

# SUPPLEMENTAL MATERIAL

Cardiac risk factors for stroke: a comprehensive Mendelian Randomization study  
Frerich et al.

## OUTLINE

Supplemental Methods.....2  
Supplemental Tables.....4  
Supplemental Figures .....5  
Checklist for Mendelian Randomization analyses.....14  
Supplemental References.....17

## SUPPLEMENTAL METHODS

All analyses were conducted in R 4.0.3<sup>1</sup> using TwoSampleMR<sup>2</sup>.

**Exposure selection.** Cardiac traits were selected as proposed by van der Ende et al.<sup>3</sup> To increase the scope of this work, we also included cardiac traits with more recent genome-wide association data available. Several search engines were addressed (PubMed, GWAS catalog, bioRxiv, medRxiv), searching for the name of each phenotype either separately or in combination with the terms “GWAS” or “GWAS new loci”. Alternative phenotype naming variants were included. We then divided all traits into groups, namely cardiovascular diseases, cardiac imaging traits, electrocardiographic (ECG) traits, and blood biomarkers. **Supplemental Table VII** lists each trait and its corresponding data sources. For increased power in the MR analysis of stroke and resting heart rate (HR), we conducted a GWAS on pulse rate in the UK biobank population (UKB Field 102, pulse rate automated reading, in beats per minute, n=388,295). When longitudinal measurements were available, we took the mean of all measurements. We restricted our GWAS to individuals of White British ancestry and excluded individuals with high heterozygosity rates, as defined by the UK Biobank. After filtering for SNPs with minor allele frequency < 0.01, we used BOLT-LMM and included sex, age at baseline, genomic principal components (PCs) 1-20, genotyping chip and assessment center as covariates.

**Instrument choice and curation.** Genetic variants were selected as instruments when satisfying the following criteria: (1) genome-wide association with the respective trait at  $p < 5 \times 10^{-8}$ , (2) association in European-ancestry individuals, in order to comply with linkage disequilibrium (LD) structures from the stroke dataset, and (3) a GWAS sample size  $\geq 5,000$ . Estimates from meta-analyses of discovery and replication studies were chosen, when available. Otherwise, METAL<sup>4</sup> was used to meta-analyze available studies and maximize statistical power. For all analyses, only independent SNPs were retained (LD  $r^2 < 0.001$  in the European 1000 Genomes Project reference panel phase 1v3, LD window=10,000kb). In case of duplicate SNPs within the same trait, only the one with the lowest association p-value was kept. Missing allele codings or allele frequencies were extracted from MEGASTROKE, as long as effect directions were unambiguous. Missing standard errors were estimated from effect estimates and p-values using the Z distribution. Odds ratio (OR) and hazard ratio (HR) effect estimates were transformed into the log-odds scale. Effects of binary traits were transformed, so that ORs of our analyses represent the average change in the outcome per doubling (2-fold increase) in the prevalence of the trait.<sup>5</sup> Non-standardized effect estimates of continuous traits were normalized by their respective study-specific phenotypic standard deviation. Prior to the analysis, beta values were plotted and visually inspected to ensure consistency of effect estimates deriving from different sources. Study-specific effect sizes (mean and standard deviation), sample sizes, as well as covariate adjustments are shown in **Supplemental Table I. Supplemental Table VIII** lists all genetic instruments.

**Outcome selection.** Stroke subtypes in the MEGASTROKE consortium were defined according to the Trial of Org 10172 in Acute Stroke Treatment (TOAST) classification.<sup>6</sup>

Subtypes in SiGN (National Institute of Neurological Disorders and Stroke-Stroke Genetics Network) were available according to both TOAST and CCS. We included the following CCS subtypes from the SiGN Stage I European subset (for definitions see **Supplemental Table II**): IS (14,300 cases/26,690 controls), CCS.c.LAS (2,207 cases), CCS.c.CES (2,811 cases), CCS.c.SVS (1,823 cases), CCS.c.UD.all (4,031 cases), CCS.c.UD.incinc (2,016 cases), CCS.c.UD.cryptcemin (1,962 cases), CCS.p.LAS (2,349 cases), CCS.p.CES (3,369 cases), CCS.p.SVS (1,966 cases), CCS.p.UD.crypt (866 cases).

**Statistical analyses.** First, we harmonized effect alleles between exposure and outcome by inferring strand directions. Palindromic SNPs were harmonized based on allele frequency and discarded if the effect allele was ambiguous (allele frequency threshold=0.5±0.015). 7 of 73 cardiac traits were excluded because all alleles were either missing in or incompatible with any stroke outcome. We chose a random-effects inverse-variance weighted (IVW) model to estimate the association between exposure and outcome. In the case of underdispersion, a fixed-effects model was used.<sup>2</sup> Details of the IVW method are described elsewhere.<sup>7</sup> To test the robustness of findings, we conducted several sensitivity analyses, namely MR-Egger intercept tests, MR-Egger<sup>8</sup>, weighted median MR<sup>9</sup>, weighted mode MR<sup>10</sup>, and MR-PRESSO (**Supplemental Tables III-V and IX**). 10,000 random drawings were used to simulate a null distribution in MR-PRESSO.  $I^2$  was calculated from the IVW models to assess between-variant heterogeneity (**Supplemental Figure V**). For multivariable MR, in case a cardiac lead SNP was not present in one of the three datasets, that particular SNP was replaced by an LD proxy SNP ( $r^2 \geq 0.8$ ). Correct associations of the alleles were validated and corrected based on European LD structures (1000G populations CEU, GBR, IBS, and TSI). Multiple testing was accounted for in all tests within each phenotype group (cardiovascular diseases, ECG traits, cardiac imaging traits, blood biomarkers, single-lead ECG traits) via the false discovery rate (FDR) with a significance level of 0.05. Statistical power and F-statistics were calculated for continuous traits as described by Burgess and Thompson, with  $\alpha=0.05$  (**Supplemental Table VI**).<sup>11</sup>

## SUPPLEMENTAL TABLES

Legends refer to Excel spreadsheets.

