erythrocytic" in the original GWAS because they did not predict diabetes and hence may be irrelevant to glycemic exposure (i.e. invalid) (21) (Set 2). Appendix 1 summarizes the choice of SNPs in different sensitivity analyses.

To rule out the possibility of false positives due to inclusion of related individuals, we also repeated the analyses including only unrelated participants.

Ethics approval

UK Biobank received ethics approval from the National Health Service National Research Ethics Service and participants provided written informed consent. No ethics approval was required for the analysis using publicly available data (CARDIoGRAMplusC4D 1000 Genomes-based GWAS).

All analyses were performed using R Version 3.3.2 (R Development Core Team, Vienna, Austria) with the R package (TwosampleMR).

Results

Among 502,642 participants in the UK Biobank, 442,698 (88%) were British White. After excluding those who have permanently withdrawn, with poor quality or missing genotype, with a mismatch between self-reported and genetic sex or ancestry, had sex chromosome aneuploidy, or extensive relatedness, 392,038 participants remained for subsequent analyses. Among these 392,038 people, 158,601 had prevalent CVD, 29,293 prevalent CAD, 9,042 prevalent stroke (3,707 ischemic and 1,655 hemorrhagic) with some participants having more than one condition according to the data available in April, 2018. Since the baseline recruitment in March, 2006, there were 2,313 CVD deaths, including 1,294 CAD and 356 stroke where the latest date of death was 16th February, 2016.

Of the 39 SNPs for HbA1c, one SNP (rs1800562, p value=0.0006) violated Hardy Weinberg equilibrium (Appendix 2) and hence was discarded. Appendix 3 shows the 38 SNPs used as genetic instruments. The F statistic for the association of the 38 SNPs on HbA1c was 142, suggesting little weak instrument bias. A few SNPs were associated with potential confounders, 2 SNPs with education (rs9818758, rs11964178), 5 SNPs with body mass index (rs8192675, rs7756992, rs17747324, rs10774625, rs1558902) and 2 SNPs with smoking (rs10774625, rs17509001) after Bonferroni correction (Appendix 2). According to GWAS Catalog or PhennoScanner, 13 SNPs were related to potential causes of CAD. Based on the information, we repeated the analyses using the two exclusion criteria for choice of SNPs. First, we excluded 15 SNPs related to potential confounders or causes of CAD, leaving 23 SNPs (Set 1). Secondly, we

additionally excluded SNPs defined as “erythrocytic” in the original GWAS (9 SNPs) amongst the 23 SNPs because they did not predict diabetes and hence may be irrelevant to glycemic exposure (i.e. invalid) (21), leaving 14 SNPs (Set 2).

Table 1 shows HbA1c was not clearly associated with CVD using all 38 SNPs (odds ratio (OR) 1.11 per %, 95% confidence interval (CI) 0.83 to 1.48). However, higher HbA1c was associated with higher CAD risk using inverse variance weighting using all 38 SNPs (OR 1.50 per %, 95% CI 1.08 to 2.11), with directionally consistent results from MR-Egger and weighted median. After excluding potentially pleiotropic SNPs or those related to confounders (23 SNPs: set 1), the positive associations remained for CAD in IVW (OR 1.47 per %, 95% CI 1.10 to 1.97), with directionally consistent results from the weighted median method. The results for CAD were most consistent across IVW, MR-Egger and weighted median when we further restricted SNPs which were non erthrocytic (set 2). The association of HbA1c with stroke and its subtypes appeared heterogeneous although these estimates had wide confidence intervals. MR-Egger intercepts suggested little evidence of directional pleiotropy in all analyses. Heterogeneity in the Wald estimates decreased after removing potentially invalid SNPs.

Table 2 shows HbA1c was positively associated with CAD in CARDIoGRAMplusC4D 1000 Genomes-based GWAS using inverse variance weighting (OR 1.52 per %, 95% CI 1.03 to 2.26), with directionally consistent estimates from sensitivity analyses including MR-Egger (OR 1.64 per %, 95% CI 0.73 to 3.71) and the weighted median (OR 1.50 per %, 95% CI 1.09 to 2.05). There was little evidence of directional pleiotropy based on the MR-Egger intercept (-0.002, p-value 0.83). Similar to the results from UK Biobank, the estimates all similarly suggested

detrimental effects of HbA1c on CAD regardless of the SNP selection. Heterogeneity in the Wald estimates decreased after removing potentially invalid SNPs.

The associations of HbA1c with CVD, CAD and stroke mortality were less clear (Appendix 4) with wide CIs, most likely due to the low mortality rate in the UK Biobank.

Similar conclusions were drawn when we restricted our analyses to unrelated participants (Appendix 5).

Discussion

To our knowledge, this is the first Mendelian randomization study using the most recently published GWAS of HbA1c applied to both the UK Biobank and CARDIoGRAMplusC4D 1000 Genomes-based GWAS, encompassing more than 700,000 participants. HbA1c was positively associated with CAD in UK Biobank which replicated in CARDIoGRAMplusC4D 1000 Genomes based GWAS, consistent with previous observational studies and an earlier Mendelian randomization study (1-3, 15). We cannot exclude HbA1c being associated with CVD. Our study is suggestive of different effects of HbA1c on stroke subtypes although the number of events in UK Biobank was not enough to allow precise estimates and should be examined further in large stroke consortiums.

Previous observational studies have consistently reported a positive association of HbA1c with CAD although they are susceptible to confounding (1-3). Randomized controlled trials targetting HbA1c reduction are difficult to interpret given the interventions, primarily on lifestyle, may have multiple effects which do not necessarily only reflect the impact of HbA1c on CAD (31, 32). Our study adds by showing that higher HbA1c is positively associated with CAD using a Mendelian randomization study in two different large studies, as well as using different analytics and SNP selections. Combining the results obtained from UK Biobank and CARDIoGRAMplusC4D did not change the conclusion (Appendix 6). Considering triangulation of the evidence from different designs with different underlying assumptions, HbA1c may be causal in the development of CAD in the general population (28). The exact mechanistic pathways remain to be elucidated.

