Life-Time Covariation of Major Cardiovascular Diseases - A 40 Year Longitudinal Study and Genetic Studies

Lars Lind, MD, PhD[1], Johan Sundström, MD, PhD[1,2], Johan Ärnlöv, MD, PhD[3,4], Martin Ingelsson, MD, PhD[5], Albert Henry, MD, MSc[6,7,8], R. Thomas Lumbers, MD, PhD[7,8,9,10], Erik Lampa, PhD[11]

1. Department of Medical Sciences, Uppsala University, Uppsala, Sweden.
2. The George Institute for Global Health, University of New South Wales, Sydney, Australia.
3. Division of Family Medicine and Primary Care, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Huddinge, Sweden.
4. School of Health and Social Sciences, Dalarna University, Falun, Sweden.
5. Department of Public Health and Caring Sciences/Geriatrics, Uppsala University, Uppsala, Sweden.
6. Institute of Cardiovascular Science, University College London, London, UK.
7. British Heart Foundation Research Accelerator, University College London, London, UK
8. Institute of Health Informatics, University College London, London, UK.
9. Health Data Research UK London, University College London, London, UK.
10. Bart's Heart Centre, St. Bartholomew's Hospital, London, UK.
11. Uppsala Clinical Research Centre (UCR), Uppsala University, Uppsala, Sweden.

Correspondence:
Lars Lind, MD, PhD
Department of Medical Sciences
Dag Hammarskjöldsv. 10 B, Uppsala Science Park
75237 Uppsala
Sweden

E-mail: lars.lind@medsci.uu.se
Phone: +46 73 0502878
Fax: +46 18 509297

Total word count: 5431

Subject terms: Myocardial Infarction; Ischemic Stroke; Heart Failure; Atrial Fibrillation; Genetic, Association Studies

ABSTRACT

Background: It is known that certain cardiovascular diseases (CVD) are associated, like atrial fibrillation (AF) and stroke. However, for other CVDs, the links and temporal trends are less studied. In this longitudinal study, we have investigated temporal epidemiological and genetic associations between different CVDs.

Methods: The ULSAM study (2322 men aged 50 years) has been followed for 40 years regarding 4 major CVDs (incident myocardial infarction (MI), ischemic stroke, heart failure (HF) and AF). For the genetic analyses, publicly available data were used.

Results: Using multi-state modelling, significant relationships were seen between pairs of all of the four investigated CVDs. However, the risk of obtaining one additional CVD differed substantially both between different CVDs and between their temporal order. The relationship between HF and AF showed a high risk ratio (RRs 24-26) regardless of the temporal order. A consistent association was seen also for MI and AF, but with a lower relative risk (RRs 4-5). In contrast, the risk of receiving a diagnosis of HF following a MI was almost twice as high as for the reverse temporal order (RRs 16 vs 9). Genetic loci linked to traditional risk factors could partly explain the observed associations between the CVDs, but pathway analyses disclosed also other pathophysiological links.

Conclusions: During 40 years, all of the four investigated CVDs were pairwise associated with each other regardless of the temporal order of occurrence, but the risk magnitude differed between different CVDs and their temporal order. Genetic analyses disclosed new pathophysiological links between CVDs.

Key Words: myocardial infarction, stroke, heart failure, atrial fibrillation, longitudinal study genetics

Non-standard Abbreviations and Acronyms:

AF = Atrial fibrillation

CHD = Coronary heart disease

CI = Confidence interval

CVD = Cardiovascular disease

CV = Cardiovascular

FDR = False discovery rate

GTEx = Genotype – tissue expression project

GWAS = genome-wide association studies

HF = Heart failure

IVW = Inverse-variance weighted

MI = Myocardial infarction

MR = Mendelian randomization

OR = Odds ratio

RR = Risk ratio

SNP = Single nucleotide polymorphism

SCF = Stem cell factor

ULSAM = Uppsala Longitudinal Study of Adult Men

INTRODUCTION

It is well known that the presence of some cardiovascular diseases (CVD) will increase the risk for other CVDs. Examples of such well-established relationships are that atrial fibrillation is a powerful risk factor for both stroke[1] and heart failure,[2] and that a myocardial infarction is a strong risk factor for future heart failure.[2] It is also known that prior CVD (except stroke) is associated with an increased risk of stroke,[3] and that heart failure and myocardial infarction are related to incident atrial fibrillation.[4] Other links between the major CVDs, myocardial infarction, stroke, heart failure and atrial fibrillation are less well-known and the strength of associations between those CVDs are not well established.

Since the major common CVDs share many traditional risk factors, such as hypertension, hyperlipidemia, obesity, diabetes and smoking, the links between pairs of these common CVDs might mainly be due to shared risk factors. It might also be that other pathophysiological pathways could be of importance.

The aim of the present study is two-fold. Firstly, to describe the temporal associations of the major CVDs; myocardial infarction, stroke, heart failure and atrial fibrillation. Secondly, to investigate shared genetics in order to investigate potential mechanisms linking these common diseases. Improved knowledge on the relationships between CVDs and mechanisms linking these common diseases might be useful in the secondary prevention following a first CVD.

For the first aim, we used data from the ULSAM study in which we have followed the development of these four major CVDs during 40 years in a population-based sample of middle-aged males.[5] We investigated the life course temporal co-morbidity between the major CVDs, myocardial infarction, stroke, heart failure and atrial fibrillation, by multi-state modeling to calculate the risks of obtaining specific major CVDs given that the individual previously had experienced another specific major CVD.

For the second aim, we used the Mendelian randomization framework to study the genetic relationships between these four major CVDs using genetic data from already published genome-wide association studies (GWAS). We also evaluated the genetic overlap between the major CVDs in terms of pathway analysis of the genetic loci overlapping between CVDs.

METHODS

The complete methods section is given in the beginning of the Supplementary material.

The study was approved by the Ethical Committee of Uppsala University, and each participant in ULSAM gave their informed consent.

The data that support the findings of this study are available from the corresponding author upon reasonable request.

RESULTS

Pairs of CVD

The incidence of the four investigated CVDs during the follow-up period were in the range from 340 (ischemic stroke) to 565 (atrial fibrillation) (Table 1 and Figure 1).

Ignoring the temporal order of the CVDs, all of the four investigated CVDs were pairwise significantly associated with each other, except myocardial infarction and atrial fibrillation (P=0.09). As could be seen in Table 1, the ORs ranged from 1.21 for myocardial infarction *vs* atrial fibrillation to 3.86 for atrial fibrillation *vs* heart failure in the unadjusted analysis. No major difference in the strength of relationships were seen when adjusting for traditional cardiovascular risk factors (Figure 2).

When related to the other three CV diseases, myocardial infarction occurred before all of the other three diseases in the majority of cases. This could clearly be seen in Figure 1 and in

Table 1. This pattern was most pronounced for heart failure, since it was quite uncommon that an individual received a diagnosis of heart failure before a diagnosis of myocardial infarction. Atrial fibrillation usually occurred before heart failure, while there was no clear temporal trend for stroke vs heart failure, or stroke vs atrial fibrillation (Table 1).