**Table I.** Study-specific effect sizes (mean and standard deviation), sample sizes, as well as covariate adjustments.

**Table II.** Definitions of stroke subtypes according to CCS.

**Table III.** Results of MR-Egger intercept tests to assess directional pleiotropy.

**Table IV.** Results from Mendelian Randomization Pleiotropy RESidual Sum and Outlier (MR-PRESSO).  $k=10,000$ .

**Table V.** MR-PRESSO raw and outlier-corrected IVW MR estimates. b: beta; se: standard error; OR (95% CI): odds ratio (95% confidence interval); pval: p-value; pval\_adj: FDR-adjusted p-value (adjusted within each phenotype group: cardiovascular diseases, ECG traits, cardiac imaging traits, blood biomarkers); t.stat: t-statistic.

**Table VI.** MR power calculations and F-statistics for continuous traits. F statistics were calculated for individual SNPs. nsnp: number of SNPs.

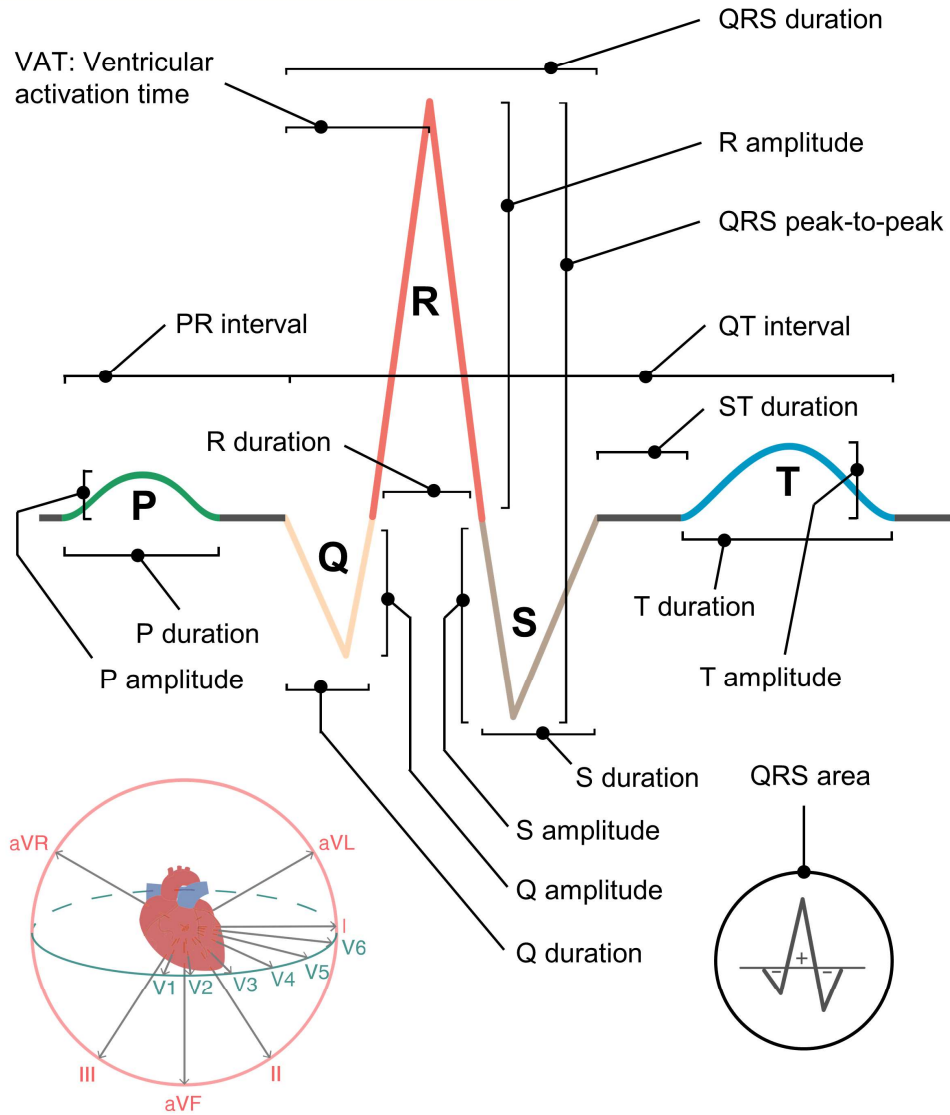
**Table VII.** Specifications of 66 cardiac traits, including abbreviations, units, and data sources.

**Table VIII.** Genetic instruments for all cardiac traits and corresponding stroke outcome values from MEGASTROKE. Non-standardized effect sizes of continuous traits were normalised by the respective study-specific phenotypic standard deviation (see **Supplemental Table I**). Effect sizes of binary traits are listed prior to transformation with  $\ln(2)$ . A1: effect allele; A2: other allele; A1freq: effect allele frequency; b: beta; se: standard error; pval: p-value.

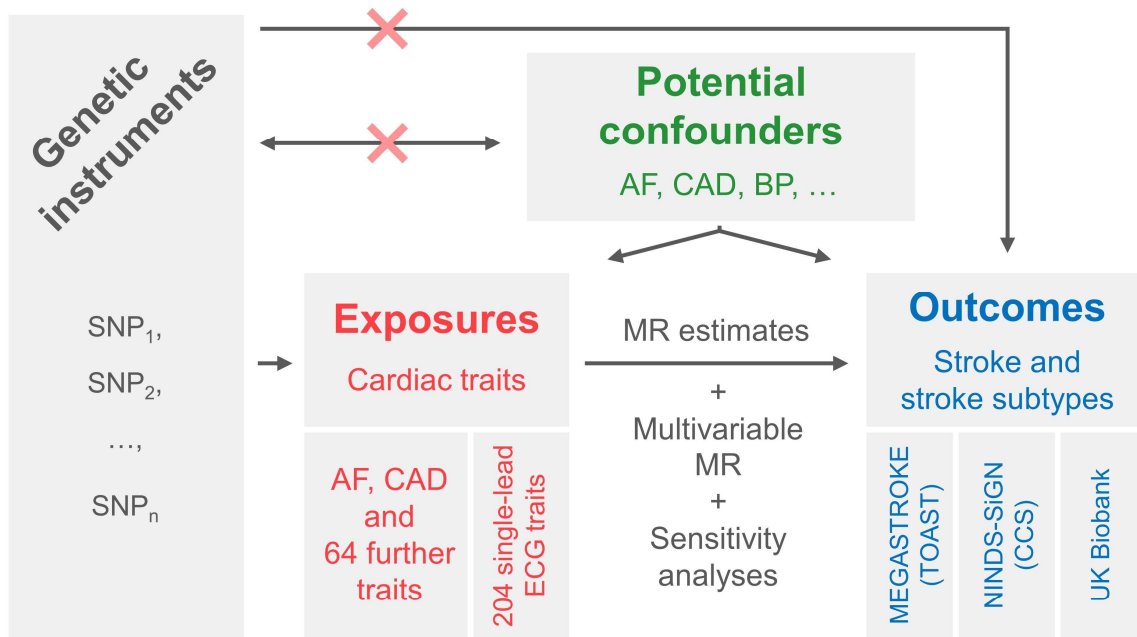
**Table IX.** MR estimates (inverse-variance weighted, MR-Egger, weighted median, simple mode, and weighted mode) between all cardiac exposures and MEGASTROKE outcomes in the primary analysis. nsnp: number of independent single-nucleotide polymorphisms; OR: odds ratio per 1-SD increase in the continuous exposures or per 2-fold increase in the prevalence of the binary exposure; OR\_lower and OR\_upper: lower and upper end of 95% confidence interval; pval: nominal p-value.

