Acute respiratory infections, cardiovascular complications, and prevention among people with raised cardiovascular risk

JENNIFER DAVIDSON

Thesis submitted in accordance with the requirements for the degree of Doctor of Philosophy of the University of London

JULY 2022

Department of Non-Communicable Disease Epidemiology

Faculty of Epidemiology and Population Health

LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

Funded by British Heart Foundation (grant FS/18/71/33938)

Research group affiliation(s): Electronic Health Records Research Group

Declaration

I, Jennifer Davidson, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, this has been indicated in the thesis. I have read and understood the School's definition of plagiarism and cheating given in the Research Degrees Handbook.

Signed:

Date: 29 July 2022

Acknowledgements

I am extremely grateful to my PhD supervisors, Charlotte Warren-Gash, Amitava Banerjee, and Liam Smeeth who have provided me constant support, encouragement, and helpful research insight over the past three and a half years.

I am also appreciative to my advisory panel, Emily Herrett, Helen McDonald, and Richard Pebody, who have also helped guide me through this journey. I would also like to extend my gratitude to the additional researchers who co-authored my research; Rutendo Muzambi, Harriet Forbes, Daniel Grint, Ian Douglas, Clémence Leyrat, Helen Strongman, and Judith Breuer. Helen Strongman and Emily Herrett supported me with adapting and enhancing the LSHTM QRISK algorithm programs I used. Rutendo Muzambi was the second reviewer in my systematic review study selection. Daniel Grint and Clémence Leyrat provided statistical guidance. Ian Douglas provided study methodology support.

Further specific thanks go to Helena Carreira for being a great office mate in the early days of my PhD journey, as well as for extracting all the CPRD data for my research. I would also thank the wonderful LSHTM Electronic Health Record research group who I have had the pleasure to learn from during my PhD, and for being an all-around lovely bunch of people. Also, to my fellow PhD students, especially Rutendo, Liang-Yu, and Jeremy, without who I would not have made it to the end.

I wish to thank the British Heart Foundation for funding this PhD and for their ongoing interest in my research.

I am so indebted to my friends and family who have supported me throughout this PhD. To Paige and Frances, the simple act of checking on how it was progressing and constantly telling me it was almost done. To my sister, Karen, the endless number of phone conversations over the last several years where you listened to me describe the latest happenings of my research and shared my excitement at each milestone and achievement, have meant so much. Huge thanks to my husband, Jorge, who supported me to give up my job in 2018 to pursue this PhD, you have been a consistent source of encouragement ever since.

Lastly, I dedicate this thesis to my mum, Margaret. I am so grateful that you were able to see me start this journey and I cherish the confidence you had that I would get to the end. I appreciate all that you did to build me into the person I am.

Abstract

Background: Although acute respiratory infections (ARIs) can lead to cardiovascular complications, the effect of underlying cardiovascular risk on the incidence of ARIs and ARI-related cardiovascular complications in people without established cardiovascular disease (CVD) is unknown. In turn, the benefit of vaccines, such as influenza vaccine, among people with raised cardiovascular risk is unmeasured.

Objectives and data sources: The objectives of this thesis were to 1) assess the validity of acute cardiovascular event diagnoses in electronic health record (EHR) data, 2) examine the association of cardiovascular risk with ARIs and ARI-related cardiovascular complications, 3) investigate the association between influenza vaccine and acute cardiovascular events by varying cardiovascular risk level, 4) examine the association of cardiovascular risk with severe COVID-9 outcomes, and 5) investigate the association of COVID-19 and acute cardiovascular events by varying cardiovascular risk level. All analysis to achieve objectives 2-5 used EHR data from the Clinical Practice Research Datalink (CPRD) and Hospital Episode Statistics (HES).

Results: Results from 81 validation studies included in the systematic review suggested EHR recorded acute cardiovascular event diagnoses have a high level of validity, but variable definitions are employed (Chapter 5). Using CPRD and HES data from 6,075,321 individuals aged 40-64 years who are not currently recommended to receive influenza vaccine, I found an increased incidence of ARI among individuals at raised cardiovascular risk (Chapter 6). I also identified a significant association between raised cardiovascular risk and ARI-related cardiovascular complication, which was higher for QRISK2 score (adjusted hazard ratio (aHR) 3.65, 95% confidence interval (CI) 3.42-3.89) than hypertension (aHR 1.98, 95% CI 1.83-2.15). Among 193,900 individuals aged 40-84 years I found a decrease in the season-adjusted relative incidence of first acute cardiovascular events occurring in the days and weeks after influenza vaccination with a tapering over time (Chapter 7). In analysis of raised cardiovascular risk and COVID-19 among 6,059,055 adults aged 40-84 years, I found elevated risk of COVID-19 death, first acute cardiovascular event, and other severe COVID-19 outcomes in those with a QRISK3 score ≥10% (Chapter 8). Hypertension was only associated with

risk of acute cardiovascular event. In a self-controlled case series analysis of 1,762 individuals with COVID-19 (Chapter 9), I identified an increased risk of first acute cardiovascular events which was greatest in the first seven days after infection (incidence ratio 7.14, 95% CI 6.06-8.41).

Conclusions: People with raised cardiovascular risk are at higher risk of ARI-related cardiovascular complications following infection, including influenza, pneumonia, and COVID-19. Raised cardiovascular risk was more strongly associated with ARI-related cardiovascular complications when cardiovascular risk was measured by QRISK2/3 score compared to hypertension alone. Addressing cardiovascular risk factors could improve outcomes after ARIs. Improved vaccine uptake could contribute to prevention of cardiovascular disease.

Contents

Abstract 5
Contents 7
List of tables 11
List of figures 13
List of abbreviations and acronyms 14
Chapter 1 Background 15
1.1. Chapter overview 15
1.2. Synopsis – published editorial 17
1.3. Cardiovascular disease 23
1.3.1. Definition and diagnosis 23
1.3.2. Incidence and prevalence 31
1.3.3. Risk factors 32
1.3.4. Prevention strategies 33
1.4. Systemic acute respiratory infections 34
1.4.1. Definitions and diagnosis 34
1.4.2. Incidence 37
1.4.3. Risk factors 37
1.4.4. Prevention and treatment 38
1.5. Cardiovascular complications of systemic acute respiratory infections 40
1.5.1. Epidemiology 40
1.5.2. Mechanistic links 42
1.5.3. Use of vaccines to prevent cardiovascular events 43
1.6. Chapter summary 46
Chapter 2 Aim and objectives 47
2.1. Thesis rationale 47
2.2. Thesis scope 47
2.2.1. Original scope 47
2.2.2. Evolution of thesis scope due to COVID-19 48
2.3. Overarching aim 48
2.4. Individual objectives 49
2.5. Organisation of thesis 49
Chapter 3 Description of data sources 51
3.1. Chapter overview 51
3.2. Use of electronic health records in observational research 51
3.3. Clinical Practice Research Datalink 53
3.3.1. Data collection 53

3.3.2. Data structure 55
3.3.3. Internal data quality 58
3.4. Hospital Episode Statistics Admitted Patient Care 58
3.4.1. Data collection 58
3.4.2. Data structure 59
3.4.3. Internal data quality 59
3.5. The Office for National Statistics mortality data 60
3.5.1. Data collection 61
3.5.2. Internal data quality 61
3.6. Second Generation Surveillance System COVID-19 positive virology results 61
3.7. COVID-19 Hospitalisations in England Surveillance System 62
3.8. Strengths and limitations of included data sources 63
3.8.1. Data quality considerations 63
3.9. Chapter summary 66
Chapter 4 Definitions 67
4.1. Principles of variable identification in electronic health record databases 67
4.2. Identifying acute respiratory infections 70
4.2.1. Validity of recording in electronic health records 70
4.2.2. Definition for thesis 70
4.3. Identifying influenza vaccination 71
4.4. Identifying COVID-19 73
4.5. Identifying established cardiovascular disease 74
4.6. Measuring cardiovascular risk 74
4.6.1. Hypertension 74
4.6.2. QRISK 75
4.7. Potential explanatory variables 78
4.7.1. Age and sex 78
4.7.2. Ethnicity 80
4.7.3. Socio-economic status 80
4.7.4. Comorbid health conditions 81
4.8. Chapter summary 82
Chapter 5 Review on the validity of acute cardiovascular outcome diagnosis 83
5.1. Chapter overview 83
5.2. Published protocol 84
5.3. Published paper 92
5.4. Deriving thesis definitions 111
5.5. Chapter summary 113

Chapter 6 Quantifying the effect of cardiovascular risk on the rates of acute respiratory infections and subsequent cardiovascular complications 114
6.1. Chapter overview 114
6.2. Published paper 115
6.3. Further methodological decisions 128
6.3.1. Detailed explanation of inclusion and exclusion criteria 128
6.3.2. Covariate selection: acute respiratory infection outcome 131
6.3.3. Covariate selection: acute respiratory infection-related cardiovascular event outcome 132
6.4. The effect of raised cardiovascular risk on cardiovascular complications independent of acute respiratory infection 133
6.4.1. Background 133
6.4.2. Methods 134
6.4.3. Results 135
6.5. Chapter summary 145
Chapter 7 The effectiveness of influenza vaccine in preventing cardiovascular events in people with raised cardiovascular risk 147
7.1. Chapter overview 147
7.2. Study design considerations 147
7.3. Submitted paper 155
7.3. Additional methods 183
7.4. Additional results 183
7.5. Chapter summary 192
Chapter 8 Investigating the effect of cardiovascular risk on the severe COVID-19 outcomes 194
8.1. Chapter overview 194
8.2. Drafted manuscript 195
8.3. Additional methods 218
8.4. Additional results 218
8.5. Chapter summary 224
Chapter 9 Investigating the association between COVID-19 and cardiovascular complications by varying cardiovascular risk 226
9.1. Chapter overview 226
9.2. Drafted manuscript 227
9.3. Additional methods 252
9.4. Additional results 252
9.5. Chapter summary 255
Chapter 10 Discussion 257
10.1. Chapter overview 257
10.2. Summary of key research findings 257

10.3. Explanation of research findings with comparison to previous literature 261
10.4. Research strengths and limitations 268
10.5. Clinical and public health implications 274
10.6. Further research directions 275
10.7. Conclusions 278
References 279
Appendices 303
Appendix 1 Example do file to create code lists 303
Appendix 2 Chapter 5 supplementary materials 304
Appendix 3 Chapter 6 supplementary material 392
Appendix 4 Chapter 7 protocol and ethics approval 425
Appendix 5 Chapter 7 supplementary material 445
Appendix 6 Chapter 8 and 9 protocol and ethics approval 461
Appendix 7 Chapter 8 supplementary material 481

List of tables

Table 1.1 Major definitions to categorise acute cardiovascular events of myocardial infarction, unstable angina, heart failure and stroke 27
Table 2.1 Thesis objectives 50
Table 3.1 Key characteristics of datasets used in this thesis 52
Table 3.2 Data files from Clinical Practice Research Datalink (CPRD) Gold and Aurum 56
Table 3.3 Data files from the Hospital Episode Statistics Admitted Patient Care 60
Table 3.4 General strengths and limitations of the data sources used in thesis research 65
Table 4.1 CPRD data used to identify influenza vaccine administered 72
Table 4.2 Process to handling status combinations on the same date 73
Table 4.3 QRISK variable definitions and source notes 78
Table 6.1 Influenza and pneumococcal vaccine indications applied to main and sensitivity analyses study populations 130
Table 6.2 Association between cardiovascular risk and acute cardiovascular event 138
Table 6.3 Association between cardiovascular risk and acute cardiovascular event by season 139
Table 6.4 Database comparison of acute cardiovascular events incidence rates and incidence rate ratios 140
Table 6.5 Association between cardiovascular risk and acute cardiovascular event in the sensitivity analysis study population* 142
Table 6.6 Association between QRISK2 score and acute cardiovascular events, by QRISK2 score identification method* 143
Table 6.7 Association between cardiovascular risk and major adverse cardiovascular event* 143
Table 6.8 Association between cardiovascular risk and acute cardiovascular event by redefined season 144
Table 7.1 Possible study design advantages and limitations 149
Table 7.2 Baseline characteristics of study population 180
Table 7.3 Incidence ratios for events in risk periods following influenza vaccination by cardiovascular risk and age 182
Table 7.4 Incidence ratios for first acute cardiovascular event in risk periods following influenza vaccination by cardiovascular risk, age group and prior acute respiratory infection 185
Table 7.5 Incidence ratios for first acute cardiovascular event in risk periods following influenza vaccination by cardiovascular risk, age group vaccine match 187
Table 7.6 Incidence ratios for first acute cardiovascular event in risk periods following influenza vaccination by cardiovascular risk with redefined season 189

Table 7.7 Incidence ratios for first acute cardiovascular event in risk periods following influenza vaccination by cardiovascular risk and age group with baseline split into pre- and post-vaccination time 190

List of figures

Figure 1.1 Hierarchy of cardiovascular disease conditions and classifications 23
Figure 1.2 Endothelial dysfunction and atherosclerosis progression 24
Figure 1.3 Overview of potential mechanistic links between acute respiratory infections and cardiovascular complications 43
Figure 4.1 Process to create codelists 69
Figure 6.1 Directed acyclic graphic depicting confounders in the association between hypertension and acute respiratory infections 131
Figure 6.2 Directed acyclic graphic depicting confounders and effect modifiers in the association between hypertension and acute cardiovascular events after acute respiratory infection 133
Figure 6.3 Directed acyclic graphic depicting confounders and mediators in the association between hypertension and major adverse cardiovascular events 134
Figure 6.4 Weekly number and rate of acute cardiovascular events, September 2008–August 2018 137
Figure 7.1 Time-updating exposure study design compared to clone and censor study design 153
Figure 7.2 Overview of study design 177
Figure 7.3 Study population flow chart 178
Figure 7.4 Incidence ratios for first acute cardiovascular events in risk periods following influenza vaccination by cardiovascular event type 179
Figure 8.1 Study design overview with 2020 England COVID-19 timeline 215
Figure 8.2 Study population flow chart 216
Figure 8.3 Fully-adjusted hazard ratios for raised cardiovascular risk effect on COVID-19 severe outcomes. QRISK3 score hazard ratios are for the effect of a score ≥10% with <10% as the reference. Hypertension hazard ratios are for the effect of having hypertension with not having hypertension as the reference. 217
Figure 9.1 Study design with exposure and baseline periods 245
Figure 9.2 Study population flow chart 246
Figure 9.3 Incidence ratios for first acute cardiovascular events in risk periods following SARS-CoV-2 by cardiovascular event type 247

List of abbreviations and acronyms

ACS	acute coronary syndrome
ABPM	ambulatory blood pressure monitoring
ACE2	angiotensin-converting enzyme 2
ARI	acute respiratory infection
BMI	body mass index
BNF	British National Formulary
BP	blood pressure
bpm	beats per minute
CHESS	COVID-19 Hospitalisations In England Surveillance System
CI	confidence interval
CKD	chronic kidney disease
COVID-19	coronavirus disease 2019
CPRD	Clinical Practice Research Datalink
CTV3	Clinical Terms Version 3
CVD	cardiovascular disease
DAG	directed acyclic graph
DALYs	disability-adjusted life years
dm+d	Dictionary Of Medicines And Devices
ECDC	European Centre For Disease Prevention And Control
ECG	electrocardiogram
eGFR	estimated glomerular filtration rate
EHR	electronic health record
EMIS	Egton Medical Information Systems
GP	general practioner
GPRD	General Practice Research Database
HBPM	home blood pressure monitoring
HDL	high-density lipoprotein cholesterol
HDU	high-dependency unit
HES	Hospital Episode Statistics
HES APC	Hospital Episode Statistics Admitted Patient Care
HF	heart failure
HF-pEF	heart failure with preserved left ventricular function
HF-rEF	heart failure with reduced left ventricular function
HR	hazard ratio
ICD-10	International Classification Of Disease 10th Revision
ICH	intracerebral haemorrhage
ICU	intensive care unit
IHD	ischaemic heart disease
ILI	influenza-like illness
IR	incidence ratio
IRR	incidence rate ratio
LDL	low-density lipoprotein
LRTI	lower respiratory tract infection
MACE	major adverse cardiovascular events
MCCD	medical certificate of cause of death
MHRA	Medicines And Healthcare Products Regulatory Agency
MI	myocardial infarction
NHS	National Health Service
NICE	National Institute For Health And Care Excellence
NIHR	National Institute For Health Research
NPV	negative predictive value
NSTEMI	non-ST-elevation myocardial infarction

ONS	Office For National Statistics
OR	odds ratio
PCR	polymerase chain reaction
PHE	Public Health England
PPV	positive predictive value
QOF	Quality And Outcomes Framework
RCT	randomised controlled trial
RSV	respiratory syncytial virus
SAH	subarachnoid haemorrhage
SARS	severe acute respiratory syndrome
SARS-CoV-2	severe acute respiratory syndrome coronavirus 2
SCCS	Self-Controlled Case Series
SES	socioeconomic status
SGSS	Second-Generation Surveillance System
SMI	severe mental illness
SNOMED CT	Systematized Nomenclature Of Medicine Clinical Terms
STEMI	ST-elevation myocardial infarction
SUS	Secondary Use Service
TC:HDL	total cholesterol to high-density lipoprotein cholesterol ratio
THIN	The Health Improvement Network
TIA	transient ischaemic attack
UI	Uncertainty Interval
UK	United Kingdom
USA	United States of America
UTS	up-to-standard
VAMP	Value-Added Information Medical Products Ltd
WHO	World Health Organization
YLD	years lived with disability
YLL	years of life lost

Chapter 1 Background

1.1. Chapter overview

This introductory chapter presents a synopsis on the cardiovascular complications of systemic acute respiratory infection (ARI) in the form of a published editorial invited by the Expert Review of Anti-infective Therapy.

Following the editorial, I set out a more detailed background on each topic component, with an overview of cardiovascular disease (CVD) and systemic ARI definition, epidemiology, and prevention. I discuss ARI in the context of the pre-coronavirus disease 2019 (COVID-19) pandemic – the topic initially planned for this thesis – after which I expand to consider COVID-19. I then

summarise the current evidence on the role ARIs play in triggering acute cardiovascular events and the use of vaccines to prevent such events.

1.2. Synopsis – published editorial

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	230019	**Title**	Ms
First Name(s)	Jennifer		
Surname/Family Name	Davidson		
Thesis Title	Acute respiratory infections, cardiovascular complications, and prevention among people with raised cardiovascular risk		
Primary Supervisor	Dr Charlotte Warren-Gash		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	Expert Review of Anti-infective Therapy		
When was the work published?	November 2019		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	Not applicable		
Have you retained the copyright for the work?*	**Yes**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I am the first author of this editorial, which my primary supervisor Charlotte Warren-Gash was invited to contribute to the journal. With supervision from Dr Warren-Gash I planned the content of the editorial and completed the initial draft and revisions based on Dr Warren-Gash's feedback.

SECTION E

Student Signature	[redacted]
Date	25/07/2022

Supervisor Signature	[redacted]
Date	26-07-2022

EXPERT REVIEW OF ANTI-INFECTIVE THERAPY
2019, VOL. 17, NO. 12, 939–942
https://doi.org/10.1080/14787210.2019.1689817

Taylor & Francis
Taylor & Francis Group

EDITORIAL

Cardiovascular complications of acute respiratory infections: current research and future directions

Jennifer A. Davidson and Charlotte Warren-Gash

Department of Non-communicable Diseases Epidemiology, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK

ARTICLE HISTORY Received 30 August 2019; Accepted 4 November 2019

KEYWORDS Acute respiratory infections; antivirals; atherosclerosis; cardiovascular disease; inflammation; influenza vaccination; thrombosis

1. Introduction

Ischemic heart disease (IHD) and stroke have been the world's leading causes of death for over 15 years [1]. Prominent 'traditional' risk factors for cardiovascular disease (CVD) such as hypertension, smoking, diabetes, and obesity have been the focus of extensive epidemiological research and subsequent public health action [2]. Additional research has focused on a more diverse range of acute triggers including; emotional stress, physical exertion, air pollution and acute infections [3]. Here we consider the effect of acute respiratory infections (ARIs).

Population-level studies in a range of geographical settings show that patterns of CVD mortality and acute cardiovascular events mirror the seasonality of some ARIs, and persists after controlling for long-term incidence trends, seasonality and environmental conditions [4,5]. While the findings from these population-level studies suggest a link between ARIs and cardiovascular complications, stronger evidence comes from individual-level observational studies.

2. Observational research studies of the association between ARIs and acute cardiovascular events

Observational research using large electronic health record databases affords large sample sizes, and their longitudinal nature allows for a self-controlled case series (SCCS) study design. In SCCS studies, cases act as their own controls, accounting for fixed confounders, during periods of non-exposure – this is particularly useful when investigating associations between transient exposures and acute events. Results from SCCS studies have generally estimated a two to six-fold increase in the risk of cardiovascular complications – particularly myocardial infarction (MI) and stroke – following an ARI [6–8]. Meta-analysis, limited to case–control studies, identified ARIs were more likely to have occurred among MI patients, with a pooled odds ratio (OR) of 2.01, 95% confidence interval (CI) 1.47–2.76 [9].

SCSS studies have split follow-up time, from ARI symptom onset or diagnosis to cardiovascular event, by days which thereby identified the length of risk. The results suggest risk lasts for up to 1 month following infection, with the highest risk in the first week after infection [6–8].

Risk is dependent on several factors. The severity of infection will impact the likelihood of cardiovascular complications; an international multicenter cohort study in patients with community acquired pneumonia (CAP) identified composite acute cardiovascular events were more likely following severe infection (OR 1.74, 95% CI 1.15–2.63), with a particularly high OR seen for MI (OR 4.33, 95% CI 1.55–12.1) [10]. Case-control and SCCS studies have largely used broad clinical definitions of ARI. Studies which use a microbiological definition identified an increased risk of cardiovascular complications after influenza virus and *Streptococcus pneumoniae* infection [7,8]. Host health will additionally modify risk; the majority of studies have been conducted among older adults with preexisting diagnosed CVD [6,8,9]. ARIs can trigger acute cardiovascular events in people without known CVD [6], although it is likely that at least some of these events occur in people with undiagnosed atherosclerosis.

3. Potential mechanisms

Systemic inflammatory processes triggered by ARIs include the release of pro-inflammatory cytokines which are key mediators of atherosclerosis and may directly impact plaque rupture through local inflammation [11]. Furthermore, infections induce pro-coagulant and hemodynamic effects which predispose to ischemia and thrombosis [11,12]. While animal models of severe pneumonia show that *S. pneumoniae* can invade the myocardium leading to cardiac injury and scarring [13], there have been few echocardiographic studies during severe pneumonia in humans. Autopsy studies suggest that myocardial injury, which is relatively rare in uncomplicated infections [14], may occur more often in fatal influenza [15].

4. The role of influenza and pneumococcal vaccinations in reducing cardiovascular risk

Both randomized controlled trials (RCTs) and observational studies demonstrate some CVD benefits of influenza vaccine. A meta-

CONTACT Jennifer A. Davidson Jennifer.Davidson@lshtm.ac.uk London School of Hygiene and Tropical Medicine, Room LG21, Keppel Street, London WC1E 7HT, UK

analysis of five small RCTs identified influenza vaccine reduced the risk of cardiovascular events within 1 year of follow-up (relative risk (RR) 0.64, 95% CI 0.49–0.84) [16]. A subgroup meta-analysis of three RCTs in patients with IHD found risk reduction predominantly in people with recent acute coronary syndrome compared to stable IHD [16]. Recent observational research studies have demonstrated similar findings; in an SCCS study among heart failure patients, influenza vaccination was associated with a lower risk of hospitalization due to CVD (incidence rate ratio 0.73, 95% CI 0.71–-0.76) [17].

The effects of pneumococcal vaccination on cardiovascular outcomes are less clear. Meta-analysis investigating the effect of pneumococcal vaccination on cardiovascular outcomes identified no RCTs, while results from observational studies were mixed [18]. In people aged ≥65 years, vaccination was associated with a lower risk of MI (RR 0.90, 95% CI 0.82–1.00), but no reduction in risk was identified among patients of all ages with high cardiovascular risk [18].

In most high-income countries universal influenza and pneumococcal vaccination is recommended for adults aged ≥60–65 years. Adults under these ages are recommended to receive vaccination when they fall into a clinical risk group who are more likely to experience complications following infection such as those with IHD [19].

5. The effect of antiviral drugs on cardiovascular events

Only a small number of studies have investigated cardiovascular complications following antiviral use. A meta-analysis of six RCTs on the effect of oseltamivir versus placebo among influenza infected adults found that antiviral use was associated with a decrease in adverse cardiac events (RR 0.49, 95% CI 0.25–0.97) [20]. Conversely, results of one RCT suggested that, compared to placebo, oseltamivir may increase QTc prolongations during treatment periods (risk difference 4.0%, 95% CI 0.71–7.30) [20]. Another meta-analysis of 11 RCTs investigating the effect of zanamivir use in influenza-infected adults found no difference (RR 0.98, 95% CI 0.50–1.91) in adverse events affecting the 'cardiovascular body system' [20].

The association between oseltamivir and cardiovascular complications also been explored in observational research. A historical cohort study among adult US military health beneficiaries with CVD identified a reduction in the incidence of recurrent cardiovascular events within 30 days among the oseltamivir treatment group (OR 0.42, 0.35–0.50) [21]. In another historical cohort study, oseltamivir use resulted in a 28% (HR 0.72, 95% CI 0.62–0.82) reduction in stroke and transient ischemic attack risk during the subsequent 6 months [22].

Antiviral drugs are recommended for influenza prophylaxis and treatment. Prophylactic use may occur if antigenic mismatch between seasonal vaccine strains and circulating strains occurs, or as post-exposure prophylaxis, particularly to control outbreaks in residential care communities. Antiviral treatment within 48 h of symptom onset during influenza seasons in certain clinical risk groups (i.e. IHD) and immunosuppressed adults aged ≥65 years is also recommended [19].

6. Future research directions

More research focused on who is at risk of cardiovascular complications triggered by ARIs is needed. Predicting future CVD, particularly in people with different combinations of co-morbid conditions, will assist in providing targeted personalized interventions. Any expansion to current vaccine recommendations resulting from new patient groups being identified as high risk for cardiovascular complications following ARI will require effectiveness and cost-effectiveness studies.

In addition, vaccine uptake among those currently recommended to receive it is suboptimal: across Europe, influenza vaccination coverage remains well below the 75% target level among at-risk groups [19]. Identifying the optimum vaccine target groups whom vaccination campaigns should be aimed at, as well as understanding health-care barriers and facilitators to uptake should be prioritized. The timing and dose of vaccines are also important research considerations. One SCCS study identified a significant reduction in the incidence of MI within 60 days of vaccination, particularly in the first 2 weeks [23]. Some RCTs have investigated the impact of high-dose influenza vaccine; compared to standard dose vaccine, use of high-dose vaccine in people with high cardiovascular risk resulted in a greater reduction of cardiovascular complications [16]. A large RCT with nearly 10,000 CVD patients is currently underway to determine whether high-dose influenza vaccine will reduce all-cause mortality and cardiovascular hospitalizations [24].

While prevention of ARIs themselves is likely to result in the greatest clinical and public health benefit, treatment during the acute phase of infection could also prevent cardiovascular complications. A better understanding of the cardiovascular effects of antivirals is needed, e.g. from RCTs with the primary outcome of cardiovascular complications. Some observational studies have investigated whether other drugs including statins, corticosteroids and antiplatelet agents may potentially reduce risk of CVD events during acute infections [25–27]. Further research is needed on the effectiveness, timing and target populations for any such treatments. Observational research suggests cardiac biomarkers provide one method to identify patients with ARI who are at high risk of cardiovascular complications [28]. This cohort study identified hospitalized CAP patients who had blood samples taken at several time points during the first 30 days of admission and showed multiple biomarkers were higher among patients who had a cardiovascular event. People with high overall vascular risk score, such as Framingham or QRISK, are another group who could be targeted for early intervention.

7. Conclusions

Understanding and addressing interactions between diseases, such as ARIs and CVD, is increasingly important as the global population ages and multimorbidity increases. Among ARIs, much focus has been given to influenza due to its severity but also because it is one of the only respiratory infections for which there is effective prevention and treatment. This focus is supported by findings from observational studies which confirm an associa-

tion between laboratory-confirmed ARIs and cardiovascular complications. However, where infections other than preventable and treatable influenza or *S. pneumoniae* result in cardiovascular complications, other approaches to avert these outcomes are required.

Funding

J A Davidson is funded by a British Heart Foundation Non-Clinical PhD Studentship (FS/18/71/33938) and C Warren-Gash is funded by a Wellcome Intermediate Clinical Fellowship (201440/Z/16/Z).

Declaration of interest

C Warren-Gash reports receiving speaker fees from Sanofi Pasteur. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

Reviewer disclosures

Peer reviewers on this manuscript have no relevant financial or other relationships to disclose.

ORCID

Jennifer A. Davidson http://orcid.org/0000-0003-1599-7504

References

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers.

1. World Health Organization. The top 10 causes of death. [Internet]. [cited 2019 Feb 11]. Available from: https://www.who.int/en/news-room/fact-sheets/detail/the-top-10-causes-of-death
2. Mahmood SS, Levy D, Vasan RS, et al. The Framingham heart study and the epidemiology of cardiovascular disease: a historical perspective. Lancet. 2014;383(9921):999–1008.
3. Mittleman MA, Mostofsky E. Physical, psychological and chemical triggers of acute cardiovascular events: preventive strategies. Circulation. 2011;124(3):346–354.
4. Blackburn R, Zhao H, Pebody R, et al. Laboratory-confirmed respiratory infections as predictors of hospital admission for myocardial Infarction and stroke: time-series analysis of English data for 2004–2015. Clin Infect Dis. 2018;67(1):8–17.
5. Imai C, Barnett A, Hashizume M, et al. The role of influenza in the delay between low temperature and ischemic heart disease: evidence from simulation and mortality data from Japan. Int J Environ Res Public Health. 2016;13(5):454.
6. Smeeth L, Thomas SL, Hall AJ, et al. Risk of myocardial infarction and stroke after acute infection or vaccination. N Engl J Med. 2004;351(25):2611–2618.
7. Warren-Gash C, Blackburn R, Whitaker H, et al. Laboratory-confirmed respiratory infections as triggers for acute myocardial infarction and stroke: a self-controlled case series analysis of national linked datasets from Scotland. Eur Respir J. 2018;51(3):1701794.
8. Kwong JC, Schwartz KL, Campitelli MA, et al. Acute myocardial infarction after laboratory-confirmed influenza infection. N Engl J Med. 2018;378(4):345–353.
• **This study shows an increase in the relative incidence of myocardial infarction within the 7 days following confirmed influenza infection.**
9. Barnes M, Heywood AE, Mahimbo A, et al. Acute myocardial infarction and influenza: a meta-analysis of case–control studies. Heart. 2015;101(21):1738–1747.
10. Aliberti S, Ramirez J, Cosentini R, et al. Acute myocardial infarction versus other cardiovascular events in community-acquired pneumonia. ERJ Open Res. 2015;1(1):00020–2015.
11. Bazaz R, Marriott HM, Francis SE, et al. Mechanistic links between acute respiratory tract infections and acute coronary syndromes. J Infect. 2013;66(1):1–17.
12. Corrales-Medina VF, Madjid M, Musher DM. Role of acute infection in triggering acute coronary syndromes. Lancet Infect Dis. 2010;10(2):83–92.
• **This review provides an overview on the mechanistic links between acute infections and acute coronary syndrome.**
13. Reyes LF, Restrepo MI, Hinojosa CA, et al. Severe pneumococcal pneumonia causes acute cardiac toxicity and subsequent cardiac remodeling. Am J Respir Crit Care Med. 2017;196(5):609–620.
14. Ison MG, Campbell V, Rembold C, et al. Cardiac findings during uncomplicated acute influenza in ambulatory adults. Clin Infect Dis. 2005;40(3):415–422.
15. Paddock CD, Liu L, Denison AM, et al. Myocardial injury and bacterial pneumonia contribute to the pathogenesis of fatal influenza B virus infection. J Infect Dis. 2012;205(6):895–905.
16. Udell JA, Zawi R, Bhatt DL, et al. Association between influenza vaccination and cardiovascular outcomes in high-risk patients. JAMA. 2013;310(16):1711.
• **Systematic review of randomized controlled trials comparing cardiovascular outcomes following influenza vaccination or control, with meta-analysis results showing the vaccine was associated with a lower risk of cardiovascular events.**
17. Mohseni H, Kiran A, Khorshidi R, et al. Influenza vaccination and risk of hospitalization in patients with heart failure: a self-controlled case series study. Eur Heart J. 2017;38(5):326–333.
18. Vlachopoulos CV, Terentes-Printzios DG, Aznaouridis KA, et al. Association between pneumococcal vaccination and cardiovascular outcomes: a systematic review and meta-analysis of cohort studies. Eur J Prev Cardiol. 2015;22(9):1185–1199.
19. European Centre for Disease Prevention and Control. Seasonal influenza vaccination and antiviral use in EU/EEA Member States. 2018. [cited 2019 Jul 16]. Available from: https://ecdc.europa.eu/en/publications-data/seasonal-influenza-vaccination-antiviral-use-eu-eea-member-states.
20. Jefferson T, Jones MA, Doshi P, et al. Neuraminidase inhibitors for preventing and treating influenza in adults and children. Cochrane Database Syst Rev. 2014 Apr 10;(4):CD008965. doi: 10.1002/14651858.CD008965.pub4
• **This systematic review of randomized controlled trials comparing cardiovascular outcomes following neuraminidase inhibitors or control, with meta-analysis results showing a reduction in cardiovascular events in the treatment group.**
21. Casscells SW, Granger E, Kress AM, et al. Use of oseltamivir after influenza infection is associated with reduced incidence of recurrent adverse cardiovascular outcomes among military health system beneficiaries with prior cardiovascular diseases. Circ Cardiovasc Qual Outcomes. 2009;2(2):108–115.
22. Madjid M, Curkendall S, Blumentals WA. The influence of oseltamivir treatment on the risk of stroke after influenza infection. Cardiology. 2009;113(2):98–107.
23. Gwini SM, Coupland CAC, Siriwardena AN. The effect of influenza vaccination on risk of acute myocardial infarction: self-controlled case-series study. Vaccine. 2011;29(6):1145–1149.
24. Vardeny O, Udell JA, Joseph J, et al. High-dose influenza vaccine to reduce clinical outcomes in high-risk cardiovascular patients: rationale and design of the INVESTED trial. Am Heart J. 2018;202:97–103.
25. Cheng -H-H, Tang -T-T, He Q, et al. Beneficial effects of statins on outcomes in pneumonia: a systematic review and meta-analysis. Eur Rev Med Pharmacol Sci. 2014;18(16):2294–2305.
26. Cangemi R, Falcone M, Taliani G, et al. Corticosteroid use and incident myocardial infarction in adults hospitalized for community-acquired pneumonia. Ann Am Thorac Soc. 2019;16(1):91–98.
27. Winning J, Reichel J, Eisenhut Y, et al. Anti-platelet drugs and outcome in severe infection: clinical impact and underlying mechanisms. Platelets. 2009;20(1):50–57.

28. Menéndez R, Méndez R, Aldás I, et al. Community-acquired pneumonia patients at risk for early and long-term cardiovascular events are identified by cardiac biomarkers. Chest. 2019;S0012–3692(19):31392–31393.
• **This observational study investigated biomarker differences between hospitalized patients with community-acquired pneumonia who subsequently had a cardiovascular event compared to those without a cardiovascular event, and shows multiple cardiac biomarkers were higher during the acute phase of pneumonia in patients who had a cardiovascular event.**

1.3. Cardiovascular disease

1.3.1. Definition and diagnosis

Cardiovascular disease (CVD) is an umbrella term for chronic and acute disorders affecting the heart and blood vessels, including all heart and circulatory diseases [1]. ***Figure 1.1*** illustrates the cardiovascular conditions I consider in this thesis.

