BMJ Open

Incidence and prevalence of cardiovascular disease in English primary care: a cross-sectional and follow-up study of the Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC)

William Hinton,[1] Andrew McGovern,[1] Rachel Coyle,[1] Thang S Han,[1,2,3] Pankaj Sharma,[1,2] Ana Correa,[1,4] Filipa Ferreira,[1] Simon de Lusignan[1,4]

To cite: Hinton W, McGovern A, Coyle R, *et al*. Incidence and prevalence of cardiovascular disease in English primary care: a cross-sectional and follow-up study of the Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC). *BMJ Open* 2018;**8**:e020282. doi:10.1136/bmjopen-2017-020282

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2017-020282).

Received 26 October 2017
Revised 19 June 2018
Accepted 23 July 2018

Check for updates

© Author(s) (or their employer(s)) 2018. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to
Professor Simon de Lusignan; s.lusignan@surrey.ac.uk

ABSTRACT

Objectives To describe incidence and prevalence of cardiovascular disease (CVD), its risk factors, medication prescribed to treat CVD and predictors of CVD within a nationally representative dataset.

Design Cross-sectional study of adults with and without CVD.

Setting The Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC) is an English primary care sentinel network. RCGP RSC is over 50 years old and one of the oldest in Europe. Practices receive feedback about data quality. This database is primarily used to conduct surveillance and research into influenza, infections and vaccine effectiveness but is also a rich resource for the study of non-communicable disease (NCD). The RCGP RSC network comprised 164 practices at the time of study.

Results Data were extracted from the records of 1 275 174 adults. Approximately a fifth (21.3%; 95% CI 21.2% to 21.4%) had CVD (myocardial infarction (MI), angina, atrial fibrillation (AF), peripheral arterial disease, stroke/transient ischaemic attack (TIA), congestive cardiac failure) or hypertension. Smoking, unsafe alcohol consumption and obesity were more common among people with CVD. Angiotensin system modulating drugs, 3-hydroxy-3-methylglutaryl-coenzyme (HMG-CoA) reductase inhibitors (statins) and calcium channel blockers were the most commonly prescribed CVD medications. Age-adjusted and gender-adjusted annual incidence for AF was 28.2/10 000 (95% CI 27.8 to 28.7); stroke/TIA 17.1/10 000 (95% CI 16.8 to 17.5) and MI 9.8/10 000 (95% CI 9.5 to 10.0). Logistic regression analyses confirmed established CVD risk factors were associated with CVD in the RCGP RSC network dataset.

Conclusions The RCGP RSC database provides comprehensive information on risk factors, medical diagnosis, physiological measurements and prescription history that could be used in CVD research or pharmacoepidemiology. With the exception of MI, the prevalence of CVDs was higher than in other national data, possibly reflecting data quality. RCGP RSC is an underused resource for research into NCDs and their management and welcomes collaborative opportunities.

Strengths and limitations of this study

- ► High data quality for demographic, clinical and social variables.
- ► Ability to report prevalence and incidence.
- ► From our rich dataset, we are able to report diagnosis and pharmacological and physiological (eg, blood pressure) data.
- ► Data collected in primary care represent only a brief abstraction from a consultation.
- ► Prevalence of myocardial infarction was lower than reported in other UK primary care datasets linked to hospital data. This might indicate that acute conditions are less well coded in primary care records.

INTRODUCTION

Cardiovascular disease (CVD) is a major cause of mortality and morbidity. Improved preventive strategies could reduce the burden of disease. There were over 17 million CVD-related deaths in 2012; CVD accounts for almost one-third of all deaths.[1] Ischaemic heart disease has topped the list of causes of years of life lost for more than a decade,[2] highlighting the shift in the global burden of disease from communicable to chronic disease.[3] Risk factors for CVD, including raised blood pressure, hypercholesterolaemia and high body mass index (BMI), are among the most important contributors to disability-adjusted life years.[4] Primary prevention of CVD is achievable through early identification and modification of 'lifestyle risk factors' and secondary prevention, through appropriate

risk reduction, which slows disease progression. Effective prevention and treatment is reliant on the identification of people at risk of or with current CVD, and systems which facilitate monitoring of management.

Large datasets provide real-world insights into the epidemiology of CVD and better use of analytics may lead to improvements in care.[5] Notwithstanding, the widespread use of computerised medical record (CMR) published research on the use of large datasets in CVD research remains limited.[6] The majority of CVD management takes place in primary care, providing opportunities for research.[7 8] In the UK, a pay-for-performance scheme (P4P), the Quality and Outcomes Framework (QOF) was introduced to incentivise general practitioner (GP) practices to achieve indicator thresholds for the management of chronic diseases,[9] which has enhanced data quality in primary care. The ubiquitous use of CMR systems in UK primary care allows ready analysis enhanced by the UK's registration-based systems, which provide an accurate denominator.[10]

The English Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC) is one of the longest established primary care sentinel networks.[11] It comprises a nationally representative sample,[12] monitoring infections and respiratory disease, particularly influenza and assessing vaccine effectiveness.[13 14 15] The RCGP RSC extended its remit into diabetes, with a focus on disparities and adherence/persistence, and has been shown to have good data quality.[16] One recent analysis has used the cardiovascular outcome data from the RCGP RSC[17] but there has not previously been a systematic assessment of CVD data quality and completeness. We describe the incidence and prevalence of CVD in the RCGP RSC cohort; including across different patient characteristics, medications and risk factors for CVD.