Our study does not provide strong evidence for the same magnitude of association of HbA1c with CVD, contrary to previous observational studies (17). This discrepancy could indicate potential confounding or selection bias in observational studies. Potentially different associations of HbA1c with CAD and other CVD subtypes is consistent with the argument that CVD subtypes have different etiologies with different contributions of each factor (33). UK Biobank had a low response rate (~5%) at recruitment, although a low response rate at recruitment does not necessarily invalidate causal inference (34). This is evident from the similar estimates obtained from both the UK Biobank and CARDIoGRAMplusC4D, which used different study designs and sampling approaches. Recruitment of generally healthier people into the UK Biobank study would also not explain the different findings for CAD and CVD. Alternatively, given the UK Biobank recruited from age 40 years to 69 years with average age of 57 years, a

different pattern of death by age from specific types of CVD related to HbA1c genetics would artifactually generate different associations of HbA1c with CVD by subtype (35), because of varying levels of left truncation from the underlying birth cohort.

Although we included more than 700,000 participants in this study and used Mendelian randomization to reduce confounding, some limitations exist. First, the validity of Mendelian randomization depends on whether the three underlying assumptions, as described in the Methods, are satisfied, i.e., the instruments predict the exposure, the instruments are not confounded and the instruments affect the outcome only via the exposure (25). In our study, we used genetic variants predicting HbA1c identified in GWAS of people of European descent to reduce weak instrument bias (indicated by the F statistics). Restricting the samples to adults mostly of European descent reduces the likelihood of confounding by population stratification. We also assessed the associations of the genetic variants with potential confounders and found little association with most confounders (Appendix 2), which would not have been possible using summary statistics from GWAS. Although we could not assess whether the genetic instruments were associated with the outcomes only via their association with HbA1c (exclusion-restriction assumption), we conducted several sensitivity analyses, such as MR-Egger and a weighted median, which have different assumptions for validity although MR-Egger has reduced statistical power. We also repeated the analyses excluding potentially pleiotropic SNPs which may violate the exclusion-restriction assumption (36). Given the consistent results for HbA1c on CAD for these different approaches with different assumptions, the association of HbA1c on CAD is likely to be causal. Repeating the analyses for CAD by sex, as a check, showed similar patterns (Appendix 7). When we repeated the CAD analyses without self-reports, the results were most

consistent excluding erthrocytic SNPs (data not shown). We have also repeated the analyses restricted to erythrocytic SNPs and found a less clear relation of HbA1c with CAD. This is expected since these SNPs did not predict diabetes and hence these SNPs are likely irrelevant to the glycemic exposure (Appendix 8) (21). Although we used one of the largest possible studies, the relatively low number of stroke cases led to imprecise estimates. The suggestive differences in the relation of HbA1c and stroke subtypes seen in our study should be examined elsewhere using large GWAS consortium or settings where stroke is more prevalent such as China (37). Lastly, we were unable to use an allele score approach which may increase statistical power as HbA1c was not available from the UK Biobank at the time this study was conducted.

Our study provides more evidence of a causal role of HbA1c in CAD. Interventions that target HbA1c reduction may be potential targets for reducing the global burden of CAD. Future studies should also clarify the impact of HbA1c on CVD subtypes, which may provide additional insight into the global distribution of CVD subtypes, such as stroke which is more prevalent in Asians.

Acknowledgement

All three authors (SLAY, SL and CMS) designed the study, wrote the analysis plan and interpreted the results. SLAY undertook analyses with feedback from SL and CMS. SLAY wrote the first draft of the manuscript with critical feedback and revisions from SL and CMS. All authors gave final approval of the version to be published. Data on coronary artery disease have been contributed by CARDIoGRAMplusC4D investigators and have been downloaded from www.CARDIOGRAMPLUSC4D.ORG. This research has been conducted using the UK Biobank Resource (Application number: 14864). Dr SL Lin (The University of Hong Kong)

provided valuable advice at the planning stage of this study. The work presented here is the responsibility of the authors and not anyone acknowledged. SLAY is the guarantor of this study. The authors have no conflict of interest, financial or otherwise.

Funding source

This study was funded by the Small Project Funding, the University of Hong Kong (#201409176231). The funder had no role in the design, analyses, interpretation of results or writing of the paper.