Multistate results

The flow of transitions from the healthy state to one of the four diseases to further other cardiovascular diseases are shown in Figure 3 for the total follow-up period. No cases were censored due to withdrawal from study or loss of follow-up. No subject was censored because of death due to non-CVD causes before the first CVD diagnosis occurred in the sample. Of the individuals who had received a first CVD diagnosis, 395 were censored due to death from non-CVD causes before receiving a second CVD diagnosis. The analysis performed above, and in this step, only constitutes the first (from CVD healthy to first CVD) and second (form first CVD to second CVD) transition, since the number of subjects in further transitions is too limited for meaningful statistical evaluation. However, the total flow of transitions from one of the four diseases to combinations of the other CVDs over the 40 years are shown in Figure 3. It could be noted that 13 subjects received a diagnosis of all four disorders.

In the multi-state modelling, taking the temporal order into account, significant relationships were seen between pairs of all of the four investigated CVDs regardless of the temporal order of the CVDs within each pair (Table 2). However, the relative risk of obtaining one CVD following another CVD differed substantially both between different CVDs and between the temporal order. For example, the relationship between heart failure and atrial fibrillation showed a very high risk ratio (RRs 24 and 26) regardless of which of the conditions that appeared first. Also for the relationship between myocardial infarction and atrial fibrillation, the risk was not dependent on the temporal order to a major extent, but in this case the relative

risks (RRs 4 to 5) were not as high as compared to the heart failure and atrial fibrillation relationship. In contrast, the risk of receiving a diagnosis of heart failure following a MI was almost double that compared to the reverse temporal order (RR 16 vs 9). The 95%CIs are given in Table 2.

When we, as a supplementary analysis, included only non-lethal CVD events (see Supplementary Table 1), the results were essentially the same as in the main analysis when also lethal CVD were included (Table 2).

Less than 1% of all transitions between the first and second CVD events occurred within 7 days, meaning that the few cases when a second event was taking place very soon after the first event were exceptions not contributing in any major degree to the multi-stage model results.

Mendelian randomization

In the Mendelian randomization part of the study, the IVW method used as the primary analysis was highly significant ($P<0.001$) for all pairwise comparisons between the four CVDs regardless of the temporal order, except for AF (atrial fibrillation)->CHD (coronary heart disease) ($P=0.31$) when all SNP instruments were used (Table 3).

Similar results were obtained in the sensitivity analysis using the weighted median method, except for stroke-> AF ($P=0.43$). Using the MR Egger method, significant relationships were only seen for CHD -> HF (heart failure), AF ->stroke, AF-> HF, but for CHD-> stroke, CHD->AF, HF->CHD, and HF-> stroke, the estimate was similar for MR Egger compared to when using IVW and weighted median.

When only SNPs showing $P>5 \times 10^{-8}$ *vs* traditional risk factors were used as instruments, the estimates were fairly similar for the three pairwise relationships where atrial fibrillation was the exposure (Supplementary Table 2, Supplementary Figure 1). Regarding the three comparisons where coronary heart disease was the exposure, a slight reduction in the estimate was seen when SNPs being related to risk factors were removed, but in all the three cases the *P*-value was still <0.05. When heart failure was the exposure, a great reduction in the estimate was seen for HF->CHD and HF->AF, but not for HF->stroke. When stroke was used as the exposure, reductions in the estimate was seen for stroke->CHD and stoke->HF, but not for stroke->AF.

Shared genetic loci

Using already published GWAS data regarding the four major CVDs, no locus showed an association *vs* all four traits at $P<5 \times 10^{-8}$ or FDR<0.05. Three loci were related to three out of the four CVDs at $P<5 \times 10^{-8}$ (*ABO* and *ATXN2* for the triplet CHD, stroke and atrial fibrillation and RP11-119H12.3 for the triplet heart failure, stroke and atrial fibrillation). Another loci were related to heart failure, stroke and atrial fibrillation using FDR<0.05 (*CDKN1A*).

Eight loci were related to two out of the four CVDs using $P<5 \times 10^{-8}$, and another 14 using the more liberal FDR<0.05 (Figure 4 and Table 4).

The pathway enrichment analysis highlighted blood group synthesis and glucose metabolism for the genes related to the triplet CHD, stroke and atrial fibrillation, and cell cycle regulation and cellular senescence for genes related to the triplet heart failure, stroke and atrial fibrillation (Supplementary Table 2).

For genes related to both CHD and stroke, cholesterol metabolism and bile acid metabolism were the top ranked pathways, but also neuronal differentiation, actin dynamics,

notch signaling and stem cell factor (SCF) signaling were enriched pathways for these two CVDs. For genes related to both coronary heart disease and atrial fibrillation, mitogen-activated protein kinase activation and interleukin signaling were the top ranked enriched pathways, while for genes related to both coronary heart disease and heart failure, mainly lipid metabolism pathways were enriched. Other combinations of two coronary heart diseases did not show any significant enriched pathways.

DISCUSSION

Principal findings

During four decades of follow-up, all of the four investigated CVDs were associated with each other regardless of the temporal order of occurrence, but the risk magnitude differed between different CVDs and their temporal order. Genetic analyses disclosed new pathophysiological links between CVDs.

Pairs of CVD

Significant relationships were seen between pairs of all of the four investigated CVDs when ignoring the temporal order, except for atrial fibrillation and myocardial infarction. It could clearly be seen that myocardial infarction most often is the first CVD to occur (Figure 1).

All of the four investigated CVDs share the traditional risk factors, although the impact of the different risk factors varied between the diseases.[5] Thus, the most likely explanation for the relationships between the four CVDs would be these shared risk factors. However, the strength of the relationships between the CVDs was only marginally affected by including the traditional risk factors as confounders in the observational part of the study (Figure 3).

Multistate results

It is well known that atrial fibrillation is a major risk factor for heart failure and stroke,[1, 2] and that heart failure often is preceded by a myocardial infarction.[2] However, the temporal order of other pairs of CVDs has not previously been extensively studied. In this study, it was clearly seen that the risk of obtaining one CVD following another differed substantially both between different CVDs and between their temporal order.

A myocardial infarction as a complication following an acute ischemic stroke has been estimated to occur in around 2% of stroke cases.[6, 7] Also stroke as a complication to acute myocardial infarction has been reported to be in the range of 2%.[8, 9] In the present study, not only events occurring during the acute phase of a disease were studied, but events occurring over four decades, and therefore the number of events reported in the present study were substantially higher.

Mendelian randomization

Also the genetic analyses performed within the Mendelian randomization framework showed relationships between most pairs of CVDs. This was also seen following removal of SNPs linked to traditional risk factors. The exception from this was the attenuation found for the relationship between heart failure and later coronary heart disease or atrial fibrillation and stroke and later CVD following removal of SNPs. It must however be noticed that these relationships still showed $P<0.05$ also following removal of risk factor associated SNPs (Figure 4).