OBJECTIVES

- To identify the proportion of adults currently registered within the RCGP RSC who have a diagnosis of CVD.
- To describe the prevalence of risk factors for CVD and medications prescribed within the RSCP RSC network.
- To compare the incidence and prevalence with other nationally reported datasets.

METHODS

Study design and study population

This is a cross-sectional study using routinely collected data extracted from general practices registered within the RCGP RSC. All adults registered at the end of the study period, 31 December 2016, were included. Data were contributed by 164 practices across England.

Data source

The Read classification is currently used to code key clinical data.[18] Diagnoses and key symptoms, examination findings, therapies, investigations and test results, and processes of care are increasingly coded. Interactions with secondary care are generally coded into a patient's primary care record.

Data quality is very important to RCGP RSC.[19 20] A sentinel system has to differentiate first, from new to follow-up of cases. So there is a lot of emphasis on making sure all consultations have a 'problem' coded, and it is so-called 'episode type' (ie, first, new or ongoing). Practices receive extensive feedback, practice visits, a monthly newsletter and more recently a dashboard refreshed weekly.[21] We are also looking at ways of automated classification of unclassified episodes.[22]

The individuals in the cohort have their identity pseudonymised. We pseudonymise National Health Service (NHS) number to protect patients' privacy—but can apply the same pseudonymisation algorithm to other datasets (eg, hospital data or cancer registries), so we can privately link patient-level data. We can, in addition, use a combination of probabilistic and deterministic methods to link data where there is no NHS number.[23]

Data collection takes place twice per week with near real-time processing

We collect data from all patients twice per week, making us one of the freshest data sources. We have the capability to characterise and monitor disease and use of therapies, and compare use according to guidance.[24] The RCGP RSC has produced a 'weekly return' of infections and respiratory disease since 1967, though over time it has expanded in terms of size, scope, sample collection and its capability with linking with other datasets. We can process large numbers of patients' data rapidly. Our weekly report involves the processing of around 1.75 million patients data to produce our weekly report of over 30 conditions in around 4 hours.

Identifying CVD

Demographic, social and clinical characteristics of the study population were identified using Read codes. Ethnicity was characterised using an established ontological approach: a combination of recorded ethnicity and information which infers ethnicity (such as language).[25] Socioeconomic status (SES) was determined using the Index of Multiple Deprivation (IMD—the official UK SES measure) from each patient's postal code,[26] which we also have the capability to map to geocoordinate.

We used the latest code recorded for each patient to report smoking status. The categories we created were: never, current and ex-smoker. For alcohol use, we included alcohol-related disease or complications. We stratified alcohol use into: not recorded, within limits (<14 units per week) or alcohol consumption with no amount specified or excess (≥14 units per week or other codes consistent with heavy drinking). We categorised BMI using WHO categories: underweight (<18.5 kg/m^2), normal (18.5–24.9), overweight (25.0–29.9), obese class 1 (30.0–34.9), obese class 2 (35.0–39.9) and obese class 3

(≥40.0).[27] For BMI and blood pressure, we used the latest value recorded. We also report missing data.

CVD cases include a diagnosis of one or more of the following conditions: coronary artery disease (including myocardial infarction and angina), atrial fibrillation (AF), peripheral arterial disease, stroke, transient ischaemic attack and congestive cardiac failure (online supplementary appendix 1, Read code list[28]). We included hypertension in our analysis of CVD.

Risk factors for CVD and disease prevalence

In addition to those described above, we measured prevalence of established CVD risk factors demonstrated to predict CVD.[29] Our analysis included latest systolic blood pressure (SBP) in excess of 140 mm Hg, chronic kidney disease (CKD, including stages 3–5) and diabetes mellitus.

Prevalence and incidence of CVD

Prevalence of CVD was described by the number of different conditions in people by age groups (<50 years, 50–59 years, 60–69 years, 70–79 years and ≥80 years). Incidence of CVD was derived for the last 5 years (2012–2016). We reported age-adjusted and gender-adjusted annual incidence of CVD using the 2011 Census for England and Wales. We also compared prevalence of CVD conditions and risk factors/comorbidities in the RCGP RSC dataset with those reported for the QOF/P4P national financially incentivised chronic disease management scheme.

CVD prescription use

We report the following cardiovascular medication prescriptions: ACE inhibitors, alpha blockers, angiotensin II receptor blockers, antiplatelet therapy (including aspirin), glycosides, calcium channel blockers, centrally acting antihypertensives, statins, fibrates, other lipid-lowering drugs, loop diuretics, nicorandil, nitrates, potassium sparing diuretics, super ventricular antiarrhythmic, thiazide and thiazide-like diuretics, vasodilator antihypertensives, and beta-blockers. We defined medication use as the recording of one or more prescriptions in the GP record.

Predictors of CVD

Logistic regression was used to identify variables that predict CVD. Variable selection was based on established risk factors for CVD. Since CVD is a composite of conditions, which present at different stages, we also explored associations between variables and the most common CVD conditions identified in this study.

Inclusion criteria

All adults, aged ≥18 years, who were registered at an RCGP RSC practice on 31 December 2016 were included in the analysis.

Exclusion criteria

Any patients who have codes suggesting they declined any form of data sharing are not analysed by RCGP RSC (approximately 2.2% of the registered population).

Statistical analysis

We used descriptive statistics calculated using the open source software, R-studio. We reported counts and percentages of crude data. Proportions were compared using χ^2 tests, while the independent samples t-test and Mann-Whitney U test were used to compare measures of central tendency (means and medians).