# SUPPLEMENTAL FIGURES

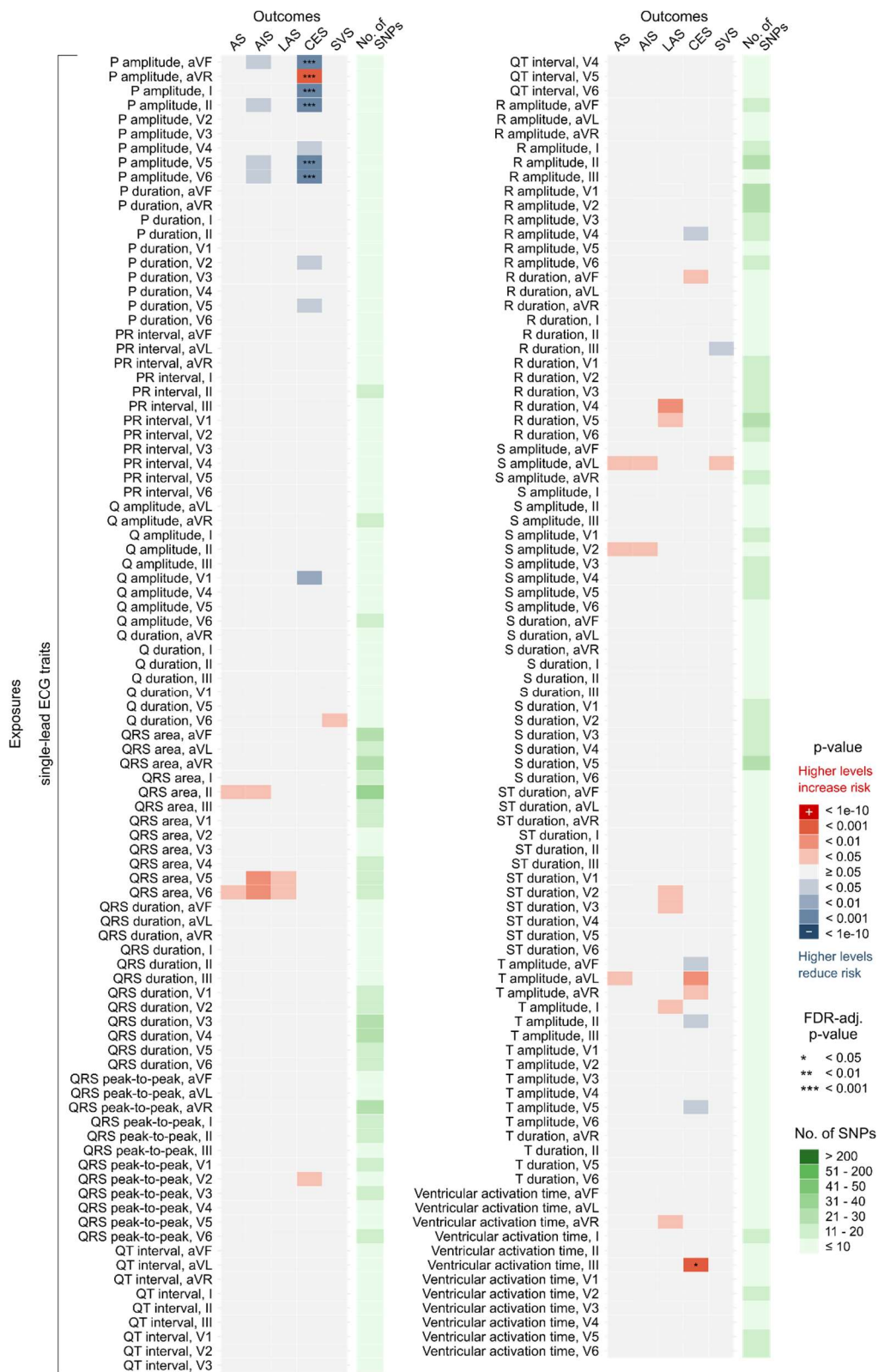
## SINGLE-LEAD ECG TRAITS



**Figure I.** 17 single-lead electrocardiographic (ECG) traits were tested for their associations with stroke and stroke subtypes in secondary analyses. The 12 ECG leads are shown on the bottom left. Genetic association derived from Norland et al.<sup>12</sup>