The Mendelian randomization approach showed relationships for pairs of CVDs both when one disease was used as the exposure and the other as outcome, as well when this order was reversed. This is in accordance with the observational data, and could be interpreted in two ways. Either is this due to etiological factors that are related to both the genetic

instrument and the outcome, as for the pair of coronary heart disease and stroke could be atherosclerosis. Or it could be due to the fact that different mechanisms are involved depending on the temporal order between the CVDs. For example, atrial fibrillation could attenuate the pumping capacity of the left ventricle leading to heart failure, while heart failure with increased filling pressures could dilate the heart, including the left atrium, which might trigger atrial fibrillation. Therefore, it is in this study hard to tell if a significant estimate in Mendelian randomization in the present setting should be interpreted as causal or not.

Shared genetic loci

Analysis of shared genetic loci between these disorders showed that most such loci were linked to common traditional risk factors, but also shared genetic loci with other functions were disclosed.

When using summary data from already published GWAS for the four major CVDs, no loci was related to all four CVDs, using either $P<5 \times 10^{-8}$ or the more liberal FDR<0.05. Four loci were related to three out of the four CVDs, and 22 loci were associated with two of the four traits using the more liberal limit of significance. Generally, most of the identified loci have previously been associated with traditional risk factors for CVD, such as hypertension, hyperlipidemia, diabetes or obesity (Table 4). This finding possibly represents that these four CVDs share the traditional risk factors, but for some loci no such known relationships *vs* traditional risk factors were found.

The SNP rs17042076, being intergenic with the nearest gene *RP11-119H12.3*, was related to heart failure, stroke, and atrial fibrillation, but has not been found in the GWASs *vs* traditional risk factors. Very little is known about the function of this locus, except that *RP11-119H12.3* is expressed in the brain, heart and testis (according to GTEx, https://gtexportal.org/home/).

Also rs3176326 in the intron of *CDKN1A* was related to the same triplet of CVDs, but not to traditional risk factors. In this case, however, other loci within this gene have been linked to hyperlipidemia and high blood pressure. Also for some of the loci being related to pairs of CVDs, associations *vs* traditional risk factors were not found between the SNP given in Table 4, but other SNPs within the gene show such relationships. This was however not found for *PLPP3*, *LRCH1*, and *BCAP29*, so it would be interesting to explore the functions of these genes more in detail to see if they could shed new light on the pathogenesis of CVDs.

For *IL6R* and *ZFP57*, no SNPs in the gene were related to traditional risk factors, but links have been published *vs* proinflammatory markers in blood as well as inflammatory diseases, possibly highlighting the role of inflammation in CVD.[10, 11]

Pathway enrichment analyses performed for the genes being related to pairs or triplet of the CVDs (Figure 4) highlighted also other pathways that might be of pathophysiological importance. For example, for loci being related to both coronary heart disease and stroke, bile acid metabolism was amongst the top ranked pathways, but also neuronal differentiation, actin dynamics, notch signaling and SCF signaling were enriched pathways for these two CVDs. Thus, while the pathway analysis brought up pathways being related to traditional risk factors, mainly lipids, also pathways not well-known for CVDs were disclosed.

Clinical implications

While connections between some CVDs, like atrial fibrillation and later stroke or heart failure is well-known for the clinician, the present study disclosed that also other CVDs are associated and the fact that one individual has a CVD diagnosis leads to an increased risk for also other CVDs. Thus, careful control of common traditional risk factors would not only lower the risk for a recurrent event, but possibly also other CVDs, although shared risk factors only explained a part of the covariation of CVDs. The remaining part of the covariation was

due to other pathophysiological mechanisms and some of those were disclosed in the present pathway analysis, which might serve as a basis for future drug discovery efforts.

Strength and limitations

The major strength of the study is the long follow-up period from 50 to 90 years of age in the population-based sample, allowing us to capture most of the CVDs that occur during a lifetime. Another strength is that we performed a longitudinal observational study and genetic studies in parallel. The obvious limitation is that we only have men of European descent in our sample, and therefore the results have to be confirmed in women, as well as in other ethnic groups. The study sample was also too small to study the transitions from two to three or four CVDs in terms of relative risk, and therefore only the transition from the first to the second CVD could be properly evaluated.

It should also be pointed out that the diagnostic criteria for the CVDs, especially myocardial infarction, has changed during the four decades of follow-up, something that might have been affecting the results. This is however unavoidable during long follow-up periods.

Another limitation is that the number of SNPs that was used as instrumental variables for heart failure and stroke were small after removal of SNPs linked to risk factors, so those estimates might not be precise and likely underpowered. However, the largest GWASs to date published for stroke and heart failure were used to select the instrumental variables.

A significant interaction with time was seen in the multi-state models meaning that the relative risks for a transition from one certain CVD to another do vary depending on the age of the subjects. Unfortunately, the sample is too small to be stratified into different age-groups, so the risk ratios presented should be regarded as “average” risk ratios over a four

decade follow-up period. Further studies in far larger samples have to be conducted in order to produce age-specific risk ratios for the transitions given in Table 2.

Table 4 gives the overlap of the loci being related to 3 or 2 of the 4 CVDs. Also, the nearest gene is given for each locus and, based on those genes, pathway analyses were conducted. It must however be acknowledged that a locus is not always associated with the expression of the nearest gene. However, many of the loci presented in Table 4 are not a significant eQTL, or are associated with expression of different genes. Therefore, we have for the sake of simplicity used the nearest gene to describe the genetic region of interest, being fully aware that the nearest gene not always is the gene of interest.

Conclusion

During 40 years, all of the four investigated CVDs were pairwise associated with each other regardless of the temporal order of occurrence, but the risk magnitude differed between different CVDs and their temporal order. Genetic analyses disclosed new pathophysiological links between CVDs.

ACKNOWLEDGEMENTS

None

SOURCES OF FUNDING

The study was funded by the Swedish Heart and Lung foundation and Uppsala University Hospital (ALF-medel).

DISCLOSURES

None of the authors had any conflict of interest regarding this study

SUPPLEMENTAL MATERIAL

- Supplementary Methods
- Supplementary Table 1
- Supplementary Figure 1

REFERENCES

1. Wolf PA, Abbott RD and Kannel WB. Atrial fibrillation as an independent risk factor for stroke: the Framingham Study. *Stroke*. 1991;22:983-8.

2. Goyal A, Norton CR, Thomas TN, Davis RL, Butler J, Ashok V, et al. Predictors of incident heart failure in a large insured population: a one million person-year follow-up study. *Circ Heart Fail*. 2010;3:698-705.

3. Wolf PA, D'Agostino RB, Belanger AJ and Kannel WB. Probability of stroke: a risk profile from the Framingham Study. *Stroke*. 1991;22:312-8.

4. Alonso A, Krijthe BP, Aspelund T, Stepas KA, Pencina MJ, Moser CB, et al. Simple risk model predicts incidence of atrial fibrillation in a racially and geographically diverse population: the CHARGE-AF consortium. *J Am Heart Assoc*. 2013;2:e000102.

5. Lind L, Sundström J, Ärnlöv J and Lampa E. Impact of Aging on the Strength of Cardiovascular Risk Factors: A Longitudinal Study Over 40 Years. *J Am Heart Assoc*. 2018;7:e007061.