As mentioned above, we used logistic regression to identify variables that predict CVD as an outcome. Three models were run, which included the following outcome variables: (1) CVD as a composite of conditions (coronary artery disease, myocardial infarction, angina, AF, peripheral arterial disease, stroke, transient ischaemic attack and congestive cardiac failure); (2) coronary artery disease (including myocardial infarction and angina) and (3) AF.

Within each model, we adjusted for the following predictor variables: age, gender, ethnicity, SES (using IMD stratified into quintiles), smoking status, alcohol consumption, BMI category, presence of diabetes (no diabetes, type 1 diabetes, type 2 diabetes) and presence of other known risk factors (uncontrolled SBP: >140 mm Hg; hypertension and CKD (stages 3–5). All variables were retained in the regression analyses irrespective of associations in bivariate analyses.

We reported ORs with 95% CIs and p values for each model parameter. Results were deemed significant if they were associated with a significance level of $p<0.05$.

Ethics

The data used for the analysis were pseudonymised at the point of extraction and encrypted prior to uploading to the Clinical Informatics Research Group secure server. Personal data were not identifiable during the analysis. This study was approved by the RCGP RSC study approval committee and was classified as a study of 'usual practice'.[30] RCGP RSC data are available to researchers and applications should be made direct to RCGP.[31]

Patient and public involvement

No patients or public were involved in the design, recruitment or conduct of this study.

RESULTS

Demographics about people with and without CVD

A total of 1 275 174 adults were registered with RCGP RSC practices. SES was identified within almost the entire adult population and three-quarters had their ethnicity recorded. The least deprived quintiles (IMD quintiles 4 and 5) were over-represented in the cohort (table 1).

The prevalence of CVD and hypertension was 21.3% (271 684) and more common among males. Smoking, unsafe drinking and obesity were more prevalent among people with CVD than without, as was white ethnicity. There was a lower proportion of Asian, mixed or other ethnicity individuals among those diagnosed with CVD by comparison to those without CVD, and the proportion of

Table 1 The characteristics of the adult population with and without CVD

Characteristic	All adults, n (%) (n=1 275 174)	Any CVD* and hypertension, n (%) (n=271 684)	No CVD, n (%) (n=1 003 490)	P values
Age mean (SD)	47.5 (18.9)	67.9 (13.7)	41.9 (16.1)	<0.001
Age median (IQR)	46.0 (31.0–62.0)	69.0 (59.0–78.0)	40.0 (29.0–53.0)	<0.001
Female	651 298 (51.1)	133 765 (49.2)	517 533 (51.6)	<0.001
Ethnicity recorded	975 924 (76.5)	219 732 (80.9)	756 192 (75.4)	
White	828 900 (65.0)	198 884 (73.2)	630 016 (62.8)	<0.001
Asian	78 603 (6.2)	9853 (3.6)	68 750 (6.9)	<0.001
Black	39 805 (3.1)	8165 (3.0)	31 640 (3.2)	<0.001
Mixed	14 955 (1.2)	1577 (0.6)	13 378 (1.3)	<0.001
Other	13 661 (1.1)	1253 (0.5)	12 408 (1.2)	<0.001
Not recorded	299 250 (23.5)	51 952 (19.1)	247 298 (24.6)	<0.001
IMD recorded	1 271 764 (99.7)	271 120 (99.8)	1 000 644 (99.7)	
IMD quintile 5 (least)	337 239 (26.4)	75 278 (27.7)	261 961 (26.1)	<0.001
IMD quintile 4	277 357 (21.8)	62 736 (23.1)	214 621 (21.4)	<0.001
IMD quintile 3	232 524 (18.2)	50 534 (18.6)	181 990 (18.1)	<0.001
IMD quintile 2	216 466 (17.0)	41 311 (15.2)	175 155 (17.5)	<0.001
IMD quintile 1 (most deprived)	208 178 (16.3)	41 261 (15.2)	166 917 (16.6)	<0.001
IMD quintile not available	3410 (0.3)	564 (0.2)	2846 (0.3)	<0.001
Smoking status recorded	1 198 579 (94.0)	263 970 (97.2)	934 609 (93.1)	
Never	514 481 (40.3)	78 045 (28.7)	436 436 (43.5)	<0.001
Current	242 599 (19.0)	38 455 (14.2)	204 144 (20.3)	<0.001
Ex-smoker	441 499 (34.6)	147 470 (54.3)	294 029 (29.3)	<0.001
Not recorded	76 595 (6.0)	7714 (2.8)	68 881 (6.9)	<0.001
Alcohol use recorded	1 025 722 (80.4)	258 187 (95.0)	767 535 (76.5)	
None	205 708 (16.1)	50 670 (18.7)	155 038 (15.4)	<0.001
Within limits	466 063 (36.5)	93 408 (34.4)	372 655 (37.1)	<0.001
Excess	353 951 (27.8)	114 109 (42.0)	239 842 (23.9)	<0.001
Not recorded	249 452 (19.6)	13 497 (5.0)	235 955 (23.5)	<0.001
BMI recorded	1 100 374 (86.3)	262 715 (96.7)	837 659 (83.5)	
BMI mean (SD)	26.5 (5.9)	41.9 (6.0)	25.8 (5.6)	<0.001
Underweight	42 258 (3.3)	3291 (1.2)	38 967 (3.9)	<0.001
Normal	450 616 (35.3)	64 395 (23.7)	386 221 (38.5)	<0.001
Overweight	360 340 (28.3)	99 941 (36.8)	260 399 (25.9)	<0.001
Obese class 1	159 638 (12.5)	59 506 (21.9)	100 132 (10.0)	<0.001
Obese class 2	57 097 (4.5)	23 328 (8.6)	33 769 (3.4)	<0.001
Obese class 3	31 018 (2.4)	12 525 (4.6)	18 493 (1.8)	<0.001
Not recorded	174 207 (13.7)	8698 (3.2)	165 509 (16.5)	<0.001

*CVD comprised any of coronary artery disease (including myocardial infarction and angina), atrial fibrillation, peripheral arterial disease, stroke and TIA and congestive cardiac failure.
BMI, body mass index; CVD, cardiovascular disease; IMD, Index of Multiple Deprivation; TIA, transient ischaemic attack.

people in the two least deprived quintiles was higher in those with CVD (table 1).

