THE LANCET Diabetes & Endocrinology

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Shah AD, Langenberg C, Rapsomaniki E, et al. Type 2 diabetes and incidence of cardiovascular diseases: a cohort study in 1·9 million people. *Lancet Diabetes Endocrinol* 2014; published online Nov 11. http://dx.doi.org/10.1016/S2213-8587(14)70219-0.

Type 2 diabetes and incidence of a wide range of cardiovascular diseases: a cohort study in 1·9 million people

Supplementary material

Table of Contents

Supplementary Methods

CALIBER program: study data sources

The CALIBER[1] (Cardiovascular disease research using Linked Bespoke studies and Electronic Records) research platform contains linked electronic health records from four data sources in England:

1. Primary care data from 225 general practices in the Clinical Practice Research Datalink (CPRD).[2] CPRD provides primary care data on demographics, ethnicity, health behaviours, diagnoses, investigations, procedures and prescriptions. Diagnoses are coded using the Read Clinical Terminology. (Read Terms are a major component of the SNOMED-CT terminology).

2. Hospital Episodes Statistics (HES), containing details of hospital admissions.[3] Diagnoses are coded using the International Statistical Classification of Diseases and Health Related Problems, 10th revision (ICD-10); interventions are coded using the Office of the Population Censuses and Surveys Classification of Interventions and Procedures (OPCS). Ethnicity as recorded during hospital attendances is also included in HES.

3. Details of acute coronary syndromes from the Myocardial Ischaemia National Audit Project registry (MINAP).[4]

4. Date and ICD-10 coded cause of death from the Office for National Statistics (ONS) death registry. The index of multiple deprivation according to the patient's area of residence was also obtained from ONS.[5]

Definition of diabetes

Patients were defined as diabetic at baseline (type 1, type 2 or uncertain type) based on coded diagnoses recorded in CPRD or HES on or before study entry (Figure S1). Lists of Read and ICD-10 diagnostic codes which convey diabetic status are curated on the CALIBER data portal (https://www.caliberresearch.org/portal/chapter/6).

The aim of this algorithm is to identify patients with type 2 diabetes. The majority of diabetic patients in this cohort have type 2 diabetes, and we are not aiming to identify all patients with type 1 diabetes (this algorithm would not be suitable for a study focusing on type 1 diabetes). Our algorithm does not use any information after study entry to classify patients, in order to reduce immortal time bias. However this bias is not eliminated completely because specific type 1 or type 2 codes may have been entered retrospectively only for patients who survived.

Our algorithm leaves many patients with an unclassified type of diabetes. However, rather than using a more complex algorithm, we carried out sensitivity analyses with different definitions (e.g. the type 2 definition or any diabetes).

Definition of endpoints

Cardiovascular phenotype definitions based on the CALIBER data sources are curated on the CALIBER data portal (www.caliberresearch.org/portal).

Stable angina was defined by a coded diagnosis in primary or secondary care of ischaemic chest pain or stable angina, a positive myocardial ischaemia test, two or more prescriptions of antianginal medication, or coronary revascularisation.

Unstable angina was defined as a primary or secondary care diagnosis of unstable angina, or an acute coronary syndrome without myocardial infarction recorded in the disease registry.

Coronary disease not further specified is a non-specific diagnosis of ischaemic or coronary heart disease in primary or secondary care that does not fall into one of the more specific categories. It was not included among the 12 diseases in the main displays of hazard ratios, but for cumulative incidence calculations it was combined with unstable angina.

Non-fatal myocardial infarction was defined as a disease registry diagnosis of an acute coronary syndrome with elevated troponin, or a primary or secondary care diagnosis of myocardial infarction.

Unheralded coronary death was death with the primary cause certified as coronary heart disease, and no prior history of cardiovascular disease. Patients with myocardial infarction who died on the day of their infarct were considered to have unheralded coronary death.

Heart failure was defined by coded diagnoses in primary care, secondary care and death certificates.

Arrhythmia or sudden cardiac death was a composite of ventricular arrhythmias, cardioversion procedures, implantable cardioverter defibrillator, and sudden cardiac death. It was defined using diagnoses and procedure codes in primary care, secondary care and death certificates.

Transient ischaemic attack was defined by coded diagnoses in primary or secondary care.

Ischaemic stroke was defined using coded diagnoses in primary care, secondary care and death certificates. Patients with a procedure code for carotid endarterectomy within 90 days of a stroke of unspecified type were considered to have ischaemic stroke.

'Stroke not further specified' is a diagnosis of stroke which does not state it is ischaemic or haemorrhagic. This clinical event was not included among the 12 diseases in the main displays of hazard ratios, but for cumulative incidence calculations it was combined with ischaemic stroke.

Subarachnoid haemorrhage was defined by coded diagnoses in primary care, secondary care and death certificates.

Intracerebral haemorrhage was defined by coded diagnoses in primary care, secondary care and death certificates.

Peripheral arterial disease includes intermittent claudication, limb ischaemia or gangrene due to atherosclerotic disease in the arteries of the legs. It was defined by coded diagnoses and procedures in primary care, secondary care and death certificates.

Abdominal aortic aneurysm was defined by coded diagnoses and procedures in primary care, secondary care and death certificates.

Sample size

We used all eligible patients in the CALIBER database for this study. We carried out a power calculation for estimating the age-adjusted hazard ratio between type 2 diabetes and haemorrhagic stroke, one of the less common endpoints, using the powerSurvEpi package in R.[6] Based on our dataset, with 1·8% of patients having type 2 diabetes and correlation coefficient between type 2 diabetes and age $\rho^2 = 0{\cdot}018$, a sample size of 1 000 000 would give 76% power to detect a hazard ratio of 1·5 with a type I error rate of 5%, and a sample size of 1 500 000 would give 90% power.

Multiple imputation

Multiple imputation was implemented using the mice[7] algorithm in the statistical package R, to replace missing values in exposure and risk factor variables. Imputation models were estimated separately for men and women and included:

1. All the baseline covariates used in the main analysis (age, quadratic age, sex, body mass index, deprivation, high density lipoprotein cholesterol, total cholesterol, systolic blood pressure, smoking status, statins and antihypertensive prescriptions);

2. Prior (between 14 and 4 years before study entry) and post (between 0 and 1 year after study entry) averages of continuous covariates in the main analysis;

3. Baseline measurements of covariates not considered in the main analysis (diastolic blood pressure, alcohol intake, white cell count, haemoglobin, creatinine, alanine aminotransferase);

4. Baseline medications (low-dose aspirin, loop diuretics, oral contraceptives and hormone replacement therapy);

5. Coexisting medical conditions (history of depression, cancer, renal disease, liver disease and chronic obstructive pulmonary disease);

6. The Nelson-Aalen hazard and the event status for each endpoint analysed in the data.[8]

Non-normally distributed variables were log-transformed for imputation and exponentiated back to their original scale for analysis. Five multiply imputed datasets were generated, and Cox models were fitted to each dataset. Coefficients were combined using Rubin's rules. The Kolmogorov-Smirnov test was used to compare the distribution of observed versus imputed log-transformed covariates.

Survival analysis and competing risks

We carried out survival analysis to describe and model the first occurrence of any cardiovascular disease. A patient's follow-up ended when they experienced one of the cardiovascular endpoints or when they were censored. Subsequent events (e.g. MI occurring after stable angina) were not analysed.