6. Alqahtani F, Aljohani S, Tarabishy A, Busu T, Adcock A and Alkhouli M. Incidence and Outcomes of Myocardial Infarction in Patients Admitted With Acute Ischemic Stroke. *Stroke*. 2017;48:2931-2938.

7. Seifi A, Carr K, Maltenfort M, Moussouttas M, Birnbaum L, Parra A, et al. The incidence and risk factors of associated acute myocardial infarction (AMI) in acute cerebral ischemic (ACI) events in the United States. *PLoS One*. 2014;9:e105785.

8. Behar S, Tanne D, Abinader E, Agmon J, Barzilai J, Friedman Y, et al. Cerebrovascular accident complicating acute myocardial infarction: incidence, clinical significance and short- and long-term mortality rates. The SPRINT Study Group. *Am J Med*. 1991;91:45-50.

9. Naderi N, Masoomi H, Mozaffar T and Malik S. Patient characteristics and comorbidities associated with cerebrovascular accident following acute myocardial infarction in the United States. *Int J Cardiol*. 2014;175:323-7.

10. Kottoor SJ and Arora RR. The Utility of Anti-Inflammatory Agents in Cardiovascular Disease: A Novel Perspective on the Treatment of Atherosclerosis. *J Cardiovasc Pharmacol Ther*. 2018;23:483-493.

11. Nguyen MT, Fernando S, Schwarz N, Tan JT, Bursill CA and Psaltis PJ. Inflammation as a Therapeutic Target in Atherosclerosis. *J Clin Med*. 2019;8:1109.

FIGURE LEGENDS

Figure 1. Cumulative incidence of the four major cardiovascular diseases, myocardial infarction (MI), ischemic stroke (Stroke), heart failure (HF) and atrial fibrillation (AF) from age 50 to 90 years.

Figure 2. Relationships between the four major cardiovascular diseases during 40 years of follow-up in the longitudinal ULSAM study given as odds ratios (OR) and 95%CI (hi/lo). No attention is paid to the temporal order of occurrence of the diseases in this evaluation. ORs are given for both an unadjusted analysis and an analysis adjusted for traditional cardiovascular risk factors at age 50. HF, heart failure; MI, myocardial infarction; AF, atrial fibrillation; Stroke, ischemic stroke.

Figure 3. Flow chart of transitions from the healthy state (E) to a first event of atrial fibrillation (AF), myocardial infarction (MI), heart failure (HF) or ischemic stroke (Stroke). Thereafter follows the transitions to two, three and four of these diseases. The thickness of the "flow" is proportional to the number of subjects.

Figure 4. Venn diagram showing an overview of the genetic overlapping associations between three or two of the four cardiovascular disease traits when FDR<0.05 was used as cut-off for associations *vs* each trait. In addition, loci in or near *RP11-119H12.3* and *CDKN1A* were related to the triplet heart failure, stroke, and atrial fibrillation. See Table 4 for further details.

Table 1. Pairwise relationships between four major cardiovascular diseases.

	Total N	Overlapping with ischemic stroke, OR (95%CI), *P*-value, before/after stroke	Overlapping with heart failure, OR (95%CI), *P*-value, before/after HF	Overlapping with atrial fibrillation, OR (95%CI), *P*-value, before/after AF
Myocardial infarction	539	n = 101 OR 1.77 (1.38-2.27) *P*=4.7e-06 69/32	n = 151 OR 3.78 (3.01-4.73) *P*=9.1e-31 124/27	n = 142 OR 1.21 (0.97-1.50) *P*=0.09 94/48
Ischemic stroke	340		n = 68 OR 1.63 (1.24-2.14) *P*=4.4e-04 44/24	n = 131 OR 2.45 (1.93-3.11) *P*=2.0e-13 62/69
Heart failure	415			n = 180 OR 3.86 (3.08-4.83) *P*=9.2e-32 71/109
Atrial fibrillation	565			

In the second column from the left the total number of cases that occurred during a 48-year follow-up are given. In the following columns, the number of cases overlapping between the diseases are given together with odds ratio (OR), 95%CI, *P*-value and how many of the overlapping cases that occurred before or after the other cardiovascular disease in the respective pair. No adjustment for cardiovascular risk factors were performed in the analyses. Stroke, ischemic stroke; HF, heart failure; AF, atrial fibrillation.

Table 2. Multi-stage modeling of cardiovascular outcomes that are followed by another cardiovascular disease (CVD). The analysis only includes non-lethal cases.

Second CVD outcome	Prior CVD	RR	95% CI lower limit	95% CI upper limit
HF	MI	16.15	12.46	20.93
	AF	26.08	19.88	34.22
	Stroke	5.86	3.00	10.74
MI	HF	8.82	5.35	14.55
	AF	4.52	3.06	6.68
	Stroke	5.80	3.66	9.21
AF	HF	24.78	17.95	34.10
	MI	5.46	4.12	7.24
	Stroke	12.74	9.15	17.74
Stroke	HF	5.57	2.36	13.99
	MI	9.14	6.66	12.56
	AF	11.54	7.97	16.72

The risk ratios (RR) are calculated as those who experience the outcome with a prior CVD *vs* those with the same outcome but without any prior CVD. All relationships showed $P<0.0001$. HF, heart failure; MI, myocardial infarction; AF, atrial fibrillation; Stroke, ischemic stroke.

Table 3. Bidirectional two-sample Mendelian randomization studies for pairwise relationships between the four major cardiovascular diseases.

CHD vs stroke (n = 41)				
Test	Estimate	95%CI lower	95%CI higher	*P*-value
MR Egger	1.10	0.87	1.37	0.41
IVW	1.19	1.14	1.23	2.2e-15
Weighted median	1.17	1.07	1.26	0.00025
Heterogeneity	Q-test = 213			<0.0001
Pleiotropy	0.0075	-0.0142	0.029	0.49
CHD vs AF (n = 39)				
Test	Estimate	95%CI lower	95%CI higher	*P*-value
MR Egger	1.04	0.90	1.19	0.57
IVW	1.05	1.03	1.09	0.00023
Weighted median	1.05	1.004	1.11	0.032
Heterogeneity	Q-test = 145			<0.0001
Pleiotropy	0.0017	-0.0113	0.015	0.79
CHD vs HF (n = 41)				
Test	Estimate	95%CI lower	95%CI higher	*P*-value
MR Egger	1.39	1.24	1.57	5.5e-08
IVW	1.34	1.30	1.39	<1.0e-25
Weighted median	1.37	1.29	1.45	<1.0e-25
Heterogeneity	Q-test = 154			<0.0001
Pleiotropy	-0.0039	-0.0151	0.007	0.50
AF vs stroke (n = 113)				
Test	Estimate	95%CI lower	95%CI higher	*P*-value
MR Egger	1.23	1.08	1.40	0.0014
IVW	1.24	1.19	1.29	1.1e-24
Weighted median	1.21	1.14	1.29	1.7e-09
Heterogeneity	Q-test = 152			0.0053
Pleiotropy	0.0004	-0.007	0.008	0.92
AF vs HF (n = 113)				
Test	Estimate	95%CI lower	95%CI higher	*P*-value
MR Egger	1.22	1.11	1.34	0.000059
IVW	1.22	1.18	1.26	<1.0e-25
Weighted median	1.19	1.13	1.25	3.00e-12
Heterogeneity	Q-test = 152			0.0066
Pleiotropy	-0.0001	-0.0056	0.005	0.98
AF vs CHD (n=113)				
Test	Estimate	95%CI lower	95%CI higher	*P*-value
MR Egger	0.90	0.78	1.03	0.11
IVW	1.02	0.98	1.06	0.31

