Supplementary Material

Methods

Descriptive statistics were presented as numbers, percentages, mean, medians and standard deviations. The patients were grouped according to their first cardiovascular diagnosis within the hospital admission ICD10 codes presented in Table 1 in the main document. If more than one diagnosis was given at the same day the A diagnosis, the primary cause of the health care contact, was used.

Charlson Comorbidity Index was calculated with the R package 'comorbidity'.[1] It is based on the Charlson score proposed by Quan et al. in 2005.[2] and includes the following comorbid conditions: acute myocardial infarction, congestive heart failure, peripheral vascular disease, cerebrovascular disease, dementia, chronic obstructive pulmonary disease [COPD], rheumatoid disease, peptic ulcer disease, mild and moderate/severe liver disease, diabetes mellitus with and without complications, hemiplegia/paraplegia, renal disease, cancer (any malignancy) and metastatic solid tumour, AIDS/HIV. ICD-8 codes were mapped to ICD-10 codes by an internal ICD-8 to ICD-10 mapping file. The diagnoses prior to the date of a patient's first cardiovascular disease, to inclusion date and end of follow-up or death were included in the Charlson Comorbidity Index calculations.

Cases and controls

In this study cases and controls were defined according to ICD10/ICD8 codes and prescription medicine from the National Patient Register and the Danish Prescription Database.[3,4] The population constitutes of participants from the Danish Blood Donor Study (n = 110 000) and patients from the Copenhagen Hospital Biobank - Cardiovascular Disease Cohort (CHB-CVDC) (n = 96 308).[5] A more detailed description of the phenotypes are presented below.

Phenotype descriptions is inspired by previously described phenotypes.[6,7]

Heart Failure

Laursen IH, *et al. BMJ Open* 2021; 11:e049709. doi: 10.1136/bmjopen-2021-049709

Hospitalization for ICD-10 code for heart failure (I50 and sub codes); or hospitalization for ICD-8 code for heart failure (42709, 42710, 42711, 42719, 78249).

Coronary Artery Disease

Hospitalization for ICD-10 code for coronary artery disease (I21, I22, I23, I24, I25); or hospitalization due to ICD-8 code for coronary artery disease (41199, 41099, 41409, 41499, 41299, 412909, 41009, 41109).

Atrial Fibrillation

Hospitalization for ICD-10 code for atrial fibrillation or atrial flutter (I48); or hospitalization for ICD-8 code for atrial fibrillation or atrial flutter (42793, 42794).

Hypertension

Hospitalization for ICD10 code for essential hypertension (I10), excluding following ICD10 codes from controls I11, I12, I13, I15; or hospitalization for ICD-8 code for essential hypertension (40009, 40299, 40199); and use of at least one antihypertensive drug, the following ATC-codes included: Renin–angiotensin system inhibitors C09; calcium channel blockers C08; b-blockers C07; diuretics C03; antiadrenergic drugs C02A, C02B, and C02C; and other antihypertensives C02DA, C02DB, C02DD, C02DG, and C02L. Controls with prescriptions of beforementioned atc codes were removed from the control group.

Cholesterol measurements

A patient's first measure of High-Density Lipoprotein (HDL), Low-Density Lipoprotein (LDL), Triglycerides (TG) and Total Cholesterol (TC) were retrieved from the Danish Laboratory Database. The following NPU codes were used: NPU01567 (HDL), NPU01568 (LDL), NPU01566 (TC) and NPU04094 (TG). The data for each NPU code used was checked for differences between laboratories.

Inclusion of previously published genetic variants

An extensive literature search was performed for each phenotype. The newest and most updated genome wide association study available, in the period this manuscript was written, of a given disease was included if it was published in English in a high impact peer reviewed journal and include populations of European ancestry. It was assumed that the most updated study was replicating previously found variants.

A genetic variant was included from the study if it was a significant independent variant and if minor allele frequency >1%. If a variant was not included in our dataset a proxy variant ($R^2 > 0.8$) was used, if possible. LDlink proxy tool with European population as reference was used to find proxy variants.[8]

Replication of included variants

For coronary artery disease, atrial fibrillation, heart failure and essential hypertension, the association between the included variants and disease were calculated using a logistic mixed model implemented by SAIGE assuming an additive genetic model and adjusted for year of birth, sex and the first 10 principal components (PCs) to correct for population structure. A variant was replicated if the effect size of the risk allele had the same direction of effect and a $P < 0.05$ (Bonferroni adjusted).

For low-density lipoprotein, high density lipoprotein, total cholesterol and triglyceride, Bolt-LMM was used to calculate associations to disease. The patients first measurement values were inverse rank normalized and adjusted for age at measurement, place of measurement, lipid lowering medications (ATC – C10 within 180 days before measurement) and sex. Residuals were used as the independent variable in association analysis adjusted for year of birth and 10 PCs. Effect sizes were calculated by regressing variants to the inverse rank normalized measurement values adjusted for sex, age at measurement, year of birth, lipid lowering medication, significant places of measurements and 10 PCs.

Effect sizes were retrieved from the included studies. Previously found effect sizes, risk allele frequencies and p-values were plotted against the obtained values from this study. Effect sizes were weighted by allele frequencies. Power calculations for each SNP were based on standard error and effect size and performed in R[9].

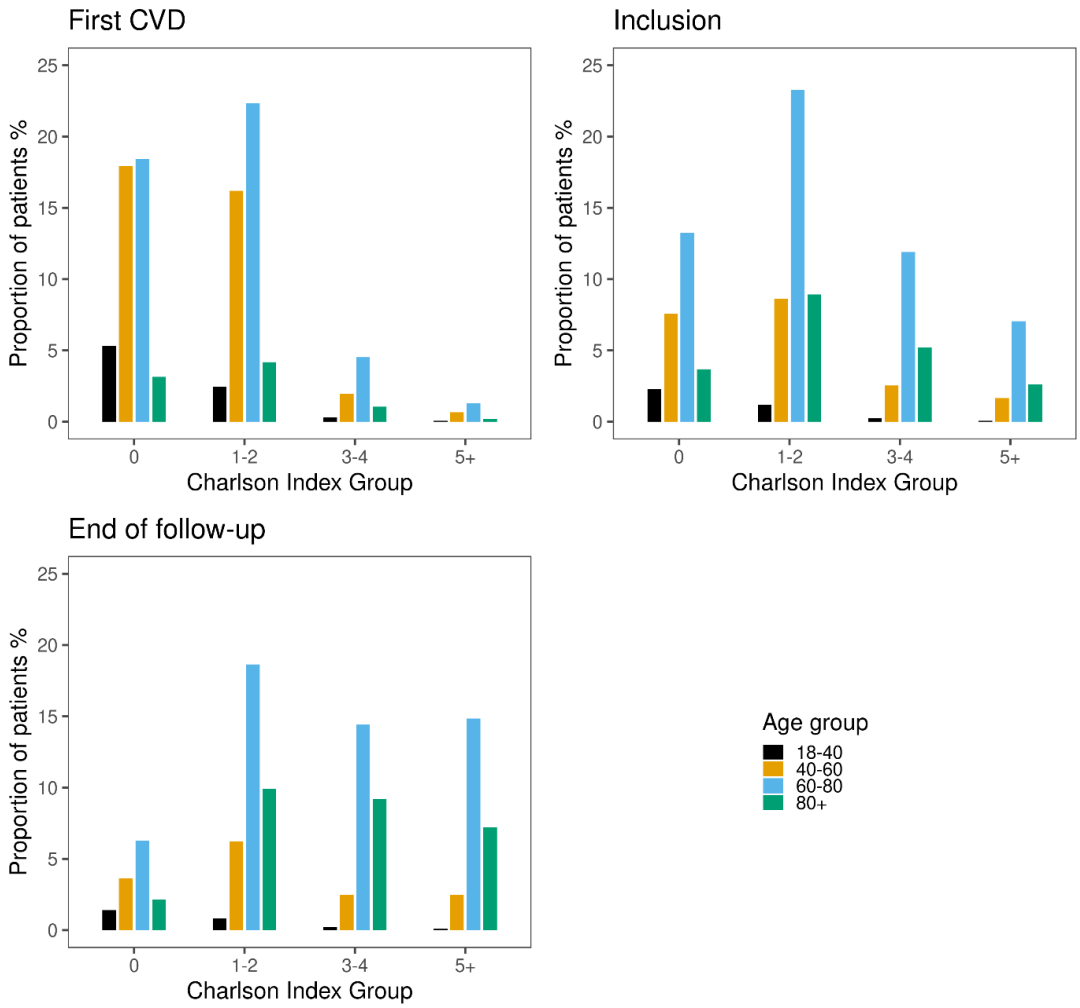
The following tools were used: PLINK 2[10,11], SAIGE 36.3.3[12], BOLT-LMM 2.3.4[13], flashpca 2.0[14], LDlink[8], Python 3 , Rstudio 1.2.1335[15].