Risk factors for CVD and disease prevalence

The most prevalent risk factor was smoking, followed by obesity, uncontrolled SBP, type 2 diabetes and CKD (table 2). Crude prevalence of smoking (active and ex-smokers) was higher in those of white and mixed ethnicity, and obesity was most prevalent in people of white and of black ethnicity. Crude prevalence of type 2 diabetes was greater in Asian and black ethnicity, and

Table 2 Cardiovascular disease profile and risk factors of the complete adult population and by ethnicity

Conditions	All adults, n (%) (n=1 275 174)	Ethnicity, n (%) White n=828 900	Asian n=78 603	Black n=39 805	Mixed n=14 955	Other n=13 661	Not known (n=299 250)
Cardiovascular risk factors							
Uncontrolled (SBP >140 mm Hg)*	217 978 (17.1)	150 779 (18.2)	5899 (7.5)	5733 (14.4)	1291 (8.6)	992 (7.3)	53 284 (17.8)
CKD (stages 3–5)†	70 965 (5.6)	53 001 (6.4)	2000 (2.5)	1613 (4.1)	350 (2.3)	212 (1.6)	13 789 (4.6)
Type 1 diabetes	5440 (0.4)	4019 (0.5)	198 (0.3)	107 (0.3)	63 (0.4)	32 (0.2)	1021 (0.3)
Type 2 diabetes	84 320 (6.6)	58 586 (7.1)	7443 (9.5)	3605 (9.1)	698 (4.7)	685 (5.0)	13 303 (4.4)
Obesity: BMI ≥30 kg/m^2	247 753 (19.4)	177 255 (21.4)	9180 (11.7)	10 624 (26.7)	2411 (16.1)	2175 (15.9)	46 108 (15.4)
Active smoker	242 599 (19.0)	164 861 (19.9)	10 366 (13.2)	5884 (14.8)	3100 (20.7)	2768 (20.3)	55 620 (18.6)
Ex-smoker	441 499 (34.6)	326 723 (39.4)	16 671 (21.2)	9172 (23.0)	3773 (25.2)	2946 (21.6)	82 214 (27.5)
Cardiovascular diseases							
Coronary artery disease	44 141 (3.5)	33 814 (4.1)	1707 (2.2)	360 (0.9)	155 (1.0)	166 (1.2)	7939 (2.7)
Atrial fibrillation	34 700 (2.7)	26 684 (3.2)	418 (0.5)	226 (0.6)	82 (0.5)	92 (0.7)	7198 (2.4)
Peripheral arterial disease	10 027 (0.8)	7812 (0.9)	188 (0.2)	111 (0.3)	24 (0.2)	22 (0.2)	1870 (0.6)
Stroke and TIA	27 222 (2.1)	20 323 (2.5)	606 (0.8)	430 (1.1)	96 (0.6)	66 (0.5)	5701 (1.9)
Congestive cardiac failure	13 907 (1.1)	10 452 (1.3)	373 (0.5)	226 (0.6)	55 (0.4)	60 (0.4)	2741 (0.9)
Hypertension (diagnosed)	234 838 (18.4)	171 825 (20.7)	8941 (11.4)	7902 (19.9)	1461 (9.8)	1117 (8.2)	43 592 (14.6)

*Includes people with and without a diagnosis of hypertension.
†Number of people with condition (% of total column population), for example, CKD (stages 3–5) in in 53 001 (6.4) people in the white ethnic group.
BMI, body mass index; CKD, chronic kidney disease; SBP, systolic blood pressure; TIA, transient ischaemic attack.

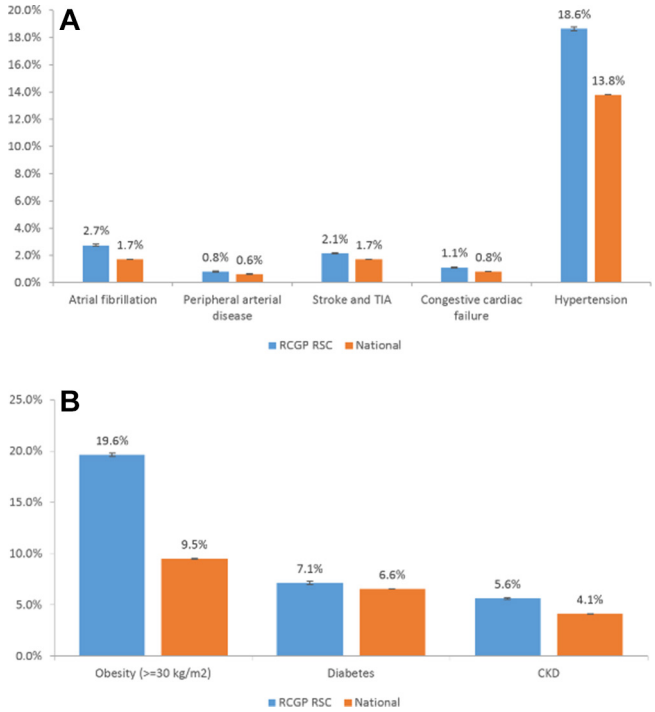

Figure 1 Prevalence of (A) cardiovascular diseases and hypertension, and (B) risk factors/comorbidities in the RCGP RSC network, compared with all-English practices, using Quality Outcomes Framework data. CKD, chronic kidney disease; RCGP, Royal College of General Practitioner; RSC, Research and Surveillance Centre; TIA, transient ischaemic attack.