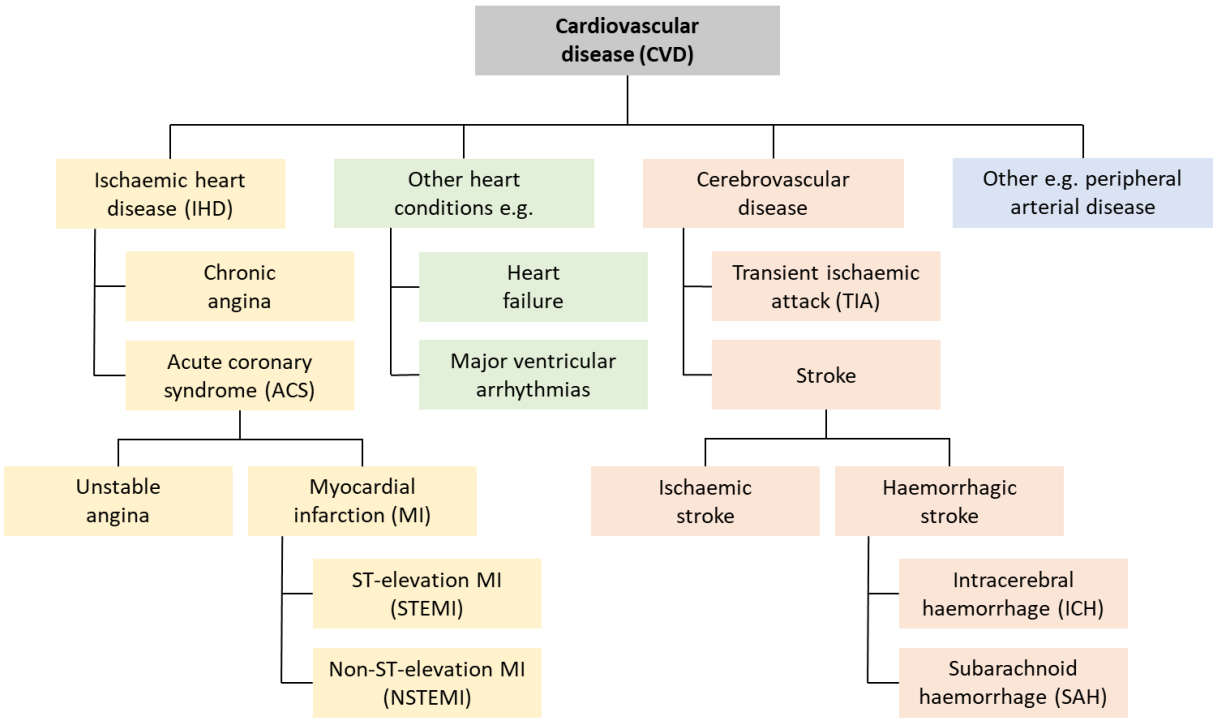


Figure 1.1 Hierarchy of cardiovascular disease conditions and classifications
Figure adapted from [2]; which is published by the Virtual Health Information Network, New Zealand under a CC BY 4.0 license.

Ischaemic heart disease (IHD) most frequently occurs due to atherosclerosis. Atherosclerosis is a lipoprotein-drive process by which arteries narrow due to atheroma (fatty material plaque) build-up on the walls limiting the flow of blood [3,4] (***Figure 1.2***). Atherosclerotic lesions develop under endothelial cells. Endothelial cells form a single layer lining blood vessels and regulate exchanges between the bloodstream and the surrounding tissues [5]. Atheromatous plaque formation occurs at specific sites of the arterial tree through intimal inflammation, necrosis, fibrosis, and calcification [4].

Most plaques remain asymptomatic causing subclinical disease [4]. Some plaques obstruct blood flow through coronary arteries leading to stable angina, and others evoke thrombosis reducing blood flow to the heart muscle and ultimately causing acute coronary syndrome (ACS) [4,6]. Unstable angina or myocardial infarction (MI) may result depending on the location and amount of blockage. An MI, commonly referred to as a heart attack, occurs when the supply of blood and oxygen to the heart muscle is severely blocked [3]. In ST-elevation MI (STEMI) the thrombus is mostly occlusive and sustained, whereas in unstable angina and non-ST-segment elevation MI (NSTEMI), the thrombus is usually incomplete and dynamic [4]. The longer the blockage occurs, the more myocardial cell necrosis (death of heart tissue) and permanent damage to the heart muscle [6].

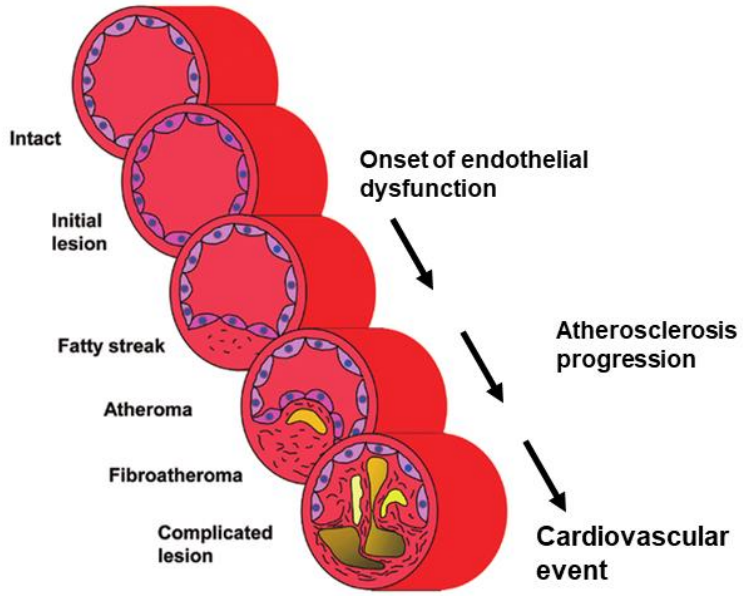


Figure 1.2 Endothelial dysfunction and atherosclerosis progression
Figure adapted from [7]; reproduced with permission from International Heart Journal Association.

For individuals who present to healthcare with a suspected MI, various clinical tools can differentiate MI from angina or non-cardiac related chest pain. The early standard MI definition included three elements; 1) a typical history of chest pain, 2) electrocardiogram (ECG) changes, and 3) the appearance of cardiac biomarkers in the blood [8]. Since then, the definition has been refined using more sensitive cardiac biomarkers [9]. ***Table 1.1*** sets out the evolution of the MI definition over time.

ECG is used to classify MI into ST-elevation MI (STEMI) and non-ST-elevation MI (NSTEMI) based on the elevation or absence of the ST segment, respectively.

The damage caused during a MI can lead to heart failure [10,11]. Heart failure is a complex syndrome in which the heart's ability to maintain blood circulation is impaired due to structural or functional deterioration of ventricular filling or ejection [12]. In addition to MI, conditions that lead to heart failure include hypertension (high blood pressure), cardiomyopathy, arrhythmias (abnormality of the heart's rhythm), or congenital heart defects. Infection, alcoholism, anaemia, or an overactive thyroid may also lead to heart failure [13]. Heart failure is a chronic condition that tends to worsen gradually over time but may have an acute start or may have periodic acute episodes. Heart failure generally starts on the left side (left ventricle) of the heart, which is the heart's main pumping chamber.

Chronic heart failure can be difficult to diagnose due to non-specific symptoms and signs. Assessment of heart failure includes symptom monitoring (such as breathlessness, fluid retention, fatigue, and light headedness) and risk factor evaluation [12]. Clinical examination for tachycardia (heart rate over 100 beats per minute), laterally displaced apex beat (suggestive of heart enlargement), heart murmurs, raised jugular venous pressure and respiratory conditions such as tachypnoea, basal crepitations, and pleural effusions are used in diagnosis [12]. During clinical diagnosis, ECG is used to quantify the heart's ejection (how well the left ventricle pumps blood with each heartbeat).

Heart failure is a clinical syndrome with different aetiologies, so it is difficult to specify with the definition evolving (***Table 1.1***). The New York Heart Association stages of heart failure represent a widely used functional classification for chronic heart failure [14]. Generally, the two main forms of left ventricular heart failure commonly specified in literature are heart failure with reduced left ventricular function (HF-rEF) and heart failure with preserved left ventricular function (HF-pEF). HF-rEF is characterised by a decrease in the ejection fraction (in most definitions to $\leq$40%), thereby resulting in insufficient amounts of oxygenated blood being pumped around the body [15]. HF-pEF occurs when the left ventricle does not relax during filling, so not able to fully fill with blood, resulting in less blood pumped around the body [15].

Arrhythmias may also occur after MI because of myocardial scarring, and ventricular arrhythmias especially can lead to sudden cardiac death. An arrhythmia is an abnormality in the heart's rhythm that can increase (tachycardia, >90 beats per minute [bpm]) or decrease (bradycardia, <50 bpm) the heart rate [16]. Arrhythmias can originate in the upper (supraventricular) or lower (ventricular) chambers of the heart. The ventricles are the heart's main pumping chambers; therefore, ventricular arrhythmias can cause severe morbidity and lead to sudden death [17,18]. Ventricular arrhythmia diagnosis is commonly made using ECG [18].

Similar to the heart, the brain needs a constant supply of blood and oxygen. Stroke is a heterogeneous clinical syndrome characterised as a neurological deficit attributed to an acute focal injury of the central nervous system by a vascular cause [19–22]. A stroke most commonly occurs when the blood supply to the brain is restricted or stopped due to a blood clot, known as ischaemic stroke or cerebral infarction [21]. Largely, blood supply becomes restricted due to atherosclerosis [4,23]. Haemorrhagic (intracerebral haemorrhage [ICH] and subarachnoid haemorrhage [SAH]) strokes can also occur when weakened blood vessels supplying the brain rupture [21,24], primarily driven by degeneration due to long-standing hypertension [25]. The symptoms of stroke, infarction or haemorrhage, last for at least 24 hours or lead to death [20,26]. If the blood supply to the brain is only temporarily interrupted, this is a transient ischaemic attack (TIA) [20].

Ischaemic stroke commonly presents as difficulty with speech and weakness on one half of the body [23]. Haemorrhagic strokes most commonly present as severe headache accompanied by vomiting [23,25]. Neuroimaging is required to confirm an ischaemic stroke, and to differentiate it from haemorrhagic strokes [23]. Analysis of cerebrospinal fluid is used to identify SAH when neuroimaging is not definitive [23]. Stroke definition has largely remained consistent over time, with criteria to distinguish ischaemic stroke, ICH and SAH (***Table 1.1***).

Table 1.1 Major definitions to categorise acute cardiovascular events of myocardial infarction, unstable angina, heart failure and stroke

Year	Defining body	Description
Myocardial infarction		
1958 & 1979	WHO [8,27]	The first standard definition of: 1. typical history of chest pain 2. ECG changes Updated to include detection of cardiac biomarkers
1985	WHO MONICA [28]	Definite event: a. definite ECG or b. symptoms typical, atypical or inadequately described with probable ECG and abnormal enzymes, or c. symptoms typical and abnormal biomarkers with ischaemic or non-codable ECG or ECG not available, or d. fatal case with the naked-eye appearance of fresh MI or recent coronary occlusion found at necropsy
2000	ESC and ACC [29]	"Universal definition": 1. typical rise and gradual fall of troponin or more rapid rise and fall of CK-MB biomarkers of myocardial necrosis with at least one of: a. ischaemic symptoms b. development of pathologic Q waves on the ECG c. ECG changes indicative of ischemia (ST-segment elevation or depression) d. coronary artery intervention (e.g., coronary angioplasty) Or 2. pathologic findings of an acute MI
2007	Global MI Task Force (endorsed by ESC/ACC/AHA/WHF [30]	Update of 2000 to split into: type 1 - spontaneous due to a primary coronary event such as plaque erosion, rupture, fissuring, or dissection type 2 - secondary due to increased oxygen demand or decreased supply, e.g., coronary artery spasm, coronary embolism, anaemia, arrhythmias, hypertension, or hypotension type 3 - sudden unexpected cardiac death, including cardiac arrest type 4 - associated with a percutaneous coronary intervention, stent thrombosis as documented by angiography or at autopsy. type 5 - associated with coronary artery bypass grafting
2012	Joint ESC/ ACC/ AHA/ WHF Task Force [31]	Update of 2007 (known as "third universal definition") to account for more sensitive biomarker assay and to include imaging diagnosis

Year	Defining body	Description
2019	Joint ESC/ ACC/ AHA/ WHF Task Force [32]	Update of 2012 (known as "fourth universal definition") to distinguish non-ischaemic myocardial injury and MI subtypes. If there is no evidence to support the presence of myocardial ischaemia, then the diagnosis is myocardial injury.
Unstable angina		
2003	AHA [33]	1. new cardiac symptoms and positive ECG findings with normal biomarkers 2. changing symptom pattern and positive ECG findings with normal biomarkers
2020	ESC [34]	1. myocardial ischaemia at rest or minimal exertion, and 2. absence of cardiomyocyte necrosis
Heart failure		
1971	Framingham criteria [35]	Diagnosis requires two major or one major and two minor criteria. Major criteria: 1.paroxysmal nocturnal dyspnea or orthopnea, 2.neck vein distension, 3.rales, 4.radiographic cardiomegaly, 5.acute pulmonary edema, 6.S3 gallop, 7.central venous pressure (>16 cm water at the right atrium), 8.circulation time ≥25 seconds, 9.hepatojugular reflux Minor criteria: 1.bilateral ankle edema, 2.nocturnal cough, 3.dyspnoea on ordinary exertion, 4.hepatomegaly, 5.pleural effusion, 6.decrease in vital capacity by 33% from maximal value recorded, 7.tachycardia
1985	Boston criteria [36]	A composite score based on points from three categories (maximum 4 points per category), diagnosis is classified as "definite" with 8-12 points, "possible" with 5-7 points, and "unlikely" ≤4. Category I history: rest dyspnea (4 points), orthopnea (4 points), paroxysmal nocturnal dyspnea (3 points), dyspnea while walking on a level area (2 points), dyspnea while climbing (1 point) Category II physical examination: heart rate abnormality (1 point if 91 to 110 beats per minute; 2 points if more than 110 beats per minute), jugular venous elevation (2 points if greater than 6 cm H2O; 3 points if greater than 6 cm H2O plus hepatomegaly or oedema), lung crackles (1 point if basilar; 2 points if more than basilar), wheezing (3 points), third heart sound (3 points) Category III chest radiography: pulmonary alveolar oedema (4 points), interstitial pulmonary oedema (3 points), bilateral pleural effusion (3 points), cardiothoracic ratio greater than 0.50 (3 points), upper zone flow redistribution (2 points)
2005	ESC [37]	Acute HF split into groups I-VI on clinical and haemodynamic characteristics. I: acute decompensated HF (de novo or as decompensation of chronic HF) with signs and symptoms of acute HF, which are mild and do not fulfil criteria for cardiogenic shock, pulmonary oedema or a hypertensive crisis II: hypertensive AHF where signs and symptoms of HF are accompanied by high blood pressure and relatively preserved left ventricular function with a chest radiograph compatible with acute pulmonary oedema

Year	Defining body	Description
		III: pulmonary oedema (verified by chest X-ray) accompanied by severe respiratory distress, with crackles over the lung and orthopnoea, with O_2 saturation usually <90% on room air before treatment IV: cardiogenic shock defined as evidence of tissue hypoperfusion induced by HF after correction of pre-load, reduced blood pressure (systolic BP <90 mmHg or a drop of mean arterial pressure>30 mmHg) or low urine output (<0.5 ml/kg/h), with a pulse rate >60 b.p.m. with or without evidence of organ congestion V: high-output failure characterised by high cardiac output, usually with high heart rate (caused by arrhythmias, thyrotoxicosis, anaemia, Paget's disease, iatrogenic or by other mechanisms), with warm peripheries, pulmonary congestion, and sometimes with low BP as in septic shock VI: right-sided HF characterised by low output syndrome with increased jugular venous pressure, increased liver size and hypotension
2013	ACC/AHA [38]	HF is a complex clinical syndrome resulting from any structural or functional impairment of ventricular filling or ejection of blood. The cardinal manifestations of HF are dyspnea and fatigue, limiting exercise tolerance, and fluid retention, leading to pulmonary or splanchnic congestion or peripheral oedema. Some patients have exercise intolerance but little evidence of fluid retention, whereas others complain primarily of oedema, dyspnea, or fatigue.
2016	ESC [39]	HF is a clinical syndrome characterised by typical symptoms (e.g., breathlessness, ankle swelling and fatigue) that may be accompanied by signs (e.g., elevated jugular venous pressure, pulmonary crackles and peripheral oedema) caused by a structural or functional cardiac abnormality, resulting in reduced cardiac output or elevated intracardiac pressures at rest or during stress.
2021	HFSA/ HFA-ESC/ JHFS [40]	Universal definition and classification. Stage A: at risk for HF but without current or prior symptoms or signs of HF and without structural or biomarker evidence of heart disease. Stage B: pre-heart failure for patients without current or prior symptoms or signs of HF but evidence of structural heart disease, abnormal cardiac function, or elevated natriuretic peptide levels. Stage C: for patients with current or prior symptoms or signs of HF caused by a structural or functional cardiac abnormality. Stage D: advanced HF for patients with severe symptoms or signs of HF at rest, recurrent hospitalisations despite guideline-directed medical therapy (GDMT), refractory or intolerant to GDMT, requiring advanced therapies such as consideration for transplant, mechanical circulatory support, or palliative care. New and revised classification of HF according to left ventricular ejection fraction (LVEF): -HF with reduced ejection fraction (HFrEF): HF with LVEF ≤40% -HF with mildly reduced ejection fraction (HFmrEF): HF with LVEF 41–49%

Year	Defining body	Description
		-HF with preserved ejection fraction (HFpEF): HF with LVEF ≥50%; -HF with improved ejection fraction (HFimpEF): HF with a baseline LVEF ≤40%, a ≥10 point increase from baseline LVEF, and a second measurement of LVEF >40%
Stroke		
1970s	WHO [41]	Rapidly developing clinical signs of focal (or global) disturbance of cerebral function, lasting ≥24 hours or leading to death, with no apparent cause other than that of vascular origin.
2013	AHA/ASA [21]	Ischaemic stroke: an episode of neurological dysfunction caused by focal cerebral, spinal, or retinal infarction. Stroke caused by intracerebral haemorrhage: rapidly developing clinical signs of neurological dysfunction attributable to a focal collection of blood within the brain parenchyma or ventricular system that is not caused by trauma. Stroke caused by subarachnoid haemorrhage: rapidly developing signs of neurological dysfunction or headache because of bleeding into the subarachnoid space, which is not caused by trauma. Stroke caused by cerebral venous thrombosis: infarction or haemorrhage in the brain, spinal cord, or retina because of thrombosis of a cerebral venous structure. Symptoms or signs caused by reversible oedema without infarction or haemorrhage do not qualify as stroke. Stroke, not otherwise specified: an episode of acute neurological dysfunction presumed to be caused by ischemia or haemorrhage, persisting ≥24 hours or until death, but without sufficient evidence to be classified as one of the above.
Transient ischaemic attack		
Unknown	Unknown	Definition referred to in sources as a sudden, focal neurologic deficit lasting <24 hours presumed to have a vascular origin and confined to an area of the brain or eye perfused by a specific artery, but no original source accredited [42].
2009	AHA/ASA [43]	A transient episode of neurological dysfunction caused by focal brain, spinal cord or retinal ischaemia, without acute infarction.

1.3.2. Incidence and prevalence

CVD has been the leading global cause of death for the last two decades and accounted for nearly 18 million deaths in 2019, with 85% due to MI and stroke [1,44,45]. The number of global CVD deaths has increased over time, from 12 million in 1990 [45]. However, CVD mortality has decreased in many high-income countries, including the United Kingdom (UK). In 2019 there were an estimated 168,000 CVD deaths in the UK [46], at a rate of 255 per 100,000 people compared with 527 per 100,000 in 1999 [47]. However, deaths are only one element of CVD impact.

The Global Burden of Disease survey estimated that in 2019, the global prevalence of CVD was 523 million (95% uncertainty interval [UI] 497-550 million) in 2019 compared with 271 million (95% UI 257-285 million) in 1990 [45,48]. In the UK, IHD and stroke prevalence have remained constant at 3–4% and 2%, respectively, over this time [49], but stroke prevalence projections suggest there may be an increase of 120% by 2035 [50]. However, incidence has decreased over time so constant or increased prevalence is likely linked to reduced case fatality [49]. One study using hospital admission data for England found 33% and 31% declines in the incidence of myocardial infarction in men and women, respectively, between 2002 and 2010 [51]. Another study using primary care records found the UK incidence of stroke decreased by 30% between 1998 and 2008 [51].

CVD results in high individuals, social and economic costs. Events such as myocardial infarction and stroke can limit the sufferers' physical health, potentially resulting in inability to work or drive, and can impact mental wellbeing. The global trends for disability-adjusted life years (DALYs), years lived with disability (YLD), and years of life lost (YLL) associated with CVD have also increased over time. In particular, YLD doubled from 17.7 million (95% UI 12.9-22.5 million) to 34.4 million (95% UI 24.9-43.6 million) between 1990 and 2019 [45]. High-income country healthcare expenditure on stroke is estimated to be 3–5% [52–54] and heart failure to be 1–2% [55]. Incident MIs result in around 100,000 hospitalisations a year in the UK [56]. In the UK, previous estimates suggest an annual expenditure of £6.8 billion for CVD [49]. The high cost is due to the acute and severe nature of conditions such as MI and stroke (estimated £4.4 million annually) as well as the need for prescription

medications to manage and reduce the risk of further cardiovascular complications among those with established CVD (estimated £1.4 million annually) [49].

1.3.3. Risk factors

The traditional risk factors for the onset of CVD have been long established, with landmark twentieth century studies such as the United States of America (USA) Framingham Heart Study [57] and UK Whitehall Study [58]. Age is the strongest predictor of CVD, with risk rising as age increases [49]. For example, the risk of stroke in people aged 75-84 years is 25 times higher than the risk in people aged 45-54 years [20]. Risk differs by sex, with CVD more common in men than women. A recent UK Biobank study found that the incidence of MI was nearly 3.5 times higher in men than women [59]. Similarly, stroke incidence is higher in men, although as women have a longer life expectancy, a higher absolute number of women have a stroke than men [20,60].

Epidemiological studies show that hypertension (high blood pressure in the arteries) is strongly associated with the development of CVD [61]. Ideal blood pressure generally falls between 90/60mmHg and 120/80mmHg, while high blood pressure is over 140/90mmHg. The global prevalence of hypertension is estimated to have doubled since 1990 from 331 (306-359) million women and 317 (292-344) million men to 626 (584-668) million women and 652 (604–698) million in 2019 [62]. However, the age-standardised prevalence has remained stable [62]. Approximately 50% of incident MI and stroke events are attributable to hypertension [63].

In addition, other major modifiable (controllable) CVD risk factors include high total and low-density lipoprotein (LDL) cholesterol, smoking, physical inactivity, unhealthy diet, increasing body mass index (BMI). Overall, the combined effects of these risk factors likely account for nearly 90% of IHD and 75% of strokes [63]. Diabetes is also associated with increased cardiovascular risk [64], with an estimated one in three adults with type 2 diabetes having concomitant CVD [65,66].

Various risk prediction models estimate individual likelihood, most commonly 10-year risk, of future CVD based on known risk factors [67]. The widely used Framingham risk score was developed in 1998 using results from the Framingham Heart Study [68], and more recent models include the European SCORE [69] and UK QRISK [70,71] algorithms. QRISK has been widely adopted in the UK, with routine use in primary care and is outlined in more detail in Chapter 4. Overall, CVD risk prediction tools can aid clinicians' targeted modifiable CVD risk factors treatment decisions and engage patients in adhering to such treatments [72].

Many CVD risk factors are interrelated; for example, obesity can lead to type 2 diabetes and hypertension [73]. Clustering of risk factors occurs among both modifiable and non-modifiable risk factors. For instance, non-White ethnic groups have a higher risk of diabetes and CVD [74] while individuals with a higher level of deprivation have high rates of obesity and CVD [75]. Observational studies suggest a low incidence of CVD events in people with the best cardiovascular health i.e., with no or very few of the aforementioned risk factors [76–78]. Overall, the presence of co- or multi-morbid conditions are risk factors for the onset of CVD and once a cardiovascular event is experienced this predisposes an individual to a higher risk of future events.

1.3.4. Prevention strategies

CVD prevention methods can be primary or secondary. Primary prevention refers to attenuation of risk to prevent the onset of disease while secondary prevention focuses on reducing further acute events or deterioration in those with existing disease [79]. CVD prevention strategies, both primary and secondary, largely involve the reversal of or treatment for the major modifiable risk factors of hypertension, cholesterol, smoking, physical inactivity, poor diet, and diabetes [80]. Such prevention strategies are the likely significant contributors to CVD incidence reduction [81,82].

Lifestyle advice regarding, for example, weight loss or reduced salt intake and antihypertensive drug treatment to lower high blood pressure are used for primary and secondary prevention. The

prescription of antihypertensive drugs and the drug used depends on ongoing blood pressure (BP) measures, age, ethnicity and co-morbidities such as type 2 diabetes [83]. However, it is estimated that large proportions (41% [38-45%] globally and 27% [22-31%] in high-income countries) of people living with hypertension are undiagnosed and among those with diagnosed hypertension, between one-fifth and one-quarter of people will continue to have uncontrolled hypertension despite treatment provision [84].

Statins are lipid-lowering agents that inhibit cholesterol synthesis and are widely prescribed in high-income countries, including the UK, to people with high LDL cholesterol as a primary and secondary CVD prevention strategy [85]. Randomised controlled trial (RCT) evidence shows that statin use reduces the risk of cardiovascular events in people with and without established CVD [86,87]. However, many studies suggest adherence issues limit the cardiovascular benefit the treatment can provide [88,89].

Antiplatelet treatment, of which aspirin has the most widespread use, is recommended for secondary CVD prevention but not recommended for primary CVD prevention as the risk of bleeding outweighs cardiovascular benefit [90,91]. Short courses of antiplatelet agents can be used to avoid further severe cardiovascular complications; for example, evidence shows that use following a TIA or a minor ischaemic stroke can prevent a subsequent more severe stroke [92].

1.4. Systemic acute respiratory infections

1.4.1. Definitions and diagnosis

ARIs may affect the upper (nasal cavity to the larynx) or lower respiratory tract (trachea to lungs). A range of microorganisms infect the respiratory tract. Generally, conditions such as the common cold, sinusitis, and tonsillitis, which affect the upper respiratory tract, result in mild illness. Conversely, many lower respiratory tract infections (LRTIs) such as influenza, bronchitis, and pneumonia result in moderate to severe clinical manifestations.

Pneumonia is a severe infection of the lung parenchyma characterised by a cough and at least one of fever, breathlessness or rapid breathing, and chest pain [93]. Pneumonia is traditionally referred to as community-acquired pneumonia (CAP) or hospital-acquired pneumonia (HAP), based on the location of acquisition. CAP is caused by various bacterial, viral, or fungal agents (the latter is less common), with *S. pneumoniae* the predominant causative agent. A review of European data found that adults with CAP *S. pneumoniae* accounted for up to 85% of infections [94].

Influenza is largely mild, with symptoms being fever, headache, myalgia, cough or sore throat, or severe in its presentation, including the development of pneumonia [93]. A significant number of infections are asymptomatic; one study identifying 75% of serologically confirmed influenza cases to be asymptomatic [95]. Human influenza epidemics are caused by influenza A and B virus strains. In temperate climates, influenza is a largely seasonal viral infection that typically coincides with winter, while in subtropical and tropical climates the virus can circulate year-round and result in two to three outbreaks a year [96,97]. Influenza A virus strains can result in more severe epidemics and are an ongoing public health threat [98]. Influenza A virus is subdivided based on combinations of its two surface glycoproteins, hemagglutinin (H) and neuraminidase (N) [99]. Influenza A epidemics recur and vary in severity due to the accumulation of point mutations in the virus surface glycoproteins (known as antigenic drift), thereby evading previously acquired immunity (from infection or vaccination) [99]. A major change in influenza A surface glycoprotein (known as antigenic shift), which results from genetic reassortment between several subtypes of influenza in one host, with pigs able to be infected by both avian and human influenza strains, can produce a novel strain [99]. Sixteen H subtypes have been identified in wildfowl, the natural reservoir for influenza A, providing a source of strain novel to humans [99]. Due to little pre-existing immunity, the transmission of a novel strain can lead to an influenza pandemic (global circulation of a novel influenza A virus). The most recent influenza pandemic occurred in 2009/10 following the circulation of a new H1N1 subtype of the influenza A virus [100].

The emergence of new non-influenza respiratory viral diseases able to infect humans is an ongoing global public health threat [101]. In the past two decades, two coronaviruses of global concern

emerged – severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV) – before the onset of the coronavirus disease 2019 (COVID-19) pandemic in 2019/20. COVID-19, due to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), clinical manifestation ranges from asymptomatic infection to cough, fever and fatigue to breathlessness and hypoxemia (below-normal blood oxygen level) with the requirement of respiratory support [102,103].

Laboratory confirmation for ARIs pre-COVID-19 was often not performed, so no causative agent was attributable to the infection [93]. The lack of confirmation stems from many ARIs only requiring primary care medical attention where testing is limited. Among individuals presenting with an ARI, clinical judgement regarding the severity and possible cause of infection will be made based on the clinical features such as signs and symptoms as well as epidemiological intelligence i.e., current outbreaks and trends [104].

A range of laboratory methods can identify ARI causative agents. For influenza, a nasopharyngeal swab or wash procedure is most commonly collected for PCR testing, although virus culture is the gold standard it is labour intensive and takes several days to obtain a result [105]. Similar to influenza, COVID-19 confirmation is based on nasopharyngeal swab and PCR testing. In the case of pneumonia, a sputum or blood culture would most commonly be used.

When confirmation is obtained, pre-COVID-19 common causes of LRTI in European populations are *Streptococcus pneumoniae*, *Haemophilus influenzae*, influenza viruses, human rhinoviruses, human coronavirus and respiratory syncytial virus [106,107]. Most infections will be self-limiting or, if bacterial in origin, easily treated with antibiotics. However, an infection can deteriorate and lead to pneumonia or other complications.

1.4.2. Incidence

LRTIs are the leading infectious disease cause of death globally [44], including in the UK, with many pneumonia-related [108]. Among people hospitalised due to CAP, 5–15% die within 30 days [109]. Due to the high hospitalisation rate among people with CAP, the associated healthcare expenditure is significant; in Europe, an estimated €6.4 billion is spent annually [108]. Measuring the exact burden and impact of LRTIs, particularly seasonal influenza, which often goes unidentified, is difficult. Estimates suggest seasonal influenza can infect up to 20% of the global population annually and is associated with 290,000–650,000 deaths globally [110–112]. In the UK, seasonal influenza levels have remained moderate in recent years, but the annual impact on healthcare remains, including high hospital bed usage [113]. With the onset of the COVID-19 pandemic, there was minimal influenza circulation during the winter of 2020/21 [114]. As of the first week of 2022, there had been more than 300 million reported cases of COVID-19 and nearly 5.5 million deaths globally [116].

Overall, ARIs are a frequent cause of primary care consultation. Over the 14 years of 1997–2011, based on primary and secondary care electronic health records (EHRs), the incidence of CAP in the UK was estimated to be nearly 8 per 1000 person-years, with half of those individuals hospitalised [116]. A time trends analysis of hospital admissions in Oxfordshire, UK found the incidence of CAP increased by 4.2% (3.6–4.8) per year between 1998 and 2008, and consequently, the increase accelerated to 8.8% (7.8–9.7) per year from 2009 to 2014 [117]. Similar CAP hospital admission incidence increases have occurred in other European countries, such as Spain (142.4 per 100,000 in 2004 to 163.87 per 100,000 in 2013 [118]) and Sweden (271 per 100,000 in 2005 to 306 per 100,000 in 2015 [119]).

1.4.3. Risk factors

LRTIs particularly impact young children (aged <5 years), older adults (≥65 years), and those with underlying health conditions. Multiple studies reported the rate of influenza-related mortality to be six

to ten times higher in people aged ≥65 years when compared to younger age groups [97,120–122]. Although age is the primary driver of severe influenza, other factors contribute. One UK study found the risk of hospitalisation due to influenza was higher across all age groups in people who had an underlying health condition, but the relative risk decreased with age (5-14 years = 5.7 [5.2-6.3], 15-44 years = 4.9 [4.1-5.1], 45-64 years = 4.5 [4.3-4.7], ≥65 years = 1.8 [1.7-2.0]) [123]. A 2012 review of CAP incidence in Europe unsurprisingly found a sharp rise as age increased, but also found the increase was particularly pronounced in men [124]. Aside from age and sex, chronic lung disease, CVD, dementia, liver or renal disease, cancer or immunosuppressive conditions increase CAP incidence and associated adverse outcomes [125,126]. COVID-19 has highlighted the increased risk of severe infection or adverse outcomes in the previously mentioned risk groups as well as those with diabetes and the severely obese [127,128].

1.4.4. Prevention and treatment

In the UK among adults, like most other high-income countries, an annual influenza vaccine and a one-off 23-valent pneumococcal polysaccharide vaccine (PPV23) are recommended for all aged ≥65 years and those aged <65 years with underlying health conditions to prevent complications of infections caused by influenza virus and *S. pneumoniae*, respectively [129,130]. Repeat PPV23 is recommended every five years for people with no spleen, splenic dysfunction or chronic renal disease [129]. Seasonal influenza vaccine is tailored annually to account for antigenic drift to have a vaccine that aims to match the virus strains predicted to circulate that year [130].

People aged ≥65 years and those with underlying health conditions are targeted for vaccination to reduce the direct morbidity and mortality attributed to the influenza virus or *S. pneumoniae* and the impact the infection may have on an individual's other health conditions. In the UK, influenza vaccine uptake among people aged ≥65 years is high at nearly 75% [131]. Influenza and pneumococcal vaccine uptake among the younger populations are suboptimal at between 10–60% (depending on the

clinical risk group), despite these groups making up a sizable proportion (approximately 60%) of the population who are recommended to receive vaccines [131].

Vaccine effectiveness (VE) studies suggest PPV23 protection wanes over time, decreasing from 48% (32%–60%) at two years post-vaccination to 15% (-3%–30%) after five or more years in those aged ≥65 years [132]. Further research suggests particularly low VE in adults with underlying health conditions [133]. The changing composition of the influenza vaccine and annual variation in circulating strains leads to varying VE. Generally, VE for the influenza vaccine is estimated to be 30-60% for different influenza A (H1N1 and H3N2) and B strains [134]. In the last 10-15 years, a mismatch between the vaccine and circulating strains occurred during the 2009/10 H1N1 pandemic and in 2014/15 due to antigenic drift in the H3N2 strain [113].

A global effort to develop effective COVID-19 vaccines resulted in several successful candidates approved for use within one year of the virus' emergence [135]. The global rollout has been variable, with widespread coverage limited mainly to high-income countries [136]. In the UK, 91% of individuals aged ≥40 years were fully vaccinated by the end of July 2021 [137], with a booster programme targeted at people aged ≥50 years started in Autumn 2021 [138]. Very high vaccine efficacy estimates for the prevention of severe illness were obtained from the initial COVID-19 RCTs [139].

Although prevention of influenza infection is most effectively accomplished by vaccination, antiviral drugs can aid targeted influenza prevention and treatment [140]. Prophylactic use in the UK may be required when adequate protection has not been provided through vaccination, for example, from an antigenic mismatch between seasonal vaccine strains and circulating strains or <14 days between vaccination and exposure to the virus, or to control an outbreak (regardless of vaccination status) in residential care communities or prisons [141]. In the UK, antiviral treatment for suspected or confirmed influenza depends on the likelihood of severe illness or influenza-related complications. Antiviral treatment should be started as soon as symptoms appear, ideally within 48 hours [141]. As symptoms can start as non-specific and mild, this may limit the effectiveness of antiviral treatment.

Antiviral treatments for COVID-19 have also been successfully created. RCT results showed 50% and 89% reductions in the risk of hospital admission or death in non-hospitalised adults with mild to moderate COVID-19 symptoms who were at high risk of severe illness and outcomes for the Merck produced Molnupiravir and Pfizer produced Paxlovid, respectively, when compared to placebos [142,143].

1.5. Cardiovascular complications of systemic acute respiratory infections

1.5.1. Epidemiology

Ecological studies have reported seasonal patterns in MI [144] and CVD deaths [145,146], which have remained over time despite decreasing CVD incidence. This seasonal variation has been identified in both the North and South hemispheres and temperate and subtropical climates [144–146]. In temperate climates, MI incidence and CVD deaths have winter peaks and summer troughs, with greater seasonal variation as proximity to the equator increases [145]. However, some studies have reported no MI seasonality [147], and there are inconsistent seasonal trends in stroke incidence [147–152].

Multiple explanatory factors may contribute to cardiovascular event seasonality, including temperature [153], air quality [154] and respiratory infections. Several ecological studies also show that the timing of influenza epidemics correspond to peaks in MI and CVD deaths [155–159]. Again, these results are from both hemispheres and temperate, subtropical and tropical climates.