Comparison of different study designs

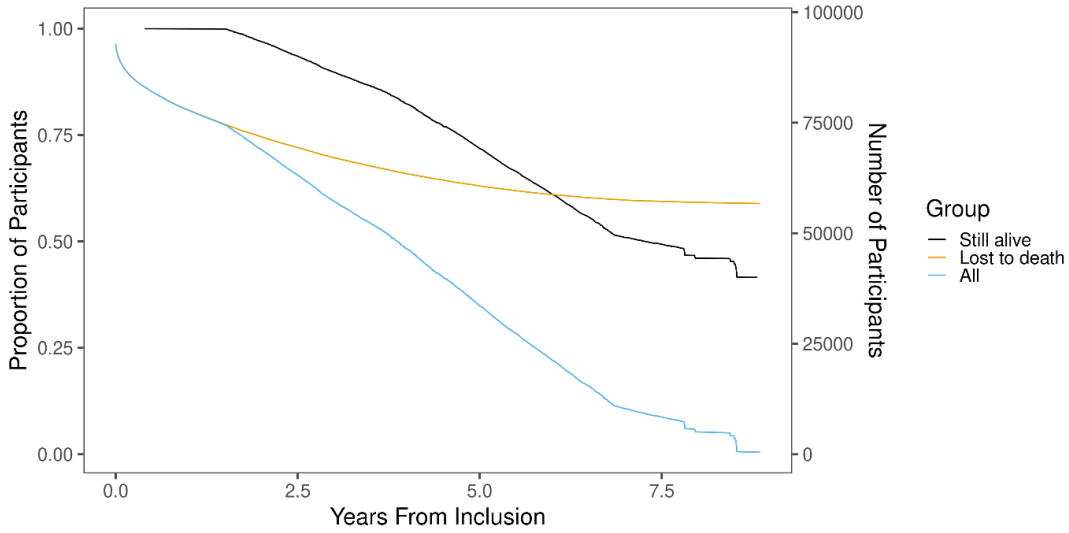
We evaluated the study designs with CHB-CVDC as the only controls and DBDS as the only controls respectively.

We employed LD Score regression to investigate residual confounding. Residual confounding was evaluated through the LD Score regression intercept and the attenuation ratio (the ratio between the intercept and the mean χ^2 statistic).[16] Genetic correlations were estimated using LD Score Regression with the 1000Genomes EUR v3 LD reference panel. Summary stats for a pre-selected list of phenotypes were retrieved from the GWAS catalog[17]. See Table 2 for a complete list.