To describe the incidence of each initial presentation over time we constructed cumulative incidence curves, taking into account the other possible initial presentations as competing events. Normal-based confidence intervals were constructed based on Greenwood's variance formula, as implemented in the R prodlim package.

For multivariable modellling we considered using Cox models (for modelling cause-specific hazards) or the Fine and Gray model (to compare cumulative incidence curves by modelling subdistribution hazards). As the aim of this study was observational epidemiology – to explore associations rather than predict risk – we considered cause specific hazard ratios to be appropriate quantities to estimate. They should be interpreted together with cumulative incidence curves but cannot be used to predict cumulative incidence. We also found the Fine and Gray model to be more computationally intensive, and it would require significant software engineering and/or computing time to apply it to the large dataset used in these analyses. We used follow-up time as the timescale for the Cox models.

Supplementary Tables

Table S1. Inclusion criteria and endpoints in recent trials in type 2 diabetes

Ref	Author, year, study	Intervention	Inclusion criteria	Exclusion criteria	Components of primary outcome				
					Nonfatal MI	Nonfatal stroke	Cardio-vascular death	Hospital-ized heart failure	Unstable angina
9	ACCORD Study Group 2008	Intensive or standard HbA1c targets	1. Type 2 diabetes 2. HbA1c ≥ 7·5% 3. Either: a. Aged 40-79 with cardiovascular disease, OR b. Aged 55-79 with anatomical evidence of significant atherosclerosis, albuminuria, left ventricular hypertrophy, or at least two of: dyslipidemia, hypertension, current smoking, obesity.	Frequent or serious hypoglycaemic events, unwillingness to do home monitoring of glucose or inject insulin, BMI > 45 kg/m^2, serum creatinine > 1·5 mg/dL, other serious illness	Yes	Yes	Yes		
10	Parving 2012 (ALTITUDE)	Aliskiren versus placebo	1. Type 2 diabetes 2. Age ≥ 35 3. Microalbuminuria, macroalbuminuria, or eGFR 30-60 mL/min/1·73m^2 and MI, stroke, heart failure or coronary artery disease	Unstable serum creatinine, NHYA class III or IV heart failure, renal artery stenosis, malignancy in the past 5 years, valvular heart disease	Yes	Yes	Yes	Yes	
11	Green 2013 (TECOS)	Sitagliptin versus placebo	1. Type 2 diabetes 2. Age ≥ 50 3. HbA1c 6·5% to 8·0% 4. History of MI, ischaemic stroke, revascularisation, carotid stenosis, or peripheral arterial disease 5. On a stable dose of insulin or oral antidiabetic drugs	Liver cirrhosis, severe hypoglycaemia, life expectancy < 2 years, eGFR < 30 mL/min/1·73m^2	Yes	Yes	Yes		Yes
12	Scirica 2013 (SAVOR-TIMI 53)	Saxagliptin versus placebo	1. Type 2 diabetes 2. HbA1c 6·5% to 12·0% 3. Either: a. Age ≥ 40 and ischaemic heart disease, peripheral vascular disease or ischaemic stroke, OR b. Age ≥ 55 (men) or 60 (women) and dyslipidemia, hypertension, or smoking	Life expectancy < 5 years, MI or stroke within the past 2 months, renal transplant or dialysis, HIV, long term oral steroid treatment, BMI > 50 kg/m^2, sustained BP > 180/100 mmHg	Yes	Yes	Yes		
13	Home 2007 (RECORD)	Rosiglitazone versus metformin and sulphonylurea	1. Type 2 diabetes 2. Aged 40-75 years 3. BMI > 25 kg/m^2 4. HbA1c 7·0% to 9·0% while on maximum dose of sulphonylurea or metformin	Use of other antidiabetic drugs, major cardiovascular event in previous 3 months, heart failure, renal impairment, uncontrolled hypertension	Yes	Yes	Yes	Yes	Yes
14	Young 2009 (DIAD)	Screening for coronary artery disease	1. Onset of type 2 diabetes at age 30 years or older with no history of ketoacidosis 2. Age 50-75 at enrollment	History of coronary disease or angina, limited life expectancy	Yes		Yes		

Abbreviations: BMI, body mass index; eGFR, estimated glomerular filtration rate; MI, myocardial infarction

Table S2. Age and sex-specific prevalence of diabetes recorded in CALIBER at study entry

The total number of patients was 2 135 617, and included patients with pregnancy or prior history of cardiovascular disease which were excluded in the main analysis.

Sex	Age group	Number of patients	Percentage of patients with diabetes mellitus at baseline (95% CI)			
			Any diabetes	Type 1 diabetes	Type 2 diabetes	Diabetes of unspecified type
Men	30–40	420 975	0·87 (0·85–0·90)	0·47 (0·45–0·49)	0·28 (0·26–0·30)	0·13 (0·12–0·14)
	40–50	215 607	2·25 (2·19–2·32)	0·50 (0·47–0·53)	1·37 (1·32–1·42)	0·38 (0·36–0·41)
	50–60	170 679	4·41 (4·32–4·51)	0·45 (0·42–0·48)	3·24 (3·16–3·33)	0·73 (0·69–0·77)
	60–70	119 507	8·31 (8·15–8·46)	0·52 (0·48–0·56)	6·32 (6·18–6·46)	1·47 (1·40–1·54)
	70–80	80 614	10·5 (10·3–10·8)	0·53 (0·48–0·58)	7·89 (7·71–8·08)	2·12 (2·03–2·23)
	≥80	38 991	9·96 (9·66–10·3)	0·39 (0·33–0·46)	7·32 (7·06–7·58)	2·25 (2·10–2·40)
Women	30–40	411 996	0·77 (0·74–0·80)	0·35 (0·33–0·37)	0·25 (0·23–0·26)	0·17 (0·16–0·19)
	40–50	198 619	1·52 (1·47–1·57)	0·34 (0·31–0·37)	0·90 (0·86–0·95)	0·28 (0·25–0·30)
	50–60	166 467	2·90 (2·82–2·98)	0·34 (0·31–0·37)	2·03 (1·96–2·10)	0·53 (0·49–0·56)
	60–70	123 553	5·79 (5·66–5·93)	0·36 (0·33–0·40)	4·37 (4·26–4·49)	1·06 (1·00–1·11)
	70–80	103 284	7·82 (7·66–7·98)	0·42 (0·38–0·46)	5·70 (5·56–5·84)	1·70 (1·62–1·78)
	≥80	85 325	7·69 (7·51–7·87)	0·31 (0·27–0·35)	5·59 (5·43–5·74)	1·79 (1·71–1·88)

Table S3. Distribution of events, time to event and age at event for initial presentation of cardiovascular disease among patients with no diabetes or type 2 diabetes