References

1. Pai JK, Cahill LE, Hu FB, et al. Hemoglobin a1c is associated with increased risk of incident coronary heart disease among apparently healthy, nondiabetic men and women. Journal of the American Heart Association 2013;**2**:e000077
2. Selvin E, Coresh J, Golden SH, et al. Glycemic control and coronary heart disease risk in persons with and without diabetes: the atherosclerosis risk in communities study. Arch Intern Med 2005;**165**:1910-1916
3. Selvin E, Steffes MW, Zhu H, et al. Glycated hemoglobin, diabetes, and cardiovascular risk in nondiabetic adults. N Engl J Med 2010;**362**:800-811
4. Sarwar N, Aspelund T, Eiriksdottir G, et al. Markers of dysglycaemia and risk of coronary heart disease in people without diabetes: Reykjavik prospective study and systematic review. PLoS Med 2010;**7**:e1000278
5. Action to Control Cardiovascular Risk in Diabetes Study Group, Gerstein HC, Miller ME, et al. Effects of intensive glucose lowering in type 2 diabetes. N Engl J Med 2008;**358**:2545-2559
6. Action to Control Cardiovascular Risk in Diabetes Study Group. Nine-Year Effects of 3.7 Years of Intensive Glycemic Control on Cardiovascular Outcomes. Diabetes Care 2016;**39**:701-708
7. Action to Control Cardiovascular Risk in Diabetes Study Group, Gerstein HC, Miller ME, et al. Long-term effects of intensive glucose lowering on cardiovascular outcomes. N Engl J Med 2011;**364**:818-828
8. Gaede P, Lund-Andersen H, Parving HH, et al. Effect of a multifactorial intervention on mortality in type 2 diabetes. N Engl J Med 2008;**358**:580-591
9. Fang HJ, Zhou YH, Tian YJ, et al. Effects of intensive glucose lowering in treatment of type 2 diabetes mellitus on cardiovascular outcomes: A meta-analysis of data from 58,160 patients in 13 randomized controlled trials. Int J Cardiol 2016;**218**:50-58
10. Marso SP, Daniels GH, Brown-Frandsen K, et al. Liraglutide and Cardiovascular Outcomes in Type 2 Diabetes. N Engl J Med 2016;**375**:311-322
11. Zinman B, Wanner C, Lachin JM, et al. Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes. N Engl J Med 2015;**373**:2117-2128
12. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". Lancet 2005;**365**:82-93
13. Azoulay L, Suissa S. Sulfonylureas and the Risks of Cardiovascular Events and Death: A Methodological Meta-Regression Analysis of the Observational Studies. Diabetes Care 2017;**40**:706-714
14. Ahmad OS, Morris JA, Mujammami M, et al. A Mendelian randomization study of the effect of type-2 diabetes on coronary heart disease. Nature communications 2015;**6**:7060
15. Ross S, Gerstein HC, Eikelboom J, et al. Mendelian randomization analysis supports the causal role of dysglycaemia and diabetes in the risk of coronary artery disease. Eur Heart J 2015;**36**:1454-1462
16. Merino J, Leong A, Posner DC, et al. Genetically Driven Hyperglycemia Increases Risk of Coronary Artery Disease Separately From Type 2 Diabetes. Diabetes Care 2017;**40**:687-693
17. Huang Y, Cai X, Mai W, et al. Association between prediabetes and risk of cardiovascular disease and all cause mortality: systematic review and meta-analysis. BMJ 2016;**355**:i5953
18. Feigin VL, Roth GA, Naghavi M, et al. Global burden of stroke and risk factors in 188 countries, during 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet Neurol 2016;**15**:913-924

19. Ramachandran A, Ma RC, Snehalatha C. Diabetes in Asia. Lancet 2010;**375**:408-418
20. Larsson SC, Scott RA, Traylor M, et al. Type 2 diabetes, glucose, insulin, BMI, and ischemic stroke subtypes: Mendelian randomization study. Neurology 2017;**89**:454-460
21. Wheeler E, Leong A, Liu CT, et al. Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. PLoS Med 2017;**14**:e1002383
22. 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
23. Collins R. What makes UK Biobank special? Lancet 2012;**379**:1173-1174
24. Nikpay M, Goel A, Won HH, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nature genetics 2015;**47**:1121-1130
25. Lawlor DA, Harbord RM, Sterne JAC, et al. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. Stat Med 2008;**27**:1133-1163
26. Tyrrell J, Jones SE, Beaumont R, et al. Height, body mass index, and socioeconomic status: mendelian randomisation study in UK Biobank. BMJ 2016;**352**:i582
27. Woodfield R, Grant I, Group UKBSO, et al. Accuracy of Electronic Health Record Data for Identifying Stroke Cases in Large-Scale Epidemiological Studies: A Systematic Review from the UK Biobank Stroke Outcomes Group. PLoS One 2015;**10**:e0140533
28. Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. Int J Epidemiol 2016;**45**:1866-1886
29. 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**:512-525
30. Bowden J, Davey Smith G, Haycock PC, et al. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet Epidemiol 2016;**40**:304-314
31. Wing RR, Lang W, Wadden TA, et al. Benefits of modest weight loss in improving cardiovascular risk factors in overweight and obese individuals with type 2 diabetes. Diabetes Care 2011;**34**:1481-1486
32. Lindstrom J, Ilanne-Parikka P, Peltonen M, et al. Sustained reduction in the incidence of type 2 diabetes by lifestyle intervention: follow-up of the Finnish Diabetes Prevention Study. Lancet 2006;**368**:1673-1679
33. Joseph P, Leong D, McKee M, et al. Reducing the Global Burden of Cardiovascular Disease, Part 1: The Epidemiology and Risk Factors. Circ Res 2017;**121**:677-694
34. Rothman KJ, Gallacher JE, Hatch EE. Why representativeness should be avoided. Int J Epidemiol 2013;**42**:1012-1014
35. Schooling CM, Au Yeung SL. "Selection Bias by Death" and Other Ways Collider Bias May Cause the Obesity Paradox. Epidemiology 2017;**28**:e16-e17
36. Bowden J, Del Greco MF, Minelli C, et al. A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. Stat Med 2017;**36**:1783-1802
37. Chen Z, Chen J, Collins R, et al. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. Int J Epidemiol 2011;**40**:1652-1666

Table 1: Association of HbA1c (%) with cardiovascular disease (CVD), coronary artery disease (CAD), stroke and its subtypes using Mendelian randomization in the UK Biobank