The association between influenza and acute cardiovascular events has been the subject of multiple systematic reviews. Pooled estimates from two meta-analyses have shown that the odds of MI were higher after a recent influenza-like illness (ILI) [160,161]. From the five studies identified in the review conducted by Warren-Gash *et al*, two case-control studies reported at least a three-times increase in odds (OR 3.8 [1.4-10.8] and 3.0 [1.1-8.2]), two case-control studies showed non-significant increases in odds (OR 1.7 [0.5-5.6] and 1.2 [0.3-4.4]), and one case-crossover study found

the relative probability of MI on the day after ILI onset compared with seven days after was 2.4 (1.7-3.4). Kwok *et al* pooled estimates from seven studies (a mixture of case-control, cohort, case-crossover and self-controlled case series) to identify a two-fold increase in odds of MI after ILI (OR 2.17 [1.68-2.80]). The review by Kwok *et al* also investigated laboratory-confirmed influenza, with a pooled OR of 1.27 (0.54-2.95) from four studies (three case-control studies and one cohort study) but with substantial heterogeneity between studies (I^2 79.9%). Barnes *et al* also investigated the relationship between laboratory-confirmed influenza and MI among case-control studies but found no significant increase in MI incidence after confirmed influenza (pooled OR 2.44 [0.83-7.20]), again there was considerable heterogeneity between studies (I^2 80.9%) [160].

The association between pneumonia and acute cardiovascular complications has been less investigated. However, a systematic review conducted in 2010 by Corrales-Medina *et al* found that after CAP, among inpatients, there was pooled incidence of 17.7% (13.9-22.2) for overall cardiac complications, 14.1% (9.3-20.6) for acute heart failure, 5.3% (3.2-8.6) for ACS and 4.7% (2.4-8.9) for arrhythmias [162]. Several of the studies in the systematic review reported risk factors associated with CAP-related cardiac complications, including older age, pre-existing congestive heart failure, severe CAP infection, and the use of insulin.

Generally, the risk of cardiovascular complications triggered by ARIs has prominently been studied in older populations and those with established CVD. Several self-controlled case series (SCCS) studies have specifically looked at the role of ARIs triggering first cardiovascular events. The SCCS studies show a consistent transient increase in the risk of MI in the 1–3 days after clinically diagnosed ARI (IR 4.95 [4.43-5.53] [163] and 4.19 [3.18-5.53] [164]) as well as laboratory-confirmed *S. pneumoniae* (5.98 [2.47-14.4]) or respiratory virus (5.59 [1.77-17.6]) [165], and stroke after ARI (3.19 [2.81-3.62]) [163], laboratory-confirmed *S. pneumoniae* (12.3 [5.48-27.7]) or respiratory virus (6.79 [1.67-27.5]) [165]. The risk decreases gradually over time with a waning of risk after one month [163,164].

Analyses focused on individuals aged <65 years or with cardiovascular risk factors, such as hypertension or diabetes, have produced mixed results [160,164–166]. Age stratified SCCS analysis

identified higher rates of first MI or stroke in days 1-7 after a laboratory-confirmed respiratory virus in patients aged <65 years compared with those ≥65 years [165]. The authors hypothesise this is due to lower vaccination rates in the younger population. Other studies have found no increased risk of cardiovascular complications following influenza/ILI in younger age groups [164,166]. These studies were underpowered to estimate the relative incidence and risk in younger age groups.

Given this existing evidence base before the COVID-19 pandemic, considering the cardiovascular complications of COVID-19 has been a focus of pandemic-related research. A SCCS and matched cohort study using EHR data from Sweden February–September 2020 found in the two weeks after COVID-19 an increased relative incidence of first MI (IR days 1-7 = 2.89 [1.51-5.55] and days 8-14 = 2.53 [1.29-4.94]), beyond which time the increase was not significant (days 15-28 = 1.60 [0.84-3.04]) [167]. In comparison, the relative incidence of ischaemic stroke remained high in the month after COVID-19 (days 1-7 = 2.97 [1.71-5.15], days 8-14 = 2.80 [1.60-4.88] and days 15-28 = 2.10 [1.33-3·32]). The matched cohort analysis conducted by the authors supported SCCS findings, with three-fold increased odds of MI (OR 3.41 [1.58-7.36]) and ischaemic stroke (3.63 [1.69-7.80]) in the two weeks after infection. A SCCS analysis of EHR data from Denmark up to mid-July 2020 found that in the 31 days after COVID-19, there was an increased relative incidence of first MI (IR 3.4 [1.2-9.7]) and stroke (6.6 [3.6-11.9]) [168].

1.5.2. Mechanistic links

Several mechanisms linking ARI to acute cardiovascular events have been postulated (**Figure *1.3***). The infectious agent may directly affect the vascular cells, or the infection may induce haemodynamic (such as increased metabolic demand, coronary vasoconstriction and hypoxemia), haemostatic (increased platelet aggregation and increased plasma viscosity), inflammatory, and pro-coagulant processes [169,170]. Systemic inflammation likely plays an important role, with infectious agents such as the influenza virus and *S. pneumoniae* able to augment the circulation of interleukin (IL)-1-β, tumour necrosis factor-α [169]. The release of such pro-inflammatory cytokines in response to an

infection can mediate atherosclerosis or, along with infection-induced localised [169], directly impact plaque rupture. Endothelial dysfunction is a key early stage of atherosclerosis caused by a range of cardiovascular risk factors such as high levels of LDL, diabetes [170,171]. During severe infection, organisms such as the influenza virus and *S. pneumoniae* [170,171].

In addition to the mechanisms outlined above, the spike protein of SARS-CoV-2 facilitates entry into host cells through the angiotensin-converting enzyme 2 (ACE2) receptor, dysregulating the renin-angiotensin system, which is important for maintaining blood pressure [102].

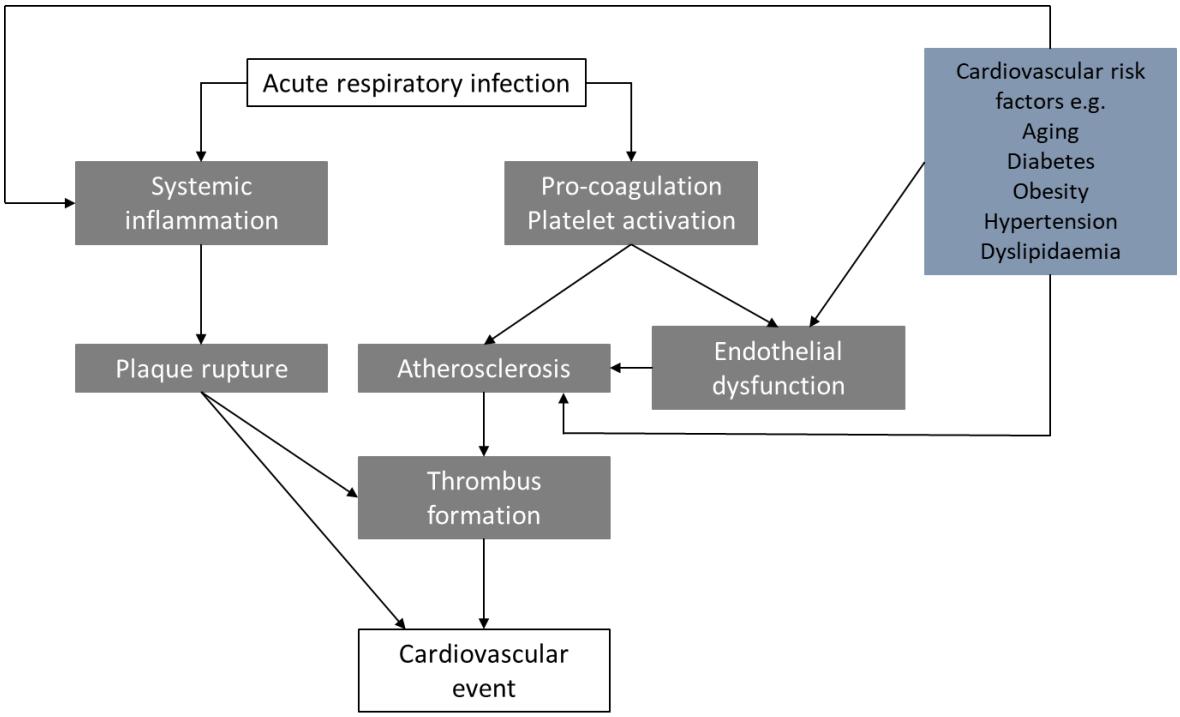


Figure 1.3 Overview of potential mechanistic links between acute respiratory infections and cardiovascular complications

1.5.3. Use of vaccines to prevent cardiovascular events

RCTs and observational studies have demonstrated the benefits of influenza and pneumococcal vaccines in people with established CVD. A 2013 systematic review identified four small efficacy trials and eight safety trials. A meta-analysis of five trials found that the influenza vaccine reduced the risk (RR 0.64 [0.48-0.86]) of composite cardiovascular outcomes within one year of vaccination

[172]. A subgroup meta-analysis with three RCTs showed this reduction predominantly occurred in people with recent ACS compared to stable IHD [172]. Another 2015 meta-analysis of four RCTs comparing influenza vaccine to placebo or no vaccination showed the vaccine's utility in secondary prevention of CVD; there was a significant reduction (RR 0.45 [0.26-0.76]) in CVD mortality among those vaccinated [173]. A further 2021 systematic review found no new efficacy RCTs [174]. However, the results from two new influenza vaccine RCTs with cardiovascular endpoints are now published [175,176]. The Influenza vaccine After Myocardial Infarction (IAMI) trial evaluated the efficacy of influenza vaccine following MI or percutaneous coronary intervention (PCI) in 2,571 participants with coronary artery disease and found that over 12 months of follow-up, the vaccine reduced the occurrence of the composite outcome of all-cause mortality, MI, or stent thrombosis (0.72, 0.52-0.99) as well as secondary endpoints CVD death (HR 0.59 [0.39-9.0]) and MI (0.86 [0.50-1.46]), although the latter was not a significant reduction [175]. The Influenza Vaccine to Effectively Stop cardio Thoracic Events and Decompensated heart failure (INVESTED) study evaluated the impact of high-dose trivalent influenza vaccine compared with standard-dose quadrivalent influenza vaccine on reducing all-cause death or cardiopulmonary hospitalisation in high-risk patients with cardiovascular disease, with no difference in all-cause death or cardiopulmonary hospitalisation identified [176]. A further efficacy trial, Influenza Vaccine to prevent adverse Vascular Events (IVVE), is also due to report results in 2022 [177].

While there are difficulties in using observational studies to investigate vaccine effectiveness due to selection bias and confounding [178] (see Chapter 7 for more detail), studies have demonstrated similar protective effects, particularly between influenza vaccine and MI [179–182]. One study found an overall 19% reduction in the rate of MI after influenza vaccination, with varying results depending on the timing of vaccination; early vaccination associated with a lower rate reduction (21%) than vaccination after mid-November (12%) [182]. Similarly, the influenza vaccine reduced stroke incidence in several observational studies [180,183,184]. In a limited number of studies, results found that the influenza vaccine was not associated with a reduction in MI or stroke [185–187], including subsequent MI [188].

In one SCCS analysis of influenza vaccination in heart failure patients, the vaccine was associated with a lower risk of hospitalisation for further cardiovascular complications (IR 0.73 [0.71-0.76]) [189]. Observational studies have also investigated the vaccine's impact on first cardiovascular events; one SCCS study reported reduced relative incidence of first MI (IR 0.87 [0.79-0.96]) and stroke (0.88 [0.80-0.97]) in the 15–28 days post-vaccination [163]. Further SCCS studies found significantly reduced relative incidence of MI (IR 0.82 [0.75-0.90]) [190] and stroke (0.83 [0.77-0.89]) [191] up to 60 days following influenza vaccination. A recent study using Norwegian EHR data compared the relative incidence of MI, stroke and pulmonary embolism during the 2009 H1N1 influenza pandemic among vaccinated adults deemed at high and low cardiovascular risk [192]. While the study found that the pandemic influenza vaccine reduced the rate of MI and stroke in those with high cardiovascular risk, increases in the rates of MI and stroke were identified among people at low cardiovascular risk. The discordant results among high and low cardiovascular risk may be due to bias. The study defined cardiovascular risk using prescriptions of anti-diabetic, anti-obesity, anti-thrombotic, pulmonary or cardiovascular medications at the time of vaccination, which was after follow-up had started. This study design would place any individual who had a MI or stroke before vaccination and therefore prescribed cardiovascular medications as high risk at the time of vaccination.

The effects of pneumococcal vaccination on cardiovascular outcomes are less clear, particularly given that *S. pneumoniae* infection and the vaccine are not seasonal, making investigating the association more complex. No RCTs were identified by the two systematic reviews conducted to date, while results from observational studies are mixed [193,194]. In people aged ≥65 years, vaccination was associated with a lower risk of MI (RR 0.90 [0.82-1.00]), but no reduction in risk was identified among patients of all ages (0.96 [0.82-1.12]) [194].

1.6. Chapter summary

- CVD is the leading cause of death globally, and while no longer the leading cause in Europe, CVD accounts for a sizeable proportion of deaths, and ageing populations contribute to increased prevalence.
- Extensive epidemiological research and subsequent public health action focused on prominent 'traditional' CVD risk factors such as hypertension, smoking, diabetes, and obesity.
- Population-level studies indicate that patterns of CVD incidence and mortality mirror the seasonality of influenza and ILI.
- Individual-level studies using EHR data and SCCS methods estimated a two- to six-fold increase in the risk of MI and stroke for up to one month after an ARI.
- Meta-analysis of case-control studies identified that people who experienced a MI had twice the odds of prior ARI than that of people who did not have a MI.
- ARIs may induce haemodynamic, inflammatory and pro-coagulant processes, which can lead to a cardiovascular event.
- RCTs and observational studies show that influenza and pneumococcal vaccines reduce the occurrence of cardiovascular complications.
- Currently, influenza and pneumococcal vaccines are recommended for use in older adults (aged ≥65 years in most countries) and those with underlying health conditions predisposing to a severe infection or adverse outcomes, including people with established CVD.
- In the UK, influenza and pneumococcal vaccines uptake are high among individuals aged ≥65 years but suboptimal in risk groups aged <65 years.
- The COVID-19 pandemic has accelerated research on the extrapulmonary manifestations of respiratory infections, including cardiovascular complications.

Chapter 2 Aim and objectives

2.1. Thesis rationale

CVD prevalence remains substantial due to an ageing population and an increase in risk factors, such as obesity and diabetes. Considerable research has demonstrated an association between systemic ARIs and transiently increased risk of acute cardiovascular events. Meanwhile, vaccine effectiveness studies show a reduction in cardiovascular events following influenza and pneumococcal vaccination. Still, to date, the research is primarily concentrated in older individuals or individuals with established CVD who are already recommended to receive the vaccines. Current primary cardiovascular prevention strategies do not consider the impact of ARIs. The role of underlying cardiovascular risk profile on ARI-related cardiovascular complications has not been widely investigated. In addition, it is uncertain whether the effectiveness of influenza vaccine against cardiovascular complications varies by cardiovascular risk status. Therefore, identifying groups not currently recommended to receive influenza or other respiratory vaccine but among who targeted vaccination could provide primary cardiovascular prevention has important public health and healthcare benefits.

EHRs provide an efficient, representative and generalisable data source to identify a large study population at risk of ARI-related cardiovascular complications and investigate vaccine effectiveness. Conducting observational studies using longitudinal EHR data to examine the effect of cardiovascular risk on ARI-related cardiovascular complications and prevention of acute cardiovascular events following vaccination will inform future RCTs and cost-effectiveness studies.

2.2. Thesis scope

2.2.1. Original scope

The original aim of my thesis was to focus on the impact cardiovascular risk has on ARI-related (particularly influenza and pneumonia) cardiovascular complications and the cardiovascular benefits

offered by influenza vaccine. To investigate ARI-related cardiovascular complications, two comparative studies in different European settings were planned; one using clinically diagnosed ARIs with available English data sources and the other using laboratory confirmed ARIs with available Danish data sources.

2.2.2. Evolution of thesis scope due to COVID-19

The emergence of COVID-19 in early 2020 and the subsequent pandemic increased attention on cardiovascular complications of severe ARIs and how cardiovascular risk factors, such as hypertension and diabetes, affect the likelihood of complications.

The analysis of laboratory-confirmed ARIs using Danish data, required travel to the host organisation, Statens Serum Institut in Denmark to access the necessary securely held datasets. However, during the pandemic, travel became impossible. Therefore, my supervisors and I decided to substitute the Danish study for a highly relevant COVID-19 project investigating the effect of underlying cardiovascular risk on outcomes associated with SARS-CoV-2 using English data.

2.3. Overarching aim

The overall aim of this research was to quantify the burden of ARI-related cardiovascular complications and to investigate the cardiovascular benefit of influenza vaccine in people with differing levels of underlying cardiovascular risk.

2.4. Individual objectives

I conducted four observational studies using EHR data. In addition, I completed a systematic review on the validity of acute cardiovascular outcomes recorded in EHRs. I have outlined the detail of each research objective in **Table *2.1***.

2.5. Organisation of thesis

The thesis follows the research paper style format, with articles incorporated into the chapters. Seven articles have been included, of which four are published and one is under review after submission for peer-review and two are shortly due to be submitted for peer-review.

The thesis comprises six further chapters, grouped into methods, results and discussion sections.

The methods section (Chapter 3-5) introduces the data sources (Chapter 3) and explains how exposures, outcomes and covariates were defined (Chapter 4). Chapter 4 considers the validity of the data used to define conditions of interest such as ARI. Acute cardiovascular outcomes are presented separately in the systematic review in Chapter 5 (thesis objective 1) which includes my published protocol and systematic review article. The Results section (Chapters 6-9) present the published/drafted research articles (thesis objectives 2-5). Chapter 10, the discussion chapter, summarises the overall findings of the research, considers its overarching strengths and weaknesses, and suggests implications for clinical practice and future research.

Table 2.1 Thesis objectives

Number	Objective	Study design	Population	Data sources	Effect measure
1	To assess the validity of acute cardiovascular diagnoses in routinely collected European EHRs	Systemic review	Adults aged ≥16 years recorded with a diagnosis of stroke, ACS or HF in any primary or secondary care EHR or gold standard data source used to validate the EHR	Any primary or secondary care European EHR published studies	Sensitivity, specificity, PPV or NPV
2	To estimate the effect of cardiovascular risk on: (i) systemic ARI, (ii) acute cardiovascular events, and (iii) acute cardiovascular events after systemic ARI	Cohort study	Adults aged 40-64 years without established CVD or a condition eligible for influenza or pneumococcal vaccination	CPRD linked to HES APC and ONS deaths	(i) and (ii) IRR (iii) HR
3	To investigate whether influenza vaccine reduces the risk of acute cardiovascular events, and if effect differs between individuals with raised and low cardiovascular risk	SCCS study	Adults aged 40-84 years with their first acute cardiovascular event in the same year as influenza vaccine was given		IR
4	To investigate the effect of cardiovascular risk on severe outcomes, including acute cardiovascular events after COVID-19	Cohort study	Adults aged 40-84 years with COVID-19	CPRD linked to HES APC, ONS deaths, SGSS and CHESS	HR
5	To quantify the relative incidence of acute cardiovascular events occurring in periods after COVID-19 to other periods	SCCS study	Adults aged 40-84 years who had COVID-19 and first acute cardiovascular		IR

HR hazard ratio, IR incidence ratio, IRR incidence rate ratio, NPV negative predictive value, PPV positive predictive value

Chapter 3 Description of data sources

3.1. Chapter overview

This chapter provides a detailed explanation of the datasets used to conduct the research presented in this thesis. I used the Clinical Practice Research Datalink (CPRD), Hospital Episode Statistics Admitted Patient Care (HES APC), and the Office for National Statistics (ONS) mortality register in all studies (Chapters 6-8). Additionally, in the COVID-19 study, I used the Second-Generation Surveillance System (SGSS) COVID-19 positive virology results and COVID-19 Hospitalisations in England Surveillance System (CHESS). **Table *3.1*** summarises the key characteristics of each dataset.

3.2. Use of electronic health records in observational research

Electronic health record (EHR) databases are derived from computerised clinical records and often contain longitudinal individual health-related data. Some EHRs are administrative; developed for financial and management purposes to allocate funding or billing of insurance claims (such as HES APC). Other EHRs collate data used to manage routine clinical care (such as CPRD).

The UK has a universal healthcare system, the National Health Service (NHS), free at the point of care for UK residents [195]. Usually, patients initially seek health provision from their primary care General Practitioners (GPs), with the vast majority of the UK population registered with a GP, who are considered the 'gatekeepers' to secondary and specialist care in the UK. This setup provides a rich source of individual-level data collected from GP consultations. After GP referral, most individuals will also use NHS secondary care services rather than private hospitals, so again captured in collated NHS secondary care data.

Table 3.1 Key characteristics of datasets used in this thesis

	Datasets used in all studies			Datasets additionally used in COVID-19 study	
	CPRD [196,197]	HES APC [198]	ONS mortality [199]	SGSS COVID-19 positive virology [200]	CHESS [200]
Type	Longitudinal primary care attendances (i.e. diagnoses, test results, prescriptions, vaccinations)	Diagnosis and procedures carried out in secondary care	Cause of death register	COVID-19 PCR positive results drawn from a larger database of laboratory positive test results	COVID-19 diagnosis in secondary care
Data collection rationale	Track clinical information for the provision of clinical care then collated for research	Monitor workload and care outcomes for payments then also used for research	Monitor death trends	Public health surveillance and linked to CPRD to facilitate COVID-19 research	Monitor COVID-19 hospitalisation and ICU/HDU admission trends and health services utilisation
Population	People registered at participating primary care practices	People with inpatient hospitalisation for any cause	People who die	People who test positive for SARS-CoV-2	People with inpatient hospitalisation for COVID-19
Geography	UK (GOLD) England (Aurum)	England	England and Wales	England	England
Data collection start	2012 (GOLD - earlier iterations 1987) 2017 (Aurum)	1997	1841, recorded with ICD-10 codes since 2001	2020	2020

An NHS number, a unique patient identifier (ID), is assigned to each individual at birth or on first registration (if born outside the UK). A unique ID, such as the NHS number, permits linkage between databases used in different settings, such as primary care, secondary care, or laboratory, to provide extensive, comprehensive, and anonymised datasets for cost-effective observational research with long-term follow-up.

The primary purpose of EHR data collection is for clinical or administrative rather than research purposes. Therefore, investigation of data completeness and accuracy must be considered as part of the research process. Many studies and systematic reviews have previously appraised the validation of specific EHRs or specific conditions recorded in EHRs. I consider the quality of diagnoses recorded in EHRs in Chapters 4 and 5.

3.3. Clinical Practice Research Datalink

The CPRD database formed the primary data source used to conduct my thesis research, and was used to identify the initial study populations for each analysis (Chapters 6-8).

3.3.1. Data collection

First established in 1987, the CPRD (formerly Value-Added Information Medical Products Ltd [VAMP] and General Practice Research Database [GPRD]) is now one of the most extensive primary care EHRs [196]. CPRD is jointly sponsored by the Medicines and Healthcare products Regulatory Agency (MHRA) and the National Institute for Health Research (NIHR). The CPRD resource comprises two databases (GOLD and Aurum) containing anonymised patient records from participating primary care practices. Practices agree to participate in data collection and optional linkage to other data sources (e.g., HES APC), but patients can opt-out. Linked data can be used to validate information recorded in CPRD (i.e.

diagnosis coding), supplement the dataset where information is missing (i.e. ethnicity or death date), or incorporate additional information not captured in CPRD (i.e. hospitalisation for an event).

Every month data are updated in CPRD GOLD and Aurum databases. GOLD was established in its current form in 2012, drawing records from practices that use the Vision clinical management system software. Aurum was rolled out in 2017 with data from practices using the Egton Medical Information Systems (EMIS) clinical management system [196,197]. In recent years the number of primary care practices using Vision has declined as practices switch to EMIS or an alternate clinical management system, TPP SystmOne, leading to a decrease in the size of GOLD and an increase in Aurum. As of November 2021, CPRD Aurum contained 40,265,295 patients with 37,566,005 eligible for linkage to HES and other datasets, and 13,283,115 currently registered and CPRD GOLD contained 20,620,714 patients with 9,270,111 eligible for linkage to HES and other datasets, and 2,591,752 currently registered [201]. GOLD draws data from GP practices in all four nations of the UK (England, Wales, Scotland, and Northern Ireland), and Aurum draws data from England. Aurum was new and not widely utilised at the start of my PhD; therefore, I used GOLD and Aurum in my first study to ensure both produced consistent results. For all remaining analyses, I only used Aurum.

The CPRD databases collect basic demographics, symptoms, diagnoses, prescriptions, tests, immunisations, and secondary or specialist care referrals. Outcomes from secondary and specialist care are also fed back to GPs for recording in their clinical management systems. The databases contain some lifestyle and anthropometric measures, including smoking status and BMI. CPRD bases socioeconomic status (SES) on the patient's residence or practice location. Specifically, in this thesis, I used the 2011 Townsend Deprivation Index to measure SES [202]. This Index is a measure of deprivation first introduced by Peter Townsend in 1987. A score based on a combination of four census variables (unemployment, non-car ownership, non-home ownership, and household overcrowding) for any census collected geographical area is calculated and used to assign a summary quintile, decile or twentile.

3.3.2. Data structure

Both CPRD Gold and Aurum split data into several files sorted by information type, with some differences between the two database structures. **Table 3.2** outlines each of the files from the databases and their key content.

CPRD data are free text and codes. Free text is not routinely provided to researchers to protect patient confidentially. GOLD and Aurum databases use different terminology for medical and non-medical codes. Diagnoses are recorded in GOLD with the Read version 2 (v2) hierarchy and in Aurum with Read v2, Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) and local EMIS codes. Read terminology was developed in the 1980s and adopted throughout the UK in the early 1990s with further iterations until the Read system's retirement in 2020 [203]. Read CTV3 hierarchy is organised into chapters (i.e., symptoms, examinations, administrative items and diagnoses) and subchapters. The initial values in a code represent high-level categories, and the following values specify further event details. Starting in 2018, NHS England switched from Read to SNOMED CT. SNOMED is an internationally recognised terminology used in more than 50 countries grouped into concepts [204]. Prescriptions are recorded in GOLD with British National Formulary (BNF) codes and in Aurum with the Dictionary of Medicines and Devices (dm+d) codes, which is part of the SNOMED CT terminology structure [197]. CPRD generate medical and product IDs for each Read CVT3 / SNOMED CT code and BNF / dm+d code, respectively.

Table 3.2 Data files from Clinical Practice Research Datalink (CPRD) Gold and Aurum

Gold files [196]		Aurum files [197]	
Name	Summary of content	Name	Summary of content
Patient	Patient-level data include: - Demographics: year of birth and sex - Major event dates (registration, death, transfer out) - Registration status - Practice ID	Patient	Patient-level data include: - Demographics: year of birth and sex - Major event dates (registration, death, transfer out) - Registration status - Practice ID
Practice	Practice-level data include: - Geographical region - Last data collection date - 'Up to standard' date (date from which practice data is of sufficient quality for research)	Practice	Practice-level data include: - Geographical region - Last data collection date
Consultation	Patient-level data include: - Date of consultation - Type of consultation	Consultation	Patient-level data include: - Date of consultation - Type of consultation
Clinical	Patient-level clinical events include: - Date of event - Symptoms experienced - Diagnosis given - Ethnicity	Observation	Patient-level medical history include: - Date of event - Symptoms experienced - Diagnosis given - Test conducted - Test result - Immunisation given - Ethnicity
Immunisation	Patient-level immunisation records include: - Immunisation type - Date of immunisation		
Test	Patient-level results include: - Test conducted - Result of test - Date of test		
Additional clinical details	Patient-level additional detail related to clinical events include: - Measurement taken - Results of measurement Examples would be blood pressure, weight or smoking habit. The file is split into “entity types”, which relates to a specific type of data. There are a total of 460 different entity types. For example, “entity type” one records information on blood pressure.	Problem	Patient-level ‘add-on’ data linked to observation file include: - Medical condition defined as being a problem - Problem significance - Problem likely duration

Referral	Patient-level referrals to secondary and specialist care include: - Date of referral - Diagnosis given - Method of referral - Referral speciality - Urgency of referral	Referral	Patient-level ‘add-on’ data linked to observation file for referrals to secondary and specialist care include: - Date of referral - Diagnosis given - Method of referral - Referral speciality - Urgency of referral
Therapy	Patient-level data on drug prescriptions and apparatus include: - Date prescription issued - BNF code - Quantity provided	Drug issue	Patient-level data on drug prescriptions and apparatus include: - Date prescription issued - BNF code - Quantity provided

Note: CPRD Gold and Aurum also both contain a file “Staff” not used in the studies reported in this thesis

3.3.3. Internal data quality

CPRD check the data received by practice and for patient records. Each practice is classified as up-to-standard (UTS) from the latest continuous data collection period of recording continuity with no temporal gaps (defined as ≥5 weeks with low event recording compared to the practice average) and when the mortality rate falls within an expected range (showing no record deletion). UTS, along with the patient's registration date, determine the start of usable follow-up for each patient [196,197]. Of note, UTS has yet to be implemented for CPRD Aurum.

At a patient level, data are subject to several consistency checks to flag the individual's entry as acceptable (passed all checks) or unacceptable (failed at least one check). The patient checks are valid gender and birth date, logically consistent and valid registration date, permanent registration at the practice, transferred out of practice reason and date must both be missing or completed. Only patients classified as acceptable are used in research.

3.4. Hospital Episode Statistics Admitted Patient Care

In addition to CPRD, I used linked data from the Hospital Episode Statistics Admitted Patient Care (HES APC) dataset for the analyses presented in this thesis.

3.4.1. Data collection

The HES database, established in 1989, includes separate datasets for APC (inpatient admissions), outpatients, accident and emergency, diagnostic imaging, and patient-reported outcomes at all NHS hospitals in England [205]. Coded hospital attendance data are inputted by trained clinical coders using patient records and discharge summaries. In turn, the care providers complete monthly submissions of the coded attendances to NHS Digital, where the data are processed as part of the Secondary Uses Service

(SUS) dataset for payment [205,206]. The Payment by Results scheme sees reimbursement to care providers based on activity level; this drives data entry and submission to SUS [198]. Several times a year, SUS consolidate data submissions to compile and update HES [207].

Eligible patient records in CPRD are linked to HES for data from 1997 onwards using an eight-step deterministic matching process based on combinations of NHS number, gender, date of birth and postcode [208].

3.4.2. Data structure

HES APC data are composed of admission episodes grouped by hospitalisation (also referred to as spells) for a single stay in hospital from admission to discharge [198,206]. Admission episodes are started and ended based on the period of care under a particular clinical team within the hospital. Each episode records a primary diagnosis and up to 20 further secondary diagnoses. Annually, between 16-17 million admission episodes are recorded in HES APC [209], although this number dropped to 12 million in 2020-21 due to reduced hospital activity and data collection using the COVID-19 pandemic [210].

HES APC diagnoses are coded using the International Classification of Disease 10th Revision (ICD-10). ICD-10 is organised into chapters by condition. **Table *3.3*** summarises the HES files I used in this thesis research, along with the key content from each file.

3.4.3. Internal data quality

HES data are released with accompanying data quality notes. The notes highlight any specific known issues within the dataset which should be considered when using the data for research [207]. One measure of HES data quality is coverage. By NHS provider, SUS reports the number of suspected missing records which is calculated as the difference between the expected and actual number of records submitted by a

specific provider. The expected number of records is estimated from on the average monthly submission of stable (no dramatic increases or decreases) periods. Additionally, SUS provides a field completeness report which reports the completion of key field by each NHS provider. The field completeness report does not consider the validity of the values submitted, but rather that key field recording is not missing. In HES APC, key fields include the primary diagnostic code and the ward specialty.

Table 3.3 Data files from the Hospital Episode Statistics Admitted Patient Care

File	Selected contents used in thesis analyses
Patient	- Patient identifier - Ethnicity
Diagnoses by episode	- Spell number (constant for all episodes in a single hospitalisation) - Episode key (identifier for each episode in a hospitalisation) - Episode start date - Episode end date - ICD-10 code - ICD-10 code position (ordering of diagnosis codes from 1-20)
Diagnoses by hospitalisation	- Spell number - Admission date - Discharge date

Note: HES APC also contains files “Procedures”, “Augmented Care”, “Critical Care”, “Maternity” and “Health Resource Group” not used in the studies reported in this thesis

3.5. The Office for National Statistics mortality data

Linkage eligible patients captured in CPRD are also linked to ONS mortality data.

3.5.1. Data collection

Mortality data contain information recorded when deaths are certified and registered [211]. Most deaths are certified by a medical practitioner, using the Medical Certificate of Cause of Death (MCCD). The certificate is taken to a registrar by an informant (i.e., a near relative of the deceased). Some deaths (i.e., unnatural or suspicious deaths) are referred to, and sometimes then investigated by, a coroner. The coroner sends information to the registrar, which is then used to register the death rather than the MCCD.

In the ONS mortality dataset, the MCCD data are largely automatically (about 20% requiring manual coding) using computer algorithms to convert text terms to corresponding ICD-10 codes. Coroner reported cause of deaths is done manually by experienced coders, as the software cannot code the free text format used by coroners [211]. ONS mortality data record date of death is recorded along with the primary cause of death and additional (up to 15) contributory causes. ONS mortality data capture all deaths registered in England and Wales [199].

3.5.2. Internal data quality

Registrars enter death registration data into an online system with some automatic validation checks [212]. When the data are uploaded to the ONS database, a series of further validation processes are used to identify inconsistencies (i.e., that the conditions on the death certificate are compatible with the sex and age of the deceased), compare cross-field data and plausibility.

3.6. Second Generation Surveillance System COVID-19 positive virology results

SGSS is used to capture routine laboratory data on infectious diseases and antimicrobial resistance from NHS and Public Health England (PHE) diagnostic laboratories in England [213]. Positive test results for a

range of notifiable organisms (including influenza virus, SARS-CoV-2 and *S. pneumoniae*) must be submitted by each laboratory to PHE through SGSS. SGSS data are collated and stored centrally within PHE.

A subset of SGSS data with positive SARS-CoV-2 polymerase chain reaction (PCR) results (with the specimen date) are linked to CPRD data to aid COVID-19 research [200]. The dataset includes results from COVID-19 pillar 1 and pillar 2 testing. Pillar 1 testing captured individuals with a clinical need to be swabbed focused on hospitalised individuals and healthcare workers. Pillar 2 included more comprehensive population testing conducted in community settings [214], with pillar 2 rolled out later when testing capacity in the UK was increased [215]. Linkage is solely based on NHS number due to the accelerated roll-out of the linked data for urgent research during the pandemic to inform patient care. The dataset includes patient and practice IDs, specimen ID, specimen date, laboratory report date, and if the patient is a care home resident.

There are known quality issues within the CPRD SGSS data. There are duplicates within the dataset with minor differences between one or two variables. CPRD report that the duplicates are due to the resubmission of updated records, mainly related to whether the sample originates from a care home. However, it is not possible to determine the order of record submission within the dataset. I required the earliest sample for my analysis, so duplicates were ignored, with only the earliest sample retained for analysis. Early in the pandemic, most SARS-CoV-2 testing was conducted in London, with other geographies testing and reporting later; this has caused some geographical bias in early data.

3.7. COVID-19 Hospitalisations in England Surveillance System

PHE established CHESS across all NHS hospitals in England at the start of the COVID-19 pandemic to collect epidemiological data on individuals with laboratory-confirmed COVID-19 who required hospitalisation. The data also identify individuals admitted to a high dependency unit (HDU) or intensive

care unit (ICU) [216]. The dataset comprises more than 90 variables, including patient and practice IDs, laboratory sampling details (with species identification for influenza and RSV in addition to SARS-CoV-2), symptom onset, date of hospital admission, comorbid conditions, type (if any) of respiratory support given, HDU/ICU admission with date, antiviral treatment and outcome details [200]. As with SGSS, CHESS linkage to CPRD data uses NHS number. The dataset captures 109 of 152 NHS Trusts in England.

3.8. Strengths and limitations of included data sources

Table *3.4* outlines the overall strengths and limitations of data sources used for my thesis research. A more detailed assessment of the data source strengths and limitations in relation to my thesis research questions is included in the Chapter 10 (Discussion).

3.8.1. Data quality considerations

Clinical practice and recording patterns may change over time, impacting trends in data and service use. In 2004 a new payment-for-performance scheme, the Quality and Outcomes Framework (QOF), was introduced to UK primary care [217,218]. QOF aimed to secure higher-quality primary care by offering financial incentives to GPs for the achievement of specific indicators [219]. Indicators have been modified and added to over time, with current indicator examples including the percentage of patients aged ≥40 years with a blood pressure measurement recorded in the preceding five years or the percentage of patients with IHD who have had an influenza vaccine [220]. Therefore, QOF encouraged GP electronic recording of certain clinical conditions (with the specific clinical codes to be used updated over time) to demonstrate the achievement of pay-for-performance indicators [218]. In turn, the quality of the data collected through CPRD improved after the introduction of QOF.