Weighted median	0.99	0.93	1.05	0.76
Heterogeneity	Q-test = 190			<0.0001
Pleiotropy	0.0077	0.0001	0.015	0.046
Stroke vs HF (n = 10)				
Test	Estimate	95%CI lower	95%CI higher	*P*-value
MR Egger	0.72	0.20	2.52	0.60
IVW	1.30	1.19	1.42	5.7e-09
Weighted median	1.27	1.11	1.46	0.00059
Heterogeneity	Q-test = 35			0.0001
Pleiotropy	0.040	-0.044	0.12	0.34
Stroke vs CHD (n = 10)				
Test	Estimate	95%CI lower	95%CI higher	*P*-value
MR Egger	4.39	0.98	19.7	0.052
IVW	1.51	1.36	1.67	<1.0e-25
Weighted median	1.31	1.10	1.57	0.0022
Heterogeneity	Q-test = 36			0.0001
Pleiotropy	-0.073	-0.17	0.029	0.15
Stroke vs AF (n = 9)				
Test	Estimate	95%CI lower	95%CI higher	*P*-value
MR Egger	1.02	0.28	3.67	0.98
IVW	1.14	1.04	1.23	0.0045
Weighted median	1.05	0.93	1.20	0.43
Heterogeneity	Q-test = 30			<0.0001
Pleiotropy	0.0078	-0.081	0.097	0.86
HF vs CHD (n = 9)				
Test	Estimate	95%CI lower	95%CI higher	*P*-value
MR Egger	4.33	0.82	23.00	0.085
IVW	2.11	1.83	2.43	<1.0e-25
Weighted median	1.76	1.22	2.53	0.0024
Heterogeneity	Q-test =78			<0.0001
Pleiotropy	-0.041	-0.13	0.051	0.38
HF vs AF (n = 9)				
Test	Estimate	95%CI lower	95%CI higher	*P*-value
MR Egger	1.98	0.58	6.65	0.26
IVW	1.58	1.42	1.76	<1.0e-25
Weighted median	1.52	1.28	1.81	1.96e-06
Heterogeneity	Q-test = 928			<0.0001
Pleiotropy	-0.012	-0.078	0.053	0.70
HF vs stroke (n = 9)				
Test	Estimate	95%CI lower	95%CI higher	*P*-value
MR Egger	1.63	0.56	4.69	0.36
IVW	1.66	1.43	1.93	2.9e-11

Weighted median	1.64	1.28	2.08	0.000062
Heterogeneity	Q-test = 23			0.0018
Pleiotropy	0.0011	-0.057	0.060	0.96

The number of SNPs used in the analysis as instrumental variable for the exposure is given as n. In this analysis, all available SNPs were used as instruments. Pleiotropy is given by the MR Egger intercept. HF, heart failure; CHD, coronary heart disease; AF, atrial fibrillation; Stroke, ischemic stroke; IVW, inverse-variance weighted meta-analysis.

Table 4. Genetic loci being shared for two or three of the four cardiovascular traits; coronary heart disease, stroke, heart failure or atrial fibrillation, at the FDR<0.05 level.

rs-number	Position (hg19)	EAF	Effective /Other allele	Nearest gene	Traits	Relations to CV risk factors according to Phenoscanner
Three traits						
rs17042076*	chr4:111652338	0.1551	T/C	*RP11-119H12.3* (intergenic)	HF, Stroke, AF	-
rs532436*	chr9:136149830	0.1859	A/G	*ABO* (intron)	CHD, Stroke, AF	Lipids, blood pressure, diabetes, trunk fat
rs4766578*	chr12:111904371	0.5229	A/T	*ATXN2* (intron)	CHD, Stroke, AF	Lipids, blood pressure, BMI, smoking
rs3176326	chr6:36647289	0.1730	A/G	*CDKN1A* (intron)	HF, Stroke, AF	-
Two traits						
rs660240*	chr1:109817838	0.1988	T/C	*CELSR2* (3'-UTR)	CHD, HF	Lipids, BMI
rs10908838*	chr1:154397984	0.5567	G/T	*IL6R* (intron)	CHD, AF	-
rs11745324*	chr5:137012171	0.2296	A/G	*KLHL3* (intron)	HF, AF	Blood pressure, pulse rate
rs55730499*	chr6:161005610	0.9245	C/T	*LPA* (intron)	CHD, HF	Lipids
rs2107595*	chr7:19049388	0.8320	G/A	*HDAC9* (intergenic)	CHD, Stroke	Blood pressure, trunk fat

rs1333043*	chr9:22106731	0.5070	A/T	*CDKN2B-AS1* (intron)	CHD, HF	Lipids
rs60212594*	chr10:75414344	0.1372	C/G	*SYNPO2L* (intron)	AF, Stroke	Blood pressure, pulse rate
rs10774624*	chr12:111833788	0.5298	A/G	*RP3-473L9.4* (intron)	CHD, Stroke	Blood pressure, weight, smoking
rs17035646	chr1:10796547	0.3529	A/G	*CASZ1* (intron)	AF, Stroke	Hypertension
rs2404716	chr1:56979076	0.2714	A/G	*PLPP3* (intron)	CHD, HF	-
rs4076834	chr2:44081627	0.9245	T/G	*ABCG8* (intron)	CHD, Stroke	Lipids
rs3820888	chr2:201180023	0.6223	T/C	*SPATS2L* (intron)	AF, HF	Pulse rate
rs12509595	chr4:81182554	0.2674	C/T	*RP11-576N17.4* (Intergenic)	CHD, AF	Hypertension, lipids
rs6909574	chr6:22606773	0.6362	A/G	*ZFP57* (Intergenic)	CHD, AF	-
rs7769954	chr6:134196381	0.7097	A/G	*TARID* (intron)	CHD, AF	Pulse rate, hypertension
rs68170813	chr7:107259721	0.2117	C/T	*BCAP29* (3’-UTR)	CHD, AF	-
rs2980858	chr8:126501177	0.3231	T/C	*RP11-136O12.2* (intron)	CHD, HF	Lipids, trunk fat
rs11000775	chr10:75539010	0.1342	C/T	*CHCHD1* (upstream)	AF, HF	Blood pressure, pulse rate
rs72841270	chr10:104642237	0.1441	G/T	*AS3MT* (Intron)	CHD, AF	Blood pressure, BMI

rs4284534	chr13:47267033	0.7624	A/G	*LRCH1* (intron)	AF, Stroke	-
rs35346340	chr15:91427872	0.3250	C/G	*FES* (splice region)	CHD, Stroke	Blood pressure
rs10405536	chr19:10756432	0.5974	A/G	*SLC44A2* (downstream)	CHD, Stroke	-

*P<5e-8. EAF, effective allele frequency; CV, cardiovascular; HF, heart failure; CHD, coronary heart disease; AF, atrial fibrillation; Stroke, ischemic stroke; BMI, body mass index.