		Inverse variance weighting with multiplicative random effects			MR-Egger				Weighted median	
Instrument	*Outcome	Odds ratio	95% CI	I^2 of Wald estimates	Odds ratio	95% CI	Intercept	p value for intercept	Odds ratio	95% CI
All SNPs (38)	CVD	1.11	0.83 to 1.48	0.89	1.12	0.64 to 1.96	0.000	0.95	1.25	1.07 to 1.47
	CAD	1.50	1.08 to 2.11	0.71	1.25	0.66 to 2.36	0.004	0.50	1.41	1.02 to 1.94
	Stroke	1.39	0.93 to 2.07	0.41	0.95	0.45 to 2.01	0.008	0.25	1.23	0.77 to 1.95
	Ischemic	1.04	0.62 to 1.75	0.17	0.49	0.19 to 1.26	0.017	0.07	0.65	0.33 to 1.27
	Hemorrhagic	1.27	0.58 to 2.81	0.20	1.25	0.28 to 5.68	0.000	0.98	0.82	0.26 to 2.59
	Others	1.84	0.99 to 3.39	0.41	1.54	0.47 to 5.03	0.004	0.73	1.42	0.65 to 3.11
	Other CVD	0.97	0.75 to 1.26	0.85	1.07	0.65 to 1.75	-0.002	0.65	1.08	0.91 to 1.30
†Excluded SNPs if associated with potential causes of CAD or confounders (23)	CVD	1.20	1.02 to 1.42	0.57	1.10	0.81 to 1.50	0.002	0.50	1.26	1.07 to 1.47
	CAD	1.47	1.10 to 1.97	0.53	0.97	0.59 to 1.60	0.011	0.07	1.27	0.89 to 1.81
	Stroke	1.47	1.04 to 2.07	0.00	0.81	0.43 to 1.52	0.016	0.04	1.22	0.76 to 1.95
	Ischemic	0.90	0.53 to 1.53	0.00	0.53	0.20 to 1.40	0.014	0.22	0.64	0.30 to 1.39
	Hemorrhagic	1.69	0.66 to 4.35	0.30	0.80	0.14 to 4.47	0.020	0.32	0.86	0.27 to 2.75
	Others	2.14	1.20 to 3.81	0.17	1.25	0.43 to 3.60	0.014	0.25	1.42	0.66 to 3.07
	Other CVD	1.06	0.90 to 1.25	0.53	1.15	0.84 to 1.56	-0.002	0.56	1.11	0.93 to 1.32
‡Excluded SNPs if classified as erthrocytic, associated with potential causes of CAD or confounders (14)	CVD	1.27	1.03 to 1.56	0.43	0.82	0.51 to 1.31	0.009	0.07	1.20	0.96 to 1.51
	CAD	2.24	1.55 to 3.25	0.39	2.10	0.80 to 5.50	0.001	0.89	1.90	1.25 to 2.88
	Stroke	1.61	0.98 to 2.65	0.00	0.88	0.25 to 3.03	0.013	0.31	1.40	0.71 to 2.74
	Ischemic	1.14	0.53 to 2.47	0.00	0.45	0.07 to 3.02	0.020	0.31	0.97	0.35 to 2.73
	Hemorrhagic	1.26	0.40 to 3.93	0.00	1.89	0.10 to 35.05	-0.009	0.77	0.72	0.16 to 3.36
	Others	2.32	0.96 to 5.59	0.25	1.14	0.12 to 10.76	0.015	0.51	1.20	0.40 to 3.61
	Other CVD	0.97	0.82 to 1.14	0.00	0.65	0.43 to 0.98	0.008	0.06	0.95	0.76 to 1.18

*Definitions of disease as below: Prevalent cardiovascular disease (CVD, defined as ICD9 401-4599, ICD10 I10-I99) and its subtype, including coronary artery disease (CAD, defined as ICD9 410-4149, ICD10 I20-I25.9); stroke (ICD9 430; 431; 434; 436, ICD10 I60, I61, I63, I64), ischemic stroke (defined as ICD9 434, 436, ICD10 I63-I64), hemorrhagic stroke (defined as ICD9 430, 431, ICD10 I60-I61). Included both cases and mortality.

† Set 1: SNP excluded if associated with potential causes of CAD or confounders based on public datasources (Phenoscanner and GWAS Catalog) and UK Biobank

‡ Set 2: SNP excluded if classified as erthrocytic based on the HbA1c GWAS or associated with potential causes of CAD or confounders based on public datasources (Phenoscanner and GWAS Catalog) and UK Biobank

Table 2: Association of HbA1c (%) with coronary artery disease (CAD) using Mendelian randomization in CARDIoGRAMplusC4D 1000 Genomes-based GWAS

		Inverse varaince weighting with multiplicative random effects			MR-Egger				Weighted median	
Instrument	Outcome	Odds ratio	95% CI	I^2 of Wald estimates	Odds ratio	95% CI	Intercept	p value for intercept	Odds ratio	95% CI
All 38 SNPs	CAD	1.52	1.03 to 2.26	0.74	1.64	0.73 to 3.71	-0.002	0.83	1.50	1.09 to 2.05
*23 SNPs (Set 1)	CAD	1.30	0.98 to 1.73	0.36	2.05	1.21 to 3.48	-0.011	0.06	1.42	1.02 to 1.95
†14 SNPs (Set 2)	CAD	1.27	0.90 to 1.78	0.14	1.79	0.75 to 4.27	-0.007	0.41	1.36	0.90 to 2.06

* Set 1: SNP excluded if associated with potential causes of CAD or confounders based on public datasources (GWAS Catalog and PhenoScanner) and UK Biobank

† Set 2: SNP excluded if classified as erthrocytic based on the HbA1c GWAS or associated with potential causes of CAD or confounders based on public datasources (GWAS Catalog and PhenoScanner) and UK Biobank

Online-only supplemental material

Appendix 1: Classification of SNP function based on HbA1c GWAS, GWAS-Catalog, and PhenoScanner (1-3)