Initial presentation of cardiovascular disease	No diabetes (107 501 events)			Type 2 diabetes (6137 events)		
	% of events	Median time to event (years)	Median age at event (years)	% of events	Median time to event (years)	Median age at event (years)
Stable angina	11·4	3·2	67·0	11·9	2·5	68·6
Unstable angina	4·9	4·5	63·1	4·0	2·8	64·7
Coronary disease not further specified	9·3	4·4	67·3	10·2	3·3	67·4
Non-fatal myocardial infarction	14·1	4·7	66·7	11·5	3·4	71·1
Unheralded coronary death	4·7	5·1	76·2	4·2	4·1	77·0
Heart failure	12·2	3·6	79·8	14·1	3·1	76·9
Arrhythmia or sudden cardiac death	3·0	5·0	67·1	1·6	4·0	68·3
Transient ischaemic attack	10·2	4·0	74·3	8·4	3·4	75·2
Ischaemic stroke	5·2	5·5	76·4	5·1	4·0	76·0
Stroke not further specified	9·4	3·4	79·7	10·3	2·5	78·3
Subarachnoid haemorrhage	1·2	4·4	57·3	0·2	2·8	74·4
Intracerebral haemorrhage	2·1	4·8	74·0	1·4	4·2	74·7
Peripheral arterial disease	9·4	4·4	70·1	16·2	3·5	71·7
Abdominal aortic aneurysm	2·8	5·0	76·2	1·0	4·5	77·8
Overall	100·0	4·2	72·0	100·0	3·2	72·6

Table S4. Cumulative incidence of initial presentation of cardiovascular diseases at age 80 for patients with type 2 diabetes or no diabetes at age 40

Cumulative incidences expressed as percentages (95% CI), taking into account risk of competing events

	Women		Men	
Cardiovascular disease initially presenting as:	**No diabetes**	**Type 2 diabetes**	**No diabetes**	**Type 2 diabetes**
Stable angina	4·5 (4·4–4·6)	10·1 (8·7–11·5)	5·7 (5·5–5·8)	9·5 (8·4–10·7)
Unstable angina or coronary disease not further specified	5·2 (5·1–5·3)	11·6 (10·0–13·2)	7·1 (6·9–7·3)	13·1 (11·7–14·5)
Non-fatal myocardial infarction	3·4 (3·3–3·5)	6·0 (4·9–7·2)	8·0 (7·8–8·1)	11·4 (9·9–12·8)
Unheralded coronary death	0·9 (0·9–1·0)	1·5 (0·9–2·0)	2·2 (2·1–2·3)	2·4 (1·9–2·9)
Heart failure	3·3 (3·1–3·4)	6·4 (5·4–7·3)	3·7 (3·6–3·9)	6·1 (5·2–6·9)
Arrhythmia or sudden cardiac death	0·9 (0·8–1·0)	0·7 (0·4–1·0)	1·7 (1·6–1·8)	1·8 (1·2–2·4)
Transient ischaemic attack	3·6 (3·5–3·7)	4·1 (3·3–4·9)	3·8 (3·7–3·9)	3·9 (3·3–4·6)
Ischaemic or unspecified stroke	3·9 (3·8–4·1)	7·3 (6·2–8·5)	4·7 (4·6–4·9)	6·1 (5·3–6·9)
Subarachnoid haemorrhage	0·6 (0·5–0·6)	0·2 (0·0–0·4)	0·3 (0·3–0·4)	0·0 (0·0–0·1)
Intracerebral haemorrhage	0·7 (0·6–0·7)	0·4 (0·2–0·6)	0·8 (0·7–0·8)	0·8 (0·5–1·1)
Peripheral arterial disease	3·2 (3·1–3·3)	9·7 (8·4–11·1)	4·5 (4·4–4·7)	11·7 (10·5–13·0)
Abdominal aortic aneurysm	0·5 (0·5–0·6)	0·1 (0·0–0·2)	1·7 (1·6–1·8)	0·6 (0·4–0·8)
All cardiovascular presentations	30·7 (30·3–31·0)	58·2 (54·9–61·4)	44·3 (43·8–44·7)	67·4 (64·4–70·4)
Death from other causes without any cardiovascular disease prior to death	14·7 (14·5–15·0)	12·4 (11·0–13·8)	16·4 (16·2–16·6)	11·9 (10·7–13·1)

Table S5. Hazard ratios obtained by analyses using different levels of adjustment

Initial presentation	**Primary analysis (adjusted for age, sex, risk factors, statins and antihypertensives)**		**Adjusted for age, sex and risk factors**		**Adjusted for age and sex only**	
	Hazard ratio (95% CI)	**P value**	**Hazard ratio (95% CI)**	**P value**	**Hazard ratio (95% CI)**	**P value**
Stable angina	1·62 (1·49–1·77)	< 0·0001	1·79 (1·64–1·95)	< 0·0001	1·93 (1·79–2·08)	< 0·0001
Unstable angina	1·53 (1·32–1·76)	< 0·0001	1·71 (1·49–1·97)	< 0·0001	1·84 (1·62–2·10)	< 0·0001
Coronary disease not further specified	1·58 (1·45–1·73)	< 0·0001	1·93 (1·76–2·11)	< 0·0001	2·11 (1·95–2·29)	< 0·0001
Non-fatal myocardial infarction	1·54 (1·42–1·67)	< 0·0001	1·50 (1·38–1·62)	< 0·0001	1·57 (1·46–1·70)	< 0·0001
Unheralded coronary death	1·43 (1·23–1·65)	< 0·0001	1·39 (1·20–1·60)	< 0·0001	1·48 (1·31–1·68)	< 0·0001
Heart failure	1·56 (1·45–1·69)	< 0·0001	1·58 (1·47–1·70)	< 0·0001	1·80 (1·68–1·93)	< 0·0001
Arrhythmia or sudden cardiac death	0·95 (0·76–1·19)	0·65	0·99 (0·80–1·24)	0·94	1·12 (0·92–1·37)	0·26
Transient ischaemic attack	1·45 (1·31–1·60)	< 0·0001	1·40 (1·27–1·55)	< 0·0001	1·42 (1·30–1·55)	< 0·0001
Ischaemic stroke	1·72 (1·52–1·95)	< 0·0001	1·68 (1·49–1·91)	< 0·0001	1·72 (1·53–1·93)	< 0·0001
Stroke not further specified	1·64 (1·48–1·81)	< 0·0001	1·64 (1·48–1·82)	< 0·0001	1·77 (1·63–1·92)	< 0·0001
Subarachnoid haemorrhage	0·48 (0·26–0·89)	0·020	0·48 (0·26–0·89)	0·020	0·45 (0·25–0·81)	0·0077
Intracerebral haemorrhage	1·28 (1·02–1·62)	0·035	1·18 (0·94–1·49)	0·15	1·20 (0·97–1·50)	0·099
Peripheral arterial disease	2·98 (2·76–3·22)	< 0·0001	3·07 (2·84–3·32)	< 0·0001	3·14 (2·94–3·35)	< 0·0001
Abdominal aortic aneurysm	0·46 (0·35–0·59)	< 0·0001	0·49 (0·38–0·63)	< 0·0001	0·52 (0·40–0·67)	< 0·0001
Other death	1·10 (1·05–1·17)	0.0004	1·01 (0·96–1·07)	0·70	1·04 (0·992–1·09)	0·11

All analyses used the complete dataset with multiple imputation of missing covariate values. 'Risk factors' comprised body mass index, deprivation, high density lipoprotein cholesterol, total cholesterol, systolic blood pressure and smoking status.