CKD in white people. The highest overall crude prevalence for CVD was in white ethnicity, particularly for AF.

Prevalence of CVD

The prevalence and number of conditions increased with age (online supplementary table S1). CVD was present in less than 1% of people under the age of 50 years, increasing to a quarter of people with at least one condition between 70 and 79 years old, and over 40% in people over the age of 80 years.

We identified a higher prevalence of CVDs from RCGP RSC data than identified by the P4P(QOF) definitions (figure 1A). We also detected a higher proportion of other CVD risk factors and comorbidities (figure 1B).

Incidence of CVD

Hypertension showed the highest incidence (2012–2016) and age-adjusted and gender-adjusted annual incidence rate followed by AF and coronary artery disease (table 3).

CVD prescription use

ACE inhibitors, statins and calcium channel blockers were the most commonly prescribed, with over half of the CVD cohort prescribed at least one of these medications (online supplementary table S2).

Predictors of CVD

When considering predictors of all types of CVD, people were more likely to have the disease if they were older, male, current or ex-smokers, hazardous drinkers or categorised as an alcoholic, or had a well-established comorbidity/risk factor (chronic kidney disease, diabetes and hypertension) for CVD (table 4). In addition, likelihood of having CVD increased with each BMI category compared with people with a normal BMI, while people in the more deprived groups (IMD quintiles 1–4) were more likely to have CVD than those in the least deprived group (IMD quintile 5), and people of non-white ethnic groups were less likely to have CVD than people of white ethnicity. Interestingly, people with uncontrolled SBP were less likely to have CVD than those with controlled SBP. However, bivariate analysis showed that people were more likely to have CVD if their SBP was uncontrolled (OR 2.09, 95% CI 2.06 to 2.12), which suggests that adjusting for other variables (eg, age and hypertension diagnosis) affected this relationship.

Similar associations were found for coronary artery disease and AF (online supplementary tables S3 and S4); however, people of Asian ethnicity were more likely to have coronary artery disease than people of white ethnicity. Current and ex-smokers and people with type 1 diabetes were less likely to have AF than non-smokers and those without diabetes, respectively.

DISCUSSION

Our results demonstrate a high prevalence of CVD, including hypertension, in a nationally representative population sample. A CVD diagnosis was recorded in

Table 3 Incidence of cardiovascular disease and hypertension in the last 5 years, and age-adjusted and gender-adjusted annual incidence in the adult population

Cardiovascular conditions	Incident cases in last 5 years, n (%)	Annual age-adjusted and gender-adjusted incidence rate per 10 000 (95% CI)
Coronary artery disease	14 033 (1.1)	22.1 (21.8 to 22.5)
Atrial fibrillation	17 913 (1.4)	28.2 (27.8 to 28.7)
Peripheral arterial disease	3817 (0.3)	6.0 (5.8 to 6.2)
Stroke/transient ischaemic attack	10 838 (0.8)	17.1 (16.8 to 17.5)
Congestive cardiac failure	7792 (0.6)	12.3 (12.0 to 12.6)
Hypertension	60 635 (4.8)	96.5 (95.7 to 97.3)

Table 4 Multivariable logistic regression for CVD

Characteristic	OR	95% CI	P values
Age	1.093	1.093 to 1.094	<0.001
Gender (male)	2.105	2.070 to 2.140	<0.001
Ethnicity			
White	1.000 (reference)		
Asian	0.849	0.810 to 0.890	<0.001
Black	0.520	0.486 to 0.557	<0.001
Mixed	0.731	0.648 to 0.825	<0.001
Missing	0.697	0.615 to 0.791	<0.001
Other	0.933	0.915 to 0.952	<0.001
IMD quintile			
IMD quintile 5 (least deprived)	1.000 (reference)		
IMD quintile 4	1.047	1.025 to 1.070	<0.001
IMD quintile 3	1.082	1.057 to 1.107	<0.001
IMD quintile 2	1.155	1.127 to 1.185	<0.001
IMD quintile 1 (most deprived)	1.230	1.199 to 1.263	<0.001
IMD quintile not available	1.034	0.873 to 1.225	0.700
Smoking status			
Never	1.000 (reference)		
Active smoker	1.565	1.526 to 1.605	<0.001
Ex-smoker	1.381	1.355 to 1.407	<0.001
Not recorded	1.056	1.005 to 1.111	0.032
Alcohol consumption			
Safe use	1.000 (reference)		
Non-drinker	1.189	1.162 to 1.216	<0.001
Hazardous use	1.217	1.195 to 1.241	<0.001
Alcoholism	1.594	1.533 to 1.657	<0.001
Not recorded	0.866	0.834 to 0.900	<0.001
BMI category			
Normal	1.000 (reference)		
Underweight	1.151	1.082 to 1.225	<0.001
Overweight	1.136	1.114 to 1.159	<0.001
Obese class 1	1.352	1.320 to 1.384	<0.001
Obese class 2	1.597	1.544 to 1.651	<0.001
Obese class 3	1.858	1.776 to 1.944	<0.001
Not recorded	0.803	0.769 to 0.839	<0.001
Diabetes			
No diabetes	1.000 (reference)		
Type 1 diabetes	2.410	2.165 to 2.683	<0.001
Type 2 diabetes	1.609	1.576 to 1.642	<0.001
Unknown/other diabetes type	2.038	1.407 to 2.951	<0.001
Other known risk factors			
Uncontrolled SBP (>140 mm Hg)	0.712	0.700 to 0.725	<0.001
Hypertension diagnosis	1.734	1.704 to 1.764	<0.001
CKD (stages 3–5)	1.653	1.620 to 1.688	<0.001