SNP	Nearby Gene	Classification according to HbA1c GWAS	From Phenoscanner and GWAS Catalog	Association with confounders in UK Biobank	Included in *Set 1 (23)	Included in †Set 2 (14)
rs1046896	*FN3KRP*	Unclassified			X	X
rs10774625	*ATXN2*	Erythrocytic	CAD, cholesterol, blood pressure	Smoking, body mass index		
rs10830963	*MTNR1B*	Glycemic			X	X
rs11248914	*ITFG3*	Erythrocytic			X	
rs11558471	*MYB*	Glycemic			X	X
rs11603334	*ARAP1*	Glycemic	Body mass index			
rs11708067	*MYO9B*	Glycemic	Birthweight			
rs11964178	*SOX30*	Erythrocytic		Education		
rs12621844	*FOXN2*	Unclassified			X	X
rs13134327	*FREM3*	Glycemic			X	X
rs1558902	*FTO*	Unclassified	Obesity related traits	Body mass index		
rs17509001	*ATAD2B*	Unclassified	Height	Smoking		
rs17533903	*MYO9B*	Erythrocytic			X	
rs17747324	*TCF7L2*	Glycemic	Body mass index	Body mass index		
rs198846	*HFE*	Erythrocytic	Blood pressure			
rs2110073	*PHB2*	Unclassified			X	X
rs2383208	*MTAP*	Glycemic			X	X
rs267738	*CERS2*	Unclassified			X	X
rs282587	*ATP11A*	Unclassified			X	X
rs3782123	*BET1L*	Unclassified			X	X
rs4607517	*GCK*	Glycemic			X	X
rs4737009	*ANK1*	Erythrocytic			X	
rs4745982	*HK1*	Erythrocytic			X	
rs4820268	*TMPRSS6*	Erythrocytic	Iron status			
rs560887	*G6PC2*	Glycemic			X	X
rs579459	*ABO*	Glycemic	CAD, cholesterol			
rs592423	*CITED2*	Erythrocytic	Triglycerides			
rs6474359	*ANK1*	Unclassified			X	X
rs6980507	*SLC20A2*	Erythrocytic			X	
rs7040409	*C9orf47*	Erythrocytic			X	
rs7616006	*SYN2*	Erythrocytic	Cholesterol			
rs7756992	*CDKAL1*	Glycemic	Birthweight, body mass index	Body mass index		
rs8192675	*SLC2A2*	Glycemic		Body mass index		
rs837763	*CDT1*	Erythrocytic			X	
rs857691	*SPTA1*	Erythrocytic			X	
rs9604573	*GAS6*	Unclassified			X	X
rs9818758	*USP4*	Unclassified	Education attainment	Education		
rs9914988	*ERAL1*	Erythrocytic			X	

*Set 1: SNP excluded if associated with potential causes of CAD or confounders based on public datasources (GWAS Catalog and PhenoScanner) and UK Biobank

†Set 2: SNP excluded if classified as erthrocytic based on the HbA1c GWAS or associated with potential causes of CAD or confounders based on public datasources (GWAS Catalog and PhenoScanner) and UK Biobank

Appendix 2: Hardy Weinberg equilibirum for 39 single nucleotide polymorphisms (SNPs) related to Hba1c and with sociodemographic and lifestyle factors in the UK Biobank

	P value						
SNP	HWE	Townsend	Age	BMI	Education	Smoking	Drinking
rs267738	0.630214	0.31202	0.425771	0.79411	0.069326	0.001038	0.768753
rs857691	0.043467	0.96335	0.536783	0.464286	0.951865	0.912496	0.46033
rs17509001	0.652415	0.282543	0.002802	0.256924	0.239933	**0.000145**	0.113411
rs12621844	0.38296	0.441702	0.996839	0.18399	0.659238	0.292595	0.677776
rs560887	0.043559	0.38083	0.693665	0.113303	0.049333	0.003194	0.939513
rs7616006	0.513255	0.847246	0.031568	0.778987	0.117231	0.899612	0.697842
rs9818758	0.404406	0.953251	0.153604	0.005425	**7.99E-10**	0.393539	0.948646
rs11708067	0.272175	0.013103	0.015906	0.00283	0.298841	0.564044	0.836277
rs8192675	0.139404	0.12566	0.383311	**1.31E-13**	0.000942	0.328327	0.092219
rs13134327	0.607805	0.013033	0.794653	0.522429	0.704403	0.996797	0.655605
rs7756992	0.2240	0.381421	0.004907	**0.00012**	0.701401	0.847097	0.544422
rs1800562	**0.000632**	**6.09E-05**	0.012101	0.136382	0.044316	0.413812	0.320652
rs198846	0.002917	0.865971	0.990811	0.186999	0.465576	0.462361	0.217603
rs11964178	0.851602	0.050435	0.937652	0.083276	**1.19E-08**	0.50254	0.527845
rs592423	0.755993	0.115987	0.220421	0.235859	0.558662	0.104069	0.588431
rs4607517	0.865322	0.230002	0.432958	0.239964	0.095068	0.464681	0.601145
rs6474359	0.402333	0.248531	0.497691	0.363764	0.652066	0.100627	0.128928
rs4737009	0.51921	0.780878	0.737799	0.465033	0.271659	0.218205	0.099371
rs6980507	0.052221	0.298746	0.174405	0.344741	0.001152	0.872464	0.276444
rs11558471	0.168743	0.624826	0.46678	0.047747	0.693509	0.265594	0.575957
rs2383208	0.840041	0.780422	0.545134	0.013188	0.801864	0.732715	0.852406
rs7040409	0.691662	0.074551	0.522984	0.216373	0.844719	0.054867	0.459586
rs579459	0.594162	0.01023	0.26882	0.230016	0.288492	0.098067	0.609528
rs4745982	0.87608	0.347514	0.14868	0.428483	0.419484	0.214274	0.978838
rs17747324	0.139331	0.803605	0.771908	**3.76E-09**	0.294054	0.988871	0.083984
rs3782123	0.379812	0.33026	0.947598	0.850885	0.134383	0.471755	0.132621
rs11603334	0.495775	0.019176	0.174708	0.016002	0.093253	0.471856	0.852732
rs10830963	0.253121	0.905395	0.216547	0.004262	0.415687	0.425279	0.214795
rs2110073	0.552226	0.101617	0.872696	0.545767	0.319253	0.849321	0.640847
rs10774625	0.775776	0.003455	0.008607	**5.32E-05**	0.173611	**2.11E-08**	0.01611
rs282587	0.27506	0.156248	0.159766	0.508878	0.736708	0.718929	0.392989
rs9604573	0.400219	0.912911	0.535763	0.202247	0.873724	0.300071	0.541628
rs11248914	0.944857	0.432639	0.870188	0.093394	0.800812	0.513249	0.253822
rs1558902	0.917891	0.017082	0.142933	**1.65E-243**	0.026188	0.012061	0.221513
rs837763	0.851763	0.215591	0.042965	0.043691	0.986355	0.556154	0.941624
rs9914988	0.130211	0.262572	0.407259	0.045498	0.384674	0.003626	0.985077
rs1046896	0.201091	0.151539	0.720657	0.11935	0.023758	0.147191	0.544566
rs17533903	0.157381	0.796032	0.475204	0.723332	0.77197	0.552368	0.297122
rs4820268	0.46793	0.017737	0.486565	0.839579	0.281602	0.46777	0.090102