Overall, the diagnoses recorded in CPRD and HES APC are extensively validated. The validity of conditions in EHRs is quantified using diagnostic accuracy test measures, comparing what is recorded in the EHR data to a recognized reference "gold" standard [221]. Gold standards are various but include manual clinician medical record review, machine learning algorithms, another database, internal text validation. The measures are the positive predictive value (PPV), the negative predictive value (NPV), the sensitivity and specificity. The PPV is the proportion of individuals with the condition in the EHR data who truly have the condition (in the reference standard). The NPV is the proportion of individuals without the condition in the EHR who truly did not have the condition. Sensitivity is the proportion of all individuals with a condition that the EHR data correctly identified. The specificity is the proportion of individuals without a condition that the EHR data correctly identified.

Several systematic reviews have evaluated the overall quality of CPRD and HES [222–224]. I consider the quality of acute respiratory infections recording in the relevant section in Chapter 4 and acute cardiovascular events in Chapter 5.

Table 3.4 General strengths and limitations of the data sources used in thesis research

Database	Strengths	Limitations
CPRD	- Size: As of November 2021, GOLD contains records from 20,620,714 patients, of which 2,591,752 were currently registered and Aurum contains records 40,265,295 patients, of which 13,283,115 were currently registered [201] - Age, sex, and ethnically representative of the UK population with coverage for all regions of England [196,197] - Internal data quality checks [196,197] and validated diagnoses [222,223] - Longitudinal, prospective, and complete data - Linkage to multiple other data sources [200]	- Missing data: QOF has helped improve recording of many health conditions, but some particularly lifestyle or anthropometric measures, are still poorly recorded [196]. Additionally, these measures are not missing at random, with better recording for patients at risk, for example blood pressure is more likely to be recorded for women of reproductive age and those with established CVD [196]. - Variation in coding practices: there are many codes, particularly SNOMED CT, for similar, or indeed the same, condition [197]
HES APC	- Size: HES captures data from all NHS secondary care providers so collates information on >16 million in-patient episodes per year [209] - Standardised coding: international coding system, ICD-10, allows for international comparison	- Missing data: ethnicity not well recorded, in comparison to age and sex [198] - Linkage: linkage between admissions within the dataset as well as to other data sources such as CPRD, highly rely on the accuracy of NHS number [198]
ONS	- Coverage: the data provides almost complete population coverage for deaths that occur in England and Wales, because death registration is a legal requirement	- Inconsistencies: because of registration delays, deaths by occurrence date can be incomplete and deaths by registration date may include those that occurred months or even years earlier [212]
COVID-19 SGSS	- Case identification: provides a method to identify individuals with laboratory-confirmed COVID-19	- Coverage: limited by testing policies early in the pandemic
CHESS	- Case identification: provides a method to identify individuals hospitalised for COVID-19	- Missing data: as this is a newly formed dataset, there is a large amount of missing data

3.9. Chapter summary

- The CPRD dataset was the primary data source used in all analyses I present in this thesis, with linked data obtained from HES APC and ONS mortality datasets. For COVID-19 research, I also used SGSS and CHESS datasets supplied by CPRD.
- CPRD GOLD and Aurum contains over 35 million individuals and is longitudinal, providing a robust resource for epidemiological analysis.
- Data in CPRD GOLD and Aurum comes directly from UK (Aurum only drawn from England) primary care practices using Vision and EMIS clinical management systems, respectively, and coded using clinical terminologies, Read and SNOMED CT.
- Aurum was new and not widely used by the LSHTM EHR Research Group when starting my PhD research. Therefore, I used both GOLD and Aurum for study 1. After ensuring results were similar across the two data sources, I only used Aurum (the larger dataset) for all remaining analyses.
- The CPRD datasets are the real-time, with monthly updates, availability of routine patient data which is broadly representative of the UK population in terms of age, sex and ethnicity.
- CPRD data is linked to other data sources. In my thesis research, I have used HES APC to identify exposures and outcomes of interest, ONS deaths, and COVID-19 data sources (SGSS and CHESS).

Chapter 4 Definitions

4.1. Principles of variable identification in electronic health record databases

EHR data analysis relies on codelists to identify variables of interest. A codelist refers to all codes indicating the patient characteristic, condition or drug prescribed. Researchers develop codelists using dictionaries that contain all codes collected within a database, such as the CPRD medical and product dictionaries or ICD-10.

Figure *4.1* outlines my decision-making process for using and creating codelists for CPRD GOLD and HES. I utilised existing codelists if they matched my requirements for a diagnosis or treatment. I sought existing codelists from other LSHTM EHR group projects (either from the online Data Compass system https://datacompass.lshtm.ac.uk/ or direct contact with group members) or non-LSHTM online repositories. I made changes to existing codelists if necessary. The first reason for making changes to existing codelists was my use of a narrower definition to increase specificity. I excluded codes based on discussions with my clinically trained supervisors (Dr Charlotte Warren-Gash, Dr Ami Banerjee and Professor Liam Smeeth). The other reason for making changes was if an older version of the CPRD database had been used to create the original codelist, so my updates included any newly added codes. I made codelist updates using the same method as new codelist creation, which I set out below.

If there was no existing codelist or the codelist did not match my requirements, I created a codelist from scratch. I searched the relevant database dictionary using a Stata do file following the core principles set out by Davé and Peterson [225] to document inclusion and exclusion criteria decisions and allow replication. An example do file is available in **Chapter 11 Appendix 1**.

In general, the steps I took to create a diagnosis codelist with the CPRD GOLD medical dictionary and HES ICD-10 dictionary were:

- Develop a list of keyword search terms (done in consultation with supervisors Dr Charlotte Warren-Gash, Dr Ami Banerjee or Professor Liam Smeeth) that identify the condition of interest and flag these in the dictionary
- Add a flag for any extra codes known to be of relevance but not picked up by search terms
- Query any uncertain terms with my supervisors
- Remove the flag from terms outside of my definition

The steps I took to create a treatment codelist with the CPRD GOLD product dictionary were:

- Develop a list of relevant BNF codes and flag these in the dictionary
- Develop a list of drug names for a keyword search (to include terms that do not have a BNF code) and flag these in the dictionary
- Remove the flag for terms that were for formulation or route of administration outside of the definition

For CPRD Aurum, existing codelists were unavailable for most conditions due to limited prior use of the new database. The steps I took to create a diagnosis codelist with the CPRD Aurum medical dictionary were:

- Merge condition GOLD codelist using Read code common to both GOLD and Aurum medical dictionaries and flag
- Run the same keyword search terms used in GOLD and flag in the dictionary
- Remove the flag from terms outside of the definition
- Identify terms that match flagged SNOMED CT concept ID
- Remove flag for matching SNOMED CT concept ID from terms outside of the definition

The steps I took to create a treatment codelist were:

- Run the same keyword search used in GOLD and flag in the dictionary

- Remove flag for terms which were for formulation or route of administration which were outside of definition

CPRD release regular new builds of Gold and Aurum data. When I developed my codelists, the latest version of Gold released to LSHTM was July 2019 and of Aurum was March 2020 – I used these builds to create or update the majority of codelists. My code lists were all published on the LSHTM Data Compass, the links to the codelists used in each project are included in the relevant chapters of this thesis.

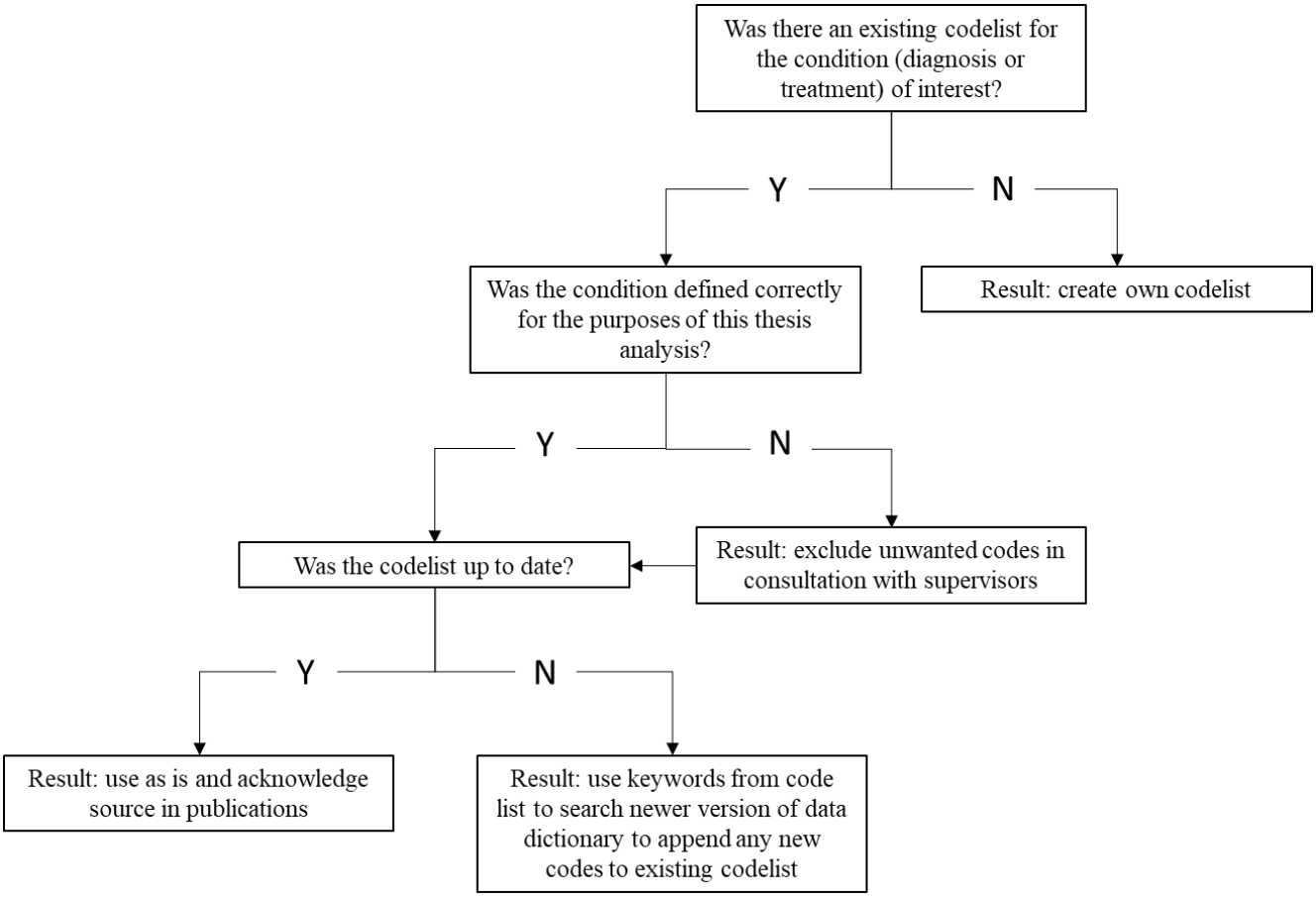


Figure 4.1 Process to create codelists

4.2. Identifying acute respiratory infections

4.2.1. Validity of recording in electronic health records

Most ARI diagnoses in primary and secondary care EHR data come from clinical judgement, i.e., based on physical signs and symptoms, without laboratory confirmation in the UK, particularly pre-COVID-19. Clinically diagnosed influenza is poorly defined [226] but syndromic definitions of influenza are often employed in public health surveillance. The European Centre for Disease Prevention and Control (ECDC) defines influenza-like illness (ILI) as the sudden onset of symptoms which include at least one of fever, malaise, headache or myalgia and at least one of cough, sore throat or shortness of breath [227]. Similarly, ECDC define ARI as the sudden onset of symptoms which include at least one cough, sore throat, shortness of breath or coryza.

Recording preferences can also influence the identification of ARIs in EHR data. An analysis of ILI recording in THIN data (another primary care EHR like CPRD) found that between 1995-2013 ILI consultations decreased over time until a spike during and following the 2009/10 H1N1 pandemic [228]. Over the time that ILI recorded consultation decreased, recording of cough and fever symptoms increased. The authors suggested that these changes reflected changes in GP recording preferences. ARI diagnoses have not been validated in CPRD or HES.

4.2.2. Definition for thesis

My definition of ARI aimed to capture infections that affected the lower respiratory tract or resulted in systemic illness. There is no standard definition for LRTIs, but many studies will include pneumonia, influenza, and bronchitis [229,230]. I, therefore, defined ARI as a clinical or confirmed diagnosis of; pneumonia, acute bronchitis, influenza, ILI, or other acute infections suggestive of lower respiratory tract involvement. Using this definition, I aimed to capture more severe infections resulting in healthcare attendance, which could plausibly induce atherosclerotic processes and result in a systemic complication

such as acute cardiovascular events. Although some ARIs will be mild and still captured in the data where individuals have a high propensity for seeking healthcare or attend for other coinciding condition.

I used both CPRD and HES to identify ARIs. My definition corresponded to a CPRD GOLD ARI codelist previously developed by my supervisor Dr Charlotte Warren-Gash [164]. I used the keywords from the existing list to run a new search in GOLD and translate the list to Aurum as set out above in Section 4.1. I manually reviewed ICD-10 chapters A and B for "Certain infectious and parasitic diseases" and J for "Diseases of the respiratory system" to identify codes that I should include, along with conducting a keyword search. I flagged influenza/ILI and pneumonia codes for separate analyses of these specific infections. ARI was both an outcome of interest and index date in study 1 (Chapter 6) and a covariate in study 2 of this thesis (Chapter 7).

4.3. Identifying influenza vaccination

Most seasonal influenza vaccines are issued in primary care, so recorded in CPRD. A small proportion of vaccines will be given by occupational health services or in pharmacies (paid vaccines), although GP practices can also record these in a patient's records where the patient makes them aware of the vaccination.

In CPRD, there are several ways to identify influenza vaccination status. In the GOLD therapy and Aurum drug issue files, automatic record generation occurs when a prescription is issued. Therefore, like other CPRD analyses, I assumed influenza vaccine product codes equated to a patient being administered the vaccine. The GOLD immunisation file records immunisation type (i.e., which vaccine) and status (i.e., given, declined). There are several code "types" in the immunisation file which correspond to the influenza vaccine. The corresponding status options are; not recorded, given, refused or advised. Records with the status of "advised" were not included. Medical codes from GOLD clinical and referral and Aurum observations files can also signify vaccination. Many of the Read / SNOMED CT codes encode

influenza vaccine terms that do not expressly state receipt of the vaccine, while some relate to the patient's consent or declining the vaccine. I classified the Read / SNOMED CT influenza vaccine terms and the corresponding codes as given, neutral or declined.

I used a combination of the three (two for Aurum) CPRD sources to assign influenza vaccine status. **Table *4.1*** outlines the method I used, and **Table *4.2*** shows how I dealt with individuals who had records for multiple statuses from the same date. In summary, when I identified a patient who had a record of being given the vaccine on the same date as a record that suggested the patient declined the vaccine, I assumed the patient had the vaccine. Using the initial CPRD GOLD and Aurum study population for my first study on cardiovascular complications of ARI, I investigated conflicts between patients recorded as given and declined influenza vaccine on the same date. Only 0.04% (257/635,598) of patients had such a conflict; this suggests my results would not be altered by assigning these patients as vaccinated. If records suggested it was unclear (neutral) if the patient had the vaccine on the same date as a record to show the patient declined the vaccine, I treated this individual as unvaccinated. If I could only assign a neutral status to the patient on a given date, I assumed the individual was vaccinated.

Table 4.1 CPRD data used to identify influenza vaccine administered

Status	**Method used to derive status**
Given	-relevant code in therapy / drug issue file -relevant type in immunisation file with status of given (GOLD only) -relevant code in clinical and referral / observation file which clearly states the vaccine was given
Declined	-relevant type in immunisation file (GOLD only) with status of refused -relevant code in clinical and referral / observational file which clearly states the vaccine was declined
Neutral	-relevant type in immunisation file (GOLD only) with no assigned status -relevant code in clinical and referral / observational file which either does not status a specific action or related to consent

Table 4.2 Process to handling status combinations on the same date

Record status	Given	Decline	Neutral
Given	Vaccinated	Vaccinated	Vaccinated
Decline	Vaccinated	Not vaccinated	Not vaccinated
Neutral	Vaccinated	Not vaccinated	Vaccinated

I obtained influenza vaccine codelists for Read / SNOMED CT codes, prescription codes, and relevant immunisation types from Helen McDonald (member of the LSHTM EHR Research Group and one of my PhD advisors).

I used influenza vaccine status as an exclusion criterion in study 1 of this thesis (Chapter 6) and as the exposure of interest in study 2 (Chapter 7).

4.4. Identifying COVID-19

I identified SARS-CoV-2 infections using SGSS and CHESS data. Individuals in either dataset had PCR-confirmed SARS-CoV-2. I also developed a codelist of clinically reported COVID-19 infections for CPRD Aurum (CPRD GOLD was not used for this study). The OpenSafely (a secure analytics platform for EHR data created to deliver urgent research during the COVID-19 pandemic) team developed several Read codelists for COVID-19 diagnosis ,which are available at https://www.opencodelists.org/. I used the OpenSafely codelists to build the keyword searches for CPRD Aurum. Codes U07.1 and U07.2 identified individuals with COVID-19 recorded in HES APC record.

COVID-19 was the exposure of interest in the final study of this thesis (Chapters 8 and 9).

4.5. Identifying established cardiovascular disease

I defined established CVD as a clinical diagnosis, major intervention or clinical reviews specific to heart disease (congenital or otherwise), heart failure, stroke or transient ischaemic attack. I excluded individuals with established CVD from all analyses, other than the descriptive summary of the incidence of COVID-19 and COVID-19 adverse outcomes (see Chapter 8 for further details).

In all studies CVD was used as an exclusion criterion. Although, in the COVID-19 incidence analysis (Chapter 8), CVD was included as a stratifying factor.

4.6. Measuring cardiovascular risk

Throughout my PhD research, I have used two measures of cardiovascular risk; hypertension and QRISK2/3 scores. In study 1 and study 3 cardiovascular risk was the exposure of interest, in study 2 and 4 it was an effect modifier.

4.6.1. Hypertension

Most hypertension diagnoses will take place in primary care with multiple BP readings needed to make a formal diagnosis. Guidance from the NICE recommends BP should be measured in both arms and repeated when the readings have a difference of >15mmHg, and where a difference persists to measure subsequent BP in the arm which had the highest reading [231]. Furthermore, the guidance recommends that if BP is measured at ≥140/90 mmHg then a second measurement should be taken during the consultation. Where the second measurement is substantially different to the first then a third should be obtained and the lowest of the second and third measurement recorded. When the recorded reading is

between 140/90 mmHg and 180/120 mmHg, NICE recommends ambulatory BP monitoring (ABPM) or, if ABPM is not suitable, home BP monitoring (HBPM) to confirm the diagnosis of hypertension.

A validation study using data from The Health Improvement Network (THIN), another primary care EHR, found hypertension prevalence was similar to estimates from the Health Survey for England when both coded diagnoses and BP readings identified hypertension [232]. Coded diagnoses alone provided an underestimate of prevalence and the inclusion of antihypertensive prescriptions provided an overestimate. However, in the non-COVID-19 analysis, I defined hypertension based on only coded diagnoses. Formal diagnosis of hypertension, as per NICE guidelines, should be a multi-step process and not only a result of a high BP reading. Additionally, coded diagnoses, unlikely BP readings, offer a practical means by which GPs can identify a risk group of patients. For example, if individuals with hypertension are at increased risk of ARI-related cardiovascular complications and targeted for influenza or pneumococcal vaccination, implementation of this approach in practice would be based on coded diagnoses. In the COVID-19 analysis, I additionally used BP readings to classify hypertension. Given the rapid onset of the pandemic, individuals undergoing hypertension diagnosis may not have a coded event yet. Additionally, the COVID-19 analysis was exploratory, considering whether cardiovascular risk was associated with severe outcomes rather than a purposed method to identify a population for public health action, i.e., influenza vaccination. I did not use antihypertensive prescriptions to identify hypertension status but did consider it in stratified analyses.

4.6.2. QRISK

Hypertension is one of the primary risk factors for future CVD; however, it is only one element of risk. QRISK uses many variables (risk factors) to assess an individual's absolute ten-year risk of CVD. QRISK was developed in 2007 using UK data from the QResearch database [233]. QResearch is a primary care

EHR, and like CPRD Aurum, derived from practices using the EMIS clinical management system. The authors designed QRISK to improve cardiovascular risk prediction within the UK population.

At the time of QRISK development, risk prediction in the UK, like most high-income countries, relied on the USA-developed Framingham score. The Framingham risk algorithm development predominantly used a white population. The authors of QRISK queried the use of the Framingham score in an ethnically diverse population. QRISK included measures of ethnicity and deprivation, not included in the Framingham score, based on local geographical census information. Comparison of the scores on the QResearch database found the Framingham algorithm overpredicted 10-year CVD risk by 35%, while the QRISK algorithm only overpredicted by 0.4% [233]. QRISK authors further validated their algorithm in THIN, producing similar results [234].

Despite the two validation studies conducted by QRISK authors, NICE continued to recommend the Framingham score for CVD risk assessment with adjustment for family history of CVD and ethnicity [235]. An external validation of the Framingham and QRISK scores using THIN data found that although QRISK underpredicted 10-year CVD risk, the underestimation (12%) was smaller than the Framingham score's overprediction (23%) [236].

QRISK2 was released in 2008 to incorporate new risk factors; type 2 diabetes, treated hypertension, rheumatoid arthritis, CKD stage 4 and 5, and atrial fibrillation [70]. Additionally, ethnicity was updated to self-reported ethnicity as recorded in the primary care data. The authors recalibrated the QRISK2 algorithm annually using the most up to data from the QResearch database. Further changes between 2008 and 2017 updated the diabetes definition to include type 1 diabetes as another variable, separated non-smokers from ex-smokers, and extended the age range covered by the score from 35-74 years to 25-84 years [71]. A recent Scottish validation study of cardiovascular risk prediction scores in people with type 2 diabetes reported QRISK2 overpredicted risk of cardiovascular events, with the median risk of 23.5% compared with an observed risk of 9.3% [237].

In 2017, the authors released QRISK3 with additional risk factors used to calculate scores [71]. A recent assessment of QRISK3 reliability found that the score performed well for population risk prediction but that there was greater uncertainty for individual prediction [238].

Table *4.3* outlines the variables included in QRISK2 and QRISK3. The improvements in the QRISK2/3 algorithms led to NICE endorsement. NICE recommend that all adults under 85 years have their CVD risk assessed using QRISK2/3 for the primary prevention of CVD [239], with statin prescription considered in those with a score of ≥10% [90]. QRISK2, and now QRISK3, was also embedded into GP clinical management systems for direct use during consultations.

The QRISK authors have never published the complete algorithms used to calculate scores, although they did share the algorithms with the external validators [236]. Using the information which the authors have published (summarised in Table 4.3), members of the LSHTM EHR group (Sarah Gadd and Emily Herrett) wrote Stata program files to calculate QRISK2 (2015 algorithm) scores for CPRD GOLD recorded patients. Following the release of QRISK3, the Stata program files were updated (by Helen Strongman and Emily Herrett) to calculate scores for the QRISK3 (2017 algorithm). I rewrote the Stata program files to run on CPRD Aurum data. In summary, the program files use codelists for each risk factor included (see **Table *4.3***) to identify a baseline (index date) score for each patient and updated scores for when a patient has a new risk factor recorded.

For non-COVID-19 analyses, I used QRISK2, which would have been the version in use during the time period analysed. Additionally, when I conducted the analysis for the first study, QRISK3 was largely unevaluated and not widely used in research yet. QRISK3 was used in the COVID-19 analysis as this was the score in use during 2020 – the time period analysed.

4.7. Potential explanatory variables

4.7.1. Age and sex

Patient files in CPRD GOLD and Aurum contain the year of birth and sex. I calculated age using the year of birth, with all patients assigned a nominal birthday of 1 July (mid-point of the year). Unless otherwise specified, I categorised age into five-year bands for analysis.

Table 4.3 QRISK variable definitions and source notes

Variable	Algorithm definition and source notes
QRISK2 and QRISK3 variables	
Age	As with all CPRD analyses, calculated from the date of birth assigned as 1 July in the individuals' year of birth, rounds age at the index to the nearest integer, and is updated each year in time updated QRISK score.
Sex	From CPRD patient file.
Townsend score	CPRD supplies Twentiles of Townsend used to assign proxy scores for the individual’s level of deprivation.
Ethnicity	Relevant Read or SNOMED CT codes recorded in the GOLD Clinical or Aurum Observation files, respectively. If two records are recorded on the same date, the program chooses the highest value (justification was that many of the discrepancies included white and another ethnic group, and therefore most likely not white).
Smoking status	Retrieves all smoking records identified by QOF Read and SNOMED CT codes in GOLD Clinical or Aurum Observation files, respectively, and, if applicable, the associated smoking status value, i.e. the number of cigarettes per day. Before the index date, the most recent record is used to define smoking status at the index date. This is updated in the time updated QRISK score each time smoking status is recorded.
Diabetes	Uses Read and SNOMED CT QOF codes from GOLD Clinical and Referral or Aurum Observational files. Categorised as type 1 or type 2. Patients with both type 1 and type 2 diabetes records are classified according to the nearest record prior to index.
Atrial fibrillation	Uses Read and SNOMED CT QOF codes from GOLD Clinical and Referral or Observational Aurum files.
Rheumatoid arthritis	Uses Read and SNOMED CT QOF codes from GOLD Clinical and Referral and Observational Aurum files.

Variable	Algorithm definition and source notes
Chronic kidney disease stage 4-5	Uses all relevant (CKD, dialysis and renal transmplant) Read and SNOMED CT codes from GOLD Clinical and Referral or Aurum Observational files.
Height & weight	Uses Read and SNOMED CT QOF codes to identify BMI records in the GOLD Clinical or Aurum Observational files, and links to value attached to record. The most recent record prior to index date is used to define BMI at the index date. This is updated in the time updated QRISK score each time BMI is recorded.
Treated hypertension	Uses Read and SNOMED CT QOF codes from GOLD Clinical and Referral or Aurum Observational files or current prescriptions for BP medications in the GOLD Therapy and Aurum Drug Issue files. A current prescription is defined as ongoing at index date, i.e., prescription date + number of days' supply >indexdate.
TC:HDL ratio	Uses codes from the GOLD Test or Observational Aurum files indicating TC:HDL, and links to value attached to record. Applies the cut-off specified in program syntax (set to 11 in my analyses). The most recent record prior to index date is used to define the ratio at the index date. This is updated in the time updated QRISK score each time TC:HDL is recorded.
Family History of CHD in relatives aged <60	Uses all relevant Read or SNOMED CT codes recorded in the GOLD Clinical or Observation Aurum files.
Systolic BP	Takes all values for records of systolic BP. The most recent record to index date is used to derive systolic BP at index date. This is updated in the time updated QRISK score each time systolic BP is recorded.
QRISK3 only variables	
Chronic kidney disease stage 3	As above for CKD stage 4-5 but additionally includes records which signify stage 3.
Measure of systolic BP variability	Standard deviation of all (minimum two) systolic BP values recorded in the five years before the index date or update date.
Migraine	Uses all relevant Read or SNOMED CT codes recorded in the GOLD Clinical and Referral or Aurum Observation files.
Corticosteroid use	At least two prescriptions for corticosteroid medications in the GOLD Therapy or Aurum Drug Issue files prior to index/update date, with the most recent prescription ≤28 days before the date.
Systemic lupus erythematosus	Uses all relevant Read or SNOMED CT codes recorded in the GOLD clinical and Referral or Aurum Observation files.
Second generation antipsychotic use	At least two prescriptions for relevant antipsychotic medications in the GOLD Therapy or Aurum Drug Issue files prior to index/update date, with the most recent prescription ≤28 days before the date.
Severe mental illness (SMI)	Uses all relevant (psychosis, schizophrenia, or bipolar affective disease) Read and SNOMED CT codes from GOLD Clinical and Referral or Aurum Observational files.
HIV or AIDS	Uses all relevant Read and SNOMED CT codes from GOLD Clinical and Referral or Observational Aurum files.

Variable	Algorithm definition and source notes
Erectile dysfunction	Uses all relevant Read and SNOMED CT codes from GOLD Clinical and Referral or Aurum Observation files or any prescription for relevant medication before index / update date.

Adapted from LSHTM EHR group QRISK standard operating procedure written by Helen Strongman and Emily Herrett

4.7.2. Ethnicity

Read and SNOMED CT codes (CPRD GOLD and Aurum, respectively) encode ethnic groups in CPRD. Ethnicity recording in primary care was boosted between 2006-2011 when completion was financially incentivised by QOF, with reporting in CPRD GOLD jumping from <30% in 2005 to over 70% in 2006 [240]. Ethnicity is also captured in the HES APC, completeness of which has been >80% since 2006 [240]. A single cleaned ethnicity is provided in the HES APC patient file.

Rohini Mathur from the LSHTM EHR Research Group developed codelists and Stata programs to identify ethnicity in CPRD GOLD and Aurum to ensure consistent assignment in CPRD studies conducted by the LSHTM EHR Research Group. The program assigns ethnicity based on the following:

1. Patients with a single usable ethnicity record in CPRD are assigned this ethnicity,
2. Patients with multiple ethnicity records in CPRD are assigned the most frequently recorded ethnicity. In the event of a tie, the program uses the most recently recorded frequent ethnicity,
3. If a patient has no recorded ethnicity in CPRD, ethnicity is taken from HES.

4.7.3. Socio-economic status

I used CPRD provided patient-level Townsend twentiles to assign socio-economic status. This classification of deprivation is used in the calculation of QRISK scores. Peter Townsend developed the

Townsend deprivation index in 1987. The score uses four indicators of; unemployment, household overcrowding, non-car ownership, and non-home ownership [241].

4.7.4. Comorbid health conditions

Health conditions were defined using relevant codelists. We assume that the absence of a code reflects the absence of a condition. In addition, as diagnosis codes alone underestimate CKD, estimated glomerular filtration rate (eGFR) based on serum creatinine using the CKD-EPI equation is routinely used to classify CKD [242]. Helen McDonald, Kate Mansfield and Angel Wong from the LSHTM EHR Research Group have written Stata programs for CPRD GOLD (Helen McDonald and Kate Mansfield) and Aurum (Angel Wong) to automate the calculation of eGFR and assign CKD stage.

4.8. Chapter summary

- EHR data analysis relies on codelists to identify variables of interest. A codelist refers to all codes indicating the patient characteristic, condition or drug prescribed. In CPRD the medical and product dictionaries are used to develop codelists.
- EHR analysis must consider data quality. Clinical practice and recording patterns change over time, which can impact trends in the data. Generally, data quality in CPRD has improved over time.
- Most ARI diagnoses in EHR data is based on clinical diagnosis without laboratory confirmation. ARI diagnoses have not been validated in CPRD or HES.
- I included CPRD and HES recorded ARIs which affected the lower respiratory tract or result in systemic illness. My definition of ARI comprised clinical or confirmed diagnosis of; pneumonia, acute bronchitis, influenza, ILI, or other acute infections suggestive of lower respiratory tract involvement.
- I used two measures of cardiovascular risk; hypertension and QRISK2/3 scores.
- In non-COVID-19 analysis, I defined hypertension based on only coded diagnoses. Coded diagnoses are a practical way primary care practices could identify a target risk population, such as those to be offered influenza vaccine.
- In the COVID-19 analysis, I used coded diagnoses and BP readings to classify hypertension.
- QRISK2/3 use many variables (risk factors) to assess an individual's absolute ten-year risk of CVD.

Chapter 5 Review on the validity of acute cardiovascular outcome diagnosis

5.1. Chapter overview

This chapter presents the systematic review I completed to investigate the validity of ACS, heart failure and stroke diagnoses in European EHRs. I published the review protocol in BMJ Open and full review in Clinical Epidemiology. Separate from the publication, in this chapter, I summarise how the systematic review findings informed my use of EHR data to identify acute cardiovascular outcomes in the original research I conducted.

5.2. Published protocol

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	230019	**Title**	Ms
First Name(s)	Jennifer		
Surname/Family Name	Davidson		
Thesis Title	Acute respiratory infections, cardiovascular complications, and prevention among people with raised cardiovascular risk		
Primary Supervisor	Dr Charlotte Warren-Gash		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	BMJ Open		
When was the work published?	October 2019		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	Not applicable		
Have you retained the copyright for the work?*	**Yes**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I am the first author of this research paper. With support of my co-authors, I designed my systematic review and drafted the protocol for publication.

SECTION E

Student Signature	[redacted]
Date	25/07/2022

Supervisor Signature	[redacted]
Date	25-07-2022

BMJ Open

Validity of acute cardiovascular outcome diagnoses in European electronic health records: a systematic review protocol

Jennifer Anne Davidson,[1] Amitava Banerjee ,[2] Rutendo Muzambi ,[1] Liam Smeeth,[1] Charlotte Warren-Gash[1]

To cite: Davidson JA, Banerjee A, Muzambi R, *et al*. Validity of acute cardiovascular outcome diagnoses in European electronic health records: a systematic review protocol. *BMJ Open* 2019;**9**:e031373. doi:10.1136/bmjopen-2019-031373

► 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-2019-031373).

Received 30 April 2019
Revised 21 August 2019
Accepted 30 September 2019

[1]Faculty of Epidemiology & Population Health, London School of Hygiene and Tropical Medicine, London, UK
[2]Farr Institute of Health Informatics Research, University College London, London, UK

Correspondence to
Jennifer Anne Davidson;
jennifer.davidson@lshtm.ac.uk

ABSTRACT

Introduction Cardiovascular diseases (CVDs) are among the leading causes of death globally. Electronic health records (EHRs) provide a rich data source for research on CVD risk factors, treatments and outcomes. Researchers must be confident in the validity of diagnoses in EHRs, particularly when diagnosis definitions and use of EHRs change over time. Our systematic review provides an up-to-date appraisal of the validity of stroke, acute coronary syndrome (ACS) and heart failure (HF) diagnoses in European primary and secondary care EHRs.

Methods and analysis We will systematically review the published and grey literature to identify studies validating diagnoses of stroke, ACS and HF in European EHRs. MEDLINE, EMBASE, SCOPUS, Web of Science, Cochrane Library, OpenGrey and EThOS will be searched from the dates of inception to April 2019. A prespecified search strategy of subject headings and free-text terms in the title and abstract will be used. Two reviewers will independently screen titles and abstracts to identify eligible studies, followed by full-text review. We require studies to compare clinical codes with a suitable reference standard. Additionally, at least one validation measure (sensitivity, specificity, positive predictive value or negative predictive value) or raw data, for the calculation of a validation measure, is necessary. We will then extract data from the eligible studies using standardised tables and assess risk of bias in individual studies using the Quality Assessment of Diagnostic Accuracy Studies 2 tool. Data will be synthesised into a narrative format and heterogeneity assessed. Meta-analysis will be considered when a sufficient number of homogeneous studies are available. The overall quality of evidence will be assessed using the Grading of Recommendations, Assessment, Development and Evaluation tool.

Ethics and dissemination This is a systematic review, so it does not require ethical approval. Our results will be submitted for peer-review publication.

PROSPERO registration number CRD42019123898

Strengths and limitations of this study

- This systematic review will comprehensively evaluate the validity of selected major cardiovascular diagnoses (stroke, acute coronary syndrome and heart failure) in electronic health record (EHR) databases used in the provision of primary and secondary clinical care in Europe by searching five bibliographic databases and two grey literature sources with no language or date restrictions.
- The Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement will be followed ensuring this systematic review provides high-quality scientific results.
- There may be heterogeneity in the results produced by our systematic review due to differences in EHR design and use between countries, in particular the relevance of our findings to countries outside of Europe requires further evaluation.