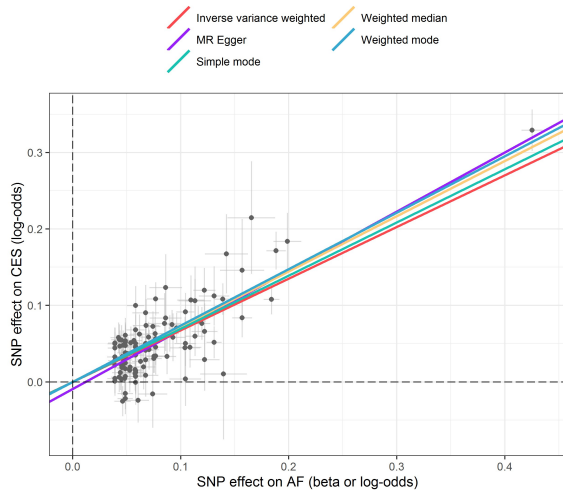
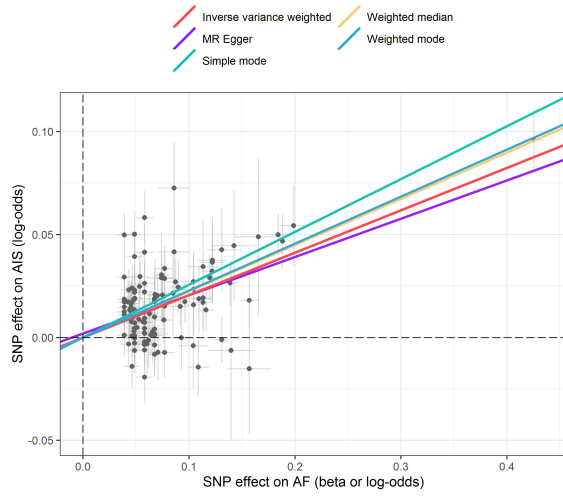
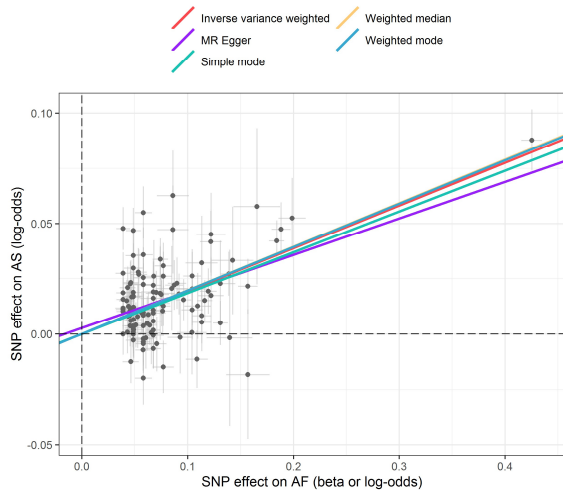


**Figure II.** Schematic representation of the study design. In primary Mendelian Randomization (MR) analyses, we investigated 66 cardiac traits for their associations with stroke and stroke subtypes. In addition, we conducted sensitivity and multivariable MR analyses to address potential violations of MR core assumptions (labeled by red crosses). Outcomes derived from MEGASTROKE (TOAST subtypes), replications were conducted in NINDS-SIGN (CCS subtypes) and the UK Biobank datasets. Secondary analyses focused on 204 single-lead ECG traits. AF indicates atrial fibrillation; CAD: coronary artery disease; ECG: electrocardiogram; BP: blood pressure; SNP: single-nucleotide polymorphism.

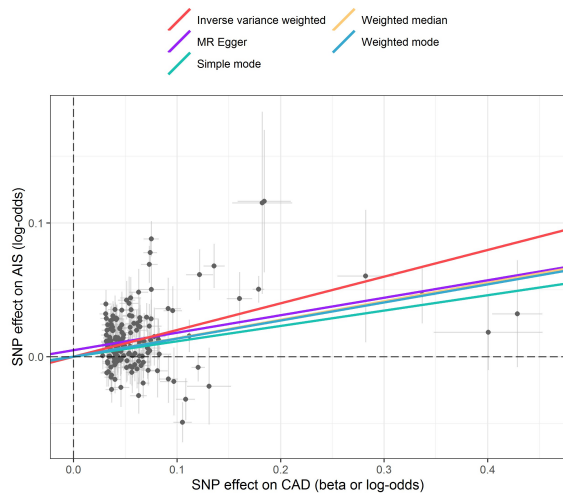
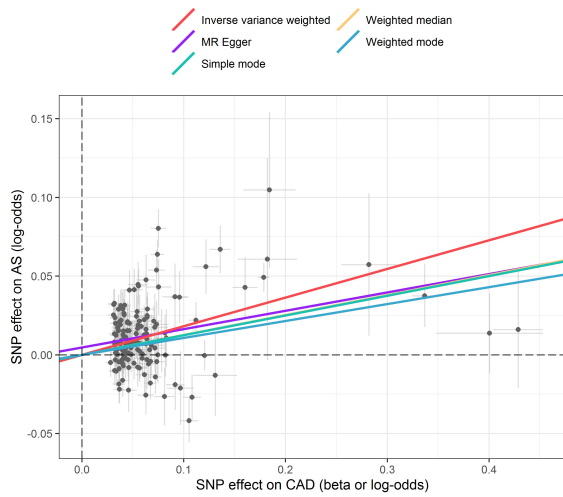


**Figure III.** Inverse-variance weighted Mendelian Randomization (IVW-MR) estimates between 204 single-lead ECG traits and stroke/stroke subtypes from the MEGASTROKE consortium. Darker colors indicate stronger effects. Asterisks indicate p-values after FDR adjustment. The green column represents the number of independent SNPs for each trait. Trait specifications are given in **Figure I**. AS: any stroke; AIS: any ischemic stroke; LAS: large-artery stroke; CES: cardioembolic stroke; SVS: small-vessel stroke.

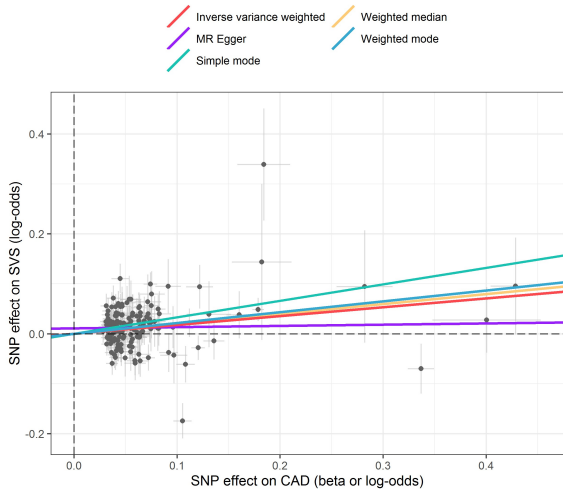
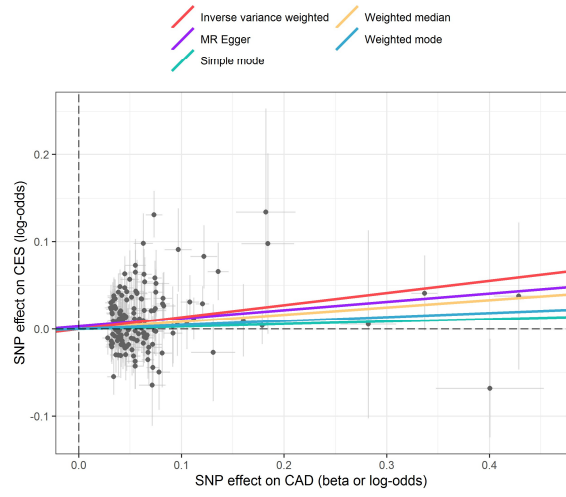
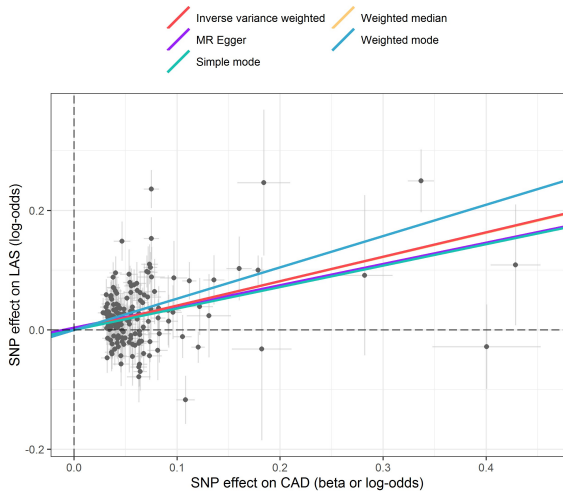
# AF