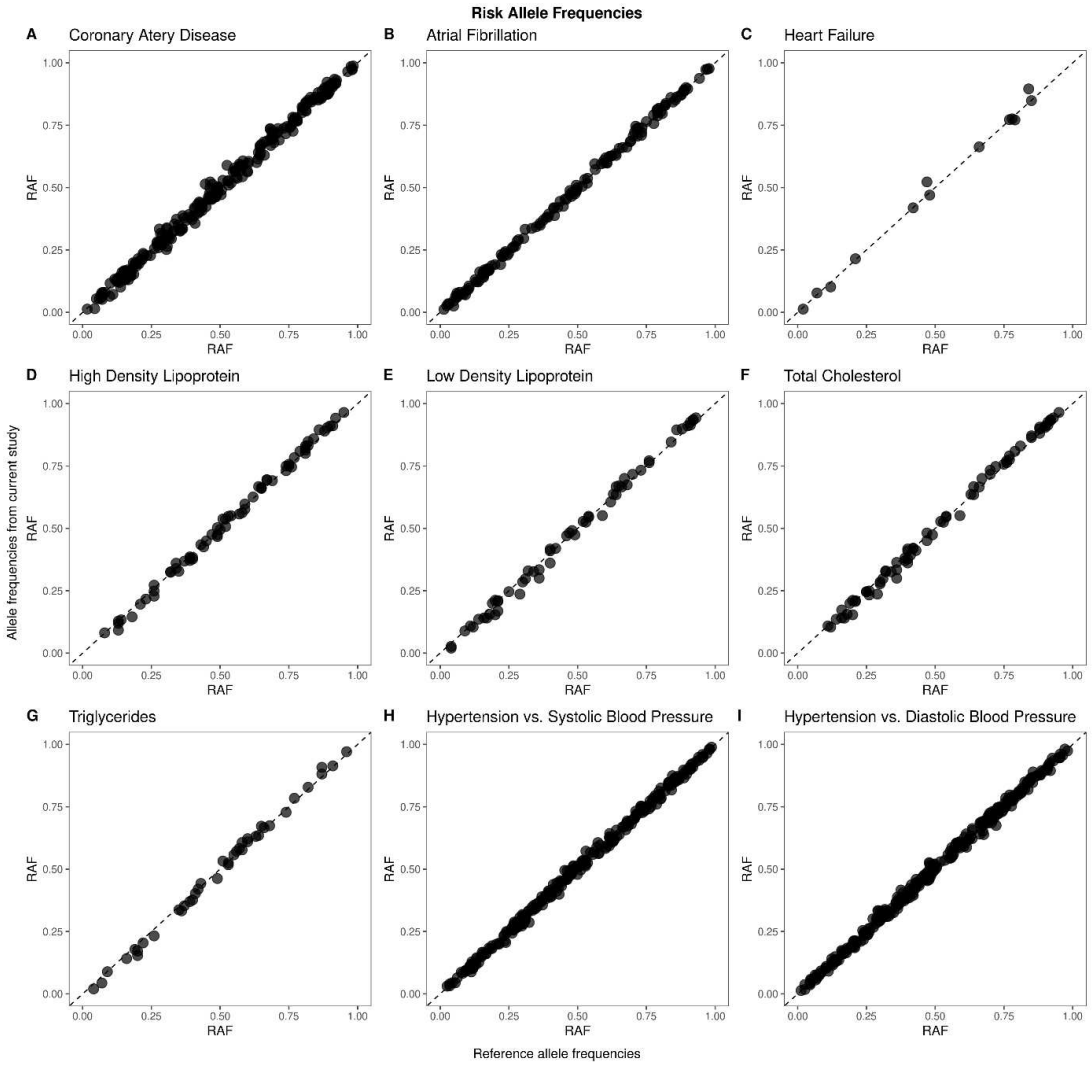
Supplementary Figures



Supplementary Figure 1: The patients in CHB-CVDC were classified by the Charlson Comorbidity Index prior to the date of the first cardiovascular diagnosis, the inclusion date and at the end of follow-up or death. The patients are stratified in age groups. CVD: cardiovascular disease

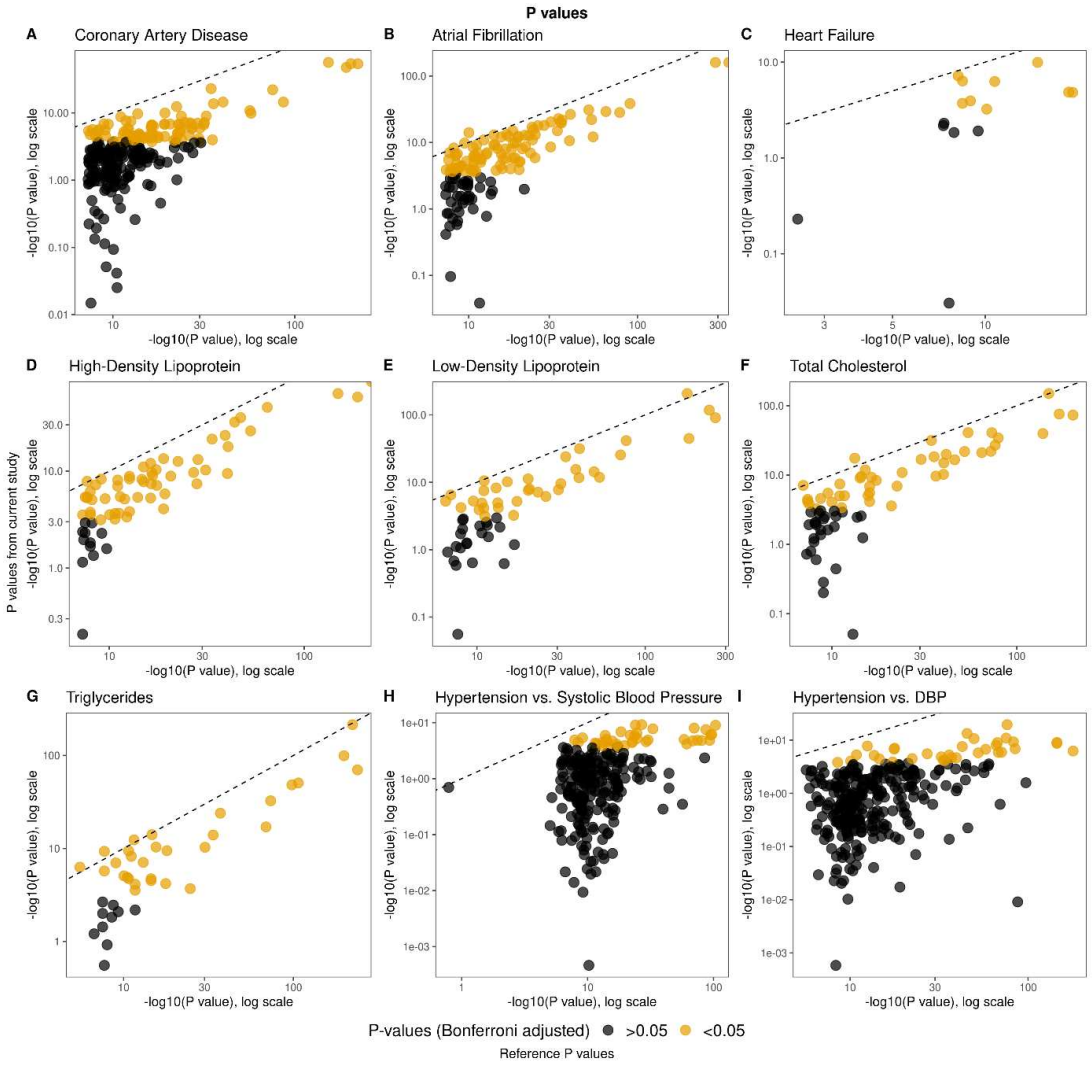


Supplementary Figure 2: The follow-up time from inclusion to either death or present was plotted. When a patient was either dead or the data extraction date (present) was reached the patient is subtracted from the total number of individuals. The black line shows the follow-up time for those still alive at present (56 259). The yellow line shows the follow-up time for those who are dead (39 539). The blue line shows the total follow-up time. More patients die within the first two years and then the curve flattens. The irregular part in the end of the survivor curve corresponds to some delays in inclusion times from the start of inclusion in 2009.



Supplementary Figure 3: This figure shows the correlation between the risk allele frequencies between the reference studies and this work. The dotted lines are the expected correlations of 1.

1



2

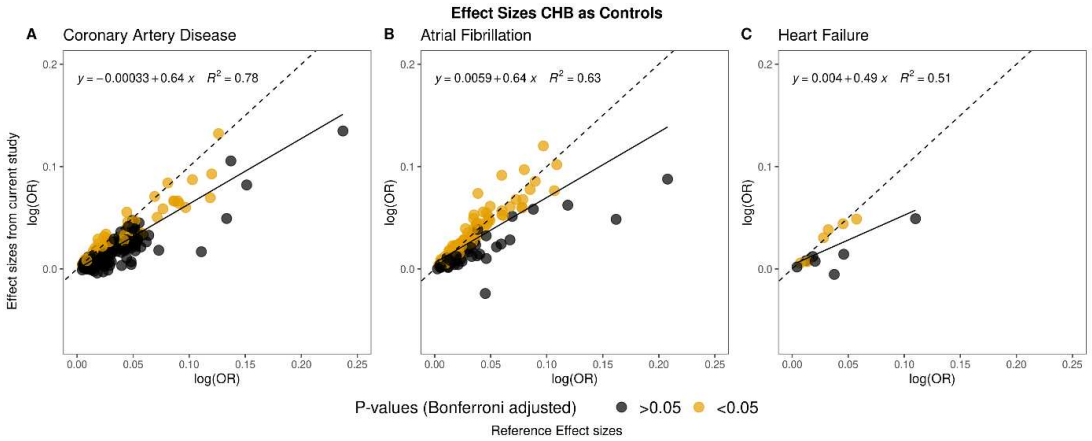
3 ***Supplementary Figure 4: For each trait the -log10(p-values) from this study were plotted against the -log(p-***