Table S6. Hazard ratios from complete case analyses

Initial presentation	Adjusted for age, sex, risk factors, statins and antihypertensives		Adjusted for age, sex and risk factors		Adjusted for age and sex only	
	Hazard ratio (95% CI)	P value	Hazard ratio (95% CI)	P value	Hazard ratio (95% CI)	P value
Stable angina	2·03 (1·59–2·60)	< 0·0001	2·10 (1·65–2·67)	< 0·0001	2·10 (1·67–2·64)	< 0·0001
Unstable angina	1·16 (0·78–1·73)	0·45	1·28 (0·87–1·88)	0·22	1·19 (0·83–1·72)	0·34
Coronary disease not further specified	1·64 (1·30–2·07)	< 0·0001	1·69 (1·35–2·13)	< 0·0001	1·75 (1·41–2·16)	< 0·0001
Non-fatal myocardial infarction	1·50 (1·17–1·93)	0·0016	1·49 (1·16–1·90)	0·0018	1·48 (1·17–1·88)	0·0011
Unheralded coronary death	2·22 (1·43–3·46)	0·0004	1·93 (1·24–3·02)	0·0038	1·73 (1·16–2·59)	0·0070
Heart failure	3·01 (2·22–4·07)	< 0·0001	2·96 (2·19–3·99)	< 0·0001	3·29 (2·48–4·35)	< 0·0001
Arrhythmia or sudden cardiac death	0·92 (0·49–1·73)	0·79	0·88 (0·47–1·63)	0·68	0·83 (0·46–1·51)	0·55
Transient ischaemic attack	1·33 (0·97–1·81)	0·076	1·27 (0·93–1·73)	0·13	1·20 (0·89–1·61)	0·23
Ischaemic stroke	1·61 (1·05–2·46)	0·028	1·57 (1·03–2·40)	0·034	1·47 (0·98–2·19)	0·060
Stroke not further specified	1·16 (0·83–1·64)	0·39	1·14 (0·81–1·60)	0·46	1·22 (0·88–1·69)	0·23
Subarachnoid haemorrhage	–	–	–	–	–	–
Intracerebral haemorrhage	1·20 (0·55–2·63)	0·66	1·15 (0·53–2·50)	0·73	0·93 (0·45–1·93)	0·85
Peripheral arterial disease	3·05 (2·40–3·88)	< 0·0001	3·16 (2·50–4·01)	< 0·0001	2·80 (2·24–3·49)	< 0·0001
Abdominal aortic aneurysm	0·30 (0·14–0·67)	0·0034	0·30 (0·14–0·65)	0·0026	0·31 (0·14–0·66)	0·0026
Other death	1·63 (1·38–1·93)	< 0·0001	1·53 (1·29–1·80)	< 0·0001	1·45 (1·24–1·69)	< 0·0001

These analyses were limited to the subset of individuals with completely recorded covariate information: 59 116 with no diabetes and 12 411 with type 2 diabetes. 'Risk factors' comprised body mass index, deprivation, high density lipoprotein cholesterol, total cholesterol, systolic blood pressure and smoking status.

Table S7. Hazard ratios from secondary analyses

Initial presentation	Patients with any diabetes versus no diabetes		Type 2 diabetes versus no diabetes, restricted to patients entering study after 2004	
	Hazard ratio (95% CI)	P value	Hazard ratio (95% CI)	P value
Stable angina	1·70 (1·57–1·83)	< 0·0001	1·62 (1·17–2·23)	0·0030
Unstable angina	1·66 (1·47–1·87)	< 0·0001	1·30 (0·86–1·95)	0·21
Coronary disease not further specified	1·73 (1·60–1·87)	< 0·0001	1·33 (1·03–1·73)	0·031
Non-fatal myocardial infarction	1·74 (1·62–1·86)	< 0·0001	1·48 (1·14–1·91)	0·0031
Unheralded coronary death	1·81 (1·60–2·04)	< 0·0001	1·93 (1·31–2·84)	0.00098
Heart failure	1·76 (1·65–1·87)	< 0·0001	2·02 (1·55–2·63)	< 0·0001
Arrhythmia or sudden cardiac death	1·16 (0·97–1·39)	0.11	1·10 (0·58–2·06)	0·78
Transient ischaemic attack	1·54 (1·42–1·68)	< 0·0001	1·24 (0·91–1·69)	0·18
Ischaemic stroke	1·72 (1·54–1·92)	< 0·0001	1·42 (1·00–2·01)	0·051
Stroke not further specified	1·80 (1·65–1·95)	< 0·0001	1·24 (0·99–1·56)	0·059
Subarachnoid haemorrhage	0·54 (0·33–0·89)	0·016	–	–
Intracerebral haemorrhage	1·40 (1·15–1·70)	0·0007	1·10 (0·53–2·27)	0·80
Peripheral arterial disease	3·33 (3·12–3·55)	< 0·0001	2·70 (2·12–3·44)	< 0·0001
Abdominal aortic aneurysm	0·49 (0·39–0·62)	< 0·0001	0·33 (0·15–0·72)	0·0056
Other death	1·39 (1·33–1·45)	< 0·0001	1·46 (1·29–1·64)	< 0·0001

Hazard ratios were adjusted for age, sex, risk factors (body mass index, deprivation, high density lipoprotein cholesterol, total cholesterol, systolic blood pressure, smoking status), and statin and antihypertensive prescription in the year before study entry.

Table S8. Hazard ratios for composite endpoints

Endpoint	Adjusted for age, sex, risk factors, statins and antihypertensives		Adjusted for age, sex and risk factors		Adjusted for age and sex only	
	Hazard ratio (95% CI)	P value	Hazard ratio (95% CI)	P value	Hazard ratio (95% CI)	P value
Cardiovascular death	1·53 (1·45–1·62)	< 0·0001	1·62 (1·53–1·71)	< 0·0001	1·72 (1·64–1·81)	< 0·0001
Non-cardiovascular death	1·29 (1·24–1·34)	< 0·0001	1·22 (1·17–1·26)	< 0·0001	1·25 (1·21–1·29)	< 0·0001
Death due to any cause	1·35 (1·31–1·39)	< 0·0001	1·31 (1·27–1·36)	< 0·0001	1·36 (1·32–1·40)	< 0·0001

All analyses used the complete dataset with multiple imputation of missing covariate values. 'Risk factors' comprised body mass index, deprivation, high density lipoprotein cholesterol, total cholesterol, systolic blood pressure and smoking status.

Supplementary Figures

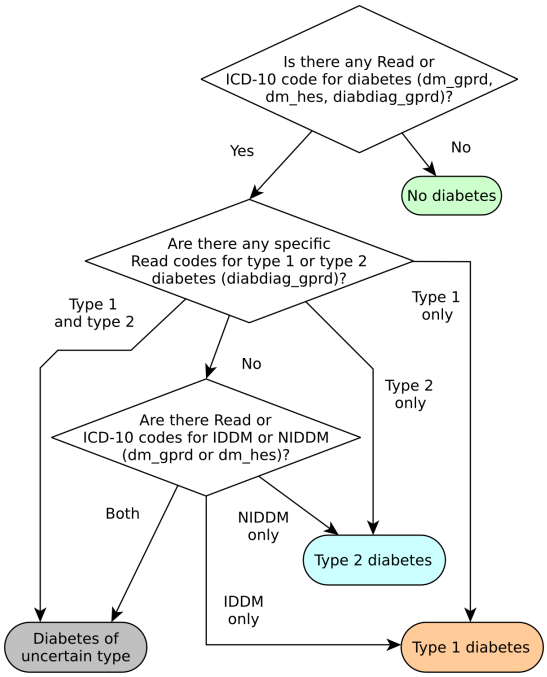
Figure S1. Patient flow diagram and classification of baseline diabetes status

The aim of this algorithm is to identify patients with type 2 diabetes. The majority of diabetic patients in this cohort have type 2 diabetes, and we are not aiming to identify all patients with type 1 diabetes (this algorithm would not be suitable for a study focusing on type I diabetes). Our algorithm does not use any information after study entry to classify patients, in order to reduce immortal time bias. However this bias is not eliminated completely because specific type 1 or type 2 codes may have been entered retrospectively only for patients who survived.