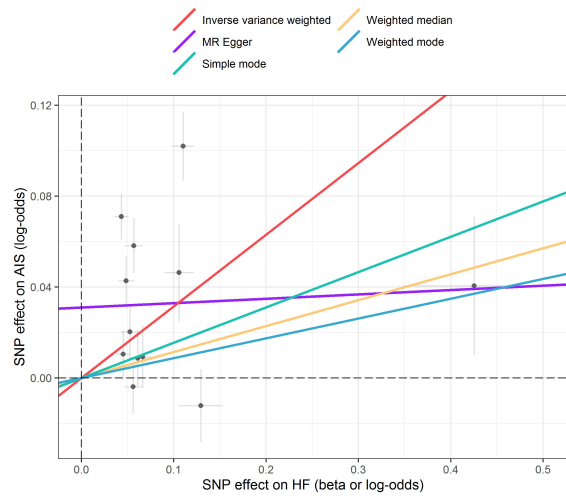
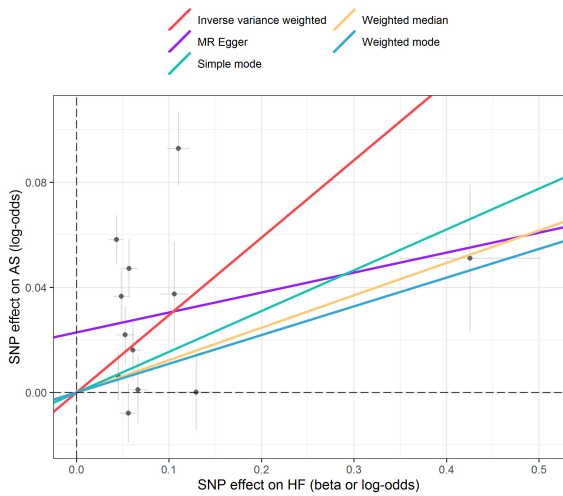
# CAD