BMI, body mass index; CKD, chronic kidney disease; CVD, cardiovascular disease; IMD, Index of Multiple Deprivation; SBP, systolic blood pressure.

more than one in five adults in our population. Regarding specific cardiovascular conditions, hypertension was the most commonly diagnosed cardiovascular condition. Coronary artery disease was diagnosed in 3.5% of individuals. This compares to previous UK studies which have demonstrated a fairly consistent prevalence of coronary artery disease at 3% in England, and 4% in Scotland, Northern Ireland and Wales.[32] Additionally, our reported prevalence of AF, 2.7%, is similar to modelling studies undertaken by Public Health England to estimate the true burden of AF in England,[33] which in itself is significantly higher than the national prevalence of AF reported by QOF/P4P. This highlights the benefit of our approach to case definition and the use of a wider range of clinical codes than used within P4P business rules. Classical risk factors for CVD were well represented within the RCGP RSC population. For example 234 838 individuals had a diagnosis of hypertension.

The RCGP RSC database is suitable for CVD research. We have defined a cohort of people with CVD (n=271 684) and identified a higher prevalence of patients than the P4P/QOF scheme: though we note QOF uses a limited list of codes. Incident cases of CVD and CVD risk factors can be identified in the cohort. We identified more obesity, our results are more closely aligned with self-reported data.[34] Our use of ontologies to improve case finding may make our approach more reproducible than studies that rely on an expert's code list.[35]

Data quality was high in the CVD cohort for demographic, clinical and social variables. These are similar to those reported in other UK primary care databases, such as Clinical Practice Research Datalink (CPRD).[36] We were able to identify uncontrolled hypertension, a risk factor for major cardiovascular events,[37 38] and highlighted an area where management could be improved through a combination of patient education and self-management and pharmacological interventions as appropriate.[39] We could identify appropriate medicines used to manage CVD and found statins prescribed at anticipated rates.[40]

The associations with ethnicity were generally those expected. We found that a higher prevalence of people of white ethnicity had CKD, while the highest prevalence for type 2 diabetes was seen in those of black or Asian

ethnicity.[41 42] Similarly a higher prevalence of hypertension was seen in people of black ethnicity.[43 44]

The age-adjusted and gender-adjusted annual incidence of CVD was highest in AF and similar to other findings that have reported rates in Europe,[45] while the incidence of stroke was higher than previously reported in other UK primary care databases (CPRD).[46] This may reflect an increase in disease in the years since these findings were reported, or enhanced case-finding strategies.

Logistic regression analyses demonstrated that established risk factors for CVD were associated with CVD in the RCGP RSC dataset. Unusually, people with uncontrolled SBP were shown to be less likely to have CVD. However, bivariate analysis confirmed that this finding was only after adjusting for other variables, and uncontrolled SBP is in fact associated with higher risk of CVD. This variable was for the latest recording only for people with or without a diagnosis of hypertension, and therefore, many people who had uncontrolled SBP are healthy.

Strengths and weaknesses

Strengths

We have presented a comprehensive assessment of the burden of CVD in a representative population sample. Our use of coding ontologies improves the accuracy of case identification. RCGP RSC has promoted high-quality data recording for over half a century.

Although RCGP data have traditionally been used for monitoring infectious diseases, the network data have recently started to be used more for non-communicable disease (NCD) research. Examples include the epidemiology of type 3c diabetes,[47] liver disease[48] and risks associated with anticoagulant use in people with AF and CKD.[17]

Weaknesses

The weakness of our data is that of all routinely collected data.[18] Data are collected for routine patient care in the context of the 10 min GP consultation. Our coded data are therefore a brief abstraction of this process.[49] We are still in the early stages of looking to ensure the robustness of real-world evidence studies.

The incidence of myocardial infarction was lower than reported UK primary care data from linked primary care, hospital and national audit datasets (CPRD).[50] This suggests there are limitations to using primary care data alone, though we have the capability to link to these same datasets.

CONCLUSIONS

We have reported the prevalence and incidence of common CVDs in the RCGP RSC dataset. Our reported prevalence of coronary artery disease and AF is comparable with other UK studies. Therefore, the RCGP RSC is ready to be more active in research and quality improvement studies including CVD. The rich clinical data available within the RCGP RSC have substantial potential utility for epidemiological research into a variety of communicable and NCD. This has recently been demonstrated using the RCGP RSC dataset to explore CVD outcomes in renal disease.[17] The combination of clinical and prescription data can be used to carry out real-world evidence studies and explore effectiveness beyond the traditional randomised control trial setting.