* Bonforroni corrected p value for Hardy Weinberg equilibrium (HWE): 0.00128; Bonforroni corrected p value for Townsend deprivation index, baseline age, body mass index, education (Degree/Non-degree/ None of the above/ Prefer not to answer), Smoking and Drinking (Current/ Never/ Not answered/ Previous): 0.0002. Bolded if less than the corrected p value.

Appendix 3: The associations of single nucleotide polymorphisms (SNPs) with Hba1c and coronary artery disease risk from corresponding genome wide association studies (GWAS) (2, 4)

			GWAS on HbA1c		CARDIoGRAMplusC4D GWAS	
SNP	Effect allele	Other allele	Beta (%)	Standard error	Log odds	Standard error
rs1046896	T	C	0.028	0.0017	0.008492	0.009929
rs10774625	G	A	0.009	0.0016	-0.06656	0.010538
rs10830963	G	C	0.02	0.002	0.020367	0.010511
rs11248914	T	C	0.014	0.0019	-0.00021	0.009571
rs11558471	A	G	0.015	0.0017	0.010826	0.010222
rs11603334	G	A	0.012	0.0021	0.004324	0.012858
rs11708067	A	G	0.013	0.0019	0.00428	0.011628
rs11964178	A	G	0.01	0.0016	0.00632	0.009554
rs12621844	T	C	0.01	0.0018	0.014706	0.00972
rs13134327	A	G	0.013	0.0017	-0.01419	0.009786
rs1558902	A	T	0.01	0.0019	0.029804	0.009617
rs17509001	C	T	0.018	0.0023	0.037037	0.013628
rs17533903	A	G	0.015	0.0022	-0.01351	0.011755
rs17747324	C	T	0.015	0.0023	0.026061	0.011381
rs198846	G	A	0.022	0.0022	-0.00461	0.013902
rs2110073	T	C	0.015	0.0028	-0.01825	0.015374
rs2383208	A	G	0.014	0.0021	-0.00944	0.012064
rs267738	T	G	0.011	0.0019	-0.01665	0.012456
rs282587	G	A	0.019	0.0027	-0.01366	0.015205
rs3782123	C	A	0.013	0.002	0.000519	0.010192
rs4607517	A	G	0.031	0.0024	0.013369	0.012455
rs4737009	A	G	0.021	0.002	0.000496	0.010974
rs4745982	T	G	0.095	0.0056	0.069028	0.022855
rs4820268	G	A	0.016	0.0017	0.014552	0.009366
rs560887	C	T	0.028	0.0018	0.008502	0.01125
rs579459	C	T	0.011	0.0019	0.072956	0.011317
rs592423	A	C	0.009	0.0017	0.005378	0.009515
rs6474359	T	C	0.044	0.0053	0.016126	0.025794
rs6980507	A	G	0.01	0.0018	-0.00693	0.00951
rs7040409	C	G	0.028	0.0037	0.011438	0.021066
rs7616006	A	G	0.01	0.0017	0.015742	0.009527
rs7756992	G	A	0.012	0.0018	0.021832	0.010057
rs8192675	T	C	0.011	0.0017	0.015985	0.010102
rs837763	T	C	0.017	0.0016	-0.00091	0.010729
rs857691	T	C	0.019	0.0019	-0.0235	0.010684
rs9604573	T	C	0.01	0.0018	0.010169	0.012491
rs9818758	A	G	0.012	0.002	-0.00968	0.013067
rs9914988	A	G	0.013	0.002	0.025984	0.010738

Appendix 4: Association of HbA1c (%) and cardiovascular mortality using Mendelian randomization in the UK Biobank

	Inverse variance weighting with multiplicative random effect			MR-Egger				Weighted median	
*Outcomes	Odds ratio	95% CI	I^2 of Wald estimate	Odds ratio	95% CI	Intercept	p value for intercept	Odds ratio	95% CI
All SNPs (38)									
CVD mortality	1.51	0.79 to 2.90	0.14	0.66	0.20 to 2.21	0.018	0.12	1.21	0.47 to 3.11
CAD mortality	1.16	0.52 to 2.57	0.00	0.34	0.08 to 1.48	0.027	0.06	0.39	0.13 to 1.17
Stroke mortality	1.93	0.40 to 9.39	0.02	5.44	0.23 to 126.45	-0.022	0.46	5.68	0.59 to 54.87
†***Excluding invalid SNPs (23)***									
CVD mortality	1.44	0.73 to 2.82	0.00	0.77	0.22 to 2.65	0.016	0.25	1.25	0.50 to 3.16
CAD mortality	0.9	0.37 to 2.18	0.00	0.42	0.08 to 2.06	0.02	0.27	0.37	0.10 to 1.38
Stroke mortality	2.15	0.28 to 16.44	0.25	9.80	0.20 to 490.17	-0.038	0.38	5.44	0.43 to 68.88
‡***Excluding invalid SNPs (14)***									
CVD mortality	1.32	0.48 to 3.65	0.08	0.36	0.03 to 4.40	0.028	0.29	1.87	0.49 to 7.11
CAD mortality	1.51	0.41 to 5.48	0.00	1.97	0.08 to 48.50	-0.006	0.86	1.31	0.23 to 7.34
Stroke mortality	0.47	0.03 to 7.77	0.22	0.19	0.00 to 263.44	0.019	0.80	0.33	0.01 to 10.08