INTRODUCTION

Rationale

Stroke, acute coronary syndrome (ACS) and heart failure (HF) are the three cardiovascular (CV) conditions, which substantially contribute to morbidity and mortality. Ischaemic heart disease followed by stroke have been the global leading causes of death for 15 years, and in 2016 accounted for 15.2 million deaths.[1] Also in 2016, worldwide more than 13 million people were estimated to have suffered a stroke[2] with healthcare expenditure on stroke estimated to be 3%–5%.[3–5] An estimated 26 million people are living with HF,[6] a chronic condition with acute episodes. HF is estimated to account for 1%–2% of healthcare expenditure in Europe and the USA.[7] Added to the complication of estimating the burden of CV conditions is changes to definitions; the fourth universal definition of myocardial infarction (MI) was issued in 2018.[8]

Increases in the incidence and prevalence of CV conditions are in part due to an ageing population,[9] but also due to modifiable risk factors, such as smoking, unhealthful diet and lack of physical exercise, and non-modifiable risk factors, such as sex and ethnicity.[10] A

range of factors, such as pollution, infections, emotional stress and physical exertion, can also trigger acute CV events particularly in those with pre-existing cardiovascular disease (CVD).[11–13]

Electronic health record (EHR) databases are derived from clinical care records and contain longitudinal patient data on diagnoses, treatment and other clinically relevant variables, such as smoking. Administrative databases were developed for financial and management purposes to allocate funding or billing of insurance claims. While both are types of computerised health-related data that have been widely used for research, they are quite distinct. In particular, the completeness and accuracy of the morbidity data may differ in the two types of data because of the very different reasons why the data were recorded in the first place. In settings where both clinical and administrative data are available, results from some studies suggest the quality of administrative data is lower.[14 15]

High-quality EHR-based research depends on correct classification of cases and non-cases. Several systematic reviews have previously appraised the validation of specific European EHRs[16–18] as well as specific conditions recorded within EHRs, including CVD.[19–23] The previous systematic reviews on the validity of CVD diagnoses included EHRs, along with administrative databases and vital registration databases. McCormick *et al* reported that the positive predictive value (PPV) of stroke diagnosis ranged from 32% to 98%, with the majority of included studies using administrative data from North America,[22] while Woodfield *et al* identified PPVs of >70% for stroke based on the results that included a greater proportion of studies from Europe,[20] where EHRs are widely used. McCormick *et al*'s review of HF diagnosis validity obtained PPVs ranging from 17% to 100% but only contained four studies outside of North America.[23]

Aim and objectives

The aim of our systematic review is to provide an up-to-date appraisal of the validity of stroke (and its subtypes), ACS (including MI and other ACS) and HF diagnoses in adults focused on European EHRs used in primary and secondary care. Our objectives are to:

1. Summarise and pool estimates of the sensitivity, specificity, PPV and negative predictive value (NPV) of stroke, ACS and HF diagnoses compared with a suitable reference standard.
2. Determine whether estimates differ by study population, validation method, data source, diagnosis and time period.

METHODS

This protocol has been prepared using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols guidelines.[24]

Eligibility criteria

We used the PICOS (Population, Intervention, Comparator, Outcomes and Study design) framework to formulate the research question and eligibility criteria for our review, but adapted this to replace 'Intervention' with 'Index test', the modification recommended for systematic reviews of diagnostic test accuracy.[25] This modification was chosen as it is the closest resemblance to the validation of EHRs.

Population

Eligible studies will include records of adults aged 16 years or older from any European primary or secondary care national or the regional EHR database. We will exclude studies that validate administrative (insurance claims or billing) databases, disease registries or vital registration systems, as well as studies that validate locally held databases. EHRs and administrative databases collect data for different purposes and may differ in accuracy, we are interested in the validity of EHRs used in clinical settings. The comprehensive data capture methods used to populate disease registries mean that these datasets are often used as the gold standard in validation of EHRs so would be unsuitable to include in our validation estimates. Vital registration systems only capture deaths so unless combined with EHRs, the data is not by itself useful in non-mortality-related research. Finally, data from locally held databases are unlikely to be captured in centralised EHRs used in research and therefore validation results are not informative for researchers.

Index test

We are interested in records with clinical codes, for example, International Classification of Primary Care (ICPC) or International Classifications of Disease (ICD), which identify a diagnosis of stroke (and its subtypes), ACS (including MI and other ACS) or HF in primary or secondary care EHRs. The ICD-9 and ICD-10 codes, we assume studies, will include (and which we look to validate) in their stroke, ACS and HF definitions are presented in table 1.

Comparator

To be included, studies must have validated against an internal or external reference standard. Eligible external reference standards include manual review of medical records, patient or clinical questionnaire, or comparison with an independent second database. Internal within database comparison includes validation against a

Table 1 Provisional list of ICD-9 and ICD-10 codes included in diagnoses of interest

Diagnosis	ICD-9	ICD-10
Stroke	430, 431, 432, 433, 434	I60, I61, I62, I63, I64
Acute coronary syndrome	410, 411	I20.0, I21, I22, I24, I49
Heart failure	428	I11.0, I13.0, I13.2, I50

ICD, International Classifications of Disease.

diagnosis algorithm or comparison of clinical codes with anonymised free text.

Outcome

Studies must either report (1) at least one of the following validation estimates; sensitivity, specificity, PPV and NPV or (2) data which allows at least validation estimates to be calculated.

Study design

We will include any type of study from any time period published in any language that includes the validation of the recording of stroke, ACS or HF diagnoses in an EHR database, regardless of if this was the main objective of the study.

Information sources

To review published and in-process citations the following databases will be searched from inception to April 2019; MEDLINE, EMBASE, SCOPUS, Web of Science and Cochrane Library. Using OpenGrey and EThOS, we will search for the relevant grey literature. Bibliographies of national EHR databases used for research will also be searched.

Search strategy

The search strategy will include subject heading terms and free text (title and abstract) for the concept of acute CV events using the synonyms of stroke, ACS and HF as well as the concepts of EHRs and validation. We will limit our search to studies conducted using European EHRs. Provisional search terms have been developed for MEDLINE (online supplementary appendix 1), and once finalised will be transcribed into corresponding searches for the other aforementioned information sources. We will also review the reference list of other relevant systematic reviews identified during the screening process as well as of articles included in our review to identify further potentially relevant studies.

Study records

Data management

Citations from the searched databases will be exported into Endnote X9. Electronic deduplication of records will be conducted, followed by manual deduplication where necessary.

Selection process

For the initial screening stage, two authors (JAD and RM) will independently review all titles and abstracts to assess whether they fulfil the eligibility criteria for inclusion. To reduce the risk of missing potentially relevant studies, reviewers will adopt a lenient approach for this first level of screening including any study that validate stroke, ACS or HF diagnoses in EHRs. Full-text articles for studies that meet the review criteria will be obtained and reviewed by the two authors (JAD and RM). The reasons for rejection of articles during the full-text screening process will be noted according to a hierarchical list: (1) could not obtain full text, (2) did not conduct validation, (3) duplicate study, (4) wrong outcome, (5) wrong index (ie, not a primary or secondary care EHR in Europe), (6) not a suitable comparator or (7) no validation estimate or insufficient data to calculate. Any discrepancies at either the initial screening or full-text screening will be discussed by the two reviewers, with a third author (CW-G) consulted when necessary.

Data collection process

To extract information for each study selected for final inclusion, data extraction tables will be piloted by the two authors (JAD and RM) for three studies with changes made, if required. We will then dual extract data from a further 10% of studies using the finalised template. If there are any significant discrepancies between the two reviewers, then we will conduct parallel data extraction for a further 10% of studies, again checking for discrepancies. This process will be repeated until no further discrepancies occur, at which stage the remaining data extraction will be completed by the single reviewer (JAD). At each stage, the third author (CW-G) will be consulted when the two reviewers cannot resolve discrepancies.

Data items

Similar to our search strategy, we will use the PICOS framework to systematise the extraction of data from each study. We will use a standardised template containing information on each of the following five domains:

1. Population: participants, age and sex, inclusion and exclusion criteria.
2. Index test: EHR country, EHR name, EHR setting (primary or secondary care), EHR coding system (ICPC, ICD, etc), EHR coverage (regional or national), diagnoses validated including whether incident or prevalent, specific diagnoses codes validated.
3. Comparator: method of validation, description of method.
4. Outcome: number of participant diagnoses planned, number of diagnoses conducted, measures of validity and raw data to calculate measures of validity.
5. Study characteristics: authors, publication year, language, study design, study period, main aim of the study (validation or not validation).

Outcomes and prioritisation

The outcome is any validation estimate of stroke (including all subtypes), ACS (MI or other ACS) or HF. The study has no secondary outcomes.

Risk of bias in individual studies

To assess bias, we will use a tailored version of the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool, which is used for assessing diagnostic accuracy studies,[26] based on the previous modifications made for assessing the validity of diagnostic coding in EHRs.[20 21] We will consider bias in each of the domains included in QUADAS-2; patient selection, index test, reference standard and flow and timing. In the context of our review,

index test translates to the clinical codes validated. We will produce a summary risk of bias figure, as well as an additional table explaining each judgement.

Two authors (JAD and RM) will independently pilot the tailored QUADAS-2 tool, assessing bias in three of the included studies. Any necessary changes will be made to the tool and dual assessment by the two reviewers will be done with the finalised tool for a further 10% of studies. If there are significant discrepancies, we will continue parallel risk of bias assessment for another 10% of studies, repeating the process until no further discrepancies occur. Assessment of the remaining studies will be completed by the single reviewer (JAD). At each stage, the third author (CW-G) will be consulted when the two reviewers cannot resolve discrepancies

Data synthesis and metabias(es)

We will describe key study characteristics and use a narrative synthesis and forest plots to summarise the validity of each of stroke, ACS and HF diagnoses in European primary and secondary care EHRs. The I^2 statistic will our guide judgements about the level of statistical heterogeneity between the studies. We will use the Cochrane's suggested guide to grade the heterogeneity as a low (0%–40%), moderate (30%–60%), substantial (50%–90%) or considerable (75%–100%) obtained from the I^2 statistic.[27] If there is the sufficient number of studies selected, we will explore the reasons for heterogeneity. We will compare heterogeneity before and after removing the studies that deemed to be at a high risk of bias overall and by subgroups of: (1) study populations, that is, specific demographic or clinical groups, (2) validation method, (3) data source, that is, primary care and secondary care EHRs, (4) specific diagnosis, that is, incident or prevalent and stroke or ACS subtype and (5) variation in validity estimates over time.

We will consider conducting meta-analyses for each CV condition to calculate pooled effect estimates for sensitivity, specificity, PPV and NPV if studies are sufficiently homogeneous. Meta-analyses would be conducted by the aforementioned subgroups. Our choice of a fixed or random effects model would also be guided by the level of heterogeneity, with random effects meta-analysis methods followed if there is substantial heterogeneity.

Confidence in cumulative evidence

Two reviewers (JAD and RM) will independently use the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) tool[28] to judge the certainty of cross-study evidence for the validity of diagnoses in EHRs and their use in research. Any discrepancies between the two reviewers' judgements will be discussed and resolved, if necessary consulting the third author (CW-G). We will examine stroke, MI and HF diagnoses in EHRs for; overall risks of bias, inconsistency, indirectness, imprecision and publication bias with the production of funnel plots. The strength of evidence will be categorised as high, moderate, low and very low. Our judgements will be presented in a summary of findings table.

Patient and public involvement

Patients and/or public were not involved in this systematic review.

DISCUSSION

This systematic review will provide an up-to-date assessment of the validity of primary and secondary care EHRs used for stroke, ACS and HF research. To our knowledge, this will be the first systematic review to focus solely on the validity of CVD diagnoses in EHRs. Previous systematic reviews have included EHRs along with administrative databases and vital registration databases. Each of these data sources has a different primary purpose, which in turn will impact the validity of the systems. One previous systematic review of MI diagnoses identified the accuracy for vital registration databases was lower (all PPVs≤59%) than the hospitalisation data (three-quarters of studies PPV >59%).[21] However, hospitalisation grouped EHRs and administrative databases together, so it is unclear if the one-quarter of studies with a PPV ≤59% differed by data source to those with higher PPVs.

Our systematic review will also serve to update several aspects covered by the previous systematic reviews validating CVD diagnoses. McCormick *et al's* 2010 review of MI diagnoses only identified three studies that validated the ICD-10 coding.[21] while Rubbo *et al's* 2014 review identified eight studies.[19] The majority of European countries implemented ICD-10 in the late 1990s. Our search run in April 2019 aims to identify more recent studies validating ICD-10 CVD diagnoses, the results of which are most relevant to today's research. In the majority of studies included in the previous systematic review of HF diagnoses, conducted in 2010, sensitivity was <69%.[23] Only one of the three included European studies reported sensitivity, this was 43%.[29] With an increase in the prevalence of HF,[6] and therefore accompanying public health research, we hypothesise that more studies validating HF diagnoses in EHRs will have been published between 2010 and 2019, the results of which will inform current HF research. Similarly, previous systematic reviews validating stroke diagnoses identified variation in accuracy by stroke subtype,[20 22] with the inclusion of up-to-date studies, we aim to analyse temporal changes in validity estimates with the assumption that more recent studies should have higher and more consistent estimates across stroke subtype. We also aim to present results for ACS other than MI, such as unstable angina which have not been included in any previous systematic review.

Our systematic review benefits from searching multiple databases with no language barriers, compared with the previous systematic reviews of CVD diagnoses, which either only searched 1–2 databases or only included English language studies. There are some limitations to our systematic review. First, by aiming to validate EHRs,

rather than broader computerised health-related datasets, we have limited our review to Europe where EHRs operate nationally or covering nationally representative populations with widespread use in research. Consequently, our findings will not be applicable to administrative databases, also commonly used in research. Second, our results will not necessarily be applicable to countries outside of Europe using EHRs, if the design and utility of the EHRs differ. Lastly, previous systematic reviews on CVD diagnoses have been unable to conduct meta-analyses due to the level of heterogeneity identified in their results. We hope that by limiting our systematic review to EHRs in Europe, many of which are set up and operated in similar ways, the level of heterogeneity between the studies will be reduced. However, we will still be limited by variation in the reference standard used and differences in the codes included in validation. Therefore, it may not be possible to conduct any meta-analysis.

Overall, our systematic review should provide useful and up-to-date findings to inform researchers on the validity of using EHRs in their research.

ETHICS AND DISSEMINATION

Important protocol amendments will be documented and a justification for deviating from the original protocol provided in a protocol addendum. The findings of this review will be submitted to a peer-reviewed journal.

Twitter Amitava Banerjee @amibanerjee1

Contributors CW-G conceived the study idea. JAD led the design of the study with the contributions from AB, LS and CW-G. JAD drafted the methods and analysis and revised the protocol following authors' comments from CW-G, AB, RM and LS. All authors approved the final version of the protocol.

Funding JAD was supported by a British Heart Foundation PhD Studentship (FS/18/71/33938).

Disclaimer The funders have no input on the protocol development and will not have influence on the conduct, analysis, interpretation or publication of the study results.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval Ethical review is not required as this study is a systematic review.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iDs
Amitava Banerjee http://orcid.org/0000-0001-8741-3411
Rutendo Muzambi http://orcid.org/0000-0003-0732-131X

REFERENCES

1 World Health Organization. The top 10 causes of death. Available: https://www.who.int/en/news-room/fact-sheets/detail/the-top-10-causes-of-death
2 GBD 2016 Stroke Collaborators. Global, regional, and national burden of stroke, 1990-2016: a systematic analysis for the global burden of disease study 2016. *Lancet Neurol* 2019;18:439–58.
3 Evers SMAA, Struijs JN, Ament AJHA, *et al*. International comparison of stroke cost studies. *Stroke* 2004;35:1209–15.
4 Saka O, McGuire A, Wolfe C. Cost of stroke in the United Kingdom. *Age Ageing* 2008;38:27–32.
5 Chevreul K, Durand-Zaleski I, Gouépo A, *et al*. Cost of stroke in France. *Eur J Neurol* 2013;20:1094–100.
6 Savarese G, Lund LH. Global public health burden of heart failure. *Card Fail Rev* 2017;03:7–11.
7 Ambrosy AP, Fonarow GC, Butler J, *et al*. The global health and economic burden of hospitalizations for heart failure: lessons learned from hospitalized heart failure registries. *J Am Coll Cardiol* 2014;63:1123–33.
8 Thygesen K, Alpert JS, Jaffe AS, *et al*. Fourth universal definition of myocardial infarction (2018). *Eur Heart J* 2019;40:237–69.
9 Izzo C, Carrizzo A, Alfano A, *et al*. The impact of aging on cardio and cerebrovascular diseases. *Int J Mol Sci* 2018;19:481.
10 World Health Organization. *Cardiovascular diseases*. World Health Organization, 2018. https://www.who.int/cardiovascular_diseases/en/
11 Strike PC, Steptoe A. Behavioral and emotional triggers of acute coronary syndromes: a systematic review and critique. *Psychosom Med* 2005;67:179–86.
12 Mustafić H, Jabre P, Caussin C, *et al*. Main air pollutants and myocardial infarction. *JAMA* 2012;307.
13. Estabragh ZR, Mamas MA. The cardiovascular manifestations of influenza: a systematic review. *Int J Cardiol* 2013;167:2397–403.
14 Laws MB, Michaud J, Shield R, *et al*. Comparison of electronic health Record-Based and Claims-Based diabetes care quality measures: causes of discrepancies. *Health Serv Res* 2018;53 Suppl 1:2988–3006.
15 Shahian DM, Silverstein T, Lovett AF, *et al*. Comparison of clinical and administrative data sources for hospital coronary artery bypass graft surgery report cards. *Circulation* 2007;115:1518–27.
16 Herrett E, Thomas SL, Schoonen WM, *et al*. Validation and validity of diagnoses in the general practice research database: a systematic review. *Br J Clin Pharmacol* 2010;69:4–14.
17 Sund R. Quality of the Finnish hospital discharge register: a systematic review. *Scand J Public Health* 2012;40:505–15.
18 Ludvigsson JF, Andersson E, Ekbom A, *et al*. External review and validation of the Swedish national inpatient register. *BMC Public Health* 2011;11:450.
19 Rubbo B, Fitzpatrick NK, Denaxas S, *et al*. Use of electronic health records to ascertain, validate and phenotype acute myocardial infarction: a systematic review and recommendations. *Int J Cardiol* 2015;187:705–11.
20 Woodfield R, Grant I, Sudlow CLM, *et al*. Accuracy of electronic health record data for identifying stroke cases in large-scale epidemiological studies: a systematic review from the UK Biobank stroke outcomes group. *PLoS One* 2015;10:e0140533.
21 McCormick N, Lacaille D, Bhole V, *et al*. Validity of myocardial infarction diagnoses in administrative databases: a systematic review. *PLoS One* 2014;9:e92286.
22 McCormick N, Bhole V, Lacaille D, *et al*. Validity of diagnostic codes for acute stroke in administrative databases: a systematic review. *PLoS One* 2015;10:e0135834.
23 McCormick N, Lacaille D, Bhole V, *et al*. Validity of heart failure diagnoses in administrative databases: a systematic review and meta-analysis. *PLoS One* 2014;9:e104519.
24 Moher D, Shamseer L, Clarke M, *et al*. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev* 2015;4:1.
25 Centre for Reviews and Dissemination University of York. *Systematic Reviews: CRD's guidance for undertaking reviews in health care*, 2009.
26 Whiting PF, Rutjes AWS, Westwood ME, *et al*. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529.
27 The Cochrane Collaboration. Cochrane Handbook for Systematic Reviews of Interventions [Internet]. Version 5. *Higgins JP, Green S, editors* 2011.
28 Atkins D, Best D, Briss PA, *et al*. Grading quality of evidence and strength of recommendations. *BMJ* 2004;328:1490.
29 Merry AHH, Boer JMA, Schouten LJ, *et al*. Validity of coronary heart diseases and heart failure based on hospital discharge and mortality data in the Netherlands using the cardiovascular registry Maastricht cohort study. *Eur J Epidemiol* 2009;24:237–47.

Please note, the published appendix for the protocol is not included in this thesis as it contains the same information as the article's published appendix.

5.3. Published paper

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	230019	**Title**	Ms
First Name(s)	Jennifer		
Surname/Family Name	Davidson		
Thesis Title	Acute respiratory infections, cardiovascular complications, and prevention among people with raised cardiovascular risk		
Primary Supervisor	Dr Charlotte Warren-Gash		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	Clinical Epidemiology		
When was the work published?	October 2020		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	Not applicable		
Have you retained the copyright for the work?*	**Yes**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I am the first author of this research paper. With support of my co-authors, I designed my systematic review, completed study identification, collated and summarised results and drafted the review for publication.

SECTION E

Student Signature	[redacted]
Date	25/07/2022

Supervisor Signature	[redacted]
Date	25-07-2022

Clinical Epidemiology

Dovepress
open access to scientific and medical research

Open Access Full Text Article

REVIEW

Validity of Acute Cardiovascular Outcome Diagnoses Recorded in European Electronic Health Records: A Systematic Review

This article was published in the following Dove Press journal:
Clinical Epidemiology

Jennifer Davidson[1]
Amitava Banerjee[2]
Rutendo Muzambi[1]
Liam Smeeth[1]
Charlotte Warren-Gash[1]

[1]Faculty of Epidemiology & Population Health, London School of Hygiene and Tropical Medicine, London, UK; [2]Institute of Health Informatics, University College London, London, UK

Background: Electronic health records are widely used in cardiovascular disease research. We appraised the validity of stroke, acute coronary syndrome and heart failure diagnoses in studies conducted using European electronic health records.

Methods: Using a prespecified strategy, we systematically searched seven databases from dates of inception to April 2019. Two reviewers independently completed study selection, followed by partial parallel data extraction and risk of bias assessment. Sensitivity, specificity, positive predictive value (PPV), and negative predictive value estimates were narratively synthesized and heterogeneity between sensitivity and PPV estimates were assessed using I^2.

Results: We identified 81 studies, of which 20 validated heart failure diagnoses, 31 validated acute coronary syndrome diagnoses with 29 specifically recording estimates for myocardial infarction, and 41 validated stroke diagnoses. Few studies reported specificity or negative predictive value estimates. Sensitivity was ≤66% in all but one heart failure study, ≥80% for 91% of myocardial infarction studies, and ≥70% for 73% of stroke studies. PPV was ≥80% in 74% of heart failure, 88% of myocardial infarction, and 70% of stroke studies. PPV by stroke subtype was variable, at ≥80% for 80% of ischaemic stroke but only 44% of haemorrhagic stroke. There was considerable heterogeneity (I^2 >75%) between sensitivity and PPV estimates for all diagnoses.

Conclusion: Overall, European electronic health record stroke, acute coronary syndrome and heart failure diagnoses are accurate for use in research, although validity estimates for heart failure and individual stroke subtypes were lower. Where possible, researchers should validate data before use or carefully interpret the results of previous validation studies for their own study purposes.

Keywords: validation, myocardial infarction, heart failure, stroke; routinely collected health data

Introduction

Ischaemic heart disease and cerebrovascular disease have been the leading causes of death globally for more than 15 years.[1] In Europe, cardiovascular disease (CVD) deaths and prevalence have decreased but remain substantial; in 2015 an estimated 85 million people had CVD including 11.3 million with new diagnoses.[2]

CVD determinants and outcomes research increasingly utilize electronic health records (EHRs). EHRs contain comprehensive longitudinal health data, extracted from primary and secondary care clinical systems, for large patient populations which provide cost-effective data for research. EHR data is mostly "structured"

Correspondence: Jennifer Davidson
Email Jennifer.Davidson@lshtm.ac.uk

submit your manuscript | www.dovepress.com
DovePress
http://doi.org/10.2147/CLEP.S265619

with diagnoses coded using, for example, the International Classification of Diseases (ICD) but can also be "unstructured" with anonymized free-text notes.[3] EHR-based research predominantly uses structured data. As the primary purpose of EHR data collection is clinical, it is essential to consider the validity of the data's use in research.

EHR use is widespread in Europe, where many countries have national healthcare systems, and several systematic reviews have previously explored the quality of specific European EHRs.[4–7] Other systematic reviews[8–12] have investigated the validity of CVD diagnoses in computerized health-related records, which included EHRs but mainly drew results from disparate claims-based systems. The previous reviews did not separate results for EHR and claims data, the quality of which may differ due to the differences in setup and collection rationale.

In our systematic review, we provide an up-to-date assessment of the validity of acute CVD diagnoses recorded in European EHRs. We defined acute CVD as heart failure (HF), acute coronary syndrome (ACS), and stroke. These high-burden conditions are key diagnoses commonly included in the composite endpoint of major adverse cardiovascular events (MACE) which is increasingly employed in both clinical trials and observational research studies.[13] We investigated whether the validity of these diagnoses differed by subtype, definition, data source, reference standard, and study population.

Methods

Protocol and Registration

Our protocol was published in October 2019[14] following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocol guidelines (PROSPERO registration number CRD42019123898).

Eligibility Criteria

We included articles that validated diagnoses in patients aged ≥16 years captured in any European primary or secondary care EHR. We excluded claims-based databases, disease registries, vital registration systems, or locally held databases. Articles needed to validate clinical codes for the diagnoses of HF, ACS, or stroke (Table 1) against a suitable internal or external reference standard. HF is most frequently a chronic condition which can deteriorate with acute exacerbations. HF may also have an acute onset, for example after an MI. The European Society of Cardiology (ESC) defines acute HF as rapid onset or worsening of symptoms and/or signs of existing HF.[15] ACS encompasses different clinical forms of myocardial ischaemia which includes myocardial infarction (MI) and unstable angina. The specific diagnosis of MI or unstable angina depends on symptoms, signs, biomarkers, and ECG and/or autopsy findings, with the definitions refined over time.[16] The diagnosis of stroke includes subtypes ischaemic stroke, intracerebral haemorrhage (ICH), and subarachnoid haemorrhage (SAH).[17] At least one validation estimate (Figure 1) or the raw data to calculate it was required.

Information Sources

We searched for eligible articles in five databases (Medline, Embase, Scopus, Web of Science, and Cochrane Library), two grey literature sources (OpenGrey and Ethos), and, where available, the bibliographies of EHR databases from the date of inception to April 2019 in any language.

Search Strategy

We searched medical subject heading terms and free-text (in the title and abstract) for the concepts of (1) CVD

Table 1 Example Clinical Codes Included for Stroke, Acute Coronary Syndrome and Heart Failure Diagnosis Definitions

Diagnosis	Subtype	ICD-10	ICD-9	ICPC
Acute coronary syndrome	Myocardial infarction	I21	410	K75
	Unstable angina	I20.0		
	Cardiac arrest	I46		
	Other acute heart disease	I24	411	
Heart failure		I50	428	K77
Stroke	Subarachnoid haemorrhage	I60	430	K90
	Intracerebral haemorrhage	I61	431, 432	
	Cerebral infarction	I63	433, 434	
	Non-specific stroke	I64	436	

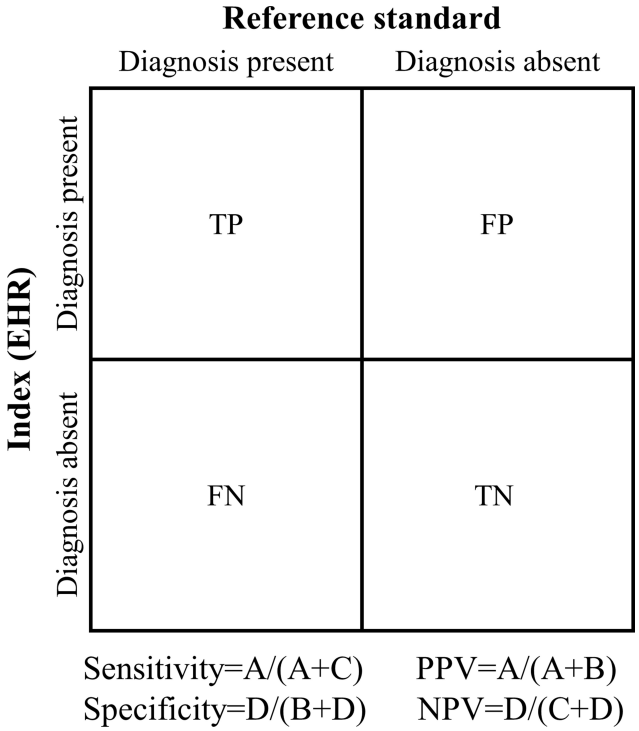


Figure 1 Illustration of validity estimates calculations.

diagnoses, (2) EHRs, (3) Europe, and (4) validation. Search terms were developed for Medline and transcribed for the remaining databases (S1 Appendix). To identify any additional articles, we checked reference lists of eligible articles and relevant systematic reviews.

Study Selection and Data Collection

Two reviewers (J.A.D. and R.M.) independently screened the titles and abstracts of all retrieved articles, followed by the full-text of articles deemed eligible in the first stage. Our published protocol details the full data collection process.[14] Briefly, we extracted data using a pre-defined template (S2 Appendix) which we piloted using dual extraction for three studies, followed by further parallel extraction for 20% of studies, and completed by a single reviewer (J.A.D.) for the remaining studies.

Risk of Bias in Individual Studies

We used a modified version of the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2)[18] tool to assess bias (S3 Appendix). As with our data extraction, two authors (J.A.D. and R.M.) piloted the tool for three studies, then independently assessed risk in a further 10% of studies, with the process completed by a single reviewer (J.A.D.).

Synthesis of Results

We synthesized results with a narrative approach, grouping studies by acute CVD diagnosis (HF, ACS or stroke) and, where possible, subgroups of interest. Subgroups were; diagnosis type, definition, data source including diagnostic position and coding system, reference standard, and study population including time period, age and sex. For studies that reported validation estimates without confidence intervals (CIs), but included raw data, we calculated 95% CIs using the Wilson method for binomial proportions. We used the I^2 statistic to assess heterogeneity between the sensitivity and positive predictive value (PPV) estimates, following the Cochrane thresholds.[19] Heterogeneity assessment did not include specificity or negative predictive value (NPV), as few studies reported these measures. To investigate sources of heterogeneity, we compared I^2 before and after removing studies at a high risk of bias and by the previously mentioned subgroups. We used the Stata metaprop command[20] to calculate I^2. Metaprop uses raw data rather than precalculated estimates; studies that reported sensitivity or PPV but not the data used to calculate were excluded from heterogeneity assessment.

Risk of Bias Across Studies

We used the Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) tool for diagnostic accuracy systematic reviews[21] to summarise cross-study quality. Evidence was categorised as "high", "moderate", "low" or "very low" quality. See S4 Appendix for the reasons we rated quality down or up.

Results

Studies Included

We identified 4595 studies, of which 218 were included in full-text review and 81 met eligibility criteria (Figure 2).

Study characteristics are summarized in S1 Table, results are displayed in S2 Table, Figures 3–8 and S1–6 Figs, additional subgroup results are described in S5 Appendix, QUADAS-2 results are in S3 Table, and our GRADE assessment is detailed in S4 Table.

Study Characteristics

The 81 included studies validated EHRs from 11 different countries, most frequently Denmark (18 studies)[22–39] and the UK (17 studies).[40–56] Validation was the primary aim of all but 10 studies.[35,36,41,48,57–62] Fourteen studies[26,27,31,63–73] validated a vital registration system or disease registry in

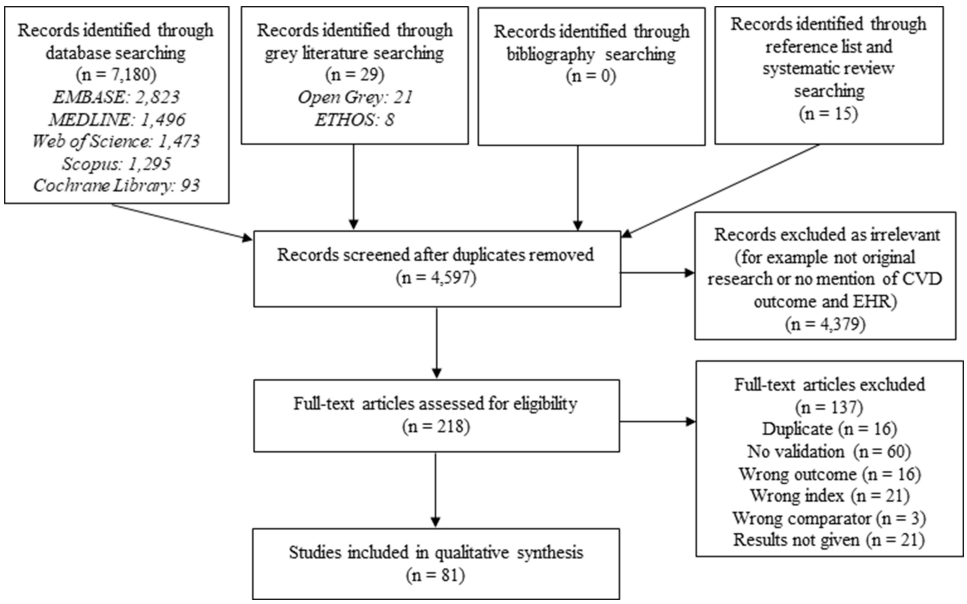


Figure 2 Flow diagram of study selection.

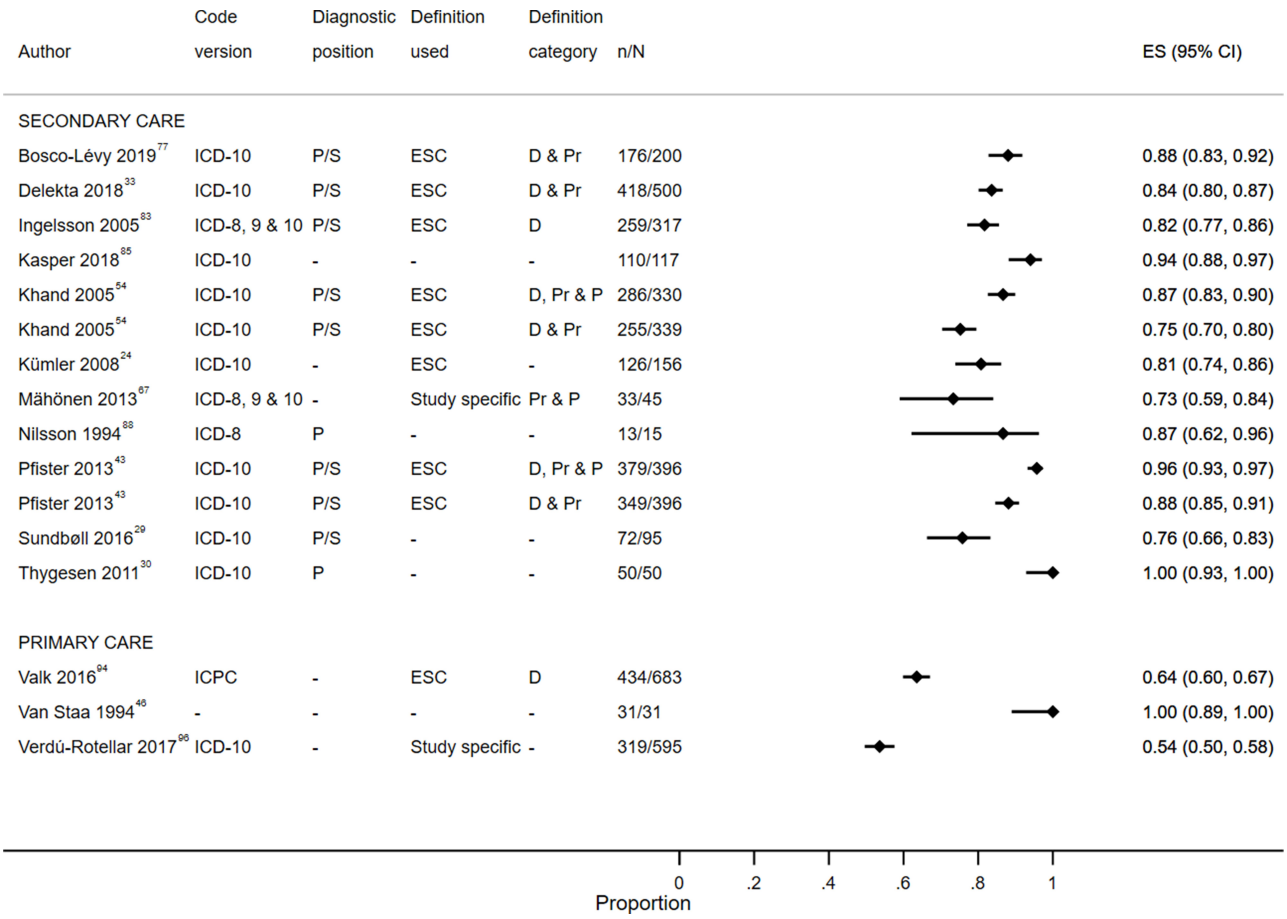


Figure 3 Positive predictive value for heart failure diagnoses from studies which reported the number of records confirmed positive and the total number of records.
Abbreviations: D, definite; D & Pr, definite and probable; D, Pr & P, definite, probable and possible; P, primary; P/S, primary or secondary.