4 ***values) from the reference studies. The axes are on logarithmic scales. The dotted lines correspond to a***

5 ***correlation of 1.***

6

8

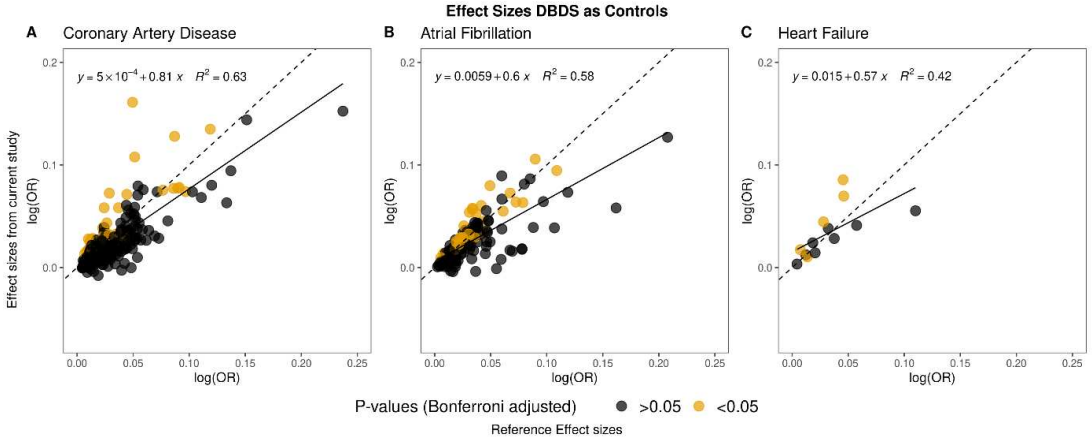


1

2 ***Supplementary Figure 5: Comparison of effect sizes between reference studies and the CHB-CVDC as controls***

3 ***study design.***

4

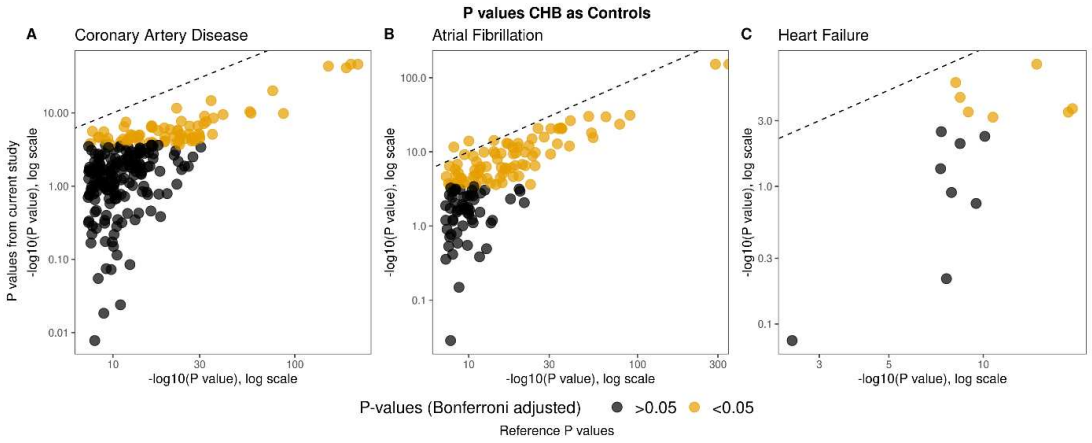


5

6 ***Supplementary Figure 6: Comparison of effect sizes between reference studies and the DBDS as controls study***

7 ***design.***

9

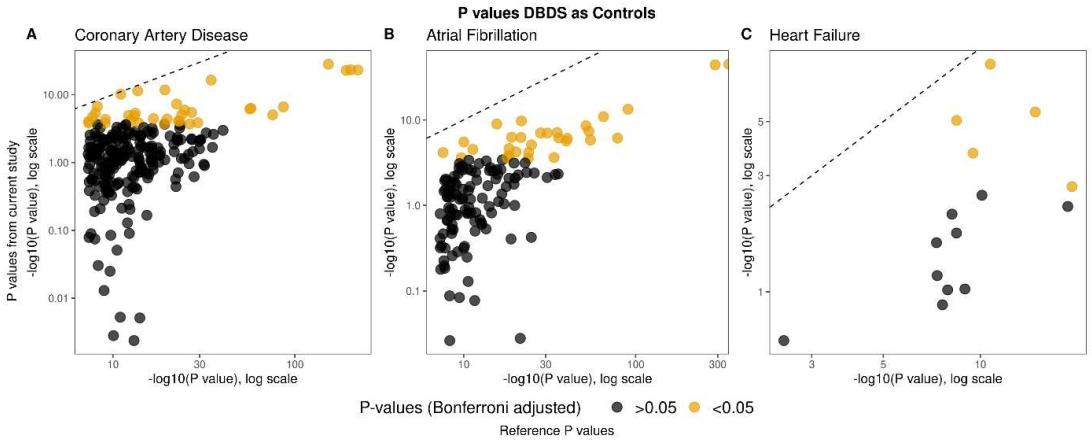


1

2 ***Supplementary Figure 7: Comparison of p-values between reference studies and the CHB-CVDC as controls***

3 ***study design.***

4



5

6 ***Supplementary Figure 8: Comparison of p-values between reference studies and the DBDS as controls study***

7 ***design.***

10

Supplementary Tables

Supplementary Table 1: Population characteristics

	ICD-10 Codes	N (% females)	Year of birth, Mean (SD)	Age at inclusion, Mean (SD)	Age at first CVD, Mean (SD)	Hemoglobin Complete %	Troponin T Complete %	Creatinine Complete %	CRP Complete %	LDL Complete %	HDL Complete %	Na^+ Complete %	K^+ Complete %
Hypertension and hypertensive cardiac diseases	I10-15	29 546 (55%)	1943.8 (13.1)	69.6 (13.2)	62.5 (14.3)	98.97 %	21.54 %	98.26 %	97.35 %	85.02 %	89.90 %	98.99 %	98.98 %
Coronary artery diseases and atherosclerosis	I20-25, I70	24 795 (36%)	1943.6 (12.7)	69.5 (12.8)	60.0 (12.4)	98.48 %	37.27 %	97.67 %	96.64 %	87.49 %	90.78 %	98.57 %	98.55 %
Lipid disorders	E78	4046 (50%)	1948.3 (13.1)	65.2 (13.2)	58.5 (13.8)	98.91 %	20.27 %	98.15 %	97.21 %	90.29 %	94.07 %	98.84 %	98.84 %
Cardiac arrhythmia	I44-49	16 208 (44%)	1945.5 (16.7)	67.9 (16.8)	60.5 (17.2)	98.65 %	22.77 %	97.80 %	96.99 %	80.81 %	85.48 %	98.53 %	98.51 %
Heart failure, cardiac valve disorders, and myocardial diseases	I50, I34-39, I05-09, I40-44	7264 (41%)	1943.9 (16.1)	69.2 (16.2)	62.5 (17.6)	98.73 %	36.65 %	97.88 %	97.94 %	82.30 %	86.38 %	98.61 %	98.61 %
Vascular disorders and aneurysms	I71-79	3121 (40%)	1942.6 (12.2)	70.3 (12.3)	62.3 (13.8)	98.08 %	31.66 %	97.18 %	97.53 %	85.61 %	88.72 %	98.17 %	98.14 %
Cerebrovascular diseases and cerebral hemorrhage	I60-69	8401 (44%)	1941.1 (12.2)	72.1 (12.4)	63.1 (13.2)	98.76 %	23.90 %	97.74 %	97.82 %	85.69 %	89.85 %	98.70 %	98.68 %
Pulmonary heart and pulmonary circulation diseases	I26-28	1521 (49%)	1942.4 (13.6)	70.7 (13.8)	59.7 (15.8)	98.75 %	31.62 %	98.09 %	97.50 %	81.53 %	87.44 %	98.69 %	98.69 %
Vascular kidney disease	N17-19	1406 (39%)	1945.2 (14.2)	67.4 (14.4)	61.0 (16.9)	98.44 %	26.53 %	97.51 %	98.08 %	79.16 %	83.85 %	98.44 %	98.36 %
Above diseases combined		**96 308 (45%)**	**1943.9 (13.9)**	**69.3 (14.0)**	**61.3 (14.5)**	**98.71 %**	**27.65 %**	**97.90 %**	**97.21 %**	**84.90 %**	**89.12 %**	**98.71 %**	**98.69 %**