Our algorithm leaves many patients with an unclassified type of diabetes. However, rather than using a more complex algorithm, we carried out sensitivity analyses with different definitions (e.g. the type 2 definition or any diabetes).

Lists of Read and ICD-10 diagnostic codes which convey diabetic status are curated on the CALIBER data portal (https://www.caliberresearch.org/portal/chapter/6).

A. CALIBER phenotype algorithm for diabetes



Consider diagnosis codes recorded in CPRD (Read) and HES (ICD-10) on or before the date on which diabetes status is to be ascertained.

B. Patient flow diagram

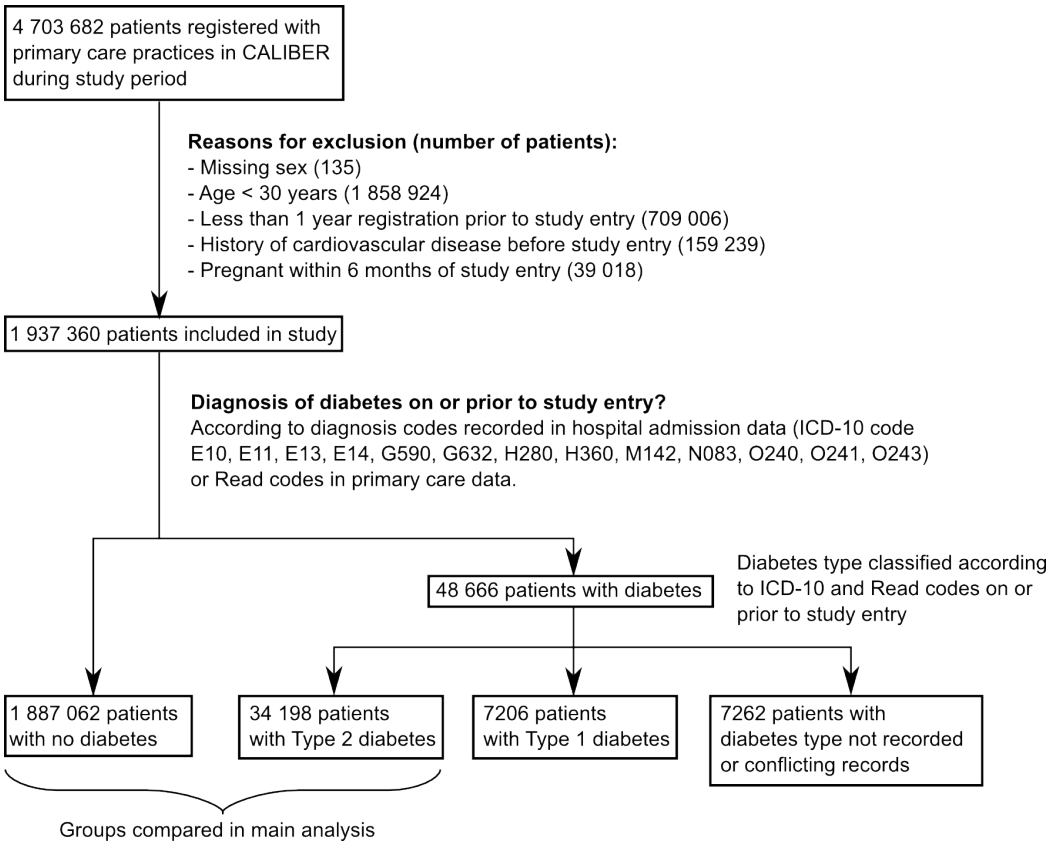


Figure S2. Association of type 2 diabetes with initial presentation of cardiovascular diseases using different adjustments

'Risk factors' adjustment includes age, sex, body mass index, deprivation, high density lipoprotein cholesterol, total cholesterol, systolic blood pressure, and smoking status. 'Treatment' includes statins and antihypertensive prescription in the year before study entry. P values *** < 0·001, ** < 0·01, * < 0·05