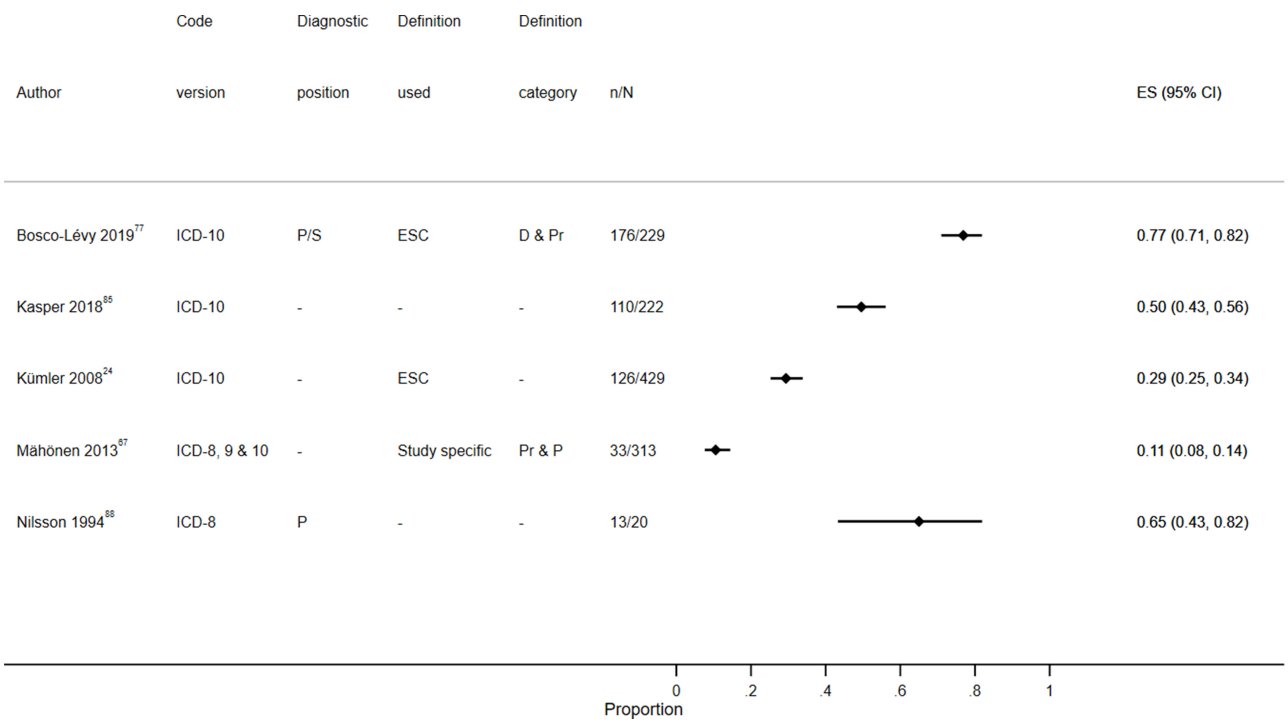


Figure 4 Sensitivity for heart failure diagnoses from studies which reported the number of records confirmed positive and the total number of records.
Abbreviations: D & Pr, definite and probable; Pr & P, probable and possible; P, primary; P/S, primary or secondary.

addition to the EHR. The records validated included data from 1969–2015. Where ICD coding was validated this covered versions 8–10. Sixty studies used medical record review as a reference standard.[22,23,25–39,42,43,45,46,49,50,54,55,57–63,69,72,74–96] Twenty studies validated HF,[24,28–30,33,43,46,54,59,65,67,77,82,83,85,88,94–97] 31 ACS[22,23,26,27,29,30,32,34,37,42,46,47,50,52,53,58,65,68–70,75,76,80,81,84,87,88,91,98–100] and 41 stroke diagnoses.[25,31,32,35,36,38–41,44,45,47–49,51,55–57,60–64,66,71–74,78,79,81,86,87,89–93,98,101,102]

Study Quality

Study quality was high for 54 (67%) of studies,[22–26,28,29,31–34,38,39,42–44,47,50,51,53,54,56,59,60,62–65,67–70,72,73,75–79,85–90,92–94,96,98–102] medium for 19 (24%) studies[27,30,35–37,46,49,52,55,57,58,61,66,74,81–84,95] and low for eight (10%) of studies.[40,41,45,48,71,80,91,97] Studies were overall at low risk of bias in patient selection (76 low, 3 unclear, 2 high), index test (71 low, 10 high), and flow and timing (78 low, 3 unclear) domains and higher risk in the reference standard domain (36 low, 28 unclear, 17 high). Generally, reference standard methods and definitions were poorly described, and on occasion the reference standard was not independent of the EHR. Risk of bias was also higher in studies which validated primary care EHRs. HF validation studies had high quality in 14 (70%) studies, medium in five (25%) and low in one (5%). For ACS validation, quality was high for 21 (68%), medium for eight (26%) and low for two (6%) studies. In stroke validation studies, quality was high for 26 (63%), medium for nine (22%) and low for six (15%) studies.

Heart Failure Study Characteristics

HF diagnoses were most extensively validated using EHR data from Denmark (five studies),[24,28–30,33] the Netherlands (four studies),[59,65,94,95] Sweden (three studies)[82,83,88] and the UK (three studies).[43,46,54] In addition, EHR data from Finland,[67] France,[77] Germany,[85] Italy[97] and Spain[96] were validated in one study each. Fourteen studies validated secondary care EHRs[24,28–30,33,43,54,59,65,67,77,83,85,88] and six studies validated primary care EHRs.[46,82,94–97] Medical record review was used as the reference standard in all but three studies.[24,65,97]

Heart Failure Validation Results

Overall

From the main validation result reported by each of the studies; sensitivity (available from nine studies)-[24,46,65,67,77,82,85,88,95] was ≥50% in six studies[46,77,82,85,88,95] but >66% (range 11–100%) in only one study,[46] PPV (19

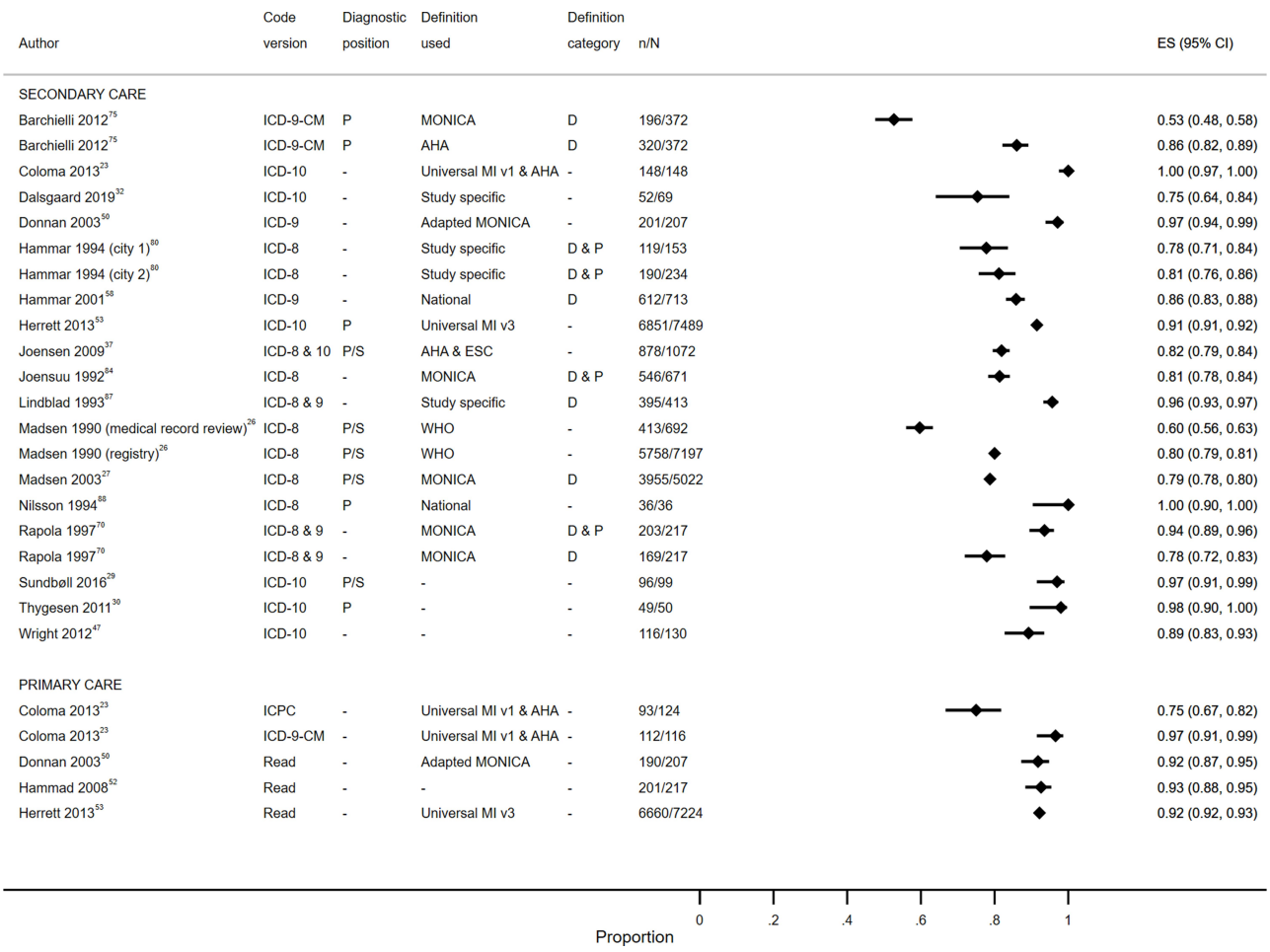


Figure 5 Positive predictive value for myocardial infarction diagnoses from studies which reported the number of records confirmed positive and the total number of records.
Abbreviations: D, definite; D & P, definite and possible; P, primary; P/S, primary or secondary; RS, reference standard.

studies)[24,28–30,33,43,46,54,59,65,67,77,83,85,88,94–97] was ≥80% (range 54–100%) in all but five studies,[29,67,94,96,97] specificity (three studies)[24,67,95] was ≥95% in all studies, and NPV (three studies)[24,67,95] was ≥84% (range 84–96%) in all studies.

Diagnosis Type

In the three studies that reported results for first diagnosis, the PPV range was 76–88%.[28,29,77] One study compared the PPV for all diagnoses (84%) to first diagnosis (80%),[28] and another study found the same PPV for first diagnosis and recurrent diagnosis (both 76%).[29]

Definition

In seven of the eight studies[24,28,33,43,54,77,83,94] which used the ESC definition,[15] the PPV was ≥80%. The study[94] with the lower PPV of 64% was the only one to validate a primary care EHR. Other studies used; both Framingham[103] and Boston[104] criteria (one study,[59] PPV 80–81%), the American College of Cardiology (ACC)/American Heart Association (AHA) definition[105] (one study,[97] PPV 55%), or study-specific definitions (three studies,[67,95,96] PPV 54–83%). An overview of the definitions used by the studies is presented in S6 Appendix.

Seven studies reported classification criteria; the PPV for definite HF ranged between 61–82%,[33,43,54,77,83] including both definite and probable HF increased the PPV to 73–88%[33,43,54,77,83,94] and the two studies which additionally included possible HF reported high PPV as 87%[54] and 96%.[43]

Diagnostic Position

Six studies[29,33,43,54,77,83] reported HF recorded in any diagnostic position (PPV 76–96%) and two studies[30,88] only included primary position (PPV 87% and 100%). Three studies,[33,77,83] which validated any position, also included breakdowns by primary (PPV 88–96%) and secondary (PPV 66–84%) positions.

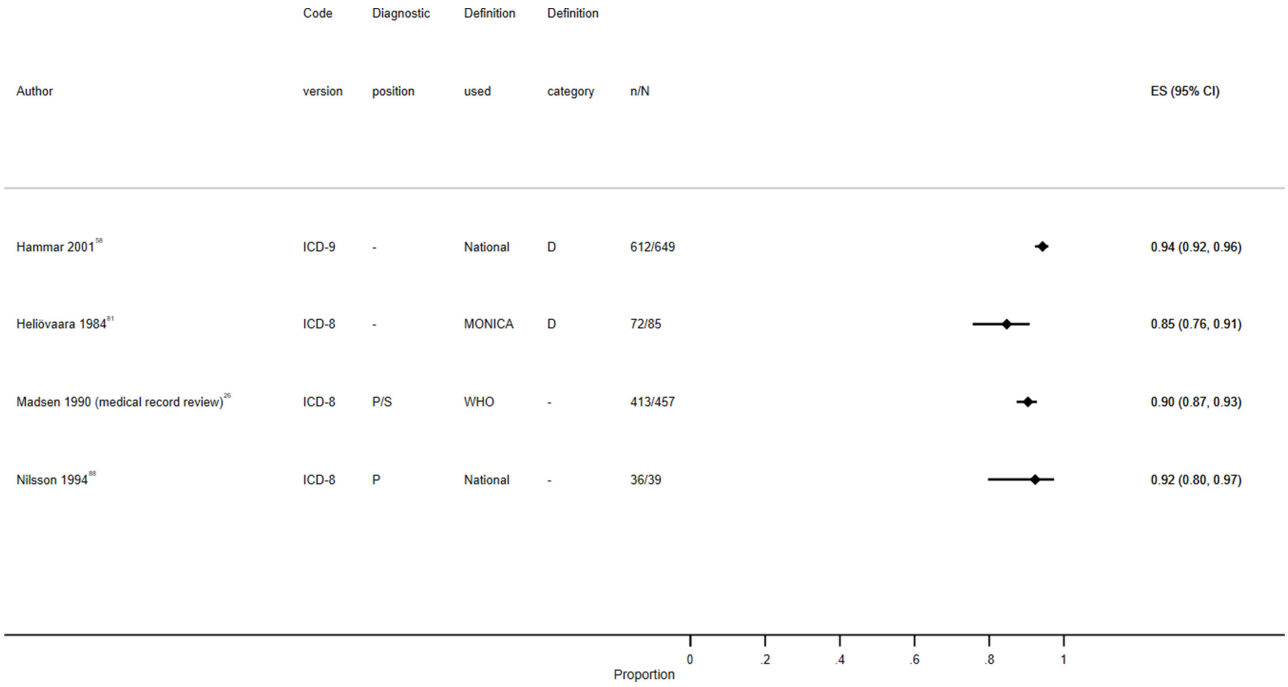


Figure 6 Sensitivity for myocardial infarction diagnoses from studies which reported the number of records confirmed positive and the total number of records.
Abbreviations: D, definite; D & P, definite and possible; P, primary; P/S, primary or secondary; RS, reference standard.

Coding System

Twelve studies validated ICD-10,[24,28–30,33,43,54,67,77,82,83,96] with all but one[83] reporting results specifically for this version of ICD (PPV 78–99%). Six studies[24,33,43,77,82,96] validated I50; two studies of primary care EHRs reported lower validity estimates (PPV 54%[96] and sensitivity 66%)[82] compared to four studies of secondary care EHRs (PPV 81–96%,[24,33,43,77] and sensitivity 29%[24] and 64%).[77] Five studies included a broader range of ICD-10 codes, all of which differed. The estimates for ICD-10 codes were no higher than those for ICD-8 (PPV 87%),[67,83,88] ICD-9 (PPV 79–97%),[59,65,67,83] or combinations of the three ICD systems (PPV 73–82%).[67,83] Two studies validated ICPC K77 in primary care EHRs (PPV 64%[94] and 83%[95]).

Acute Coronary Syndrome Study Characteristics

Similar to HF, ACS diagnoses were most frequently validated using EHR data from Denmark (nine studies),-[22,23,26,27,29,30,32,34,37] followed by Finland (seven studies),[68–70,81,84,99,100] the UK (six studies)[42,46,47,50,52,53] and Sweden (4 studies).[58,80,87,88] Two studies validated data in each of Italy,[23,75] the Netherlands,[23,65] and Spain,[91,98] and a final study used data from France.[76] Twenty-six of the studies validated a secondary care EHR,[22,26,27,29,30,32,34,37,42,47,58,65,68–70,75,76,80,81,84,87,88,91,98–100] three studies validated both a primary and secondary care EHR[23,50,53] and two studies validated a primary care EHR.[46,52]

Four studies[22,37,68,76] presented overall ACS results, of which one study[68] included an additional breakdown for MI and two studies[37,76] included unstable angina and MI, one of which also included cardiac arrest.[37] A further two studies[29,65] did not report results for ACS overall but did include both unstable angina and MI. The remaining 25 studies solely validated MI diagnoses.[23,26,27,30,32,34,42,46,47,50,52,53,58,69,70,75,80,81,84,87,88,91,98–100]

Acute Coronary Syndrome Validation Results

Overall

For ACS, three studies[33,37,76] reported one main PPV (range 66–87%), while results presented by Pajunen et al[68] were broken down by age, sex and time period, with sensitivity of 66–87% and PPV of 63–86%.

Diagnosis Type

The PPV for unstable angina varied; with low values of 20%[76] and 27.5%[37] in two studies and higher values of 78%[65] and 88%[29] in the other two studies. Sensitivity was only reported by one study,[65] at 53%. For MI, the main validation result for sensitivity (11 studies)[26,27,34,42,46,50,58,65,81,88,98] was ≥80% in

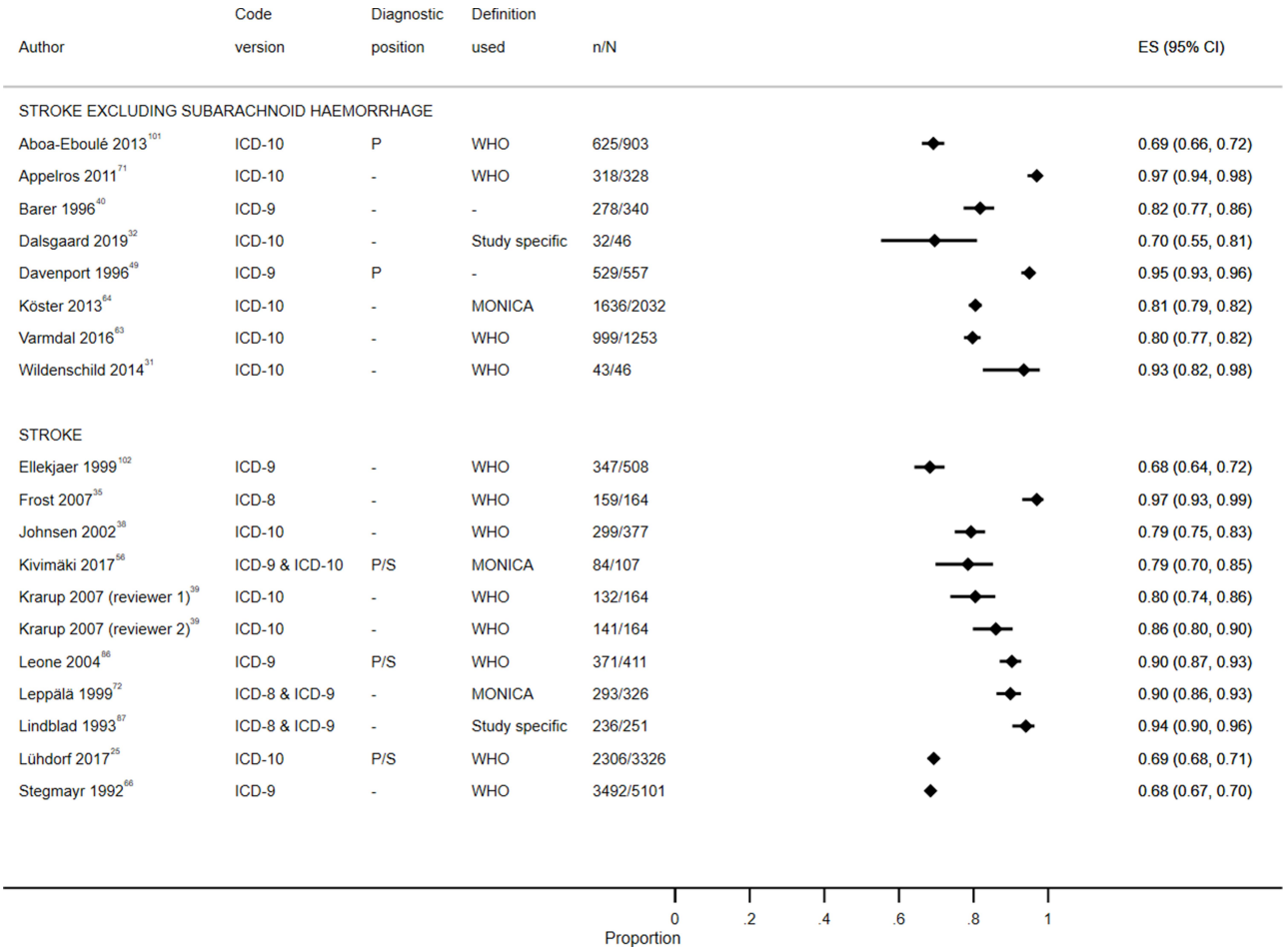


Figure 7 Positive predictive value for stroke diagnoses recorded in secondary care EHRs from studies which reported the number of records confirmed positive and the total number of records.
Abbreviations: ES, effect size; P, primary; P/S, primary or secondary.

all but one study[42] (range 56–97%), and six[26,27,34,58,88,98] >90%. PPV (24 studies)[23,26,27,29,30,32,34,37,42,46,47,50,52,53,58,65,70,75,76,80,84,87,88,98] was ≥80% (range 42–100%) in all but three studies[27,32,34] with 12[23,29,30,42,50,52,53,65,87,88,98] ≥90%. Three studies[34,42,98] reported specificity (range 93–100%) and two-[34,98] included NPV (range 82–100%).

Four studies[29,32,37,84] reported the PPV for first MI, with estimates of 75–97%, and one study[29] also included recurrent MI with a PPV of 88% compared to 97% for first MI.

Definition

Varying MI definitions were used (S6 Appendix). Most frequently (nine studies)[26,27,50,70,75,81,84,99,100] the World Health Organization (WHO) Monitoring trends and determinants in cardiovascular disease (MONICA) definition[106] was used, with variable PPV estimates of 53–96% obtained. Two studies compared MONICA to another MI definition; one[75] showed MONICA-defined definite MI had a substantially lower PPV than AHA/ESC-defined[16] definite MI (53% vs 86%), while the other[84] also showed a lower PPV for MONICA compared to "normal clinically defined MI" but with a smaller difference (81% vs 89%). One further study used the AHA/ESC definition[37] (PPV 82%). The universal definition[107] was used in a study[23] which included EHR data from three countries, with PPVs of 75–100%. Three studies used the third universal definition,[108] one[76] of which combined it with the earlier universal definition (PPV 85%). In another[53] PPVs of 92% with obtained for the primary and secondary care EHRs validated. The third[34] validated MI diagnoses recorded for patients with drug-eluting coronary stents, the PPV was 42% for all admission and 73% for acute admissions.

Diagnostic Position

Of the 10 studies which reported the diagnostic position used to validate MI diagnoses, five[26,27,29,34,68] used any diagnostic position (PPV 42–97%) and five[30,75,76,88,98] primary position (PPV 53–100%). One study[27] which

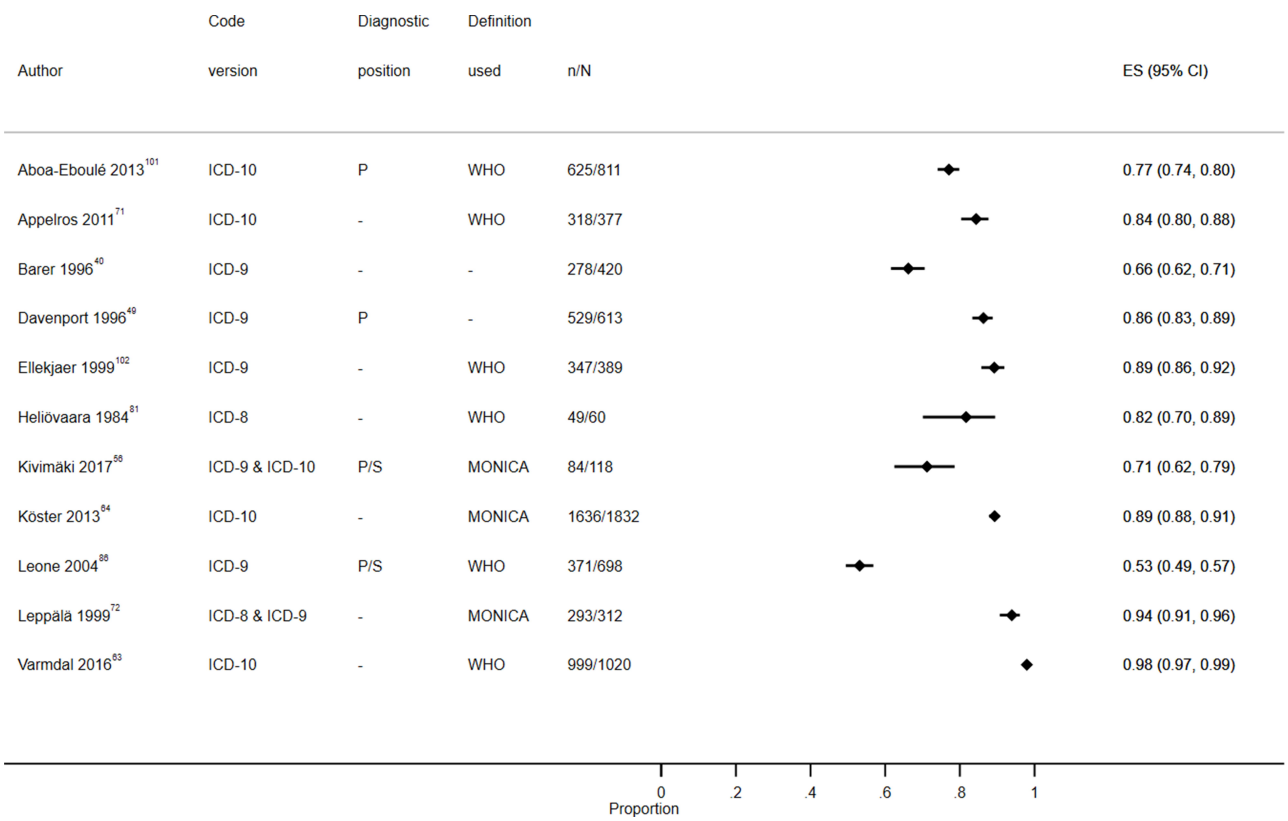


Figure 8 Sensitivity for stroke diagnoses recorded in secondary care EHRs from studies which reported the number of records confirmed positive and the total number of records.
Abbreviations: ES, effect size; P, primary; P/S, primary or secondary.

validated any position (PPV 79%) also included a breakdown by primary position (PPV 80%) and another study[29] included breakdowns by primary (PPV 99%) and secondary positions (PPV 80%).

Coding System

Ten studies validated ICD-10 coded MI, eight reported results specifically for ICD-10.[23,29,30,32,34,47,53,76] Four studies validated ICD-10 I21 with PPV ≥85% (range 42–100%)[23,29,34,76] in all but one.[34] Two studies included I21-I23 and reported high PPVs of 92%[53] and 98%;[30] however, the latter study was small in size (50 patients). One study validated I21-I22 (PPV 89%)[47] and another I21-I24 (PPV 75%).[32] The estimates for ICD-10 codes were no higher than those for ICD-8 (PPV 79–100%),[26,27,80,84,88] ICD-9 (86–100%),[42,50,58,65,75,98] or combinations of three ICD systems (PPV 82–96%).[37,87] Of the studies to validate data in primary care, one[23] included IPCI K75 code (PPV 75%) and three[50,52,53] validated Read coding in the UK (PPV 91–93%).

Reference Standard

The PPV for MI diagnoses varied between 53–100% when medical record review was the reference standard (20 studies)[22,23,26,29,30,32,37,42,46,50,58,69,70,75,76,80,84,87,88,91] and 89–93% when a registry was used.[26,27,53,68,98–100] One study[34] used medical record review after comparing EHR and registry results (PPV 42%). Two studies used a GP questionnaire (PPV 89% and 93%),[47,52] and one study used a local cardiology database (PPV 97%).[65]

Stroke Study Characteristics

Stroke diagnoses were most frequently validated in UK EHRs, with 10 studies conducted,[40,41,44,45,47–49,51,55,56] followed by Denmark (seven studies),[25,31,32,35,36,38,39] Sweden (5 studies)-[60,64,66,71,87] and Italy (4 studies).[74,86,90,93] Data from Finland,[72,73,81] France,[78,79,101] Norway,[63,89,102] and Spain[62,91,98] were validated in three studies each. A further two studies validated EHR data from the Netherlands[57,61] and one from the Czech Republic.[92] All but three studies[41,44,48] validated secondary care EHRs.

Twenty-eight studies presented validation estimates for overall stroke (including both ischaemic and haemorrhagic).[25,31,32,35,38–41,44,45,48,49,56,60,63,64,66,71–73,81,86,87,91,92,98,101,102] Ischaemic stroke was assessed in 18 studies,[25,32,38,39,47,57,62,72–74,78,79,86,90,92,93,101,102] in all but four studies[62,74,79,90] this was done as a subgroup analysis after validating overall stroke. Similarly, haemorrhagic stroke was assessed by 21 studies; two reported results for overall haemorrhagic stroke[32,51] with this the main focus of one study,[51] 17 studies reported results for ICH as a subgroup analysis[25,38,39,47,51,55,57,72,73,78,86,87,89,92,93,101,102] and 18 studies reported results for SAH[25,36,38,39,47,51,55,61,72,73,78,81,86,87,89,92,93,102] with this being the main result in two studies.[36,61]

Stroke Validation Results

Overall

For overall stroke, sensitivity (15 studies)[31,40,45,49,56,63,64,71,73,81,86,91,98,101,102] was ≥80% (range 33–97%) in seven studies[49,63,64,71,73,81,102] and ≥70% in 11 studies. PPV (27 studies)[25,31,32,35,38–41,45,48,49,56,60,63,64,66,71–73,81,86,87,91,92,98,101,102] was ≥80% (range 20–97%) in 19 studies.[31,35,39–41,45,48,49,60,63,64,71,72,81,86,87,92,98] Nine of the studies[31,32,40,49,60,63,64,71,101] did not include codes to validate SAH, three of which had stated this in their inclusion criteria.[40,71,101] Excluding these studies did not affect the sensitivity (53–89%) or PPV (68–97%). Specificity and NPV, reported by five studies, were 99–100%[49,56,63,98] other than one study[31] which obtained a specificity of 96% and NPV of 72%.

Diagnosis Type

Three studies[56,64,101] included first and recurrent overall stroke with sensitivity from 71–89% and PPV 69–81%, while three studies[32,71,73] also included only first stroke for which sensitivity was 85–89% and PPV 70–97%.

For ischaemic stroke, the main sensitivity reported (6 studies)[74,79,81,86,90,102] was ≥66% in all but one[86] study (range 37–82%). Fourteen studies[25,32,38,47,57,62,72,74,78,79,86,90,92,102] included one main PPV of 66–96%. One study[101] classified results separately for cardiac embolism, large artery atherosclerosis, lacunar infarct and ischaemic stroke of other aetiology. Sensitivity and PPV were highest in the cardiac embolism classification (83% and 87%, respectively) and lowest for other aetiology (67% and 35%, respectively). For ICH, the main sensitivity reported was 59–98% (4 studies)[73,86,101,102] and main PPV 55–96% (15 studies).[25,38,39,47,51,55,57,72,73,78,86,87,92,101,102] The sensitivity of SAH diagnoses was 35–92% (4 studies)[73,81,86,102] and PPV was 42–96% (18 studies).[25,36,38,39,47,51,55,61,72,73,78,81,86,87,89,92,93,102]

Definition

Stroke was defined in 22 of the 41 studies, 13[25,31,35,38,39,63,66,71,81,86,90,92,101,102] used the WHO definition[109] (sensitivity 53–97%[63,71,86,101,102] and PPV 68–97%),[25,35,38,39,63,66,71,81,86,92,101,102] seven[56,60,62,64,72,74,93] used MONICA[110] (sensitivity 71–89%[56,64] and PPV 79–92%),[56,60,64,72] and two[32,87] defined stroke specifically for their study (PPV 70% and 91%). The stroke definitions used are summarized in S6 Appendix.

Diagnostic Position

For overall stroke diagnoses recorded in any diagnostic positions, sensitivity ranged from 53–97%[56,63,86] and PPV from 69–90%.[25,56,63,86] In comparison, results only for primary position were 67–86% for sensitivity and 69–95% for PPV.[49,63,73,98,101]

Coding System

Thirteen studies validated ICD-10 (PPV 20–97%,[31,32,38,39,45,47,55,60,63,64,71,78,92] sensitivity 76–97%).[45,63,64,71,101] Four studies[31,63,64,71] which excluded SAH from the stroke definition validated ICD-10 I61, I63 and I64 (sensitivity 89–97% and PPV 79–97%). Aboa-Eboule et al[101] additionally included G46 in their definition (sensitivity 77% and PPV 69%) while Dalsgaard et al[32] validated I61-I65 (PPV 70%). In comparison, Holmqvist et al[60] only included I61 and I63, and obtained PPV estimates of 92% and 89% in people with and without rheumatoid arthritis, respectively. Three studies[38,39,92] which included SAH in the stroke definition validated I60, I61, I63 and I64 (PPV 79–86%) and one[45] additionally included I62 (PPV 96%). The estimates for ICD-10 codes were no higher than those for ICD-8 codes (sensitivity 82%),[81] ICD-9 (PPV 20–95%,[40,49,66,86,91,93,98,102] sensitivity 33–89%),[40,49,86,91,98,102] or combinations of three ICD systems (PPV 79–97%,[35,72,73,87] sensitivity 71–85%).[73]

Seven studies validated ICD-10 I63 for ischaemic stroke diagnosis (PPV 78–96%).[25,32,38,47,78,79,92] One study[73] used a broad (ICD-9433, 434, 436 and ICD-10 I63, I64) and narrow range of codes (ICD-9433, 434 and ICD-10 I63) to define ischaemic stroke, with similar sensitivity (82% vs 81%) and PPV (84% vs 83%). One other study[74] reported results by ICD-9 codes 443*1 and 434*1 (PPV 86% and 90%, respectively). Six studies[25,38,55,78,89,92] validated ICD-10 I61, with another two[39,101] presumed to have also validated this code, for ICH (PPV 66–96%) and a further three

studies[86,93,102] validated ICD-9431 (PPV 71–78%). For SAH, eight studies[25,38,39,47,55,78,89,92] validated ICD-10 I60 with PPV >90% in half of the studies (range 46–96%), four studies[61,86,93,102] validated ICD-9430 (PPV 42–95%), one study[81] validated ICD-8430 (PPV 85%) and two studies[72,87] validated both versions for 430 (PPV 78–79%).

Reference Standard

In the 17 studies[25,31,32,35,38,39,45,55,56,60,63,72,79,86,87,91,92] which used medical record review as the reference standard to validate overall stroke diagnoses, the PPV was ≥79% (range 20–97%) in all but four studies.[25,31,32,91] A further eight studies used a registry reference standard (PPV 88–97%).[40,64,66,71,73,98,101,102]

Heterogeneity

We were able to assess the heterogeneity between the main PPV reported in; 14 studies with 16 estimates of HF (I^2=97.0%), 18 studies with 26 estimates of MI (I^2=98.5%), and 19 studies with 20 estimates of stroke (I^2=97.9%) diagnoses. Additionally, we assessed heterogeneity between the main sensitivity for; six studies of HF (I^2=98.6%), four of MI (I^2=74.3%), and 11 of stroke (I^2=98.8%) diagnoses. Heterogeneity between the estimates was considerable, at more than >95% in all cases other than sensitivity estimates for MI. Furthermore, heterogeneity remained considerable after removal of studies at a high risk of bias.

Overall Strength of Evidence

GRADE showed that cross-study quality was very low for all HF outcomes (sensitivity and PPV in secondary care EHRs and PPV in primary care EHRs), low for MI sensitivity and PPV in secondary care EHRs and moderate for PPV in primary care EHRs, and very low for stroke sensitivity in secondary care EHRs and PPV in primary care EHRs and moderate for PPV in secondary care EHRs.

Discussion

Summary of Findings

Our systematic review suggests that the sensitivity of coded data in European EHRs for HF diagnoses is low at ≤66% in all but one study. There was also wide variation in stroke sensitivity estimates, with only half of studies ≥80%, although three-quarters were ≥70%. The sensitivity of ACS was higher at ≥80% in the vast majority of studies. The majority of studies which validated ACS diagnosis did so specifically for MI.

The PPV of all diagnoses was ≥80% in the majority of studies; two-thirds for HF (nearly three-quarters for secondary care EHRs), nearly three-quarters for MI, and 70% of stroke validation studies. Where subtypes were validated, PPV was ≥80% for four-fifths of ischaemic stroke diagnoses but only 44% of ICH and SAH diagnoses.

The specificity and NPV were also high where available (three HF studies, three MI studies and five stroke studies). However, as most studies only included patients with the diagnosis of interest recorded in the EHR and reference standard, the results presented were mostly limited to sensitivity and PPV.