N: Number of cases, SD: Standard Deviation, CRP: C-Reactive Protein, LDL: Low Density Lipoprotein, HDL: High Density Lipoprotein, Na^+: Sodium ion, K^+: Potassium ion

Supplementary Table 2: Charlson Comorbidity Index

Charlson index calculated at first cardiovascular diagnosis

sex	0	1-2	3-4	>=5	total
Counts	43 130	43 506	7520	2152	96 308

	M	22 287	24 845	3956	1185
	F	22 843	18 661	3564	967
% of total		45	45	8	2
	M	52	57	53	55
	F	53	43	47	45

Charlson index calculated at inclusion

	sex	0	1-2	3-4	>=5	total
Counts		25 745	40 361	19 070	10 913	96 089
	M	13 213	22 102	10 735	6616	
	F	12 532	18 259	8335	4297	
% of total		27	42	20	11	
	M	51	55	56	61	
	F	49	45	44	39	

Charlson index calculated at the end of follow up

	sex	0	1-2	3-4	>=5	total
Counts		12 977	34 375	25 300	23 656	96 308
	M	6878	18 258	14 045	14 427	
	F	6099	16 117	11 255	9229	
% of total		13	36	26	25	
	M	53	53	56	61	
	F	47	47	44	39	

Supplementary Table 3: An overview of diagnosis assigned prior to the first cardiovascular disease. If a patient has received multiple diagnosis within a ICD10 chapter they were only counted as one diagnosis. E.g 5 465 have a digestive disorder prior to a cardiac arrythmia diagnosis as their first cardiovascular disease.

First assigned cardiovascular diagnosis	-	Certain conditions originating in the perinatal period	Certain infections and parasitic diseases	Codes for special purposes	Congenital malformations, deformations and chromosomal abnormalities	Diseases of the blood and blood-forming organs and certain disorders involvning the immune mechanism	Diseases of the circulatory system included in CHB-CVDC	Diseases of the circulatory system included in CHB-CVDC diagnosed the same day
Cardiac arrythmia	0	236	2187	133	688	837	16208	2039
Cerebrovascular diseases and cerebral haemorrhage	0	16	855	45	266	296	8401	2672
Heart failure, cardiac valve disorders and myocardial diseases	0	118	1058	43	467	474	7264	1639
Hypertensive cardiac diseases	<10	139	3867	273	1049	1722	29546	2176
Ischemic heart diseases and atherosclerosis	0	80	2435	176	726	711	24795	4319
Lipid disorders	0	42	547	52	160	188	4046	1185
Pulmonary heart diseases	0	<10	202	<10	64	88	1521	232
Vascular disorders and aneurysms	0	12	359	17	127	129	3121	307
Vascular kidney diseases	0	12	394	13	107	215	1406	323

Supplementary Table 3 continued:

First assigned cardiovascular diagnosis	Diseases of the circulatory system not included in CHB-CVDC	Diseases of the digestive system	Diseases of the eye and adnexa, Diseases of the ear and mastoid process	Diseases of the genitourinary system	Diseases of the musculoskeletal system and connective tissue	Diseases of the nervous system	Diseases of the respiratory system	Diseases of the skin and subcutaneous tissue
Cardiac arrythmia	1751	5465	4294	4973	6296	2145	4258	1836
Cerebrovascular diseases and cerebral haemorrhage	682	2635	2252	2495	3033	1941	1556	898
Heart failure, cardiac valve disorders and myocardial diseases	1083	2589	2183	2175	2876	973	2188	878
Hypertensive cardiac diseases	3177	11929	8869	11674	13411	4777	7100	3914
Ischemic heart diseases and atherosclerosis	2180	7983	5281	6903	9721	2809	4803	2825
Lipid disorders	442	1665	1138	1492	1976	1015	913	635
Pulmonary heart diseases	428	505	346	506	604	189	524	164
Vascular disorders and aneurysms	356	1128	829	920	1308	401	665	399
Vascular kidney diseases	191	597	442	675	608	218	499	233

Supplementary Table 3 continued:

First assigned cardiovascular diagnosis	Endocrine, nutritional and metabolic diseases	External causes of morbidity and mortality	Factors influencing health status and contact with health services	Injury, poisoning and certain other consequences of external causes	Mental and behavioural disorders	Neoplasms	Pregnancy, childbirth and the puerperium	Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
Cardiac arrythmia	2452	194	11302	10193	2253	4158	1963	6539
Cerebrovascular diseases and cerebral haemorrhage	1286	78	5487	4910	1305	1873	887	3121
Heart failure, cardiac valve disorders and myocardial diseases	1232	66	5254	4440	1019	1783	663	3050
Hypertensive cardiac diseases	8609	574	21820	18895	4786	10230	4765	12924
Ischemic heart diseases and atherosclerosis	3223	184	15421	14076	2905	4890	2417	8150
Lipid disorders	1584	78	3129	2728	742	1254	742	1820
Pulmonary heart diseases	198	15	935	868	206	427	189	566
Vascular disorders and aneurysms	391	19	2180	1924	409	745	297	1110
Vascular kidney diseases	493	21	1032	887	280	449	140	694

Supplementary Table 4: Patients in CHB-CVDC were grouped according to their first cardiovascular disease. The number of patients with diagnoses within the other cardiovascular diseases were calculated. If a patient has received multiple diagnoses within a disease group, they were only counted as one diagnosis. E.g. 2 403 patients receive a cardiac arrythmia diagnosis as their first cardiovascular disease and later in life one or more diagnoses within the group cerebrovascular diseases and cerebral haemorrhage.