Author affiliations
[1]Section of Clinical Medicine & Ageing, Department of Clinical & Experimental Medicine, University of Surrey, Guildford, UK
[2]Institute of Cardiovascular Research, Royal Holloway University of London, Egham, UK
[3]Department of Endocrinology, Ashford and St Peter's NHS Foundation Trust, Chertsey, UK
[4]Royal College of General Practitioners, Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC), London, UK

Acknowledgements Rachel Byford and Julian Sherlock, SQL developers, database management and data extraction, and the participating practices and patients for providing the data for this cohort. Collaboration with the GP computer system suppliers—EMIS, InPractice Systems and TPP; Apollo systems; Public Health England and our other funders and collaborators.

Contributors WH led the analysis and drafting of the paper. AM and AC supported the statistical analysis of the paper. SdL led the development of the cohort, conceived the idea and contributed to the drafting of the paper. RC, TSH, PS and FF also contributed to the preparation and reviewing of the manuscript.

Funding This study was funded internally by the Department of Clinical and Experimental Medicine at the University of Surrey.

Competing interests None declared.

Patient consent Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The RCGP RSC data set can be accessed by researchers, approval is on a project-by-project basis (www.rcgp.org.uk/rsc). Ethical approval by an NHS Research Ethics Committee is needed before any data release/other appropriate approval. Researchers wishing to directly analyse the patient-level pseudonymised data will be required to complete information governance training and work on the data from the secure servers at the University of Surrey. Patient-level data cannot be taken out of the secure network. We encourage interested researchers to attend the short courses on how to analyse primary care data/RCGP RSC data offered twice a year.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

REFERENCES

1. World Health Organization. Cardiovascular diseases (CVDs). http://www.who.int/mediacentre/factsheets/fs317/en/ (accessed 17 May 2017).
2. GBD 2015 Mortality and Causes of Death Collaborators. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* 2016;388:1459–544.
3. Lozano R, Naghavi M, Foreman K, *et al*. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012;380:2095–128.
4. GBD 2015 Risk Factors Collaborators. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* 2016;388:1659–724.
5. Rumsfeld JS, Joynt KE, Maddox TM. Big data analytics to improve cardiovascular care: promise and challenges. *Nat Rev Cardiol* 2016;13:350–9.