*Defintions of disease based on primary cause of death as below: cardiovascular (CVD) mortality , defined as ICD10 I10-I99; coronary artery disease (CAD) mortality, defined as ICD10 I20-I25.9); stroke mortality, defined as ICD10 I60, I61, I63, I64)

†Set 1: SNP excluded if associated with potential causes of CAD or confounders based on public datasources (GWAS Catalog and PhenoScanner) and UK Biobank

‡Set 2: SNP excluded if classified as erthrocytic based on the HbA1c GWAS or associated with potential causes of CAD or confounders based on public datasources (GWAS Catalog and PhenoScanner) and UK Biobank

Appendix 5: Association of HbA1c (%) with cardiovascular disease (CVD), coronary artery disease (CAD), stroke and its subtypes using Mendelian randomization in the UK Biobank, excluding participants who were related

		Inverse variance weighting with multiplicative random effects			MR-Egger				Weighted median	
Instrument	*Outcome	Odds ratio	95% CI	I^2 of Wald estimates	Odds ratio	95% CI	Intercept	p value for intercept	Odds ratio	95% CI
All SNPs (38)	CVD	1.11	0.82 to 1.47	0.83	1.23	0.70 to 2.14	-0.002	0.65	1.27	1.05 to 1.53
	CAD	1.45	1.03 to 2.04	0.58	1.40	0.73 to 2.69	0.001	0.90	1.68	1.16 to 2.42
	Stroke	1.35	0.86 to 2.11	0.28	1.29	0.55 to 3.03	0.001	0.90	1.18	0.62 to 2.25
	Ischemic	1.20	0.62 to 2.33	0.21	0.66	0.19 to 2.27	0.013	0.27	0.78	0.30 to 1.99
	Hemorrhagic	1.01	0.43 to 2.38	0.00	0.77	0.15 to 3.82	0.006	0.69	0.51	0.13 to 2.03
	Others	1.71	0.86 to 3.39	0.28	2.79	0.75 to 10.41	-0.011	0.40	1.97	0.77 to 5.08
	Other CVD	0.97	0.74 to 1.29	0.80	1.11	0.65 to 1.89	-0.003	0.57	1.10	0.91 to 1.32
†Excluded SNPs if associated with potential causes of CAD or confounders (23)	CVD	1.22	1.01 to 1.47	0.50	1.20	0.85 to 1.70	0.000	0.90	1.27	1.05 to 1.53
	CAD	1.43	1.05 to 1.94	0.36	1.11	0.63 to 1.95	0.007	0.31	1.35	0.92 to 1.97
	Stroke	1.52	0.95 to 2.43	0.18	1.01	0.43 to 2.36	0.011	0.27	1.28	0.67 to 2.45
	Ischemic	1.03	0.51 to 2.09	0.13	0.65	0.18 to 2.36	0.012	0.41	0.82	0.30 to 2.28
	Hemorrhagic	1.58	0.60 to 4.11	0.00	0.40	0.07 to 2.27	0.036	0.08	0.58	0.15 to 2.30
	Others	2.11	0.99 to 4.50	0.26	2.31	0.55 to 9.73	-0.002	0.88	2.13	0.80 to 5.67
	Other CVD	1.09	0.90 to 1.31	0.45	1.19	0.84 to 1.68	-0.002	0.54	1.22	1.00 to 1.49
‡Excluded SNPs if classified as erthrocytic, associated with potential causes of CAD or confounders (14)	CVD	1.36	1.12 to 1.65	0.00	0.98	0.61 to 1.57	0.007	0.16	1.30	1.01 to 1.67
	CAD	1.91	1.29 to 2.82	0.16	2.47	0.91 to 6.71	-0.005	0.59	1.91	1.18 to 3.09
	Stroke	1.59	0.81 to 3.13	0.19	0.91	0.16 to 5.09	0.012	0.50	2.60	1.02 to 6.61
	Ischemic	1.36	0.47 to 3.96	0.20	0.56	0.04 to 8.37	0.019	0.50	2.34	0.56 to 9.77
	Hemorrhagic	1.68	0.40 to 7.00	0.04	1.28	0.03 to 50.95	0.006	0.88	1.18	0.17 to 8.31
	Others	1.78	0.61 to 5.17	0.23	1.13	0.07 to 17.80	0.01	0.73	1.03	0.26 to 4.15
	Other CVD	1.10	0.90 to 1.34	0.00	0.75	0.46 to 1.23	0.008	0.13	1.06	0.81 to 1.39

*Definitions of disease as below: Prevalent cardiovascular disease (CVD, defined as ICD9 401-4599, ICD10 I10-I99) and its subtype, including coronary artery disease (CAD, defined as ICD9 410-4149, ICD10 I20-I25.9); stroke (ICD9 430; 431; 434; 436, ICD10 I60, I61, I63, I64), ischemic stroke (defined as ICD9 434, 436, ICD10 I63-I64), hemorrhagic stroke (defined as ICD9 430, 431, ICD10 I60-I61). Included both cases and mortality.