SUPPLEMENTAL MATERIAL

Life-Time Covariation of Major Cardiovascular Diseases - A 40 Year Longitudinal Study and Genetic Studies

Lars Lind, Johan Sundström, Johan Ärnlöv, Martin Ingelsson, Albert Henry, R. Thomas Lumbers, Erik Lampa

CONTENTS

SUPPLEMENTARY METHODS

Sample

Between 1970-1974, all men aged 50 years living in the City of Uppsala, Sweden, were invited to a Health screening program for cardiometabolic risk factors (the Uppsala Longitudinal Study of Adult Men, ULSAM). The participation rate was 82 % resulting in a sample of 2322 investigated individuals. The study was approved by the Ethics Committee of Uppsala University and each participant gave their informed consent. This cohort has been followed for four decades and physically re-investigated at 60, 70, 77, 82, 88 and 92 years of age.

Traditional risk factors

The examination at age 50 has been described in detail previously.[1] Blood samples for fasting concentrations were drawn in the morning after an overnight fast. Cholesterol and triglyceride concentrations in serum, and high density lipoprotein (HDL) were assayed by enzymatic techniques. Low density lipoprotein (LDL)-cholesterol was calculated by Fiedewald's formula. Fasting blood glucose was determined by an oxidase method. Supine systolic and diastolic blood pressures were measured twice in the right arm after 10 minutes rest, and means were calculated. Information on current smoking was collected by a questionnaire.

CVD diagnosis

Date and cause of death were obtained from the Swedish Cause of Death Register. Date and cause of hospitalization were obtained from the Swedish Hospital Discharge Register in all individuals. We evaluated four major cardiovascular diseases; acute myocardial infarction

(International Classification of Diseases [ICD-8] code 410, ICD- 9 code 410, or ICD-10 code I20), ischemic stroke (ICD-8 codes 431, 433-436, ICD-9 code 431, 433-436, ICD-10 code I63-I66), heart failure (ICD-8 codes 427.00, 427.10, 428.99, ICD-9, 428 and ICD-10 code I50 as well as hypertensive heart disease with heart failure (I11.0 [ICD-10]), and atrial fibrillation (ICD-8 code 427.9, ICD-9 427D and ICD-10 I48). In addition, we noted if atrial fibrillation was present at ECG at the physical re-examinations. Regarding myocardial infarction, stroke and atrial fibrillation these diagnoses have been evaluated as accurate in the Swedish registers,[2] while the diagnosis of heart failure is less valid.[3] Therefore, all cases of heart failure were validated by use of medical records by an experienced clinician (L.L.). Data on CVD diagnosis and mortality were available in all subjects during the follow-up period.

Statistics

The calculations for the first aim was performed in two steps. In step one, a pair-wise cross-tabulation was performed for the four CVDs and it was noted which of the diseases within each pair that occurred first. An odds ratio (OR) was calculated by logistic regression for the covariation of the CVDs. This was done both unadjusted and adjusted for the traditional risk factors, systolic blood pressure, body mass index (BMI), diabetes, HDL and LDL-cholesterol and smoking.

In step two, a multi-state model was set up according to Figure 3. Only non-lethal cases were used in the following analysis. The states were not distinguished by order of subsequent CVDs, i.e. an individual experiencing a myocardial infarction first and then a stroke ends up in the same combined state as an individual who first experience a stroke and then an myocardial infarction. Transition rates between the states were modeled with Poisson regression on time-split data, where age was used as the time scale and follow-up was divided into one year intervals in which the rates were assumed to be constant. Details of this approach is described

elsewhere.[4] A time gap of at least one day between the first and second CVD events was demanded, and that these two events should take place during separate hospitalization periods. The rate ratios given are based on a comparison of estimated rates of experiencing a second CVD *vs* estimated rates of experiencing the same CVD but without a prior CVD. The multi-state analyses were made using R (version 3.6.1; R Foundation for Statistical Computing, Vienna, Austria) and the Epi package. The calculations for the second aim was also performed in two steps. The R code used for the calculation in the multi-state models are given at the end of the supplementary material.

In step three, bidirectional, two-sample Mendelian randomization (MR) studies were performed using already published GWASs for the four major CVDs (CARDIoGRAMplusC4D for coronary heart disease,[5] HERMES for heart failure,[6] METASTROKE for ischemic stroke,[7] and the recent large GWAS on atrial fibrillation by Roselli et al[8]).

As instrument for the exposure, only independent genetic variants (SNPs) with $P<5 \times 10^{-8}$ were used. Independency of these variants was evaluated by the clump command in the package MRbase in R (3.6.1). Thereafter, the causal estimate was calculated by both inverse-variance weighted meta-analysis (IVW) and MR Egger, as well as the weighted-median method. Outlier genetic instruments with an IVW estimate greater than twice of the next best/worst instruments were removed from the analysis if the *P*-value for heterogeneity was less than 0.0001.

To investigate the impact of known shared risk factors, first an MR analysis with all available independent SNP instruments was performed and thereafter SNPs showing $P<5 \times 10^{-8}$ for association with traditional risk factors (blood pressure, diabetes, obesity, LDL- and HDL-cholesterol, triglycerides and smoking) were removed following an evaluation of the published GWAS studies from DIAGRAM,[9] GIANT,[10] Global Lipids Consortium,[11] as well

as GWAS analyses of UK biobank from Neal Lab (http://www.nealelab.is/uk-biobank), as well as own analyses.[12] In the MR analyses, IVW was regarded as the primary analysis, whereas MR Egger and the weighted median method were used as sensitivity analyses.

In step four, we used the already published GWAS data for the four major CVDs to search for overlapping genetic associations. First, only independent genetic loci with $P<5 \times 10^{-8}$ were evaluated regarding overlap between the traits. Independency (LD<0.001) of these variants was evaluated by the clump command in the package MRbase in R (3.6.1). Secondly, in a predefined step also associations with FDR<0.05 (the Benjamani-Hochberg method) were investigated. The overlapping loci (nearest gene) were subjected to pathway enrichment analysis using the Reactome software (https://reactome.org/PathwayBrowser/), which performs a hypergeometric distribution test producing a probability score corrected for FDR. Step one, three and four used STATA14 (Stata inc, College Station, TX, USA) for calculations if not stated otherwise.

R-code used for the calculations of the multi-stage models.

The following multi-state model was used to analyse the study data. It is assumed that the data are structured so that each row corresponds to one individual being in exactly one state as well as the entry and exit times. In our data, the entry and exit times are named ”start” and “stop”, respectively. We also have each individual’s birth date (bdate) and use age as the time scale. The state an individual has entered at each row is contained in the variable ”starts” and “states” holds the state to which an individual exits to.