First assigned cardiovascular diagnosis	Cardiac arrythmia	Cerebrovascular diseases and cerebral haemorrhage	Heart failure, cardiac valve disorders and myocardial diseases	Hypertensive cardiac diseases	Ischemic heart diseases and atherosclerosis	Lipid disorders	Pulmonary heart diseases	Vascular disorders and aneurysms	Vascular kidney diseases
Cardiac arrythmia	0	2403	4630	6610	4274	2334	747	1245	1421
Cerebrovascular diseases and cerebral haemorrhage	2897	0	2126	6203	2931	4050	431	1205	1050
Heart failure, cardiac valve disorders and myocardial diseases	3187	1077	0	3448	2844	1634	419	877	1078
Hypertensive cardiac diseases	6751	4951	5523	0	7110	6419	1367	2877	3821
Ischemic heart diseases and atherosclerosis	7778	4257	8639	13777	0	10581	1328	4696	2907
Lipid disorders	699	672	637	2128	1062	0	149	383	352
Pulmonary heart diseases	674	276	574	945	638	331	0	194	212
Vascular disorders and aneurysms	1055	654	967	2081	1911	944	206	0	493
Vascular kidney diseases	558	255	432	989	467	251	80	207	0

Supplementary Table 5: Top 100 diagnosis in CHB-CVDC

	ICD10 code	Number of patients with diagnosis	Medical condition
1	I109	64 247	Essential hypertension, unspecified
2	I489	27 315	Atrial fibrillation and flutter, unspecified
3	E780	26 581	Hypercholesterolemia
4	I259	22 305	Chronic ischemic heart disease, unspecified
5	I209	21 409	Angina pectoris, unspecified
6	I509	18 294	Heart failure, unspecified
7	I251	12 825	Atherosclerotic heart disease
8	I649	11 272	Apoplexia cerebri, unspecified
9	I639	11 076	Cerebral infarction, unspecified
10	I219	10 383	Acute myocardial infarction, unspecified
11	I694	9972	Sequelae of stroke, not specified as haemorrhage or infarction
12	I489B	8639	Atrial fibrillation and atrial flutter, unspecified
13	I252	8611	Old myocardial infarction
14	I214	8486	Acute subendocardial myocardial infarction
15	I200	7801	Unstable angina
16	I500	7536	Congestive heart failure
17	N189	7256	Chronic kidney disease, unspecified
18	I350	7242	Aortic (valve) stenosis
19	I702	6920	Atherosclerosis of arteries of extremities
20	E785	6864	Hyperlipidaemia, unspecified
21	I739A	6122	Intermittent claudication
22	I471	5356	Supraventricular tachycardia
23	N179	5158	Acute renal failure, unspecified
24	I480	4917	Paroxysmal atrial fibrillation
25	I213	4182	Acute transmural myocardial infarction of unspecified site
26	N199	3964	Unspecified kidney failure
27	I501	3829	Left ventricular failure
28	I482	3423	Chronic atrial fibrillation
29	I693	3216	Sequelae of cerebral infarction
30	I269	2982	Pulmonary embolism without mention of acute cor pulmonale
31	I709	2967	Generalized and unspecified atherosclerosis
32	I340	2931	Mitral (valve) insufficiency
33	I739C	2929	Peripheral vascular disease, unspecified +
34	I714	2919	Abdominal aortic aneurysm, without mention of rupture
35	I702A	2651	Atherosclerotic gangrene
36	I499	2622	Cardiac arrhythmia, unspecified
37	I442	2473	Atrioventricular block, complete
38	I10	2294	Essential (primary) hypertension
39	I351	2288	Aortic (valve) insufficiency
40	I460	2167	Cardiac arrest with successful resuscitation

41	I269A	2155	Pulmonary embolism NOS
42	I119	2121	Hypertensive heart disease without (congestive) heart failure
43	I469	2108	Cardiac arrest, unspecified
44	I159	2107	Secondary hypertension, unspecified
45	I479	2094	Paroxysmal tachycardia, unspecified
46	I619	1979	Intracerebral haemorrhage, unspecified
47	I208	1977	Other forms of angina pectoris
48	I493	1890	Ventricular premature depolarization
49	I489BB	1876	Atrial fibrillation and atrial flutter, unspecified
50	I210	1860	Acute transmural myocardial infarction of anterior wall
51	I258	1762	Other forms of chronic ischaemic heart disease
52	I489A	1670	Atrial fibrillation and atrial flutter, unspecified
53	I652	1609	Occlusion and stenosis of carotid artery
54	I495	1579	Sick sinus syndrome
55	I481	1477	Persistent atrial fibrillation
56	I359	1469	Aortic valve disorder, unspecified
57	I249	1465	Acute ischaemic heart disease, unspecified
58	I472	1454	Ventricular tachycardia
59	I420	1442	Dilated cardiomyopathy
60	I211B	1350	Acute transmural myocardial infarction of inferior wall
61	I429	1332	Cardiomyopathy, unspecified
62	I211	1319	Acute transmural myocardial infarction of inferior wall
63	I110	1304	Hypertensive heart disease with (congestive) heart failure
64	I210B	1162	Acute transmural myocardial infarction of anterior wall
65	I491	1143	Atrial premature depolarization
66	N185	1137	Chronic kidney disease, stage 5
67	I739	1086	Peripheral vascular disease, unspecified
68	I609	1061	Subarachnoid haemorrhage, unspecified
69	I633	1047	Cerebral infarction due to thrombosis of cerebral arteries
70	N180	1025	Hypertensive renal disease
71	I719	969	Aortic aneurysm of unspecified site, without mention of rupture
72	I743	948	Embolism and thrombosis of arteries of lower extremities
73	I495B	946	Tachycardia-bradycardia syndrome
74	I255	919	Ischaemic cardiomyopathy
75	I201	856	Angina pectoris with documented spasm
76	I712	853	Thoracic aortic aneurysm, without mention of rupture
77	I389	811	Endocarditis, unspecified
78	I501C	791	Left ventricular failure, pulmonary congestion
79	I443	784	Other and unspecified atrioventricular block
80	E782	768	Mixed hyperlipidaemia
81	E789	767	Disorder of lipoprotein metabolism, unspecified
82	I270	751	Other adrenocortical overactivity
83	I48	744	Atrial fibrillation and flutter
84	I279	740	Pulmonary heart disease, unspecified

85	I691	721	Sequelae of intracerebral haemorrhage
86	I708	705	Atherosclerosis of other arteries
87	I441	678	Atrioventricular block, second degree
88	I21	673	Acute myocardial infarction
89	E780B	651	Familial hypercholesterolaemia
90	I632	649	Cerebral infarction due to unspecified occlusion or stenosis of precerebral arteries
91	I472A	627	Ventricular tachycardia
92	I612	623	Intracerebral haemorrhage in hemisphere, unspecified
93	I501B	615	Left ventricular failure, Oedema of lung
94	I470	598	Re-entry ventricular arrhythmia
95	I352	589	Aortic (valve) stenosis with insufficiency
96	N183	585	Chronic kidney disease, stage 3
97	I471R	577	Supraventricular tachycardia, atrioventricular [AV]: re-entrant (nodal) [AVNRT]
98	I728	567	Aneurysm and dissection of other specified arteries
99	N184	566	Chronic kidney disease, stage 4
100	N178	561	Other acute renal failure