Both PPV and NPV are impacted by disease prevalence, with lower estimates for rare conditions.[111] Our systematic review focused on Europe, drawing studies from 11 countries. Age-standardized prevalence of CVD in these countries is between 5000–6500 per 100,000, other than the Czech Republic (~8700 per 100,000) which only contributed one study.[2] Therefore, prevalence differences should have limited impact on our comparison of validity estimates between geographies. The prevalence of CVD increases with age, but we did not find any systematic difference in results between studies with younger or older populations.

The low sensitivity of HF diagnoses we identified is consistent with a previous systematic review validating HF diagnoses in administrative data, which identified three European studies.[11] Twelve more studies have since been published and included in our review. These more recent findings, however, do not suggest any improvement in the quality of data over time. This is perhaps unsurprising given the range of clinical aetiology and presentation. The high proportion of studies we found to have a PPV of <80% for stroke diagnoses appeared more substantial than in previous systematic reviews.[9,12] We identified 15 new studies which were not included in these previous reviews.[25,32,45,51,56,57,61–63,74,78,89,91,92,98] Our results for sensitivity and PPV of MI diagnoses are consistent with previous reviews,[8,10] and identified five[29,32,34,76,98] new MI validation studies with variable results.

There was substantial heterogeneity between the sensitivity and PPV estimates for all three acute CVD diagnoses. Heterogeneity was likely because studies differed in multiple ways; for example, even among studies which used medical record review as the reference standard, differences in study time period impacted upon the ICD version used. The heterogeneity caused by variable methods was highlighted in previous systematic reviews of

atrial fibrillation and dementia diagnoses recorded in routine health data.[112,113]

Defining Diagnosis in the EHR

We were most interested in the results of ICD-10 validation, as this is the latest ICD coding system which is widely used in Europe and elsewhere. In McCormick et al's[10] review of MI diagnoses in administrative data, the authors noted a lack of ICD-10 validation with only three studies identified, whereas our review identified 10. Nevertheless, even within ICD-10, combinations of codes used, and therefore their validity, differed, which highlights the importance of tailoring codes to each research question. Codes are arguably even more important when using other, more complex coding systems such as Read codes, which are used in UK primary care data and can generate vast numbers of codes for every clinical condition.

Defining Diagnosis in the Reference Standard

There is no single recommended gold standard to determine the validity of EHR data.[114] Nearly three-quarters (74%) of studies used medical records; more frequently for HF diagnoses (85%) than ACS (71%) or stroke (68%). This difference may be due to availability of MI and stroke registries, used in 26% and 22% of studies, respectively. No differences in the performance of the reference standard methods were discernable, probably due to heterogeneity.

Criteria to define CVD, especially MI, have been refined over time, driven by the development of more sensitive and specific biomarkers, and more precise imaging techniques.[100] However, we did not identify any temporal trends in the accuracy of MI recording, again likely due to overall study heterogeneity.

When validating HF, which can vary in clinical aetiology and presentation, clarity on the criteria used to define, with explicit classification of acute and chronic HF along with ejection fraction would benefit understanding of results.

Comparing and Combining Data Sources

Only 14 (17%) studies validated primary care systems, more than half of which were in the UK. Using primary care EHRs may be beneficial for research into conditions such as HF which are frequently managed in primary care; in our study, 30% of HF EHR validation studies used primary care data, compared to 16% for ACS and 7% for stroke studies. For acute severe conditions resulting in hospitalization, secondary care records should be the most reliable data source. Where possible, the use of linked data to increase the ascertainment of acute CVD events should be considered.

Implications for Future Research

EHR-based research is a growing field – widely used in observational analyses and increasingly employed in trials.[115] Researchers should consider the level of validity necessary for their own CVD outcome definition. When a composite outcome, such as MACE, is used researchers may need to address differing sensitivity in the individual components of the outcome. In studies which investigate CVD incidence, a sensitive definition is particularly important. For example, EHR data are being used for rapid COVID-19 pandemic analyses such as; the impact the virus has in those with CVD, CVD as an outcome after infection with the virus, and excess death estimates.[116] It is important that these rapid analyses consider the validity of the data and definitions used. Conversely, in a pragmatic trial recruitment, a specific definition is likely more important than a sensitive one.

Strengths and Limitations

Our systematic review provides a comprehensive and up-to-date evaluation of the validity of acute CVD diagnoses in European EHRs, conducted without language or time restrictions using a broad search strategy. Two independent reviewers performed our study selection, and native speaking collaborators translated foreign language articles. Similar to other systematic reviews of validation studies, we repurposed the QUADAS-2 risk of bias tool developed for diagnostic test accuracy. Additionally, we followed the diagnostic test accuracy GRADE methodology to assess the overall evidence base.

Our work is not without limitations. Firstly, only one reviewer completed full data extraction and risk of bias assessment due to resource constraints, although a sample of 20% of studies had data dual extracted. Secondly, we limited our study to Europe, so theoretically our results are only generalizable to European countries. All previous systematic reviews[8–12] on the validity of acute CVD diagnoses included both EHRs and claim-based systems, while most studies included in each of these reviews were from North America. From these existing reviews, it was unclear if the validity of EHRs differed to claims-based datasets, which reflect payments related to medical care given. Despite this, we obtained similar results to the

previous reviews. Thirdly, our review focused on acute CVD events so excluded results from studies that validated broader diagnoses of ischaemic heart disease or cerebrovascular disease, which again limits generalizability to these specific conditions.

Recommendations

For ACS and stroke diagnoses, most sensitivity and PPV results were reasonably high, providing confidence in the use of European EHR data for research into these conditions. However, there was considerable heterogeneity between studies. Sensitivity for HF diagnoses was low, and our GRADE assessment found very low quality for all HF outcomes. For studies of HF, we strongly recommend either validating the definition or referring to existing validation studies to develop the case definition. New validation studies of HF diagnoses should report whether the diagnoses validated are for acute or chronic presentation and HF with reduced ejection fraction or preserved ejection fraction. These principles are also applicable to future ACS and stroke validation studies. Identifying specific stroke subtypes can be difficult; analysis of all stroke subtypes combined is preferable.

Conclusions

Our review on the accuracy of HF, ACS and stroke diagnoses in European EHRs should guide researchers in their selection of data sources and CVD definitions for epidemiological studies. Generally, the data assessed was of reasonable quality. However, it is difficult to summarize validity given the heterogeneity between studies. Where possible, researchers should validate data before use or carefully interpret the results of previous validation studies to consider the impact validity has on research findings. Additionally, the use of linked data will bolster quality.

Acknowledgments

We thank Hanne-Dorthe Emborg and Elisabeth Bondesson for their translations of Danish and Swedish language articles.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

J.A.D. is funded by a British Heart Foundation Non-Clinical PhD Studentship (FS/18/71/33938) and C.W.G is funded by a Wellcome Intermediate Clinical Fellowship (201440/Z/16/Z). The funders had no role in study design, data collection and analysis, preparation of the manuscript, or the decision to publish.

Disclosure

Charlotte Warren-Gash reports grants from Wellcome, grants from British Heart Foundation, during the conduct of the study; personal fees from Sanofi Pasteur, outside the submitted work. The authors declare no other potential conflicts of interest for this work.

References

1. World Health Organization. The top 10 causes of death. https://www.who.int/en/news-room/fact-sheets/detail/the-top-10-causes-of-death. Accessed February 11, 2019.
2. European Heart Network. CVD statistics 2017; 2017. Available from: http://www.ehnheart.org/cvd-statistics/cvd-statistics-2017.html. Accessed October 9, 2019..
3. Denaxas SC, Morley KI. Big biomedical data and cardiovascular disease research: opportunities and challenges. *Eur Hear J - Qual Care Clin Outcomes*. 2015;1(1):9–16. doi:10.1093/ehjqcco/qcv005
4. 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
5. Sund R. Quality of the Finnish hospital discharge register: a systematic review. *Scand J Public Health*. 2012;40(6):505–515. doi:10.1177/1403494812456637
6. Ludvigsson JF, Andersson E, Ekbom A, et al. External review and validation of the Swedish national inpatient register. *BMC Public Health*. 2011;11:450. doi:10.1186/1471-2458-11-450
7. Schmidt M, Schmidt SAJ, Sandegaard JL, Ehrenstein V, Pedersen L, Sørensen HT. The Danish National Patient Registry: a review of content, data quality, and research potential. *Clin Epidemiol*. 2015;7:449–490. doi:10.2147/CLEP.S91125
8. Rubbo B, Fitzpatrick NK, Denaxas S, et al. Use of electronic health records to ascertain, validate and phenotype acute myocardial infarction: a systematic review and recommendations. *Int J Cardiol*. 2015;187:705–711. doi:10.1016/j.ijcard.2015.03.075
9. Woodfield R, Grant I, Sudlow CLM; UK Biobank Follow-Up and Outcomes Working Group. Accuracy of electronic health record data for identifying stroke cases in large-scale epidemiological studies: a systematic review from the UK biobank stroke outcomes group. *PLoS One*. 2015;10(10):e0140533. doi:10.1371/journal.pone.0140533
10. McCormick N, Lacaille D, Bhole V, Avina-Zubieta JA, Guo Y. Validity of myocardial infarction diagnoses in administrative databases: a systematic review. *PLoS One*. 2014;9(3):e92286. doi:10.1371/journal.pone.0092286

11. McCormick N, Lacaille D, Bhole V, Avina-Zubieta JA. Validity of heart failure diagnoses in administrative databases: a systematic review and meta-analysis. *PLoS One*. 2014;9(8):e104519. doi:10.1371/journal.pone.0104519
12. McCormick N, Bhole V, Lacaille D, Avina-Zubieta JA. Validity of diagnostic codes for acute stroke in administrative databases: a systematic review. *PLoS One*. 2015;10(8):e0135834. doi:10.1371/journal.pone.0135834
13. Poudel I, Tejpal C, Rashid H, Jahan N. Major adverse cardiovascular events: an inevitable outcome of ST-elevation myocardial infarction? A literature review. *Cureus*. 2019;11(7). doi:10.7759/cureus.5280
14. Davidson JA, Banerjee A, Muzambi R, Smeeth L, Warren-Gash C. Validity of acute cardiovascular outcome diagnoses in European electronic health records: a systematic review protocol. *BMJ Open*. 2019;9(10). doi:10.1136/bmjopen-2019-031373
15. Ponikowski P, Voors AA, Anker SD, et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: the task force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) developed with the special contribution. *Eur Heart J*. 2016;37(27):2129–2200.
16. Luepker RV, Apple FS, Christenson RH, et al. Case definitions for acute coronary heart disease in epidemiology and clinical research studies. *Circulation*. 2003;108(20):2543–2549. doi:10.1161/01.CIR.0000100560.46946.EA
17. Sacco RL, Kasner SE, Broderick JP, et al. An updated definition of stroke for the 21st century: a statement for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*. 2013;44(7):2064–2089. doi:10.1161/STR.0b013e318296aeca
18. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529. doi:10.7326/0003-4819-155-8-201110180-00009
19. The Cochrane Collaboration. Cochrane handbook for systematic reviews of interventions. Version 5. (Higgins JP, Green S, eds.); 2011. Available from: https://handbook-5-1.cochrane.org/. Accessed September 9, 2020.
20. Nyaga VN, Arbyn M, Aerts M. Metaprop: a Stata command to perform meta-analysis of binomial data. *Arch Public Health*. 2014;72(1):1–10. doi:10.1186/2049-3258-72-39
21. Schünemann HJ, Oxman AD, Brozek J, et al. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ*. 2008;336:1106–1110. doi:10.1136/bmj.39500.677199.ae
22. Bork CS, Al-Zuhairi KS, Hansen SM, Delekta J, Joensen AM. Accuracy of angina pectoris and acute coronary syndrome in the Danish National Patient Register. *Dan Med J*. 2017;64(5).
23. Coloma PM, Valkhoff VE, Mazzaglia G, et al. Identification of acute myocardial infarction from electronic healthcare records using different disease coding systems: a validation study in three European countries. *BMJ Open*. 2013;3(6):e002862. doi:10.1136/bmjopen-2013-002862
24. Kümler T, Gislason GH, Kirk V, et al. Accuracy of a heart failure diagnosis in administrative registers. *Eur J Heart Fail*. 2008;10(7):658–660. doi:10.1016/j.ejheart.2008.05.006
25. Lühdorf P, Overvad K, Schmidt EB, Johnsen SP, Bach FW. Predictive value of stroke discharge diagnoses in the Danish National Patient Register. *Scand J Public Health*. 2017;45(6):630–636. doi:10.1177/1403494817716582
26. Madsen M, Balling H, Eriksen LS. [The validity of the diagnosis of acute myocardial infarction in 2 registries: the Heart Registry compared to the National Patient Registry]. *Ugeskr Laeger*. 1990;152(5):308–314. Danish.
27. Madsen M, Davidsen M, Rasmussen S, Abildstrom SZ, Osler M. The validity of the diagnosis of acute myocardial infarction in routine statistics: a comparison of mortality and hospital discharge data with the Danish MONICA registry. *J Clin Epidemiol*. 2003;56(2):124–130. doi:10.1016/S0895-4356(02)00591-7
28. Mard S, Nielsen FE. Positive predictive value and impact of misdiagnosis of a heart failure diagnosis in administrative registers among patients admitted to a university hospital cardiac care unit. *Clin Epidemiol*. 2010;2:235–239. doi:10.2147/CLEP.S12457
29. Sundbøll J, Adelborg K, Munch T, et al. Positive predictive value of cardiovascular diagnoses in the Danish National Patient Registry: a validation study. *BMJ Open*. 2016;6(11):e012832. doi:10.1136/bmjopen-2016-012832
30. Thygesen SK, Christiansen CF, Christensen S, Lash TL, Sørensen HT. The predictive value of ICD-10 diagnostic coding used to assess Charlson comorbidity index conditions in the population-based Danish National Registry of Patients. *BMC Med Res Methodol*. 2011;11:83. doi:10.1186/1471-2288-11-83
31. Wildenschild C, Mehnert FW, Thomsen R, et al. Registration of acute stroke: validity in the Danish Stroke Registry and the Danish National Registry of Patients. *Clin Epidemiol*. 2013;6:27. doi:10.2147/CLEP.S50449
32. Dalsgaard E-M, Witte DR, Charles M, Jørgensen ME, Lauritzen T, Sandbæk A. Validity of Danish register diagnoses of myocardial infarction and stroke against experts in people with screen-detected diabetes. *BMC Public Health*. 2019;19(1):228. doi:10.1186/s12889-019-6549-z
33. Delekta J, Hansen S, AlZuhairi K, Bork C, Joensen A. The validity of the diagnosis of heart failure (I50.0-I50.9) in the Danish National Patient Register. *Dan Med J*. 2018;65(4):A5470.
34. Egholm G, Madsen M, Thim T, et al. Evaluation of algorithms for registry-based detection of acute myocardial infarction following percutaneous coronary intervention. *Clin Epidemiol*. 2016;8:415–423. doi:10.2147/CLEP.S108906
35. Frost L, Andersen LV, Vestergaard P, Husted S, Mortensen LS. Trend in mortality after stroke with atrial fibrillation. *Am J Med*. 2007;120(1):47–53. doi:10.1016/j.amjmed.2005.12.027
36. Gaist D. Risk of subarachnoid haemorrhage in first degree relatives of patients with subarachnoid haemorrhage: follow up study based on national registries in Denmark. *BMJ*. 2000;320(7228):141–145. doi:10.1136/bmj.320.7228.141
37. Joensen AM, Jensen MK, Overvad K, et al. Predictive values of acute coronary syndrome discharge diagnoses differed in the Danish National Patient Registry. *J Clin Epidemiol*. 2009;62(2):188–194. doi:10.1016/J.JCLINEPI.2008.03.005
38. Johnsen SP, Overvad K, Sørensen HT, Tjønneland A, Husted SE. Predictive value of stroke and transient ischemic attack discharge diagnoses in The Danish National Registry of Patients. *J Clin Epidemiol*. 2002;55(6):602–607. doi:10.1016/S0895-4356(02)00391-8
39. Krarup L-H, Boysen G, Janjua H, Prescott E, Truelsen T. Validity of stroke diagnoses in a National Register of Patients. *Neuroepidemiology*. 2007;28(3):150–154. doi:10.1159/000102143
40. Barer D, Ellul J, Watkins C. Correcting outcome data for case mix in stroke medicine. *BMJ*. 1996;313(7063):1005–1006. doi:10.1136/bmj.313.7063.1005c
41. Cook M, Baker N, Lanes S, Bullock R, Wentworth C, Michael Arrighi H. Incidence of stroke and seizure in Alzheimer's disease dementia. *Age Ageing*. 2015;44(4):695–699. doi:10.1093/ageing/afv061
42. McAlpine R, Pringle S, Pringle T, Lorimer R, MacDonald TM. A study to determine the sensitivity and specificity of hospital discharge diagnosis data used in the MICA study. *Pharmacoepidemiol Drug Saf*. 1998;7(5):311–318. doi:10.1002/(SICI)1099-1557(199809/10)7:5<311::AID-PDS371>3.0.CO;2-O

43. Pfister R, Michels G, Wilfred J, Luben R, Wareham NJ, Khaw KT. Does ICD-10 hospital discharge code I50 identify people with heart failure? A validation study within the EPIC-Norfolk study. *Int J Cardiol*. 2013;168(4):4413–4414. doi:10.1016/j.ijcard.2013.05.031
44. Ruigómez A, Martín-Merino E, Rodríguez LAG. Validation of ischemic cerebrovascular diagnoses in the health improvement network (THIN). *Pharmacoepidemiol Drug Saf*. 2010;19(6):579–585. doi:10.1002/PDS.1919
45. Sansom LT, Ramadan H. Stroke incidence: sensitivity of hospital data coding of acute stroke. *Int J Stroke*. 2015;10(6):E70. doi:10.1111/ijs.12577
46. Van Staa TP, Abenhaim L. The quality of information recorded on a UK database of primary care records: a study of hospitalizations due to hypoglycemia and other conditions. *Pharmacoepidemiol Drug Saf*. 1994;3(1):15–21. doi:10.1002/pds.2630030106
47. Wright FL, Green J, Canoy D, Cairns BJ, Balkwill A, Beral V. Vascular disease in women: comparison of diagnoses in hospital episode statistics and general practice records in England. *BMC Med Res Methodol*. 2012;12(1):161. doi:10.1186/1471-2288-12-161
48. Zhou EH, Gelperin K, Levenson MS, Rose M, Hsueh YH, Graham DJ. Risk of acute myocardial infarction, stroke, or death in patients initiating olmesartan or other angiotensin receptor blockers – a cohort study using the clinical practice research datalink. *Pharmacoepidemiol Drug Saf*. 2014;23(4):340–347. doi:10.1002/pds.3549
49. Davenport RJ, Dennis MS, Warlow CP. The accuracy of Scottish Morbidity Record (SMR1) data for identifying hospitalised stroke patients. *Health Bull (Raleigh)*. 1996;54(5):402–405.
50. Donnan PT, Dougall HT, Sullivan FM. Optimal strategies for identifying patients with myocardial infarction in general practice. *Fam Pract*. 2003;20(6):706–710. doi:10.1093/fampra/cmg614
51. Gaist D, Wallander M-A, González-Pérez A, García-Rodríguez LA. Incidence of hemorrhagic stroke in the general population: validation of data from The Health Improvement Network. *Pharmacoepidemiol Drug Saf*. 2013;22(2):176–182. doi:10.1002/pds.3391
52. Hammad TA, McAdams MA, Feight A, Iyasu S, Dal Pan GJ. Determining the predictive value of Read/OXMIS codes to identify incident acute myocardial infarction in the General Practice Research Database. *Pharmacoepidemiol Drug Saf*. 2008;17(12):1197–1201. doi:10.1002/pds.1672
53. 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(may203):f2350–f2350. doi:10.1136/bmj.f2350
54. Khand AU, Shaw M, Gemmel I, Cleland JGF. Do discharge codes underestimate hospitalisation due to heart failure? Validation study of hospital discharge coding for heart failure. *Eur J Heart Fail*. 2005;7(5):792–797. doi:10.1016/j.ejheart.2005.04.001
55. Kirkman MA, Mahattanakul W, Gregson BA, Mendelow AD. The accuracy of hospital discharge coding for hemorrhagic stroke. *Acta Neurol Belg*. 2009;109(2):114–119.
56. Kivimäki M, Batty GD, Singh-Manoux A, Britton A, Brunner EJ, Shipley MJ. Validity of cardiovascular disease event ascertainment using linkage to UK hospital records. *Epidemiology*. 2017;28(5):735–739. doi:10.1097/EDE.0000000000000688
57. Ekker MS, Verhoeven JI, Vaartjes I, van Nieuwenhuizen KM, Klijn CJM, de Leeuw FE. Stroke incidence in young adults according to age, subtype, sex, and time trends. *Neurology*. 2019;92(21):e2444–e2454. doi:10.1212/WNL.0000000000007533
58. Hammar N, Alfredsson L, Rosén M, Spetz CL, Kahan T, Ysberg AS. A national record linkage to study acute myocardial infarction incidence and case fatality in Sweden. *Int J Epidemiol*. 2001;30(Suppl 1):S30–4. doi:10.1093/ije/30.suppl_1.S30
59. Heerdink ER, Leufkens HG, Herings RM, Ottervanger JP, Stricker BH, Bakker A. NSAIDs associated with increased risk of congestive heart failure in elderly patients taking diuretics. *Arch Intern Med*. 1998;158(10):1108–1112. doi:10.1001/archinte.158.10.1108
60. Holmqvist M, Gränsmark E, Mantel Ä, et al. Occurrence and relative risk of stroke in incident and prevalent contemporary rheumatoid arthritis. *Ann Rheum Dis*. 2013;72(4):541–546. doi:10.1136/annrheumdis-2012-201387
61. Nieuwkamp DJ, Vaartjes I, Algra A, Rinkel GJE, Bots ML. Risk of cardiovascular events and death in the life after aneurysmal subarachnoid haemorrhage: a nationwide study. *Int J Stroke*. 2014;9(8):1090–1096. doi:10.1111/j.1747-4949.2012.00875.x
62. Vila-Corcoles A, Satue-Gracia E, Ochoa-Gondar O, et al. Incidence and lethality of ischaemic stroke among people 60 years or older in the region of Tarragona (Spain), 2008–2011. *Rev Neurol*. 2014;59(11):490–496.
63. Varmdal T, Bakken IJ, Janszky I, et al. Comparison of the validity of stroke diagnoses in a medical quality register and an administrative health register. *Scand J Public Health*. 2016;44(2):143–149. doi:10.1177/1403494815621641
64. Köster M, Asplund K, Johansson Å, Stegmayr B. Refinement of Swedish Administrative Registers to monitor stroke events on the national level. *Neuroepidemiology*. 2013;40(4):240–246. doi:10.1159/000345953
65. Merry AHH, Boer JMA, Schouten LJ, et al. Validity of coronary heart diseases and heart failure based on hospital discharge and mortality data in the Netherlands using the cardiovascular registry Maastricht cohort study. *Eur J Epidemiol*. 2009;24(5):237–247. doi:10.1007/s10654-009-9335-x
66. Stegmayr B, Asplund K. Measuring stroke in the population: quality of routine statistics in comparison with a population-based stroke registry. *Neuroepidemiology*. 1992;11(4–6):204–213. doi:10.1159/000110933
67. Mähönen M, Jula A, Harald K, et al. The validity of heart failure diagnoses obtained from administrative registers. *Eur J Prev Cardiol*. 2013;20(2):254–259. doi:10.1177/2047487312438979
68. Pajunen P, Koukkunen H, Ketonen M, et al. The validity of the Finnish Hospital Discharge Register and causes of death register data on coronary heart disease. *Eur J Cardiovasc Prev Rehabil*. 2005;12(2):132–137.
69. Pietilä K, Tenkanen L, Mänttäri M, Manninen V. How to define coronary heart disease in register-based follow-up studies: experience from the Helsinki Heart Study. *Ann Med*. 1997;29(3):253–259. doi:10.3109/07853899708999343
70. Rapola JM, Virtamo J, Korhonen P, et al. Validity of diagnoses of major coronary events in national registers of hospital diagnoses and deaths in Finland. *Eur J Epidemiol*. 1997;13(2):133–138. doi:10.1023/A:1007380408729
71. Appelros P, Terént A. Validation of the Swedish inpatient and cause-of-death registers in the context of stroke. *Acta Neurol Scand*. 2011;123(4):289–293. doi:10.1111/j.1600-0404.2010.01402.x
72. Leppälä JM, Virtamo J, Heinonen OP. Validation of stroke diagnosis in the National Hospital Discharge Register and the Register of Causes of Death in Finland. *Eur J Epidemiol*. 1999;15(2):155–160. doi:10.1023/A:1007504310431
73. Tolonen H, Salomaa V, Torppa J, et al. The validation of the Finnish Hospital Discharge Register and causes of death register data on stroke diagnoses. *Eur J Cardiovasc Prev Rehabil*. 2007;14(3):380–385. doi:10.1097/01.hjr.0000239466.26132.f2

74. Baldereschi M, Balzi D, Di Fabrizio V, et al. Administrative data underestimate acute ischemic stroke events and thrombolysis treatments: data from a multicenter validation survey in Italy. *PLoS One*. 2018;13(3):e0193776. doi:10.1371/journal.pone.0193776
75. Barchielli A, Balzi D, Naldoni P, et al. Hospital discharge data for assessing myocardial infarction events and trends, and effects of diagnosis validation according to MONICA and AHA criteria. *J Epidemiol Community Health*. 2012;66(5):462–467. doi:10.1136/jech.2010.110908
76. Bezin J, Girodet P-O, Rambelomanana S, et al. Choice of ICD-10 codes for the identification of acute coronary syndrome in the French hospitalization database. *Fundam Clin Pharmacol*. 2015;29(6):586–591. doi:10.1111/fcp.12143
77. Bosco-Lévy P, Duret S, Picard F, et al. Diagnostic accuracy of the International Classification of Diseases, Tenth Revision, codes of heart failure in an administrative database. *Pharmacoepidemiol Drug Saf*. 2019;28(2):194–200. doi:10.1002/pds.4690
78. Giroud M, Hommel M, Benzenine E, Fauconnier J, Béjot Y, Quantin C. Positive predictive value of French hospitalization discharge codes for stroke and transient ischemic attack. *Eur Neurol*. 2015;74:92–99. doi:10.1159/000438859
79. Haesebaert J, Termoz A, Polazzi S, et al. Can hospital discharge databases be used to follow ischemic stroke incidence? *Stroke*. 2013;44(7):1770–1774. doi:10.1161/STROKEAHA.113.001300
80. Hammar N, Larsen FF, de Faire U. Are geographical differences and time trends in myocardial infarction incidence in Sweden real? Validity of hospital discharge diagnoses. *J Clin Epidemiol*. 1994;47(6):685–693. doi:10.1016/0895-4356(94)90216-x
81. Heliövaara M, Reunanen A, Aromaa A, Knekt P, Aho K, Suhonen O. Validity of hospital discharge data in a prospective epidemiological study on stroke and myocardial infarction. *Acta Med Scand*. 1984;216(3):309–315. doi:10.1111/j.0954-6820.1984.tb03809.x
82. Hjerpe P, Merlo J, Ohlsson H, Bengtsson Boström K, Lindblad U. Validity of registration of ICD codes and prescriptions in a research database in Swedish primary care: a cross-sectional study in Skaraborg primary care database. *BMC Med Inform Decis Mak*. 2010;10(1). doi:10.1186/1472-6947-10-23
83. Ingelsson E, Ärnlöv J, Sundström J, Lind L. The validity of a diagnosis of heart failure in a hospital discharge register. *Eur J Heart Fail*. 2005;7(5):787–791. doi:10.1016/j.ejheart.2004.12.007
84. Joensuu T, Näyhä S. Reliability of hospital discharge diagnoses of acute myocardial infarction. *Scand J Soc Med*. 1992;20(2):85–86. doi:10.1177/140349489202000204
85. Kaspar M, Fette G, Güder G, et al. Underestimated prevalence of heart failure in hospital inpatients: a comparison of ICD codes and discharge letter information. *Clin Res Cardiol*. 2018;107(9):778–787. doi:10.1007/s00392-018-1245-z
86. Leone MA, Capponi A, Varrasi C, Tarletti R, Monaco F. Accuracy of the ICD-9 codes for identifying TIA and stroke in an Italian automated database. *Neurol Sci*. 2004;25(5):281–288. doi:10.1007/s10072-004-0355-8
87. Lindblad U, Råstam L, Ranstam J, Peterson M. Validity of register data on acute myocardial infarction and acute stroke: the Skaraborg Hypertension Project. *Scand J Soc Med*. 1993;21(1):3–9. doi:10.1177/140349489302100102
88. Nilsson AC, Spetz CL, Carsjö K, Nightingale R, Smedby B. [Reliability of the hospital registry. The diagnostic data are better than their reputation]. *Lakartidningen*. 1994;91(7):598,603–605. Swedish.
89. Øie LR, Madsbu MA, Giannadakis C, et al. Validation of intracranial hemorrhage in the Norwegian Patient Registry. *Brain Behav*. 2018;8(2):e00900. doi:10.1002/brb3.900
90. Rinaldi R, Vignatelli L, Galeotti M, Azzimondi G, de Carolis P. Accuracy of ICD-9 codes in identifying ischemic stroke in the General Hospital of Lugo di Romagna (Italy). *Neurol Sci*. 2003;24(2):65–69. doi:10.1007/s100720300074
91. Rodrigo-Rincon I, Martin-Vizcaino MP, Tirapu-Leon B, Zabalza-Lopez P, Abad-Vicente FJ, Merino-Peralta A. Validity of the clinical and administrative databases in detecting post-operative adverse events. *Int J Qual Health Care*. 2015;27(4):267–275. doi:10.1093/intqhc/mzv039
92. Sedova P, Brown RD, Zvolsky M, et al. Validation of stroke diagnosis in the National Registry of Hospitalized Patients in the Czech Republic. *J Stroke Cerebrovasc Dis*. 2015;24(9):2032–2038. doi:10.1016/j.jstrokecerebrovasdis.2015.04.019
93. Spolaore P, Brocco S, Fedeli U, et al. Measuring accuracy of discharge diagnoses for a region-wide surveillance of hospitalized strokes. *Stroke*. 2005;36(5):1031–1034. doi:10.1161/01.STR.0000160755.94884.4a
94. Valk MJ, Mosterd A, Broekhuizen BDL, et al. Overdiagnosis of heart failure in primary care: a cross-sectional study. *Br J Gen Pract*. 2016;66(649):e587–e592. doi:10.3399/bjgp16X685705
95. van Doorn S, Brakenhoff TB, Moons KGM, et al. The effects of misclassification in routine healthcare databases on the accuracy of prognostic prediction models: a case study of the CHA2DS2-VASc score in atrial fibrillation. *Diagnostic Progn Res*. 2017;1(1). doi:10.1186/s41512-017-0018-x
96. Verdú-Rotellar JM, Frigola-Capell E, Alvarez-Pérez R, et al. Validation of heart failure diagnosis registered in primary care records in two primary care centres in Barcelona (Spain) and factors related. A cross-sectional study. *Eur J Gen Pract*. 2017;23(1):107–113. doi:10.1080/13814788.2017.1305104
97. Gini R, Schuemie MJ, Mazzaglia G, et al. Automatic identification of type 2 diabetes, hypertension, ischaemic heart disease, heart failure and their levels of severity from Italian General Practitioners' electronic medical records: a validation study. *BMJ Open*. 2016;6(12):e012413. doi:10.1136/bmjopen-2016-012413
98. Bernal JL, Barrabés JA, Íñiguez A, et al. Clinical and administrative data on the research of acute coronary syndrome in Spain. Minimum basic data set validity. *Rev Española Cardiol (English Ed)*. 2019;72(1):56–62. doi:10.1016/j.rec.2018.01.026
99. Mähönen M, Salomaa V, Brommels M, et al. The validity of hospital discharge register data on coronary heart disease in Finland. *Eur J Epidemiol*. 1997;13(4):403–415. doi:10.1023/A:1007306110822
100. Palomäki P, Miettinen H, Mustaniemi H, et al. Diagnosis of acute myocardial infarction by MONICA and FINMONICA diagnostic criteria in comparison with hospital discharge diagnosis. *J Clin Epidemiol*. 1994;47(6):659–666. doi:10.1016/0895-4356(94)90213-5
101. Aboa-Eboulé C, Mengue D, Benzenine E, et al. How accurate is the reporting of stroke in hospital discharge data? A pilot validation study using a population-based stroke registry as control. *J Neurol*. 2013;260(2):605–613. doi:10.1007/s00415-012-6686-0
102. Ellekjaer H, Holmen J, Krüger O, Terent A. Identification of incident stroke in Norway: hospital discharge data compared with a population-based stroke register. *Stroke*. 1999;30(1):56–60. doi:10.1161/01.STR.30.1.56
103. Ho KKL, Pinsky JL, Kannel WB, Levy D. The epidemiology of heart failure: the Framingham Study. *J Am Coll Cardiol*. 1993;22(SUPPL. 4):6–43. doi:10.1016/0735-1097(93)90455-A
104. Carlson KJ, Lee DCS, Goroll AH, Leahy M, Johnson RA. An analysis of physicians' reasons for prescribing long-term digitalis therapy in outpatients. *J Chronic Dis*. 1985;38(9):733–739. doi:10.1016/0021-9681(85)90115-8
105. Yancy CW, Jessup M, Bozkurt B, et al. 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College Of Cardiology Foundation/American Heart Association Task Force On Practice Guidelines. *Circulation*. 2013;128(16). doi:10.1161/CIR.0b013e31829e8776

106. Tunstall-Pedoe H, Kuulasmaa K, Amouyel P, Arveiler D, Rajakangas AM, Pajak A. Myocardial infarction and coronary deaths in the World Health Organization MONICA project: registration procedures, event rates, and case-fatality rates in 38 populations from 21 countries in four continents. *Circulation*. 1994;90(1):583–612. doi:10.1161/01.cir.90.1.583
107. Thygesen K, Alpert JS, White HD. Universal definition of myocardial infarction. *Circulation*. 2007;116(22):2634–2653. doi:10.1161/CIRCULATIONAHA.107.187397
108. Thygesen K, Alpert JS, Jaffe AS, Simoons ML, Chaitman BR, White HD. Third universal definition of myocardial infarction. *Circulation*. 2012;126(16):2020–2035. doi:10.1161/CIR.0b013e31826e1058
109. Aho K, Harmsen P, Hatano S, Marquardsen J, Smirnov VE, Strasser T. Cerebrovascular disease in the community: results of a WHO collaborative study. *Bull World Health Organ*. 1980;58(1):113–130.
110. Thorvaldsen P, Kuulasmaa K, Rajakangas A-M, Rastenyte D, Sarti C, Wilhelmsen L. Stroke trends in the WHO MONICA project. *Stroke*. 1997;28(3):500–506. doi:10.1161/01.STR.28.3.500
111. Brenner H, Gefeller O. Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence. *Stat Med*. 1997;16(9):981–991. doi:10.1002/(SICI)1097-0258(19970515)16:9<981::AID-SIM510>3.0.CO;2-N
112. Yao RJR, Andrade JG, Deyell MW, Jackson H, McAlister FA, Hawkins NM. Sensitivity, specificity, positive and negative predictive values of identifying atrial fibrillation using administrative data: a systematic review and meta-analysis. *Clin Epidemiol*. 2019;11:753–767. doi:10.2147/CLEP.S206267
113. McGuinness LA, Warren-Gash C, Moorhouse LR, Thomas SL. The validity of dementia diagnoses in routinely collected electronic health records in the United Kingdom: a systematic review. *Pharmacoepidemiol Drug Saf*. 2019;28(2):244–255. doi:10.1002/pds.4669
114. Nissen F, Quint JK, Morales DR, Douglas IJ. How to validate a diagnosis recorded in electronic health records. *Breathe*. 2019;15(1):64–68. doi:10.1183/20734735.0344-2018
115. Hemingway H, Asselbergs FW, Danesh J, et al. Big data from electronic health records for early and late translational cardiovascular research: challenges and potential. *Eur Heart J*. 2018;39(16):1481–1495. doi:10.1093/eurheartj/ehx487
116. Banerjee A, Pasea L, Harris S, et al. Estimating excess 1-year mortality from COVID-19 according to underlying conditions and age in England: a rapid analysis using NHS health records in 3.8 million adults. *medRxiv*. 2020. doi:10.1101/2020.03.22.20040287

The appendices published with the Clinical Epidemiology article are included in **Chapter 11 Appendix 2**.