First we load the Epi package and create a Lexis object:

```
library("Epi")

lc <- Lexis(entry = list(age = start - bdate),
            exit = list(age = stop - bdate),
            entry.status = starts,
            exit.status =  states,
            id =  pat, data = data)
```

We then split the follow-up time for each individual into one-year intervals and stack the time-split data:

```
lc_split <- splitLexis(lc, breaks = 40:100)
lc_stack <- stack.Lexis(lc_split)
```

We model age using regression splines and set the knots at the 5^{th}, 35^{th}, 65^{th} and 95^{th} percentiles of the distribution of ages at which an event occurred:

```
a.kn <- with(subset(lc_stack, lex.Fail),
             quantile(age + lex.dur, c(0.05, 0.35, 0.65, 0.95)))
```

We then fit the Poisson model:

```
fit <- glm(cbind(lex.Fail, lex.dur) ~ -1 + Ns(age, knots = a.kn) + lex.Tr,
           family = poisreg, data = lc_stack)
```

The Rate Ratios for the transitions of interest could be obtained in many ways. One of them is to use the multcomp package:

```
library("multcomp")

## Set up the contrast matrix
L <- matrix(rep(0, 12 * length(coef(fit))), ncol = length(coef(fit)))
colnames(L) <- names(coef(fit))
L[1, "lex.TrMI->HF+MI"] <- 1
L[1, "lex.TrE->HF"] <- -1
L[2, "lex.TrAF->AF+HF"] <- 1
L[2, "lex.TrE->HF"] <- -1
L[3, "lex.TrStroke->HF+Stroke"] <- 1
L[3, "lex.TrE->HF"] <- -1
L[4, "lex.TrHF->HF+MI"] <- 1
L[4, "lex.TrE->MI"] <- -1
L[5, "lex.TrAF->AF+MI"] <- 1
L[5, "lex.TrE->MI"] <- -1
L[6, "lex.TrStroke->MI+Stroke"] <- 1
L[6, "lex.TrE->MI"] <- -1
L[7, "lex.TrHF->AF+HF"] <- 1
L[7, "lex.TrE->AF"] <- -1
L[8, "lex.TrMI->AF+MI"] <- 1
L[8, "lex.TrE->AF"] <- -1
L[9, "lex.TrStroke->AF+Stroke"] <- 1
L[9, "lex.TrE->AF"] <- -1
L[10, "lex.TrHF->HF+Stroke"] <-
1
L[10, "lex.TrE->Stroke"] <- -1
L[11, "lex.TrMI->MI+Stroke"] <-
1
L[11, "lex.TrE->Stroke"] <- -1
L[12, "lex.TrAF->AF+Stroke"] <-
1
L[12, "lex.TrE->Stroke"] <- -1

rownames(L) <- c("MI_HF", "AF_HF", "Stroke_HF",
                 "HF_MI", "AF_MI", "Stroke_MI",
```

```
g <- glht(fit, linfct = L)
ci <- confint(g, calpha = univariate_calpha())
```

ci contains the log-RR with corresponding confidence intervals.

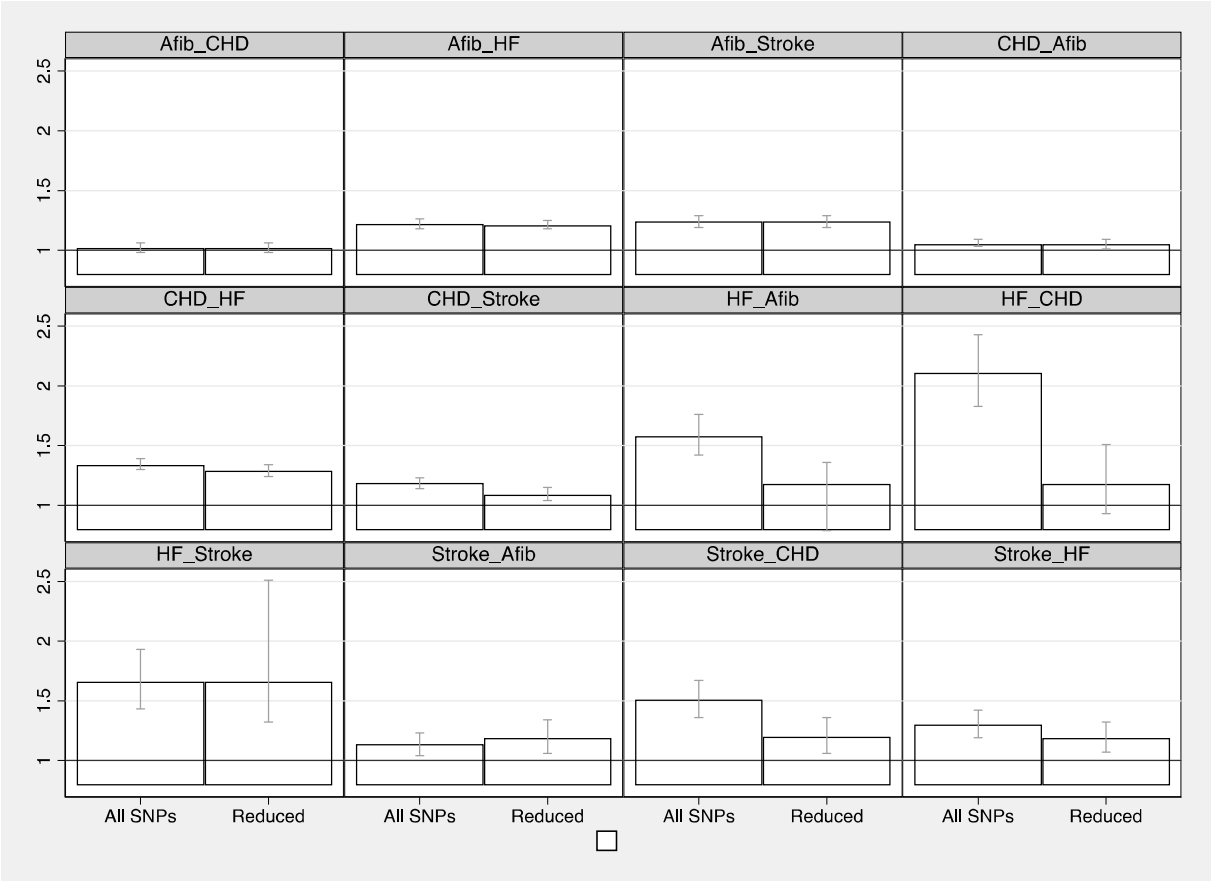
Supplementary Table 1. Bidirectional two-sample Mendelian randomization studies for pairwise relationships between the four major cardiovascular diseases.