6. Kim J. Big data, health informatics, and the future of cardiovascular medicine. *J Am Coll Cardiol* 2017;69:899–902.
7. de Lusignan S, Metsemakers JF, Houwink P, *et al*. Routinely collected general practice data: goldmines for research? A report of the European Federation for Medical Informatics Primary Care Informatics Working Group (EFMI PCIWG) from MIE2006, Maastricht, The Netherlands. *Inform Prim Care* 2006;14:203–9.
8. Mavaddat N, Mant J. Primary care research and clinical practice: cardiovascular disease. *Postgrad Med J* 2010;86:696–703.
9. Langdown C, Peckham S. The use of financial incentives to help improve health outcomes: is the quality and outcomes framework fit for purpose? A systematic review. *J Public Health* 2014;36:251–8.
10. de Lusignan S, van Weel C. The use of routinely collected computer data for research in primary care: opportunities and challenges. *Fam Pract* 2006;23:253–63.
11. Fleming DM. Weekly returns service of the royal college of general practitioners. *Commun Dis Public Health* 1999;2:96–100.
12. Correa A, Hinton W, McGovern A, *et al*. Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC) sentinel network: a cohort profile. *BMJ Open* 2016;6:e011092.
13. Queenan JA, Williamson T, Khan S, *et al*. Representativeness of patients and providers in the canadian primary care sentinel surveillance network: a cross-sectional study. *CMAJ Open* 2016;4:E28–32.
14. Fleming DM, Schellevis FG, Paget WJ. Health monitoring in sentinel practice networks: the contribution of primary care. *Eur J Public Health* 2003;13(3 Suppl):80–4.
15. Fleming DM, Miles J. The representativeness of sentinel practice networks. *J Public Health* 2010;32:90–6.
16. McGovern A, Hinton W, Correa A, *et al*. Real-world evidence studies into treatment adherence, thresholds for intervention and disparities in treatment in people with type 2 diabetes in the UK. *BMJ Open* 2016;6:e012801.
17. Kumar S, de Lusignan S, McGovern A, *et al*. Ischaemic stroke, haemorrhage, and mortality in older patients with chronic kidney disease newly started on anticoagulation for atrial fibrillation: a population based study from UK primary care. *BMJ* 2018;360:k342.
18. de Lusignan S. Codes, classifications, terminologies and nomenclatures: definition, development and application in practice. *Inform Prim Care* 2005;13:65–9.
19. Liaw ST, Rahimi A, Ray P, *et al*. Towards an ontology for data quality in integrated chronic disease management: a realist review of the literature. *Int J Med Inform* 2013;82:10–24.
20. Liaw ST, Pearce C, Liyanage H, *et al*. An integrated organisation-wide data quality management and information governance framework: theoretical underpinnings. *Inform Prim Care* 2014;21:199–206.
21. Pathirannehelage S, Kumarapeli P, Byford R, *et al*. Uptake of a dashboard designed to give realtime feedback to a sentinel network about key data required for influenza vaccine effectiveness studies. *Stud Health Technol Inform* 2018;247:161–5.
22. Smith N, Livina V, Byford R, *et al*. Automated differentiation of incident and prevalent cases in primary care computerised Medical Records (CMR). *Stud Health Technol Inform* 2018;247:151–5.
23. de Lusignan S, Navarro R, Chan T, *et al*. Detecting referral and selection bias by the anonymous linkage of practice, hospital and clinic data using Secure and Private Record Linkage (SAPREL): case study from the evaluation of the Improved Access to Psychological Therapy (IAPT) service. *BMC Med Inform Decis Mak* 2011;11:61.
24. Fleming DM, Durnall H. Ten lessons for the next influenza pandemic-an English perspective: a personal reflection based on community surveillance data. *Hum Vaccin Immunother* 2012;8:138–45.
25. Tippu Z, Correa A, Liyanage H, *et al*. Ethnicity recording in primary care computerised medical record systems: an ontological approach. *J Innov Health Inform* 2017;23:799.
26. Gov.UK. Department for communities and local government the english indices of deprivation. 2015 https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015 (accessed 17 May 2017).
27. World Health Organization. Physical status: the use and interpretation of anthropometry. Report of a WHO Expert Committee, WHO Technical Report Series 85. 1995 http://apps.who.int/iris/bitstream/10665/37003/1/WHO_TRS_854.pdf (accessed 17 May 2017).
28. McGovern A, Feher M, Munro N, *et al*. Sodium-Glucose Co-Transporter-2 (SGLT2) inhibitors: comparing trial and real world use (Study Protocol). *Diabetes Ther* 2017;8:355–63.
29. Hippisley-Cox J, Coupland C, Vinogradova Y, *et al*. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. *BMJ* 2008;336:1475–82.
30. Health Research Authority. Defining research table - October 2017. 2017 http://www.hra-decisiontools.org.uk/research/docs/DefiningResearchTable_Oct2017-1.pdf (accessed 6th Feb 2018).
31. Royal College of General Practitioners (RCGP). Research and Surveillance Centre (RSC). www.rcgp.org.uk/rsc (accessed 16 Jun 2018).
32. Bhatnagar P, Wickramasinghe K, Wilkins E, *et al*. Trends in the epidemiology of cardiovascular disease in the UK. *Heart* 2016;102:1945–52.
33. Public Health England. Atrial Fibrillation prevalence estimates in England: application of recent estimates of AF in Sweden. 2015 https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/644869/atrial_fibrillation_AF_briefing.pdf (accessed 22 Feb 2018).
34. NHS Digital. Health survey for England. 2015 http://www.content.digital.nhs.uk/catalogue/PUB22610 (accessed 3 Aug 2017).
35. Liyanage H, Liaw ST, Kuziemsky C, *et al*. Ontologies to improve chronic disease management research and quality improvement studies - a conceptual framework. *Stud Health Technol Inform* 2013;192:180–4.
36. Herrett E, Gallagher AM, Bhaskaran K, *et al*. Data resource profile: clinical practice research datalink (CPRD). *Int J Epidemiol* 2015;44:827–36.
37. Lewington S, Clarke R, Qizilbash N, *et al*. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet* 2002;360:1903–13.
38. Turnbull F, Neal B, Ninomiya T, *et al*. Effects of different regimens to lower blood pressure on major cardiovascular events in older and younger adults: meta-analysis of randomised trials. *BMJ* 2008;336:1121–3.
39. NICE. *Hypertension in adults: diagnosis and management. CG127*. London, 2016.
40. Murphy C, Bennett K, Fahey T, *et al*. Statin use in adults at high risk of cardiovascular disease mortality: cross-sectional analysis of baseline data from The Irish Longitudinal Study on Ageing (TILDA). *BMJ Open* 2015;5:e008017.
41. Dreyer G, Hull S, Aitken Z, *et al*. The effect of ethnicity on the prevalence of diabetes and associated chronic kidney disease. *QJM* 2009;102:261–9.
42. Riste L, Khan F, Cruickshank K. High prevalence of type 2 diabetes in all ethnic groups, including Europeans, in a British inner city: relative poverty, history, inactivity, or 21st century Europe? *Diabetes Care* 2001;24:1377–83.
43. Cappuccio FP, Cook DG, Atkinson RW, *et al*. Prevalence, detection, and management of cardiovascular risk factors in different ethnic groups in south London. *Heart* 1997;78:555–63.
44. Primatesta P, Bost L, Poulter NR. Blood pressure levels and hypertension status among ethnic groups in England. *J Hum Hypertens* 2000;14:143–8.
45. Zoni-Berisso M, Lercari F, Carazza T, *et al*. Epidemiology of atrial fibrillation: European perspective. *Clin Epidemiol* 2014;6:213–20.
46. Lee S, Shafe AC, Cowie MR. UK stroke incidence, mortality and cardiovascular risk management 1999-2008: time-trend analysis from the General Practice Research Database. *BMJ Open* 2011;1:e000269.
47. Woodmansey C, McGovern AP, McCullough KA, *et al*. Incidence, demographics, and clinical characteristics of diabetes of the exocrine pancreas (Type 3c): a retrospective cohort study. *Diabetes Care* 2017;40:1486–93.
48. Williams R, Alexander G, Armstrong I, *et al*. Disease burden and costs from excess alcohol consumption, obesity, and viral hepatitis: fourth report of the Lancet Standing Commission on Liver Disease in the UK. *Lancet* 2018;391:1097–107.
49. Kumarapeli P, de Lusignan S. Using the computer in the clinical consultation; setting the stage, reviewing, recording, and taking actions: multi-channel video study. *J Am Med Inform Assoc* 2013;20:e67–e75.
50. Herrett E, Shah AD, Boggon R, *et al*. Completeness and diagnostic validity of recording acute myocardial infarction events in primary care, hospital care, disease registry, and national mortality records: cohort study. *BMJ* 2013;346:f2350.