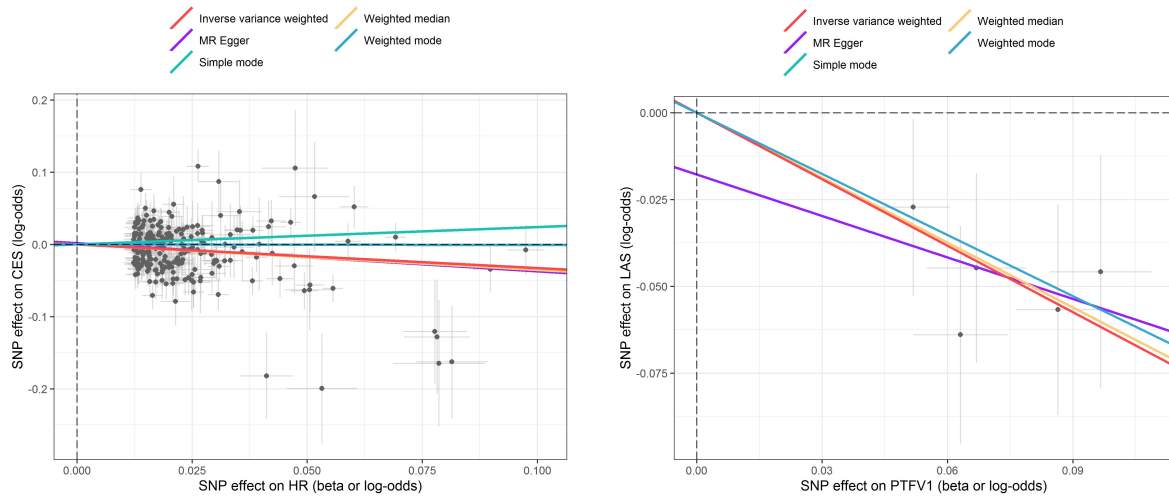




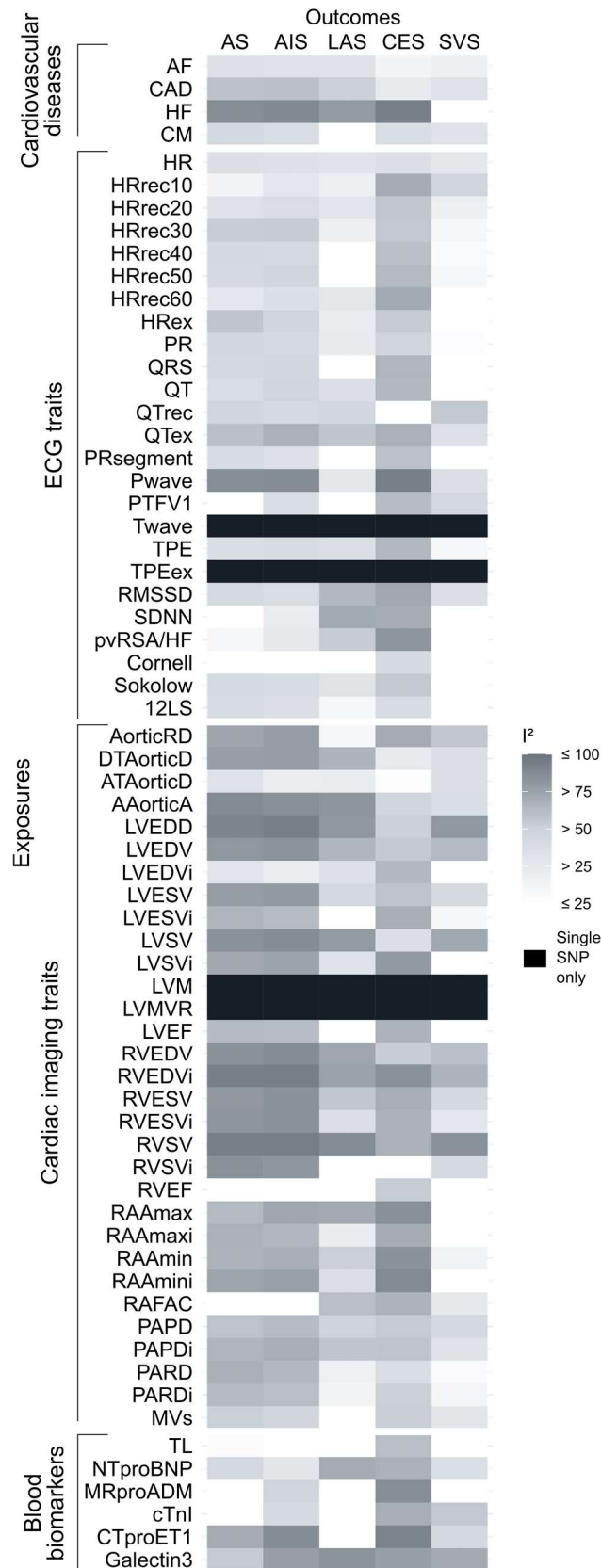


HF

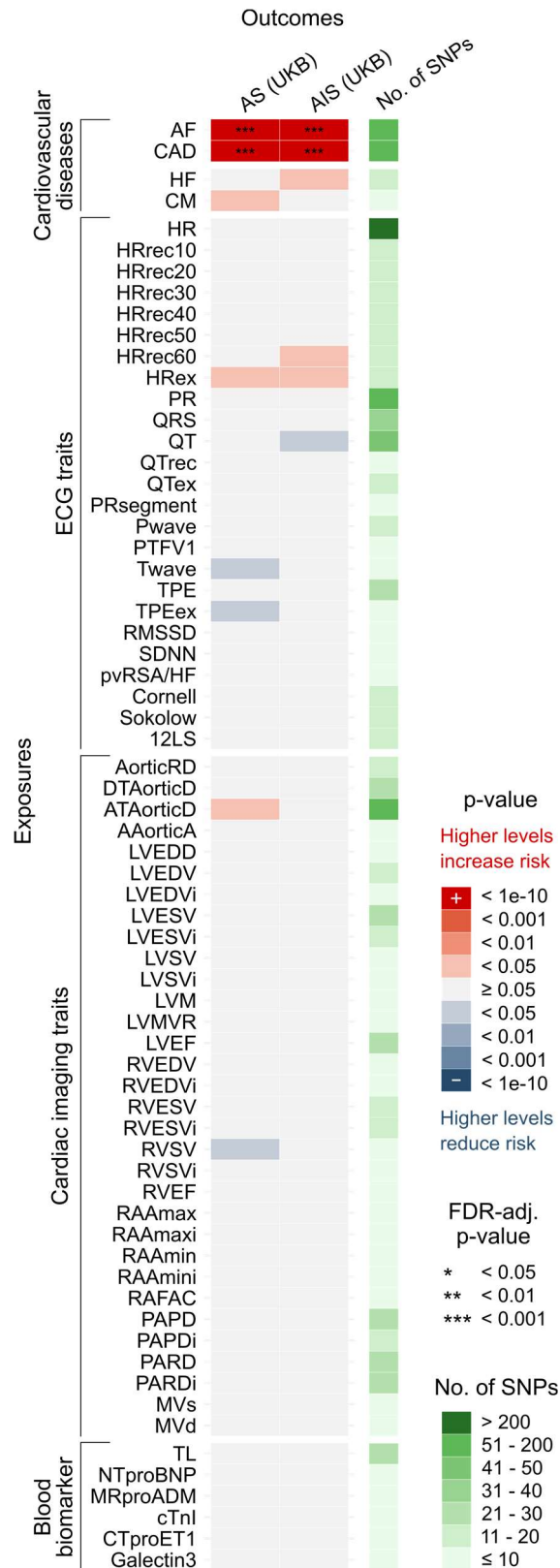




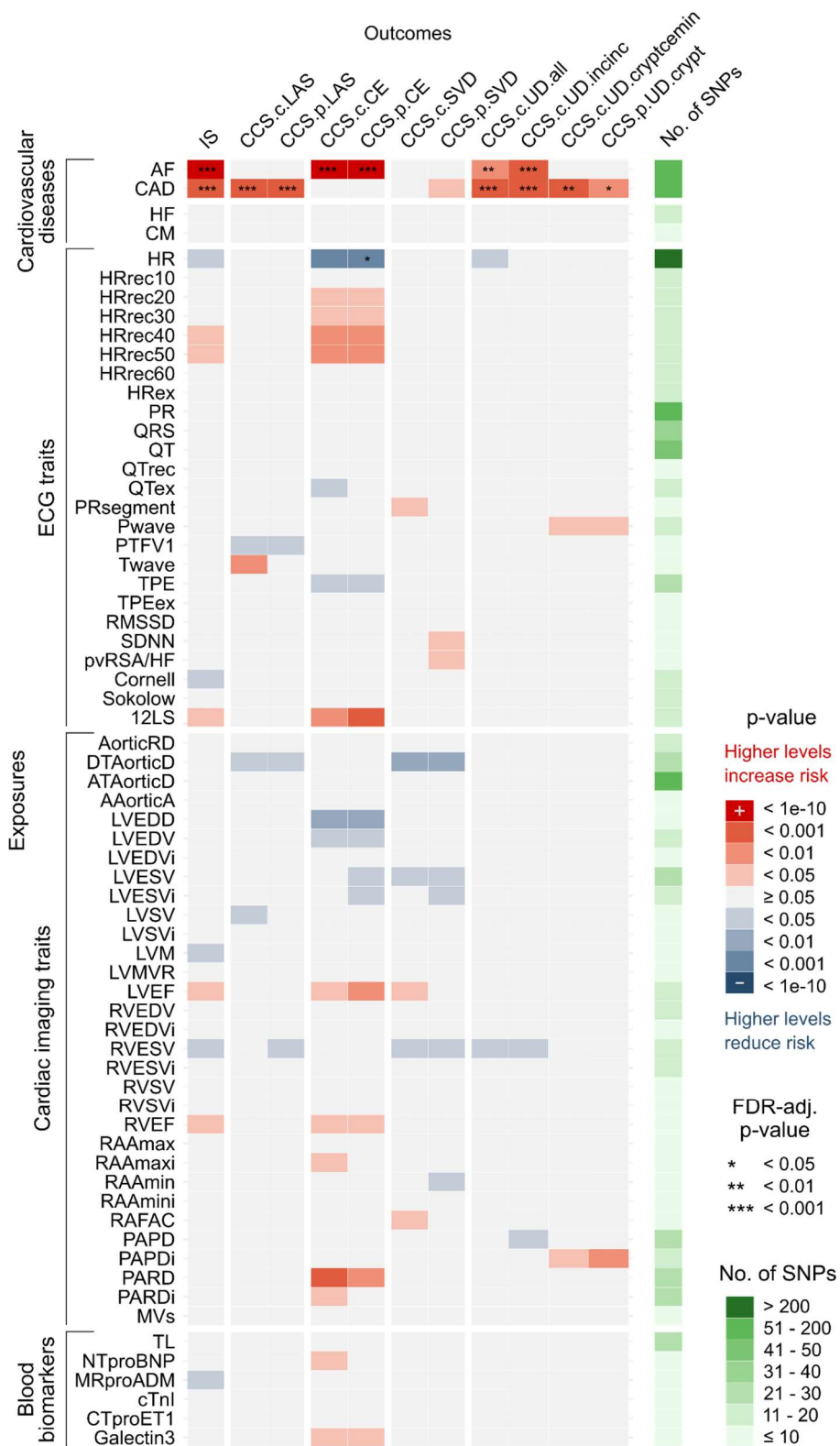
**Figure IV.** Scatter plots of the main results (FDR<0.05). Each dots represent a SNP and its association with the respective cardiac trait (exposure; x-axis), and stroke or stroke subtype (outcome; y-axis). Effect sizes are given on the log-odds scale (for binary traits) or beta scale (for continuous traits). Grey lines indicate 95% confidence intervals. Colored lines represent linear regressions according to different MR methods (red: inverse-variance weighted; yellow: weighted median; purple: MR-EGGER; blue: weighted mode; cyan: simple mode). AF: atrial fibrillation; CAD: coronary artery disease; HF: all-cause heart failure; PTFV1: P-wave terminal force in V1; HR: resting heart rate; SNP: single-nucleotide polymorphism. AS: any stroke; AIS: any ischemic stroke; LAS: large-artery stroke; CES: cardioembolic stroke; SVS: small-vessel stroke.



**Figure V.** Heterogeneity in MR estimates. Color shadings indicate  $I^2$  values (from white = 0 to dark = 100). Black fields indicate analyses with a single independent SNP. Trait specifications are given in **Figure 1**. AS: any stroke; AIS: any ischemic stroke; LAS: large-artery stroke; CES: cardioembolic stroke; SVS: small-vessel stroke.



**Figure VI.** Inverse-variance weighted Mendelian Randomization (IVW-MR) estimates between 66 cardiac traits and any stroke (AS) / any ischemic stroke (AIS) in the UK Biobank (UKB). Darker colors indicate stronger effects. Asterisks indicate p-values after FDR adjustment. The green column represents the number of independent SNPs for each trait. Trait specifications are given in **Figure 1**.



**Figure VII.** Inverse-variance weighted Mendelian Randomization (IVW-MR) estimates between 66 cardiac traits and NINDS-SiGN CCS-defined stroke subtypes. Darker colors indicate stronger effects. Asterisks indicate p-values after FDR adjustment. The green column represents the number of independent SNPs for each trait. **Figure 1** specifies all cardiac traits. **Supplemental Table II** specifies CCS-defined stroke subtypes.

# CHECKLIST FOR MENDELIAN RANDOMIZATION ANALYSES

Here, we provide answers to the checklist for MR studies from Burgess et al.<sup>13</sup>