CHD *vs* Stroke (n=27)				
Test	Estimate	95%CI lower	95%CI higher	*P*-value
MR Egger	1.10	0.86	1.39	0.44
IVW	1.09	1.04	1.15	0.00090
Weighted median	1.16	1.05	1.28	0.0036
Heterogeneity	Q-test=81			<0.0001
Pleiotropy	-0.0007	-0.022	0.021	0.94
CHD *vs* AFib (n=27)				
Test	Estimate	95%CI lower	95%CI higher	*P*-value
MR Egger	0.95	0.78	1.15	0.61
IVW	1.05	1.01	1.09	0.015
Weighted median	1.05	0.99	1.11	0.11
Heterogeneity	Q-test=95			<0.0001
Pleiotropy	0.0093	-0.0082	0.027	0.29
CHD *vs* HF (n=27)				
Test	Estimate	95%CI lower	95%CI higher	*P*-value
MR Egger	1.32	1.16	1.50	0.000031
IVW	1.29	1.24	1.34	<1.0e-25
Weighted median	1.36	1.28	1.45	<1.0e-25
Heterogeneity	Q-test=38			0.048
Pleiotropy	-0.0021	-0.0137	0.009	0.72
AFib *vs* Stroke (n=107)				
Test	Estimate	95%CI lower	95%CI higher	*P*-value
MR Egger	0.21	0.075	0.34	0.0020
IVW	1.24	1.19	1.29	0
Weighted median	0.19	0.13	0.26	2.48e-09
Heterogeneity	Q-test=150			0.0039
Pleiotropy	0.0005	-0.0069	0.008	0.89
AFib *vs* HF (n=107)				
Test	Estimate	95%CI lower	95%CI higher	*P*-value
MR Egger	0.20	0.10	0.30	0.000074
IVW	1.21	1.18	1.25	0
Weighted median	0.17	0.12	0.23	2.68e-11
Heterogeneity	Q-test=149			0.0050
Pleiotropy	-0.0002	-0.0059	0.005	0.92
AFib *vs* CHD (n=107)				
Test	Estimate	95%CI lower	95%CI higher	*P*-value
MR Egger	0.90	0.79	1.03	0.13
IVW	1.02	0.98	1.06	0.44

Weighted median	0.99	0.92	1.05	0.71
Heterogeneity	Q-test=180			0.0001
Pleiotropy	0.0072	-0.0004	0.015	0.062
Stroke *vs* HF (n=7)				
Test	Estimate	95%CI lower	95%CI higher	*P*-value
MR Egger	0.40	0.10	1.64	0.20
IVW	1.19	1.07	1.32	0.0014
Weighted median	1.12	0.95	1.32	0.17
Heterogeneity	Q-test=15			0.0077
Pleiotropy	0.073	-0.020	0.16	0.12
Stroke *vs* CHD (n=7)				
Test	Estimate	95%CI lower	95%CI higher	*P*-value
MR Egger	1.52	0.56	4.22	0.40
IVW	1.20	1.06	1.36	0.0045
Weighted median	1.17	0.98	1.39	0.080
Heterogeneity	Q-test=3.3			0.64
Pleiotropy	-0.016	-0.083	0.051	0.63
Stroke *vs* AFib (n=6)				
Test	Estimate	95%CI lower	95%CI higher	*P*-value
MR Egger	0.81	0.14	4.69	0.81
IVW	1.19	1.06	1.34	0.0029
Weighted median	1.06	0.90	1.24	0.48
Heterogeneity	Q-test=15			0.0016
Pleiotropy	0.026	-0.093	0.14	0.66
HF *vs* CHD (n=4)				
Test	Estimate	95%CI lower	95%CI higher	*P*-value
MR Egger	No convergence			
IVW	1.18	0.93	1.51	0.16
Weighted median	1.22	0.84	1.78	0.30
Heterogeneity	No convergence			
Pleiotropy	No convergence			
HF *vs* AFib (n=5)				
Test	Estimate	95%CI lower	95%CI higher	*P*-value
MR Egger	No convergence			
IVW	1.18	0.99	1.42	0.068
Weighted median	1.04	0.79	1.36	0.77
Heterogeneity	No convergence			
Pleiotropy	No convergence			
HF *vs* Stroke (n=4)				
Test	Estimate	95%CI lower	95%CI higher	*P*-value
MR Egger	1.42	0.01	272	0.89
IVW	1.66	1.32	2.07	7.63e-06

Weighted median	1.82	1.32	2.51	0.00025
Heterogeneity	Q-test=13			0.0012
Pleiotropy	0.0079	-0.25	0.27	0.95

The number of SNPs used in the analysis as instrumental variable for the exposure is given as n. In this analysis, SNPs being related to traditional CV risk factors have been removed. CHD, coronary heart disease; AFib; atrial fibrillation; HF, heart failure; Stroke, ischemic stroke; MR Egger, Mendelian randomization Egger; IVW, inverse-variance weighted meta-analysis.



Supplementary Figure 1. Bidirectional Mendelian randomization analysis of the four major cardiovascular diseases. The first disease in the X-Y notation is the exposure. Odds ratio and 95%CI are shown. All SNPs denotes when all available SNPs are used as the genetic instrument for the exposure. Reduced is used when all SNPs related to the traditional cardiovascular risk factors were removed. Afib, atrial fibrillation; CHD, coronary heart disease; HF, heart failure; Stroke, ischemic stroke.

REFERENCES

1. Ärnlöv J, Ingelsson E, Sundström J and Lind L. Impact of body mass index and the metabolic syndrome on the risk of cardiovascular disease and death in middle-aged men. *Circulation*. 2010;121:230-6.

2. Merlo J, Lindblad U, Pessah-Rasmussen H, Hedblad B, Rastam J, Isacsson SO, et al. Comparison of different procedures to identify probable cases of myocardial infarction and stroke in two Swedish prospective cohort studies using local and national routine registers. *Eur J Epidemiol*. 2000;16:235-43.

3. Ingelsson E, Ärnlöv J, Sundström J and Lind L. The validity of a diagnosis of heart failure in a hospital discharge register. *Eur J Heart Fail*. 2005;7:787-91.

4. Carstensen BP, M. Using Lexis objects for multi-state models in R. *Journal of Statistical Software*. 2011;38:1-18.

5. CARDIoGRAMplusC4D Consortium, Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet*. 2013;45:25-33.

6. Shah S, Henry A, Roselli C, Lin H, Sveinbjörnsson G, Fatemifar G, et al. Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. *Nat Commun*. 2020;11:163.

7. Malik R, Chauhan G, Traylor M, Sargurupremraj M, Okada Y, Mishra A, et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet*. 2018;50:524-537.

8. Roselli C, Chaffin MD, Weng LC, Aeschbacher S, Ahlberg G, Albert CM, et al. Multi-ethnic genome-wide association study for atrial fibrillation. *Nat Genet*. 2018;50:1225-1233.

9. Scott RA, Scott LJ, Mägi R, Marullo L, Gaulton KJ, Kaakinen M, et al. An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. *Diabetes*. 2017;66:2888-2902.

10. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015;518:197-206.

11. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet*. 2013;45:1274-1283.

12. Lind L. Genetic Determinants of Clustering of Cardiometabolic Risk Factors in U.K. Biobank. *Metab Syndr Relat Disord*. 2020;18:121-127.