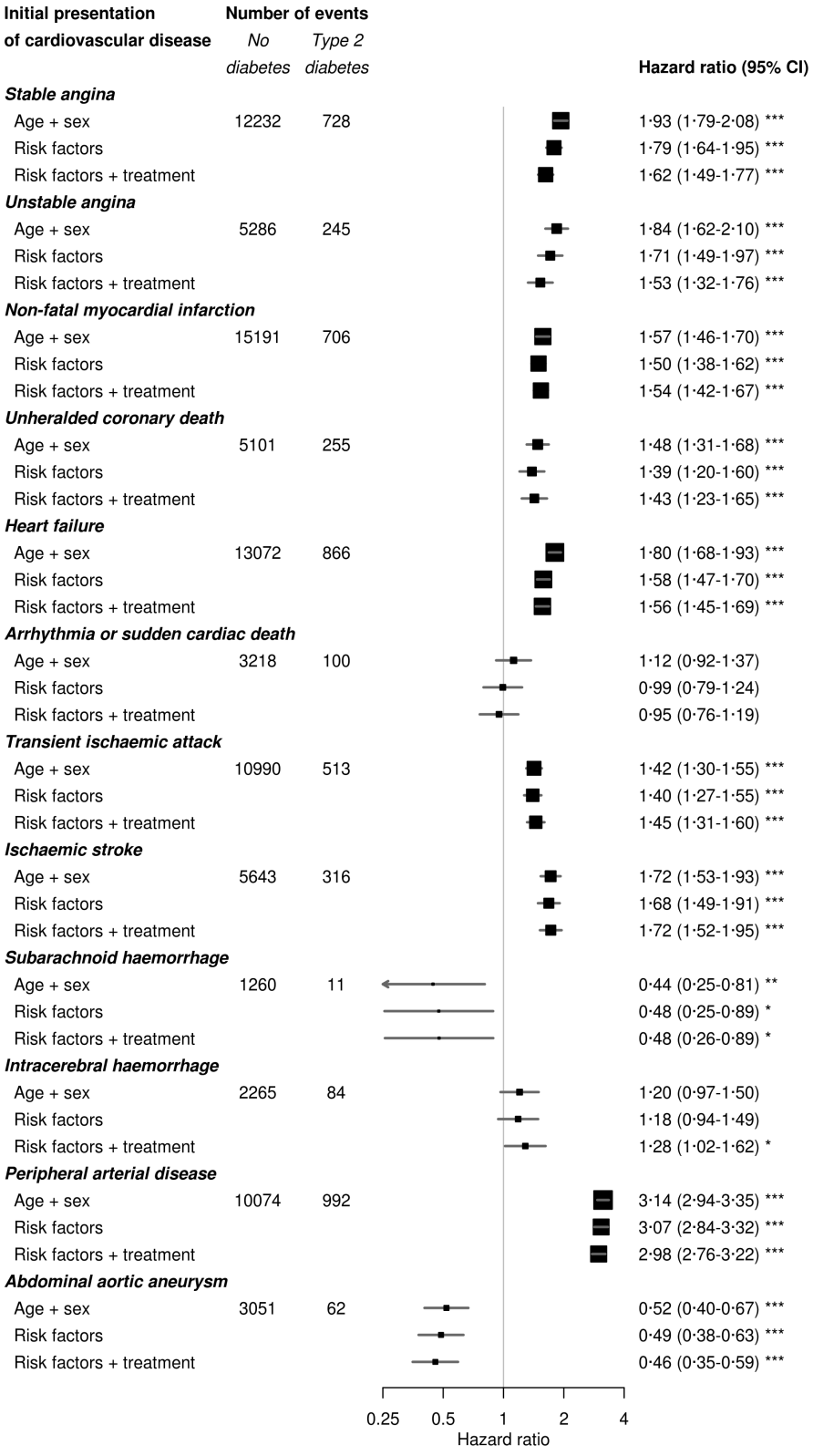


Figure S3. Adjusted hazard ratios for type 2 diabetes and initial presentations of cardiovascular diseases by age group

Hazard ratios by age group for the association of different initial presentations of cardiovascular disease with type 2 diabetes, adjusted for sex, body mass index, deprivation, high density lipoprotein cholesterol, total cholesterol, systolic blood pressure, smoking status, statins and antihypertensive prescriptions. P values *** < 0·001, ** < 0·01, * < 0·05

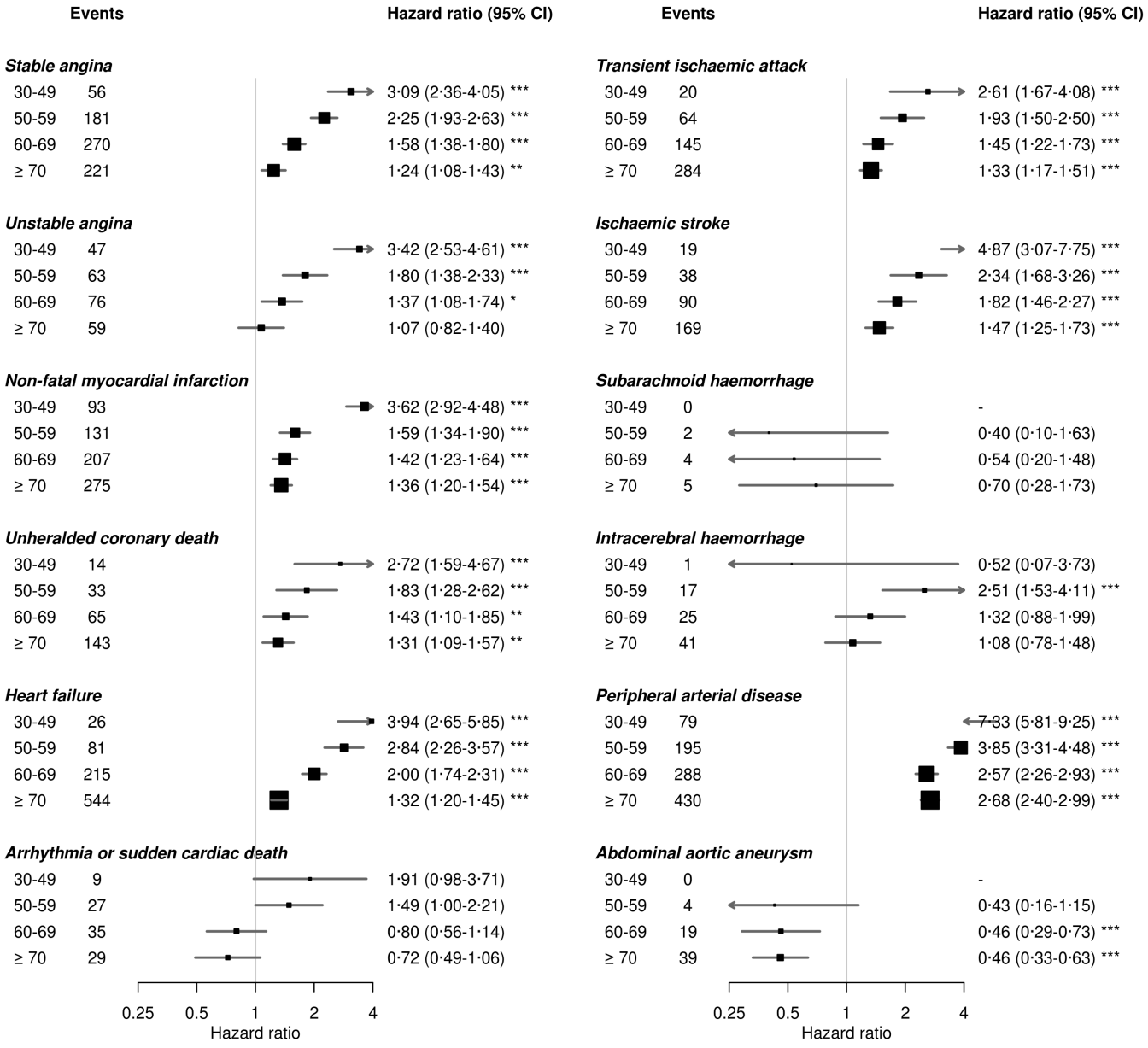


Figure S4. Association of type 2 diabetes with initial presentation of cardiovascular diseases by level of glycaemic control.

Hazard ratios for initial presentations of cardiovascular diseases associated with type 2 diabetes with different levels of HbA1c (mean in mmol/mol within 3 years of baseline), compared with no diabetes, adjusted for age, sex, body mass index, deprivation, high density lipoprotein cholesterol, total cholesterol, systolic blood pressure, smoking status, statins and antihypertensive prescriptions. P values *** < 0·001, ** < 0·01, * < 0·05