1. *What is the primary hypothesis of interest? What is the motivation for using Mendelian randomization? What is the scope of the investigation? What and how many primary analyses?*

*Using Mendelian Randomization we sought to investigate potential causal associations of cardiac traits with stroke and stroke subtypes.*

*Recently published genome-wide association studies (GWAS) for a wide range of cardiac traits offer an opportunity to comprehensively assess causal relationships with stroke.*

*Our primary work covers 66 cardiac traits based on 80 publicly available GWAS summary statistics (Supplemental Table VII). Primary analyses were conducted between each cardiac trait and any stroke (AS; 40,585 cases), any ischemic stroke (AIS; 34,217 cases), and TOAST-defined subtypes (large-artery stroke [LAS]: 4,373 cases; cardioembolic stroke [CES]: 7,193 cases; small-vessel stroke [SVS]) from the MEGASTROKE consortium.*

*Further analyses included multivariable and sensitivity analyses, as well as the investigation of additional outcomes (stroke in the UK Biobank, alternative stroke definitions in SiGN; see Supplemental Figures VI and VII).*

*Lastly, in secondary analyses we investigated 204 single-lead ECG traits for their associations with stroke (Supplemental Figures I and III).*

2. *Data sources*

*What type of Mendelian randomization investigation is this? One-sample or two-sample? Sample overlap? Summarized data or individual-level data? Drawn from sample population? Relevance to applied research?*

*The primary analysis of this study is a two-sample summarized-data MR approach. The genetic data derived from heterogeneous samples, including population-based studies such as the UK Biobank.*

*Exposure and outcome samples overlapped in several analyses, in particular in analyses based on large genetics consortia such as CHARGE. We mentioned this in the discussion section, as it can bias our estimates<sup>11</sup>.*