† Set 1: SNP excluded if associated with potential causes of CAD or confounders based on public datasources (Phenoscanner and GWAS Catalog) and UK Biobank

‡ Set 2: SNP excluded if classified as erthrocytic based on the HbA1c GWAS or associated with potential causes of CAD or confounders based on public datasources (Phenoscanner and GWAS Catalog) and UK Biobank

Appendix 6: *Estimates for HbA1c on coronary artery disease (CAD) from UK Biobank and CARDIoGRAMplusC4D 1000 Genomes-based genome wide association studies (GWAS) meta-analyzed using inverse variance weighting

Instrument	Outcome	Odds ratio	95% CI	I^2
All 38 SNPs	CAD	1.51	1.17 to 1.95	0.00
†23 SNPs (Set 1)	CAD	1.38	1.12 to 1.69	0.00
††14 SNPs (Set 2)	CAD	1.68	0.96 to 2.93	80.0%

*Additive random effect model for the estimates derived from 14 SNPs. Otherwise, fixed effect model was used
†Set 1: SNP excluded if associated with potential causes of CAD or confounders based on public datasources (GWAS Catalog and PhenoScanner) and UK Biobank
†† Set 2: SNP excluded if classified as erthrocytic based on the HbA1c GWAS or associated with potential causes of CAD or confounders based on public datasources (GWAS Catalog and PhenoScanner) and UK Biobank

Appendix 7: Association of HbA1c (%) and coronary artery disease (CAD)* using Mendelian randomization in the UK Biobank, stratified by sex

Outcome: CAD	Inverse variance weighting with multiplicative random effect			MR-Egger				Weighted median	
	Odds ratio	95%CI	I^2 of Wald estimate	Odds ratio	95%CI	Intercept	p value for intercept	Odds ratio	95%CI
Men									
All SNPs (38)	1.49	1.01 to 2.19	0.68	1.17	0.56 to 2.42	0.005	0.45	1.26	0.85 to 1.87
†***Excluding invalid SNPs (23)***	1.46	1.01 to 2.12	0.57	0.91	0.47 to 1.74	0.012	0.10	1.18	0.76 to 1.82
‡***Excluding invalid SNPs (14)***	2.47	1.51 to 4.05	0.49	2.17	0.60 to 7.82	0.003	0.83	2.45	1.42 to 4.22
Women									
All SNPs (38)	1.53	1.03 to 2.27	0.36	1.44	0.67 to 3.06	0.001	0.85	1.13	0.70 to 1.81
†***Excluding invalid SNPs (23)***	1.47	1.03 to 2.09	0.00	1.10	0.58 to 2.11	0.007	0.31	1.12	0.67 to 1.89
‡***Excluding invalid SNPs (14)***	1.81	1.05 to 3.12	0.14	1.97	0.48 to 8.09	-0.002	0.90	2.25	1.10 to 4.61

*Coronary artery disease defined as ICD9 410-4149, ICD10 I20-I25.9. Included both cases and mortality.

† Set 1: SNP excluded if associated with potential causes of CAD or confounders based on public datasources (GWAS Catalog and PhenoScanner) and UK Biobank

‡ Set 2: SNP excluded if classified as erthrocytic based on the HbA1c GWAS or associated with potential causes of CAD or confounders based on public datasources (GWAS Catalog and PhenoScanner) and UK Biobank

Appendix 8: Association of HbA1c (%) with cardiovascular disease (CVD), coronary artery disease (CAD), stroke and its subtypes using Mendelian randomization in the UK Biobank and CARDIoGRAMplusC4D genome wide association study (GWAS) (CAD only), only using 15 genetic instruments classified as erythrocytic based on the previous GWAS (2)

	Inverse variance weighting with multiplicative random effects			MR-Egger				Weighted median	
*Outcomes	Odds ratio	95%CI	I^2 of Wald estimate	Odds ratio	95%CI	Intercept	P value for intercept	Odds ratio	95%CI
CVD	0.94	0.53 to 1.67	0.94	1.29	0.52 to 3.23	-0.008	0.40	1.26	1.04 to 1.51
CAD	1.09	0.61 to 1.95	0.81	0.97	0.38 to 2.51	0.003	0.76	0.91	0.65 to 1.28
Stroke	1.18	0.61 to 2.25	0.56	1.01	0.35 to 2.92	0.004	0.72	1.07	0.61 to 1.90
Ischemic stroke	0.77	0.34 to 1.75	0.36	0.61	0.16 to 2.32	0.006	0.67	0.65	0.28 to 1.54
Hemorrhagic stroke	1.90	0.52 to 6.86	0.41	0.77	0.10 to 5.79	0.024	0.28	0.98	0.27 to 3.61
Other stroke	1.51	0.63 to 3.61	0.42	1.79	0.43 to 7.53	-0.004	0.77	1.46	0.61 to 3.48
Other CVD	0.92	0.56 to 1.51	0.92	1.34	0.62 to 2.90	-0.01	0.24	1.31	1.09 to 1.59
CVD mortality	1.53	0.66 to 3.55	0.00	0.95	0.25 to 3.61	0.012	0.38	1.23	0.41 to 3.73
CAD mortality	0.66	0.21 to 2.04	0.06	0.28	0.05 to 1.59	0.023	0.23	0.39	0.09 to 1.63
Stroke mortality	7.67	0.75 to 78.57	0.07	21.32	0.41 to 1116.57	-0.025	0.54	20.0	0.91 to 437.93
CAD (CARDIoGRAMplusC4D)	1.25	0.66 to 2.36	0.79	2.24	0.77 to 6.53	-0.013	0.21	1.91	1.23 to 2.95

*Definitions of disease as below: Prevalent cardiovascular disease (CVD, defined as ICD9 401-4599, ICD10 I10-I99) and its subtype, including coronary artery disease (CAD, defined as ICD9 410-4149, ICD10 I20-I25.9); stroke (ICD9 430; 431; 434; 436, ICD10 I60, I61, I63, I64), ischemic stroke (defined as ICD9 434, 436, ICD10 I63-I64), hemorrhagic stroke (defined as ICD9 430, 431, ICD10 I60-I61). Included both cases and mortality.

References

1. Staley JR, Blackshaw J, Kamat MA, et al. PhenoScanner: a database of human genotype-phenotype associations. Bioinformatics 2016;**32**:3207-3209
2. Wheeler E, Leong A, Liu CT, et al. Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. PLoS Med 2017;**14**:e1002383
3. Welter D, MacArthur J, Morales J, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res 2014;**42**:D1001-1006
4. Nikpay M, Goel A, Won HH, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nature genetics 2015;**47**:1121-1130