Supplementary Table 6: Overview of causes of death in CHB-CVDC according to the Danish Registry of Causes of Death

Cause of death	***n***	***% of dead***
Cancer	*13715*	*35.8*
Heart disease	*6478*	*16.9*
Respiratory disease	*3937*	*10.3*
Other diseases in the circulatory system	*3097*	*8.1*
Diseases of the digestive system	*2019*	*5.3*
Endocrine, nutritional, and metabolic diseases	*1427*	*3.7*
Unknown medical information	*1185*	*3.1*
Mental and behavioural disorders	*1102*	*2.9*
Diseases of the nervous system	*946*	*2.5*
Accidents	*917*	*2.4*
Certain infectious and parasitic diseases	*885*	*2.3*
Diseases of the genitourinary system	*799*	*2.1*
Symptoms, signs, and abnormal clinical and laboratory findings, not elsewhere classified	*644*	*1.7*
Neoplasms	*288*	*0.8*
Diseases of the musculoskeletal system and connective tissue	*262*	*0.7*
Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	*212*	*0.6*
Suicide	*175*	*0.5*
Diseases of the skin and subcutaneous tissue	*69*	*0.2*
Congenital malformations, deformations, and chromosomal abnormalities	*65*	*0.2*
Other causes	*27*	*0.1*

Supplementary Table 7: Overview of databases and variables

Databases	**Variables**
The National Patient Registry	ICD10/ICD8, SKS codes. Patient type, hospital code, department code, admission date, admission time, discharge date, discharge time, days of treatment, mode of admission, reason for contact, mode of discharge, speciality, action diagnosis, accident code, accident code counterpart, accident code case, accident code activity, accident code place, accident code mechanism, accident code traffic, outpatient date, diagnosis, type of diagnosis, additional diagnosis, diagnosis modification, SKS procedure (operation), SKS procedure classification (operation), additional code (operation), procedure hospital (operation), procedure department (operation), procedure day (operation), procedure hour (operation), procedure minute (operation), SKS procedure (diagnosis and treatment), SKS procedure classification (diagnosis and treatment), additional code (diagnosis and treatment), procedure hospital (diagnosis and treatment), procedure department (diagnosis and treatment), procedure day (diagnosis and treatment), procedure hour (diagnosis and treatment), procedure minute (diagnosis and treatment), and other similar variables that are needed for these classifications.
The Danish Registry of Causes of Death	Date of death, cause of death, place of death.
Central Person Registry (CPR)	Status (dead, emigrated or living) and also relatedness between cases.
The Danish National Prescription Registry	Medicine bought in pharmacies with prescription
BigTempHealth	Validate diagnoses and drug prescriptions. Include data from electronic health records (EHRs) and laboratory test results in the phenotypic characterization.
Danish Agency of Labour Market and Recruitment	Working history and occupation
Health Care Statistics Registry	Speciality, service code, year, month, number of services to access risk factors for heart diseases. This also includes data on dental diseases and treatments to access dental risk factors for heart disease.
The Danish Laboratory Database	Patient CPR, sampling date, sampling time, analysis code, Laboratory ID-CODE, value, unit, result type, reference interval upper limit, reference interval lower limit, and NPU code.
The Copenhagen GP Laboratory database	Electrocardiography data
The Regions imaging data, including data in IntelliSpace, PACS and in Xeroviewer	Echocardiography, computer tomography (CT), Magnetic Resonans (MR), imaging, coronary arteriography, and nuclear imaging.

The Regions ECG and Holter database; Kardia	
National Clinical Registries (RKKB):	
Danish Anesthesia Database	Information on blood pressure, height, weight, smoking, co-morbidity, ID, date, department, hospital, region, indication, operation information, complications, ASA score, and other related variables
Danish Stroke Registry	Disease classification, information on smoking history, cardiac disease, comorbidity, ID, date, department, hospital, region, alcohol, AK treatment, diabetes, bleeding or infraction, hypertension, treatment, outcome, height, weight, and other related variables
Danish Heart Registry	Disease classification, indication, operation type, left ventricular ejection fraction (LEVF), coronary artery pathology, height, weight, smoking, diabetes history, package, treatment, ID, date, department, hospital, region, complications, EURO score, lung disease, cerebrovascular disease, previous cardiac surgery, creatinine, endocarditis, pulmonal hypertension, angina, left ventricular dysfunction, and other related variables
Danish Heart Failure Registry	Disease classification, disease severity measures, co-morbidity, life style factors, treatments, ID, date, department, hospital, region alcohol, BMI, diabetes, hypertension, tobacco, ejection fraction, NYHA group, COPD, and other related variables
Danish Heart Disease Rehabilitation Database	Disease classification, cardiovascular disease history, co-morbidity, LVEF, lifestyle factors, cardiac risk factors, current disease measures including CSS and NYHA classification, waist circumference, current treatment, current laboratory values, HRQoL score, ID, date, department, hospital, region, treatment, liver values, CK, BMI, depression, GFR, diabetes values, lung function, exercise test, blood pressure, blood sugar values incl. HbA1c, lipids, HRQL, ejection fraction, height, weight, and other related variables.
Danarrest and the Danish Registry of Cardiac Arrest	Cardiac cause of death, ID, date, department, hospital, region and similar data as for the other databases
Atrial Fibrillation Database	Information on ID, date, department, hospital, region and similar data as for the other databases
Danish Pacemaker and ICD register	Information on ID, date, department, hospital, region, devices, indication, and similar data as for the other databases
Danish Ablation Database	Information on ID, date, department, hospital, region, indication, conduction disorder, operation, and similar data as for the other databases

The PATS Database	Includes data on patients that has undergone invasive cardiac procedures until 2016 – we will include data on ID, date, department, hospital, region, indication, operation procedures, risk factors, laboratory values, co-morbidity and organ functions, and similar data as for the other databases
The Database for Familial Hypercholesterolemia	Information on ID, date, department, hospital, region, indication, dyslipidemia disorder, genetics, and similar data as for the other databases
Progeny	The database holds data on clinical and genetic findings in families with inherited cardiac diseases including pedigrees. Information on ID, date, department, hospital, region, indication for assessment, findings, and similar data as for the other databases
Danish Obesity Surgery Database	Information on ID, date, department, hospital, region, diabetes, height, weight, diabetes type, hypertension, lipids, depression, asthma, COPD, PCO, joint complaints, tobacco, alcohol, HRQL, treatment, and other related variables
Danish Registry for Chronic Obstructive Lung Diseases	Information on ID, date, department, hospital, region, BMI, co-morbidity, lung function measures, diagnosis, tobacco, tobacco intervention, and other related variables
Sleep Apnea Database	Information on ID, date, department, hospital, region, diagnosis, treatment
KARBASE	Information on ID, date, department, hospital, region, diagnosis, intervention, death, date of death, date of operation, co-morbidity, BMI, tobacco, alcohol, and other related variables

Supplementary Table 8: P-values and confidence intervals of the effect sizes regression lines.