3. *Selection of genetic variants – how were the genetic variants chosen? Single or multiple gene regions?*

- a. *Biological rationale*

- b. *GWAS analysis? If so, what dataset? What was the p-value threshold? Clumping?*

- c. *Were genetic variants excluded from the analysis? Associations with pleiotropic pathways?*

- d. *How else was the validity of genetic variants as instrumental variants assessed?*

*Genetic variants from multiple gene regions were selected based on their association with cardiac traits. Details regarding the selection can be found in the Supplemental Methods. In brief, genetic variants were selected as instruments when satisfying the following criteria: (1) genome-wide association with the respective trait at  $p < 5 \times 10^{-8}$ , (2) association in European-ancestry individuals, in order to comply with linkage disequilibrium (LD) structures from the stroke dataset, and (3) a GWAS sample size  $\geq 5,000$ . For all analyses, only independent SNPs were retained (LD  $r^2 < 0.001$  in the European 1000 Genomes Project reference panel phase*

*Iv3, LD window=10,000kb). No variants were excluded based on potential pleiotropic associations with other traits.*

*To address the validity of genetic variants, several sensitivity analyses were conducted (MR-Egger intercept test, MR-Egger, weighted median MR, weighted mode MR, and MR-PRESSO). We also reported the between-variant heterogeneity ( $I^2$ ).*

4. *Variant harmonization? (for two-sample analyses)*

*Was it checked that genetic variants were appropriately orientated across the datasets?*

*We harmonized effect alleles between exposure and outcome by inferring strand directions. Palindromic SNPs were harmonized based on allele frequency and discarded if the effect allele was ambiguous (allele frequency threshold= $0.5 \pm 0.015$ ).*

5. *Primary analysis*

*What was the primary analysis? What was the statistical method? How implemented? Multiple testing?*

*The primary analysis was conducted between 66 cardiac traits and AS, AIS, and TOAST-defined subtypes LAS, CES, and SVS. Results are shown in Figure 2, based on a random-effects inverse-variance weighted (IVW) model, implemented via TwoSampleMR<sup>2</sup>. In the case of underdispersion, a fixed-effects model was used.*

*Multiple testing was accounted for in all tests within each phenotype group (cardiovascular diseases, ECG traits, cardiac imaging traits, blood biomarkers, single-lead ECG traits) via the false discovery rate (FDR) with a significance level of 0.05.*

6. *Supplementary analyses and*

7. *Sensitivity analyses*

*What analyses were performed to support and assess the validity of the primary analysis? For example: stricter criteria for variant selection, assess heterogeneity, robust methods, subgroup analysis, positive/negative control, 'leave-one-out' analyses, colocalization (single gene region)*

*Multivariable MR was used to test whether the observed associations were attenuated by the effects of known stroke risk factors. For this purpose, summary-level data for AF (both permanent and paroxysmal), coronary artery disease (CAD; 122,733 cases/424,528 controls), and systolic blood pressure (SBP;  $n=1,006,863$ ) were acquired. Results are shown in Figure 3.*

*$I^2$  was calculated from the IVW models to assess between-variant heterogeneity (Supplemental Figure V).*

*The following sensitivity analyses were conducted: MR-Egger intercept test, MR-Egger, weighted median MR, weighted mode MR, and MR-PRESSO. Results of these can be found in Supplemental Tables III, IV, V, and IX.*

8. *Data presentation*

*How are data and results presented to allow readers to assess the analysis and assumptions? For example: scatter plot, forest/funnel/radial plot,  $R^2/F$  statistics, comparison of methods, power*

*For clarity, we presented our extensive primary results as a heatmap, depicting significance and directionality of all 330 associations (Figure 2). Main findings ( $FDR < 0.05$ ) and*

*corresponding multivariable analyses were presented as forest plots, including ORs, 95% CIs, and p-values (Figure 3).*

*In addition, we presented scatter plots of our main univariable findings (Figure IV in the Supplement), as well as R<sup>2</sup>/F statistics and power calculations of continuous exposures (Supplemental Table VI).*

9. *Interpretation*

*How have results been interpreted, particularly any numerical estimates?*

*We were cautious when interpreting results in the light of causal inference, in particular numerical results. Please refer to the discussion section.*



## SUPPLEMENTAL REFERENCES

1. The R Core Team. R: A Language and Environment for Statistical Computing. 2018. [www.r-project.org/](http://www.r-project.org/). Accessed May 01, 2021.
2. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, Laurin C, Burgess S, Bowden J, Langdon R, et al. The MR-base platform supports systematic causal inference across the human phenome. *Elife*. 2018;7:1–29.
3. Van Der Ende MY, Said MA, Van Veldhuisen DJ, Verweij N, Van Der Harst P. Genome-wide studies of heart failure and endophenotypes: Lessons learned and future directions. *Cardiovasc. Res*. 2018;114:1209–1225.
4. Willer CJ, Li Y, Abecasis GR. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26:2190–2191.
5. Burgess S, Labrecque JA. Mendelian randomization with a binary exposure variable: interpretation and presentation of causal estimates. *Eur. J. Epidemiol*. 2018;33:947–952.
6. Adams H., Bendixen B., Kappelle L., Biller J, Love B., Gordon D., Marsh E. Classification of Subtype of Acute Ischemic Stroke. *Stroke*. 1993;23:35–41.
7. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol*. 2013;37:658–665.
8. Bowden J, Smith GD, Burgess S. Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression. *Int. J. Epidemiol*. 2015;44:512–525.
9. 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;40:304–314.
10. Hartwig FP, Smith GD, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int. J. Epidemiol*. 2017;46:1985–1998.
11. Burgess S, Thompson SG. Mendelian Randomization: Methods for Causal Inference Using Genetic Variants. Chapman and Hall/CRC; 2021.
12. Norland K, Sveinbjornsson G, Thorolfsdottir RB, Davidsson OB, Tragante V, Rajamani S, Helgadottir A, Gretarsdottir S, van Setten J, Asselbergs FW, et al. Sequence variants with large effects on cardiac electrophysiology and disease. *Nat. Commun*. 2019;10:1–10.
13. Burgess S, Davey Smith G, Davies NM, Dudbridge F, Gill D, Glymour MM, Hartwig FP, Holmes M V., Minelli C, Relton CL, et al. Guidelines for performing MR investigations. *Wellcome Open Res*. 2020;4:186.