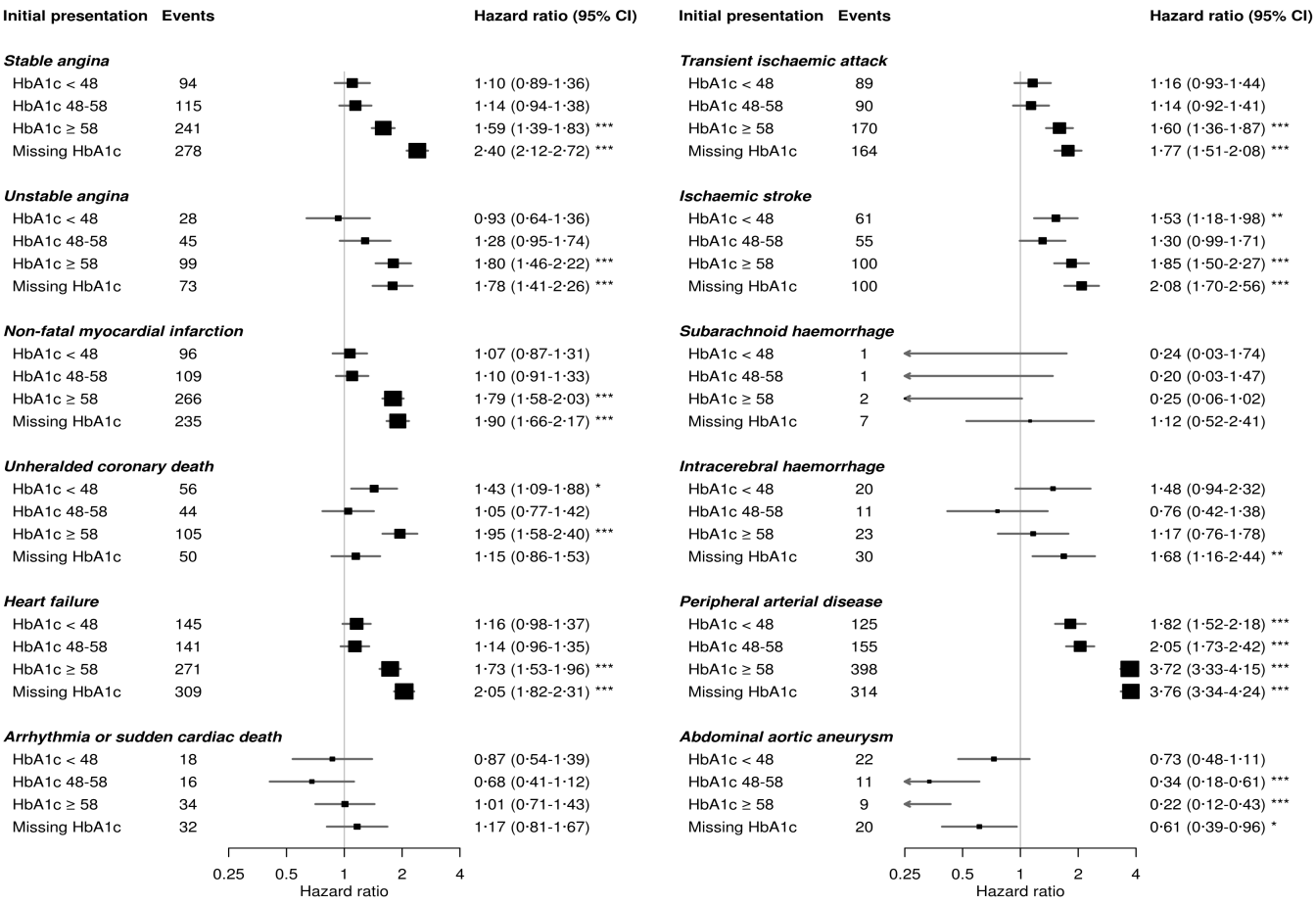


Figure S5. Association of type 2 diabetes with initial presentation of cardiovascular diseases using different subsets of the data sources.

Hazard ratios adjusted for age, sex, body mass index, deprivation, high density lipoprotein cholesterol, total cholesterol, systolic blood pressure, smoking status, statins and antihypertensive prescriptions. P values *** < 0·001, ** < 0·01, * < 0·05

Initial presentation of cardiovascular disease	Number of events: *No diabetes*	*Type 2 diabetes*	Hazard ratio (95% CI)
Stable angina			
Using all data sources (main analysis)	12232	728	1·62 (1·49-1·77) ***
Ignoring primary care	1204	50	1·46 (1·08-1·97) *
Unstable angina			
Using all data sources (main analysis)	5286	245	1·53 (1·32-1·76) ***
Ignoring primary care	9092	471	1·70 (1·54-1·88) ***
Non-fatal myocardial infarction			
Using all data sources (main analysis)	15191	706	1·54 (1·42-1·67) ***
Ignoring primary care	14316	683	1·58 (1·46-1·72) ***
Heart failure			
Using all data sources (main analysis)	13072	866	1·56 (1·45-1·69) ***
Ignoring primary care	6582	572	2·15 (1·96-2·36) ***
Using death endpoints only	1835	127	1·59 (1·27-1·98) ***
Arrhythmia or sudden cardiac death			
Using all data sources (main analysis)	3218	100	0·95 (0·76-1·19)
Ignoring primary care	3505	112	0·95 (0·78-1·17)
Transient ischaemic attack			
Using all data sources (main analysis)	10990	513	1·45 (1·31-1·60) ***
Ignoring primary care	3437	175	1·63 (1·39-1·92) ***
Ischaemic stroke			
Using all data sources (main analysis)	5643	316	1·72 (1·52-1·95) ***
Ignoring primary care	7547	427	1·75 (1·57-1·94) ***
Using death endpoints only	896	48	1·61 (1·15-2·24) **
Subarachnoid haemorrhage			
Using all data sources (main analysis)	1260	11	0·48 (0·26-0·89) *
Ignoring primary care	1579	14	0·48 (0·28-0·84) **
Using death endpoints only	607	6	0·42 (0·18-0·94) *
Intracerebral haemorrhage			
Using all data sources (main analysis)	2265	84	1·28 (1·02-1·62) *
Ignoring primary care	2498	86	1·14 (0·91-1·43)
Using death endpoints only	1274	46	1·03 (0·75-1·42)
Peripheral arterial disease			
Using all data sources (main analysis)	10074	992	2·98 (2·76-3·22) ***
Ignoring primary care	3296	297	2·92 (2·56-3·33) ***
Using death endpoints only	330	7	0·60 (0·27-1·30)
Abdominal aortic aneurysm			
Using all data sources (main analysis)	3051	62	0·46 (0·35-0·59) ***
Ignoring primary care	2638	47	0·40 (0·30-0·54) ***
Using death endpoints only	1855	30	0·34 (0·23-0·50) ***