	Entire cohort with DBDS		**CHB-CVDC as controls**		**DBDS as controls**	
Phenotype	p-value	2.5%:97.5%	p-value	2.5%:97.5%	p-value	2.5%:97.5%
CAD	<2.2*10-16	0.76:0.88	<2.2*10-16	0.60:0.68	<2.2*10-16	0.73:0.89
AF	<2.2*10-16	0.65:0.83	<2.2*10-16	0.56:0.72	<2.2*10-16	0.52:0.69
HF	0.01079	0.14:0.83	0.0062	0.17:0.80	0.0161	0.13:1.01
HDL	<2.2*10-16	0.81:0.95				
LDL	<2.2*10-16	0.61:0.70				
TC	<2.2*10-16	0.66:0.75				
TG	<2.2*10-16	0.88:1.1				
SBP	<2.2*10-16	0.09:0.12				
DBP	<2.2*10-16	0.13:0.18				

Supplementary Table 9: Comparison of different control groups

Atrial Fibrillation

	Variants with concordant direction of effect	Replicated/Total	Replicated/Power to Replicate	Cases	Controls
CHB as cases, DBDS and CHB as controls. As presented in Table 2	137/140 (98%)	96/140 (69%)	96/109 (88%)	30 229	157 669
CHB as cases and controls	139/140 (99%)	83/140 (59%)	83/102 (81%)	30 152	65 870
CHB as cases, DBDS as controls	136/140 (97%)	30/140 (21%)	30/54 (56%)	30 152	91 280

Coronary Artery Disease

	Variants with concordant direction of effect	Replicated/Total	Replicated/Power to Replicate	Cases	Controls
CHB as cases, DBDS and CHB as controls. As presented in Table 2	236/241 (98%)	90/241 (37%)	90/137 (66%)	33 746	154 311
CHB as cases and controls	233/241 (97%)	71/241 (29%)	71/113 (63%)	37 862	69 824
CHB as cases, DBDS as controls	232/241 (96%)	35/241 (15%)	35/72 (49%)	38 561	104 011

Heart Failure

	Variants with concordant direction of effect	Replicated/Total	Replicated/Power to Replicate	Cases	Controls
CHB as cases, DBDS and CHB as controls. As presented in Table 2	14/15 (93%)	9/15 (60%)	9/10 (90%)	21 443	167 068
CHB as cases and controls	14/15 (93%)	7/15 (47%)	7/9 (78%)	21 421	74601
CHB as cases, DBDS as controls	15/15 (100%)	5/15 (33%)	5/6 (83%)	21 421	91892

Supplementary Table 10: Residual confounding

Phenotype	Design	Intercept	Intercept SE	Ratio
AF	Entire cohort with DBDS	1,0957	0,0087	0,3366
AF	CHB-CVDC	1,0356	0,0079	0,1566
AF	DBDS control	1,1030	0,0088	0,5366
CAD	Entire cohort with DBDS	1,1163	0,0092	0,3995
CAD	CHB-CVDC	1,0303	0,0085	0,1887
CAD	DBDS control	1,1638	0,0099	0,5091
HF	Entire cohort with DBDS	1,0705	0,0073	0,5313
HF	CHB-CVDC	1,0203	0,0065	0,3414
HF	DBDS control	1,0977	0,0079	0,5709

Supplementary Table 11: Genetic Correlation

Phenotype	Design	Comparison study	rg	se	p
AF	Entire cohort with DBDS	Nielsen 2018	1,0778	0,0417	2,76E-147
AF	CHB-CVDC	Nielsen 2018	0,9647	0,035	1,83E-167
AF	CHB-CVDC case / DBDS control	Nielsen 2018	NA	NA	NA
CAD	Entire cohort with DBDS	Cardiogramplusc4d	1,0452	0,0675	4,32E-54
CAD	CHB-CVDC	Cardiogramplusc4d	0,9249	0,0667	1,12E-43
CAD	CHB-CVDC case / DBDS control	Cardiogramplusc4d	NA	NA	NA
HF	Entire cohort with DBDS	Shah 2020	0,9808	0,1027	1,26E-21
HF	CHB-CVDC	Shah 2020	-0,8248	0,1289	1,57E-10
HF	CHB-CVDC case / DBDS control	Shah 2020	NA	NA	NA

Supplementary references

1 Gasparini A. comorbidity: An R package for computing comorbidity scores. *JOSS* 2018;**3**:648. doi:10.21105/joss.00648

2 Quan H, Sundararajan V, Halfon P, *et al.* Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care* 2005;**43**:1130–9. doi:10.1097/01.mlr.0000182534.19832.83

3 Schmidt M, Schmidt SAJ, Sandegaard JL, *et al.* The Danish National Patient Registry: a review of content, data quality, and research potential. *Clin Epidemiol* 2015;**7**:449–90. doi:10.2147/CLEP.S91125

4 Kildemoes HW, Sørensen HT, Hallas J. The Danish National Prescription Registry. *Scand J Public Health* 2011;**39**:38–41. doi:10.1177/1403494810394717

5 Hansen TF, Banasik K, Erikstrup C, *et al.* DBDS Genomic Cohort, a prospective and comprehensive resource for integrative and temporal analysis of genetic, environmental and lifestyle factors affecting health of blood donors. *BMJ Open* 2019;**9**:e028401. doi:10.1136/bmjopen-2018-028401

6 Aragam KG, Chaffin M, Levinson RT, *et al.* Phenotypic Refinement of Heart Failure in a National Biobank Facilitates Genetic Discovery. *Circulation* 2019;**139**:489–501. doi:10.1161/CIRCULATIONAHA.118.035774

7 Krogager ML, Torp-Pedersen C, Mortensen RN, *et al.* Short-term mortality risk of serum potassium levels in hypertension: a retrospective analysis of nationwide registry data. *Eur Heart J* 2017;**38**:104–12. doi:10.1093/eurheartj/ehw129

8 Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* 2015;**31**:3555–7. doi:10.1093/bioinformatics/btv402

9 R Core Team. R: A Language and Environment for Statistical Computing. 2020.https://www.R-project.org/

10 Purcell S, Neale B, Todd-Brown K, *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;**81**:559–75. doi:10.1086/519795

11 Chang CC, Chow CC, Tellier LC, *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015;**4**. doi:10.1186/s13742-015-0047-8

12 Zhou W, Nielsen JB, Fritsche LG, *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet* 2018;**50**:1335–41. doi:10.1038/s41588-018-0184-y

13 Loh P-R, Tucker G, Bulik-Sullivan BK, *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* 2015;**47**:284–90. doi:10.1038/ng.3190

14 Abraham G, Qiu Y, Inouye M. FlashPCA2: principal component analysis of Biobank-scale genotype datasets. *Bioinformatics* 2017;**33**:2776–8. doi:10.1093/bioinformatics/btx299

15 RStudio Team. RStudio: Integrated Development for R. 2020.https://www.rstudio.com/

16 Bulik-Sullivan BK, Loh P-R, Finucane H, *et al.* LD Score Regression Distinguishes Confounding from Polygenicity in Genome-Wide Association Studies. *Nat Genet* 2015;**47**:291–5. doi:10.1038/ng.3211

17 Buniello A, MacArthur JAL, Cerezo M, *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* 2019;**47**:D1005–12. doi:10.1093/nar/gky1120