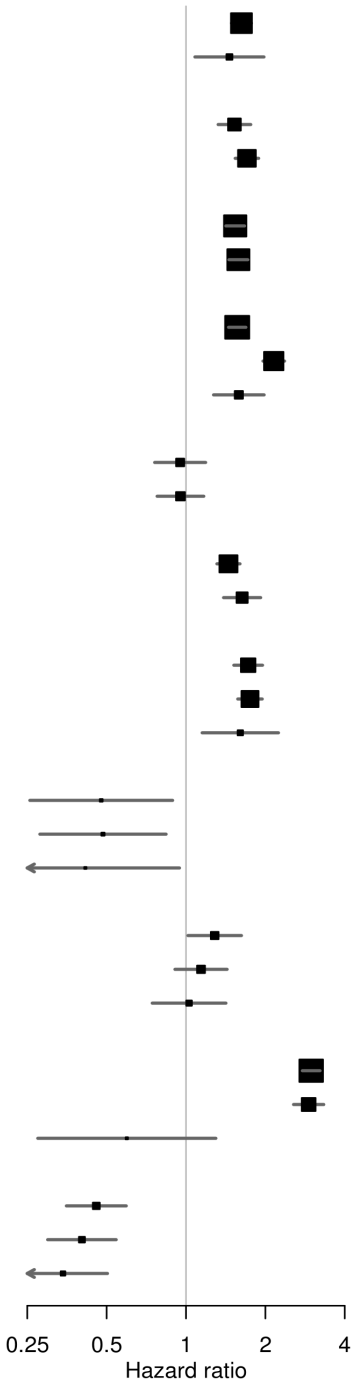
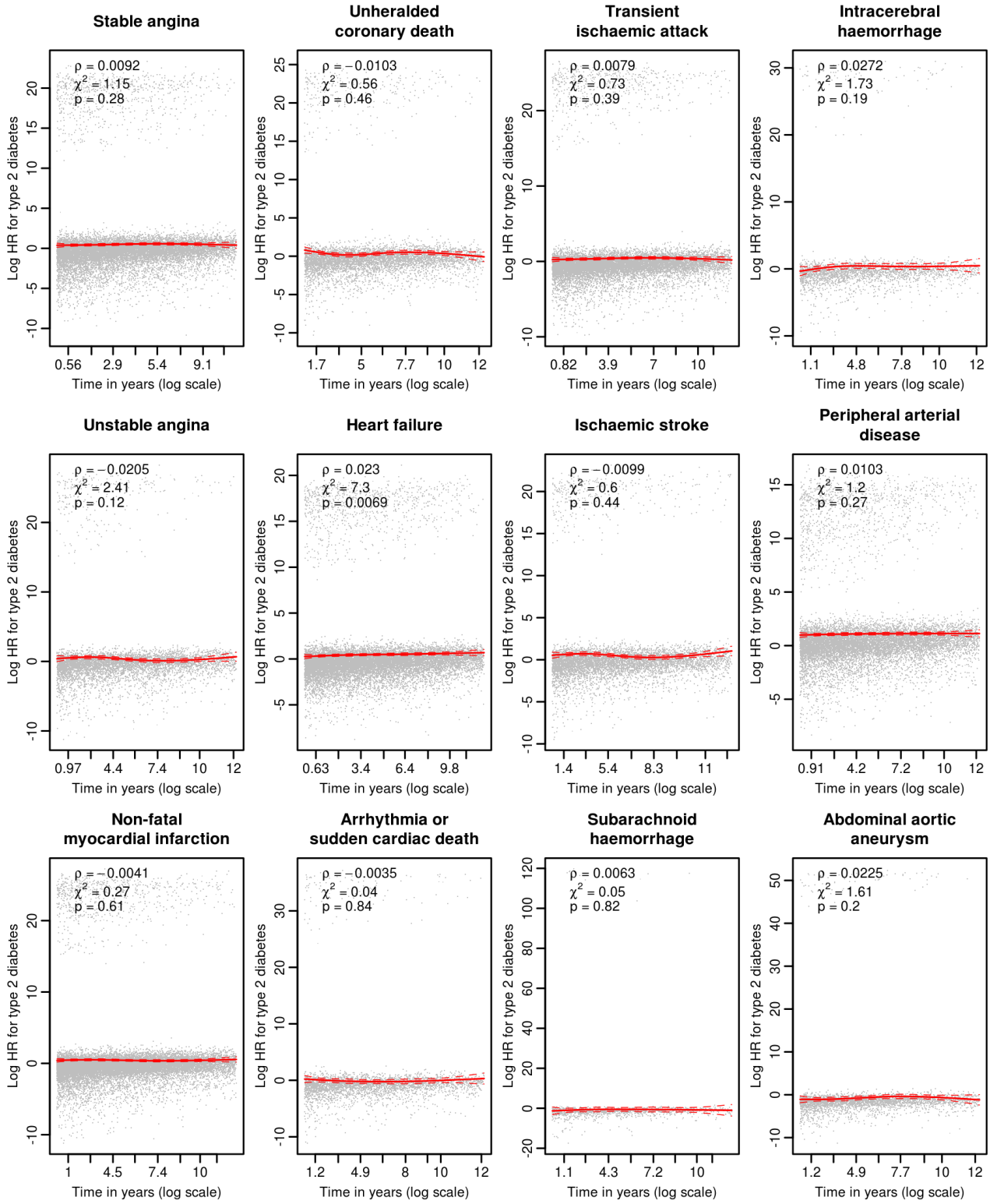


Figure S6. Schoenfeld residuals and tests for proportional hazards

Scaled Shoenfeld residuals for type 2 diabetes versus no diabetes, adjusted for age, sex, body mass index, deprivation, high density lipoprotein cholesterol, total cholesterol, systolic blood pressure, smoking status, statins and antihypertensive prescriptions. The lines show the estimate and 95% confidence interval for the beta coefficient for type 2 diabetes over time. The top left corner of each graph contains ρ, χ^2 and p value for the correlation between transformed survival time and scaled Schoenfeld residuals. For clarity, points plotted on the graph are for a random sample of 20 000 patients rather than all 1.9 million.



References

1. Denaxas SC, George J, Herrett E, Shah AD, Kalra D, Hingorani AD, et al. Data resource profile: cardiovascular disease research using linked bespoke studies and electronic health records (CALIBER). *Int J Epidemiol.* 2012; 41: 1625-38.

2. Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. Br J Clin Pharmacol. 2010;69(1):4-14. doi: 10.1111/j.1365-2125.2009.03537.x.

3. Hospital Episode Statistics. Health and Social Care Information Centre, Department of Health, UK. http://www.hscic.gov.uk/hes. Accessed January 15, 2014.

4. Herrett E, Smeeth L, Walker L, Weston C; MINAP Academic Group. The Myocardial Ischaemia National Audit Project (MINAP). Heart. 2010;96(16):1264-7. doi: 10.1136/hrt.2009.192328.

5. English indices of deprivation 2010: technical report. Department for Communities and Local Government, 2011. ISBN 9781409829225. https://www.gov.uk/government/publications/english-indices-of-deprivation-2010-technical-report. Accessed January 15, 2014.

6. Qiu W, Chavarro J, Lazarus R, Rosner B, Ma J (2012). powerSurvEpi: Power and sample size calculation for survival analysis of epidemiological studies. R package version 0.0.6. http://CRAN.R-project.org/package=powerSurvEpi

7. van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. *Stat Methods Med Res.* 2007;16:219-242.

8. White IR, Royston P. Imputing missing covariate values for the Cox model. *Stat Med.* 2009; 28: 1982-1998.

9. The Action to Control Cardiovascular Risk in Diabetes Study Group. Effects of Intensive Glucose Lowering in Type 2 Diabetes. *N Engl J Med.* 2008; 358:2545-255. doi: 10.1056/NEJMoa0802743.

10. Parving HH, Brenner BM, McMurray JJ, et al. Cardiorenal end points in a trial of aliskiren for type 2 diabetes. *N Engl J Med*. 2012;367(23):2204-13. doi: 10.1056/NEJMoa1208799.

11. Green JB, Bethel MA, Paul SK, et al. Rationale, design, and organization of a randomized, controlled Trial Evaluating Cardiovascular Outcomes with Sitagliptin (TECOS) in patients with type 2 diabetes and established cardiovascular disease. *Am Heart J*. 2013 Dec;166(6):983-989.e7. doi: 10.1016/j.ahj.2013.09.003.

12. Scirica BM, Bhatt DL, Braunwald E, et al. Saxagliptin and cardiovascular outcomes in patients with type 2 diabetes mellitus. *N Engl J Med*. 2013;369(14):1317-26. doi: 10.1056/NEJMoa1307684.

13. Home PD, Pocock SJ, Beck-Nielsen H. et al. Rosiglitazone evaluated for cardiovascular outcomes—an interim analysis. *N Engl J Med*. 2007;357(1):28-38.

14. Young LH, Wackers FJT, Chyun DA. Cardiac Outcomes After Screening for Asymptomatic Coronary Artery Disease in Patients With Type 2 Diabetes The DIAD Study: A Randomized Controlled Trial. *JAMA*. 2009;301(15):1547-1555. doi:10.1001/jama.2009.476.