5.4. Deriving thesis definitions

The sensitivity and PPV findings from my systematic review show EHR recording of ACS was reasonably high. Stroke was similarly high overall, but sensitivity and PPV were variable by stroke subtype. Sensitivity for HF diagnoses was lower. There was considerable heterogeneity between studies. The studies included differed in many ways, as explained in the paper, particularly the populations and the codes included in diagnoses definitions.

I considered the systematic review findings in my study definitions and interpretation of results. Four studies validated ICD-10 codes in secondary care data in England [243–246]. One study reported a PPV of 96% for I50 recorded heart failure [244]. Two studies reported estimates for MI with high PPVs of 92% for I21-I23 and 89% for I21-I22 [243,245]. Two studies validated stroke, one of which combined I60-I64 with a PPV of 96% [246] while the other separately validated I60 for SAH, I61 for ICH and I63 for ischaemic stroke with PPVs of 96%, 78% and 86%, respectively [245].

Given the small number of studies that validated HES ICD-10 codes of interest and the overall variability in codes validated, I used broad definitions in my analysis to capture ACS, heart failure and stroke. In defining ACS, I included unstable angina (I20.0 and I24.0), MI (I21, including those further specified as I21.X) or other acute events (I24, I24.8 and I24.9). I defined heart failure using I50, including those specified as I50.1 or I50.9. I included all stroke subtypes in my definition, with codes G46 (including all G46.X), I60 (all I60.X), I61 (all I61.X), I62 (all I62.X), I63 (all I63.X) and I64. In addition, in my composite outcome of acute cardiovascular events, I included TIA and acute limb ischaemia as the mechanisms linking respiratory infections and ACS, heart failure, stroke, TIA and acute limb ischaemia would likely be similar across all the outcomes. TIA was defined coding codes G45 (including all G45.X)

and acute limb ischaemia using I70.2 (including all I70.2X), I70.8 2 (including all I70.8X) and I70.9 (including all I70.9X).

Eight studies included in my systematic review validated UK primary care EHR data from CPRD or THIN [243,247–253]. Although the sample was small, one study validated HF with a PPV of 100% [251], three studies validated MI diagnoses with PPV estimates of 85-93% and one sensitivity of 89% [243,248,253]. Finally, four studies validated stroke diagnoses with PPV estimates of 78-91% [247,249,250,252]. These results allowed me to be confident in identifying acute cardiovascular events using CPRD code lists.

In COVID-19 analysis, I redefined the definition of acute cardiovascular events from ACS, heart failure, stroke, TIA and acute limb ischaemia to ACS, heart failure, ischaemic stroke, and major ventricular arrhythmia. I additionally included major ventricular arrhythmia in the definition due to the reported increase in arrhythmia diagnosis post-COVID-19 [254]. See Chapter 8 for further detail. All analyses included CVD deaths in the composite acute cardiovascular event outcome. My first thesis analysis (Chapter 6) also included a sensitivity analysis definition of major adverse cardiovascular events (MACE) which included the most severe cardiovascular outcomes of MI, stroke, heart failure and CVD death.

5.5. Chapter summary

- To inform my definition of acute cardiovascular events and specific outcomes of heart failure, ACS and stroke, I carried out a systematic review on the validity of EHR recorded diagnoses.
- The review found 81 studies, of which 20 validated heart failure diagnoses, 31 validated acute coronary syndrome diagnoses with 29 specifically recording estimates for myocardial infarction, and 41 validated stroke diagnoses.
- The codes studies included to validate heart failure, ACS and stroke definitions varied. Still, overall sensitivity was ≤66% in all but one heart failure study, ≥80% for 91% of MI studies, and ≥70% for 73% of stroke studies. PPV was ≥80% in 74% of heart failure, 88% of MI, and 70% of stroke studies. PPV by stroke subtype was variable, at ≥80% for 80% of ischaemic stroke but only 44% of haemorrhagic stroke.
- Given the variability in definitions and accuracy of coded data, I included a broad range of acute cardiovascular events in my outcome definition; ACS (MI and unstable angina), heart failure, stroke, TIA and acute limb ischaemia.

Chapter 6 Quantifying the effect of cardiovascular risk on the rates of acute respiratory infections and subsequent cardiovascular complications

6.1. Chapter overview

In this chapter, I address thesis objective 2. I used CPRD GOLD and Aurum data to identify a cohort of patients without established CVD or a health condition making them eligible for influenza vaccination. I classified the individuals to be at low or raised cardiovascular risk, using diagnosed hypertension and QRISK2 score, to estimate the effect of raised cardiovascular risk on i ARI, ii acute cardiovascular events, and iii acute cardiovascular events after ARI.

The chapter begins with the research paper published, including supplementary materials, in Lancet Digital Health which addressed study objectives i and iii, as outlined above. I then outline detailed methodological decisions for all study objectives beyond the published article's inclusion scope. Finally, I detail the methods and results specific to the study objective ii.

6.2. Published paper

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	230019	**Title**	Ms
First Name(s)	Jennifer		
Surname/Family Name	Davidson		
Thesis Title	Acute respiratory infections, cardiovascular complications, and prevention among people with raised cardiovascular risk		
Primary Supervisor	Dr Charlotte Warren-Gash		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	Lancet Digital Health		
When was the work published?	December 2021		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	Not applicable		
Have you retained the copyright for the work?*	**Yes**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I am the first author of this research paper. With support of my co-authors, I designed the study, conducted the analysis, drafted and revised the manuscript.

SECTION E

Student Signature	[redacted]
Date	25/07/2022

Supervisor Signature	[redacted]
Date	25-07-2022

Risk of acute respiratory infection and acute cardiovascular events following acute respiratory infection among adults with increased cardiovascular risk in England between 2008 and 2018: a retrospective, population-based cohort study

Jennifer A Davidson, Amitava Banerjee, Liam Smeeth, Helen I McDonald, Daniel Grint, Emily Herrett, Harriet Forbes, Richard Pebody, Charlotte Warren-Gash

Summary

Background Although acute respiratory infections can lead to cardiovascular complications, the effect of underlying cardiovascular risk on the incidence of acute respiratory infections and cardiovascular complications following acute respiratory infection in individuals without established cardiovascular disease is unknown. We aimed to investigate whether cardiovascular risk is associated with increased risk of acute respiratory infection and acute cardiovascular events after acute respiratory infection using 10 years of linked electronic health record (EHR) data in England.

Methods In this retrospective, population-based cohort study we used EHRs from primary care providers registered on the Clinical Practice Research Datalink (CPRD) GOLD and Aurum databases in England. Eligible individuals were aged 40–64 years, did not have established cardiovascular disease or a chronic health condition that would make them eligible for influenza vaccination, were registered at a general practice contributing to the CPRD, and had linked Hospital Episode Statistics Admitted Patient Care data in England from Sept 1, 2008, to Aug 31, 2018. We classified cardiovascular risk on the basis of diagnosed hypertension and overall predicted cardiovascular risk, estimated by use of the QRISK2 risk-prediction tool (comparing a score of ≥10% [increased risk] with a score of <10% [low risk]). Using multivariable Poisson regression models, we calculated incidence rate ratios (IRRs) for systemic acute respiratory infection. Among individuals who had an acute respiratory infection, we used multivariable Cox regression to calculate hazard ratios (HRs) for the risk of acute cardiovascular events within 1 year of infection.

Findings We identified 6 075 321 individuals aged 40–64 years with data in the CPRD and linked data in the Hospital Episode Statistics Admitted Patient Care database between Sept 1, 2008, and Aug 31, 2018. Of these individuals, 4 212 930 (including 526 480 [12·5%] with hypertension and 607 087 [14·4%] with a QRISK2 score of ≥10%) were included in the assessment of the incidence of acute respiratory infection. After adjusting for confounders (age, sex, ethnicity, socioeconomic status, body-mass index, alcohol consumption, smoking status, and consultation frequency in the hypertension analysis; and alcohol consumption and consultation frequency in the QRISK2 analysis), the incidence of acute respiratory infection was higher in individuals with hypertension than those without (IRR 1·04 [95% CI 1·03–1·05]) and higher in those with a QRISK2 score of 10% or higher than in those with a QRISK2 score of less than 10% (1·39 [1·37–1·40]). Of the 442 408 individuals who had an acute respiratory infection, 4196 (0·9%) had an acute cardiovascular event within 1 year of infection. After adjustment (for age, sex, ethnicity, socioeconomic status, body-mass index, alcohol consumption, and smoking status in the hypertension analysis; and for alcohol consumption in the QRISK2 analysis), hypertension (HR 1·98 [95% CI 1·83–2·15]) and a QRISK2 score of 10% or higher (3·65 [3·42–3·89]) were associated with a substantially increased risk of acute cardiovascular events after acute respiratory infection.

Interpretation People with increased cardiovascular risk but without diagnosed cardiovascular disease, measured by diagnosed hypertension or overall predicted cardiovascular risk, could benefit from influenza and pneumococcal vaccine prioritisation to reduce their risk of both acute respiratory infection and cardiovascular complications following an acute respiratory infection.

Funding British Heart Foundation and the Wellcome Trust.

Lancet Digit Health 2021; 3: e773–83

Department of Non-Communicable Disease Epidemiology (J A Davidson MSc, Prof L Smeeth FRCGP, E Herrett PhD, H Forbes PhD, C Warren-Gash PhD), **Department of Infectious Disease Epidemiology** (H I McDonald PhD, D Grint PhD), **and National Institute for Health Research Health Protection Research Unit in Immunisation** (Prof L Smeeth, H I McDonald), **London School of Hygiene & Tropical Medicine, London, UK; Institute of Health Informatics, University College London, London, UK** (Prof A Banerjee DPhil); **Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK** (H Forbes); **National Infection Service, Public Health England, London, UK** (R Pebody PhD)

Correspondence to: Ms Jennifer Davidson, Department of Non-Communicable Disease Epidemiology, London School of Hygiene & Tropical Medicine, London WC1E 7HT, UK **jennifer.davidson@lshtm.ac.uk**

Introduction

The COVID-19 pandemic has expedited research on cardiovascular complications following systemic acute respiratory infection. Before the pandemic, observational studies showed that acute respiratory infections increased the risk of myocardial infarction and stroke. In self-controlled case-series using large electronic health record (EHR) datasets, risk of myocardial infarction and stroke

Research in context

Evidence before this study

Before the COVID-19 pandemic, self-controlled case-series estimated a 2–6 times transient increase in the risk of myocardial infarction and stroke following a range of clinically diagnosed or laboratory-confirmed acute respiratory infections. These studies did not explore whether these effects were modified by underlying cardiovascular risk. We searched PubMed using the search terms "*hypertensi*" OR "cardiovascular risk" AND "influenza" OR "pneumonia" OR "respiratory infection" AND "cardiovascular event" OR "myocardial infarction" OR "acute coronary syndrome" OR "stroke" OR "heart failure" for primary research studies and reviews investigating the effect of underlying cardiovascular risk on cardiovascular complications after acute respiratory infection published in any language, from database inception to Jan 25, 2021. Of 421 studies identified, three presented estimates for the effect of cardiovascular risk on cardiovascular complications after acute respiratory infection. One prospective cohort study of individuals with community-acquired pneumonia from five North American medical centres in 1991–94 showed that arterial hypertension was associated with an increased odds of cardiac complications (new or worsening heart failure, new or worsening arrhythmias, or myocardial infarction) within 30 days of a diagnosis of community-acquired pneumonia. The second study used clinical records from Beijing, China, to explore risk factors for cardiovascular complications after hospitalisation for community-acquired pneumonia in 2013–15, and found that patients hospitalised with community-acquired pneumonia who had cardiovascular complications had a significantly higher prevalence of hypertension than those without cardiovascular complications. The final study used UK primary care electronic health records (EHRs) to identify individuals who had a myocardial infarction or stroke in 1995–2004 and matched (on year of birth, sex, primary care practice, and calendar time) controls. The authors found that individuals who had an acute respiratory infection in the previous month had an increased odds of having a first myocardial infarction or stroke, regardless of background cardiovascular risk. The effect of underlying cardiovascular risk on cardiovascular complications after acute respiratory infection therefore remains unclear. We also did a search of PubMed on Sept 28, 2021, using the search terms "*hypertensi*" OR "cardiovascular risk" AND "influenza" OR "pneumonia" OR "respiratory infection". We searched for research articles or reviews published in any language between database inception and Jan 1, 2021. Of the 2175 studies identified, one presented estimates for the effect that cardiovascular risk has on the incidence of acute respiratory infections itself. The study used UK Biobank data to show that hypertension was independently associated with an increased risk of acute lower respiratory infections, particularly pneumonia.

Added value of this study

Using primary and secondary care EHRs from more than 4·2 million individuals in England, we found that the incidence of acute respiratory infection, particularly pneumonia, was higher in those with increased cardiovascular risk, defined as diagnosed hypertension or a QRISK2 10-year risk score of 10% or more. In addition, individuals with increased cardiovascular risk also had an elevated risk of acute cardiovascular events after an acute respiratory infection. This risk was more pronounced when QRISK2 score, which incorporates multiple factors associated with cardiovascular disease onset, was used rather than a diagnosis of hypertension alone. Therefore, QRISK2 provides a practical method to identify individuals at risk of cardiovascular complications after acute respiratory infection and to thereby prevent early-onset cardiovascular disease.

Implications of all the available evidence

Our analyses and the available evidence to date indicate that individuals with increased cardiovascular risk have a higher risk of acute respiratory infection, and that these infections are more likely to trigger acute cardiovascular events than in those without increased cardiovascular risk. These findings highlight the importance of managing and reducing cardiovascular risk to reduce acute respiratory infections and their consequences. Individuals with increased cardiovascular risk have not been considered at high risk of SARS-CoV-2 infection or COVID-19 complications in policy decisions; however, hypertension has been associated with an increased risk of severe outcomes in most analyses. Individuals with increased cardiovascular risk are not typically targeted for seasonal influenza or pneumococcal vaccines, and have not been prioritised in the roll-out of COVID-19 vaccines. The prevention or treatment of acute respiratory infection in individuals with increased cardiovascular risk, as well as in those with established cardiovascular disease, could reduce the risk of cardiovascular events.

was elevated by 2–6 times in the days following clinically diagnosed acute respiratory infection, with the risk remaining elevated for up to 1 month.[1,2] A range of organisms, including *Streptococcus pneumoniae* and the influenza virus, are known to trigger cardiovascular events.[3,4]

Observational studies and the few randomised controlled trials (RCTs) published to date show that pneumococcal and influenza vaccines reduce the risk of cardiovascular complications when compared with placebo or no vaccine.[5–7] A meta-analysis of secondary prevention RCTs found that influenza vaccination reduced cardiovascular mortality by 55%.[7] Pneumococcal vaccination reduced the odds of acute coronary syndrome by 17% in a meta-analysis of observational studies conducted among individuals aged 65 years or older.[5]

In Europe, people with underlying cardiovascular risk but without established cardiovascular disease are not typically recommended to receive seasonal influenza or pneumococcal vaccines.[8–10] Public Health England

recommends a one-time polysaccharide pneumococcal vaccine and seasonal influenza vaccines for adults aged 65 years or older and for those aged younger than 65 years with specific underlying health conditions, including chronic heart disease.[9,10]

An association between high blood pressure and cardiovascular complications has been observed in individuals as young as 40 years of age.[11] Blood pressure is an essential predictor of cardiovascular risk, but it is only one component. Cardiovascular risk scores such as the QRISK2[12] are increasingly used to predict an individual's likelihood, most commonly the 10-year risk, of future cardiovascular disease on the basis of multiple factors. The National Institute for Health and Care Excellence recommends the use of the QRISK2 score for assessment and management of cardiovascular disease.[13]

Little is known about the effect of cardiovascular risk on the risk of acute respiratory infection; however, a recent UK Biobank study showed that high blood pressure increased the risk of acute respiratory infection, particularly pneumonia and influenza.[14] Similarly, the role of cardiovascular risk in the association between acute respiratory infection and cardiovascular complications has only been sporadically studied. We identified two previous studies that reported an association between hypertension and cardiovascular complications after pneumonia.[15,16] Another previous study investigated the association between overall cardiovascular risk and myocardial infarction and stroke after acute respiratory infection.[17] This previous study reported an increased risk of myocardial infarction and stroke after acute respiratory infection, regardless of cardiovascular risk level. Cardiovascular risk was derived from associations identified in univariable analysis (angina, hypertension, hyperlipidaemia, diabetes, chronic heart disease among first degree relatives, peripheral vascular disease, smoking status, and previous stroke for myocardial infarction, and hypertension, diabetes, peripheral vascular disease, smoking status, previous myocardial infarction, and urinary tract infection for stroke) rather than a pre-existing risk score.[17]

In this study, we aimed to investigate how cardiovascular risk, defined by diagnosed hypertension or QRISK2 score, is associated with the risk of acute respiratory infection and acute cardiovascular events after acute respiratory infection using 10 years of linked EHR data in England.

Methods

Study design and data sources

We did a retrospective cohort study using routinely collected primary care data from the Clinical Practice Research Datalink (CPRD). The CPRD collects anonymised coded data from general practices in England through the GOLD database using the Vision clinical management system and Aurum database using the EMIS clinical management system.[18,19] The databases include more than 35 million individuals, representative of the UK population in terms of age, sex, and ethnicity. The coded data collected include diagnoses, prescriptions, immunisations, and basic demographic characteristics.

Consenting general practices have patient records linked to other data sources. We used linked secondary care data from the Hospital Episode Statistics Admitted Patient Care (HES APC) database, data on deaths from the Office for National Statistics (ONS), and individual-level Townsend deprivation index scores. HES APC data contain information on diagnoses made and procedures done for all National Health Service (NHS) admissions in England, coded with the International Classification of Diseases version 10 (ICD-10).[20] The ONS mortality data contain information on the date and cause of death, also coded with the ICD-10.

The CPRD Independent Scientific Advisory Committee approved this study (19_209). The CPRD provided data on relevant HES APC, ONS, and Townsend index variables for the study population. The London School of Hygiene & Tropical Medicine provided ethical approval (17894).

The study protocol and data analysis plan are available in the appendix (pp 21–31).

See Online for appendix

Participants

Eligible individuals were aged 40–64 years, not already recommended for seasonal influenza vaccination according to current UK guidelines,[9] and registered at a general practice contributing to the CPRD (via GOLD or Aurum) with linked HES APC data in England from Sept 1, 2008, to Aug 31, 2018. This time period covered the duration of QRISK2 use. We started follow-up in September as this is when primary care practices identify individuals eligible for the seasonal influenza vaccine. We selected individuals aged 40–64 years to include those who had an increased likelihood of having a high cardiovascular risk or acute cardiovascular events, or both, but who were younger than the age cutoff for universal influenza vaccination. We defined the start of follow-up as Sept 1, 2008, the individual's 40th birthday, or the research standard CPRD date (12 months after current registration in the Aurum dataset, and the latest of 12 months after the current registration or the practice research standard in the GOLD dataset), whichever came first.

We excluded individuals who had established cardiovascular disease, received a previous pneumococcal vaccine, received an influenza vaccine within the previous 12 months, or a chronic health condition making them eligible for influenza vaccination, recorded in the CPRD at baseline. We defined conditions recommended for influenza vaccination as chronic liver disease, chronic respiratory disease, stage 3–5 chronic kidney disease, diabetes, asplenia or other splenic dysfunction, chronic neurological conditions, severe obesity (ie, a body-mass index [BMI] of ≥40 kg/m²), or an

immunosuppressive condition.[9] Full exclusion criteria are provided in the appendix (p 2).

Individual patient consent was not required to collect the original underlying data because data were generated during routine clinical encounters. Consent was given at the general practice level for anonymised primary care data to be used for research through contributing to the CPRD. Patients can advise their general practice if they wish to opt out of data collection. For our research we have obtained the data from the CPRD following their approval process.

Outcomes and variables

We used the whole study population to investigate the incidence of acute respiratory infection. We ended follow-up at diagnosis of cardiovascular disease or a chronic condition with which an individual would be recommended for influenza vaccination, receipt of pneumococcal or influenza vaccination, death, transfer out of general practice, the date of last data collection from general practice, the individual's 65th birthday, or Aug 31, 2018, whichever came first (appendix p 3).

We defined acute respiratory infection as a clinical or confirmed diagnosis of pneumonia, acute bronchitis, influenza, influenza-like illness, or other acute infections suggestive of lower respiratory tract involvement in the CPRD or HES APC record. Selected diagnostic codes were based on the code lists used in a previous study.[2] We did not include symptoms in our definition. In further analyses, we analysed the incidence of influenza or influenza-like illness and pneumonia separately. For each individual, we grouped acute respiratory infection records within 28 days into a single episode (see appendix [p 3] for full details).

Within the study population, individuals with an acute respiratory infection were followed up from diagnosis to investigate the incidence and risk of acute cardiovascular events after this infection. Follow-up for this study population ended at the occurrence of an acute cardiovascular event, death, transfer out of general practice, the date of last data collection from general practice, 1 year after diagnosis of the acute respiratory infection, or Aug 31, 2018, whichever came first (appendix p 3).

For our main analysis, we used a broad definition for acute cardiovascular events of acute coronary syndrome (myocardial infarction and unstable angina), left ventricular heart failure, stroke or transient ischaemic attack, acute limb ischaemia, or cardiovascular death. We included diagnoses recorded in the CPRD or HES APC, with the codes used informed by previous studies,[21] and cardiovascular deaths (ICD-10 codes I00–I99) recorded by the ONS. In further analyses, we also assessed each cardiovascular condition separately.

We considered two measures of cardiovascular risk: diagnosed hypertension and QRISK2 score. QRISK2 is a prediction algorithm that estimates an individual's 10-year risk of cardiovascular disease.[12] The risk factors included and the process we used to calculate QRISK2 scores are outlined in the appendix (p 4). Briefly, a score is calculated using a range of risk factors, such as age, sex, ethnicity, socioeconomic status, comorbid health conditions, BMI, blood pressure reading, and smoking status. QRISK2 is not widely used outside the UK; therefore, we also included hypertension as a pragmatic definition of cardiovascular risk. To ensure that we included only individuals with persistent and diagnosed hypertension, we used coded CPRD diagnoses with no time limit.

Individuals were classified as having increased cardiovascular risk (ie, diagnosed hypertension or a QRISK2 score of ≥10%) or not (no hypertension or a QRISK2 score of <10%) at baseline, and their risk was updated if hypertension was recorded or new measures relevant to the QRISK2 algorithm resulted in a change in QRISK2 score from less than 10% to 10% or higher during follow-up.

We considered demographics, lifestyle factors, and primary care consultation frequency in analyses of hypertension and acute respiratory infection. The demographic features were age (5-year bands of 40–44, 45–49, 50–54, 55–59, and 60–64 years), sex (male and female), race or ethnicity (White, Black, south Asian, and mixed or other), and socioeconomic status (individual-level Townsend score data grouped into quintiles, ranging from least deprived [quintile 1] to most deprived [quintile 5]). The lifestyle factors were baseline alcohol consumption (heavy drinking [defined as either a recorded intake of >42 units per week or a diagnostic code suggestive of alcohol addiction or excessive alcohol consumption] or no known heavy drinking), smoking status (current smoker, never smoker, or former smoker), and BMI (underweight [BMI <18·5 kg/m²], normal [BMI 18·5–24·9 kg/m²], overweight [BMI 25·0–29·9 kg/m²], or obese [BMI 30·0–39·9 kg/m²]). Consultation frequency was derived from the number of in-person or telephone consultations in the year before baseline. For analyses involving QRISK2 scores, we included only alcohol consumption and consultation frequency, as all other factors are included in the QRISK2 algorithm. We identified these included covariates using CPRD data, with data on race or ethnicity additionally collected from the HES APC database.

Our analyses of acute cardiovascular events after acute respiratory infection included the aforementioned demographic and lifestyle factors. We stratified results by statin, antihypertensive, or antiplatelet prescriptions recorded in the CPRD in the year before infection. Our rationale for covariate and effect-modifier selection is explained in the appendix (p 4).

Statistical analysis

We pooled individual-level data from the CPRD GOLD and Aurum databases. We calculated age-stratified

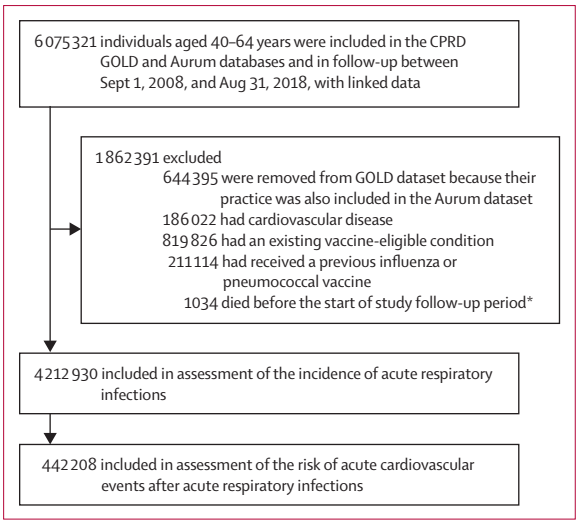


***Figure 1*: Study profile**
CPRD=Clinical Practice Research Datalink. *Identified from linked Office for National Statistics mortality data.

	Cohort (n=4 212 930)
Age, years	
40–44	1 920 369 (45·6%)
45–49	782 897 (18·6%)
50–54	612 202 (14·5%)
55–59	490 619 (11·6%)
60–64	406 843 (9·7%)
Sex	
Male	2 226 561/4 212 898 (52·9%)
Female	1 986 337/4 212 898 (47·1%)
Race or ethnicity	
White	3 242 107/3 702 718 (87·6%)
South Asian	194 931/3 702 718 (5·3%)
Black	154 270/3 702 718 (4·2%)
Mixed or other	111 410/3 702 718 (3·0%)
Townsend quintile	
1 (least deprived)	1 004 670/4 207 605 (23·9%)
2	905 691/4 207 605 (21·5%)
3	825 679/4 207 605 (19·6%)
4	739 187/4 207 605 (17·6%)
5 (most deprived)	732 378/4 207 605 (17·4%)
BMI category*†	
Underweight	51 002/3 447 604 (1·5%)
Normal weight	1 449 683/3 447 604 (42·0%)
Overweight	1 274 701/3 447 604 (37·0%)
Obese	672 218/3 447 604 (19·5%)
Smoking status*	
Non-smoker	1 686 919/4 082 791 (41·3%)
Current smoker	1 076 707/4 082 791 (26·4%)
Former smoker	1 319 165/4 082 791 (32·3%)
Alcohol consumption*	
No known heavy drinking	3 477 336/3 674 494 (94·6%)
Heavy drinking‡	197 158/3 674 494 (5·4%)

Data are n (%) or n/N (%). BMI=body-mass index. *Closest measure before the start of follow-up. †Underweight was defined as a BMI of <18·5 kg/m², normal weight as a BMI of 18·5–24·9 kg/m², overweight as a BMI of 25·0–29·9 kg/m², and obese as a BMI of 30·0–39·9 kg/m². ‡Defined as either a recorded intake of more than 42 units per week or a diagnostic code suggestive of alcohol addiction or excessive alcohol consumption.

***Table 1*: Baseline demographic and lifestyle characteristics**

acute respiratory infection incidence rates by cardiovascular risk. We used multivariable Poisson regression to calculate incidence rate ratios (IRRs). To account for multiple episodes of acute respiratory infection per individual, the random-effects models included time-updated exposure by age and cardiovascular risk. We initially adjusted for age and sex before adjusting for additional covariates (age, sex, ethnicity, socioeconomic status, BMI, alcohol intake, smoking status, and consultation frequency in the hypertension analysis; and alcohol consumption and consultation frequency in the QRISK2 analysis). A complete case-analysis approach was used in multivariable analyses. We did not do multiple imputation because data in CPRD are unlikely to be missing at random. Among individuals with an acute cardiovascular event within 1 year of acute respiratory infection, we summarised (using medians [IQRs]) the time between the infection and event. To assess the effect of cardiovascular risk on acute cardiovascular events after acute respiratory infection, we used multivariable Cox proportional hazards regression finely adjusted for time under follow-up. We used robust standard errors and initially adjusted for age and sex before adjusting for additional covariates (age, sex, ethnicity, socioeconomic status, BMI, alcohol consumption, and smoking status in the hypertension analysis; and alcohol intake in the QRISK2 analysis). In all analyses, for individuals who entered and exited follow-up on the same day, we added 1 day to ensure all individuals contributed at least 1 day of follow-up. Additionally, we separately added an interaction term to our Cox random-effects models for statin, antihypertensive, or antiplatelet prescriptions, and compared the results to models without interaction.

We did five prespecified sensitivity analyses. First, we validated results obtained from combined individual-level GOLD and Aurum data by analysing each database separately before combining them with a random-effects meta-analysis. Between-database heterogeneity was assessed by use of the I^2 statistic. Second, we excluded only individuals eligible for both pneumococcal and influenza vaccines (see appendix p 2).[9,10] Third, we repeated the analyses restricted to individuals with a QRISK2 score recorded by their general practitioner and collected in the CPRD from Jan 1, 2015, to Dec 31, 2017 (see appendix p4 for further detail of why this restricted time period was used), to compare results with those of our calculated QRISK2 scores and ensure consistent

	Number of events	Incidence per 1000 person-years (95% CI)	Crude IRR (95% CI)	Age and sex-adjusted IRR (95% CI)	Fully adjusted* IRR (95% CI)
Acute respiratory infection					
Hypertension	77 674	40·3 (40·0–40·7)	1·38 (1·36–1·39)	1·33 (1·32–1·34)	1·04 (1·03–1·05)
No hypertension	508 473	29·1 (29·0–29·2)	1 (ref)	1 (ref)	1 (ref)
QRISK2 ≥10%	81 662	43·8 (43·4–44·2)	1·52 (1·50–1·53)	NA	1·39 (1·37–1·40)
QRISK2 <10%	504 485	28·8 (28·7–28·9)	1 (ref)	NA	1 (ref)
Influenza or influenza-like illness					
Hypertension	12 050	6·3 (6·1–6·4)	1·14 (1·11–1·16)	1·25 (1·22–1·27)	0·98 (0·96–1·00)
No hypertension	95 589	5·5 (5·4–5·5)	1 (ref)	1 (ref)	1 (ref)
QRISK2 ≥10%	10 010	5·4 (5·3–5·5)	0·96 (0·94–0·98)	NA	0·88 (0·86–0·90)
QRISK2 <10%	97 629	5·6 (5·5–5·6)	1 (ref)	NA	1 (ref)
Pneumonia					
Hypertension	4479	2·3 (2·2–2·4)	1·59 (1·53–1·65)	1·32 (1·27–1·38)	1·12 (1·07–1·16)
No hypertension	26 589	1·5 (1·5–1·5)	1 (ref)	1 (ref)	1 (ref)
QRISK2 ≥10%	6476	3·5 (3·4–3·6)	2·60 (2·52–2·69)	NA	2·32 (2·25–2·40)
QRISK2 <10%	24 592	1·4 (1·4–1·4)	1 (ref)	NA	1 (ref)

Total person-years per 1000 years of follow-up available was 1926·2 for hypertension, 17 467·9 for no hypertension, 1865·1 for a QRISK2 score of 10% or higher, and 17 529·1 for a QRISK2 score of less than 10%. Likelihood ratio test p values for all comparisons were less than 0·0001. IRR=incidence rate ratio. NA=not applicable. *Hypertension models were adjusted for age, sex, race or ethnicity, socioeconomic status, body-mass index, alcohol consumption, smoking status, and consultation frequency; and QRISK2 models were adjusted for alcohol consumption and consultation frequency.

Table 2: **Association between cardiovascular risk and acute respiratory infection**

results were obtained. Fourth, we included only major adverse cardiovascular events of myocardial infarction, heart failure, stroke, and cardiovascular death in the cardiovascular event outcome. Finally, we repeated the analysis of acute cardiovascular events after acute respiratory infection excluding individuals who received pneumococcal or influenza vaccines during follow-up.

All statistical analyses were done using Stata, version 16.

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

6 075 321 individuals aged 40–64 years included in the GOLD and Aurum databases were in follow-up and had linked data in the HES APC between Sept 1, 2008, and Aug 31, 2018, of whom 4 212 930 eligible individuals (773 362 from the GOLD and 3 439 568 from the Aurum database) were included in the final cohort (figure 1). Individuals were followed up for a median of 3·9 years (IQR 1·6–7·6), 2 226 561 (52·9%) of 4 212 898 with available data on sex were men, 1 986 337 (47·1%) were women, and the median age was 46 years (IQR 40–53; table 1). Demographic and lifestyle characteristics of individuals in the GOLD and Aurum datasets were similar, except that higher proportions of individuals in the Aurum dataset were non-White (399 150 [13·0%] of 3 060 297 *vs* 61 461 [9·6%] of 642 421), residing in more deprived regions (627 333 [18·3%] of 3 434 620 *vs* 105 045 [13·6%] of 772 985), and current or former smokers (2 002 856 [60·1%] of 3 334 866 *vs* 393 016 [52·6%] of 747 925) than in the GOLD dataset (appendix p 5).

526 480 (12·5%) of 4 212 930 individuals had diagnosed hypertension: 347 418 at baseline and a further 179 062 during follow-up. 607 087 (14·4%) individuals had a QRISK2 score of 10% or higher: 402 594 at baseline, and a further 204 493 had a score that increased to 10% or higher during follow-up. 239 184 (5·7%) individuals had both hypertension and a QRISK score of 10% or higher.

586 147 episodes of acute respiratory infection were recorded among 442 408 individuals: 107 639 episodes of influenza or influenza-like illness and 31 068 episodes of pneumonia. The incidence of all-cause acute respiratory infection and pneumonia increased with age, with the exception of the incidence of acute respiratory infection among those with a QRISK2 score of 10% or higher, for which there was no age trend, and the incidence of influenza or influenza-like illness decreased with age (appendix p 18). The incidence of acute respiratory infection was higher among individuals with hypertension (40·3 infections per 1000 person-years [95% CI 40·0–40·7]) and a QRISK2 score of 10% or higher (43·8 infections per 1000 person-years [43·4–44·2]) than among those without hypertension (29·1 infections per 1000 person-years [29·0–29·2]) and a QRISK2 score of less than 10% (28·8 infections per 1000 person-years [28·7–28·9]; table 2).

In the unadjusted analysis, there was an increased incidence of acute respiratory infection among individuals with hypertension compared with those without hypertension (IRR 1·38 [95% CI 1·36–1·39]) and similarly for individuals with a QRISK2 score of 10% or higher compared with those with a QRISK2 score of less than 10% (1·52 [1·50–1·53]). After adjustment for confounders, the association between hypertension and acute respiratory infection was substantially reduced (IRR 1·04 [95% CI 1·03–1·05]), but the association between a QRISK2 score of 10% or higher and acute respiratory infection remained (1·39 [1·37–1·40]; table 2). The increased incidence of pneumonia among individuals with hypertension compared with those without hypertension (1·12 [1·07–1·16]), and among individuals with a QRISK2 score of 10% or higher compared with those with a QRISK2 score of less than 10% (2·32 [2·25–2·40]) was more pronounced than for all-cause acute respiratory infection (table 2). For influenza and influenza-like illness, the incidence rate ratio was lower among individuals with hypertension than among those without hypertension (0·98 [0·96–1·00]), and lower among individuals with a QRISK2 score of 10% or higher than among those with a QRISK2 score of less than 10% (0·88 [0·86–0·90]; table 2).

The sensitivity analyses, in which we analysed CPRD GOLD and Aurum datasets separately (appendix p 6), redefined the study population (appendix p 7), and compared recorded and calculated QRISK2 scores (appendix p 8), showed similar results to those of the primary analysis.

Among the 442 408 individuals (with 526 800 acute respiratory infection episodes) who had an acute respiratory infection, 4169 (0·9%) had an acute cardiovascular event within 1 year. 1606 (38·5%) of these individuals had pneumonia (appendix p 10).

985 (11·2 events per 1000 person-years [95% CI 10·5–12·0]) acute cardiovascular events occurred in individuals with hypertension, 3184 (5·8 events per 1000 person-years [5·6–6·0]) in those without hypertension, 1526 (17·5 events per 1000 person-years [16·7–18·5]) in individuals with a QRISK2 score of 10% or higher, and 2643 (4·8 events per 1000 person-years [4·6–5·0]) in those with a QRISK2 score of less than 10% (table 3). The number of acute cardiovascular events that occurred in individuals after influenza or influenza-like illness and pneumonia, according to cardiovascular risk, is shown in the appendix (pp 9–10).

2106 (50·5%) of the 4169 acute cardiovascular events occurred within 30 days of an acute respiratory infection (appendix p 19), but the median time interval was slightly longer in individuals with hypertension (median 41 days [IQR 1–168]) or a QRISK2 score of 10% or higher (31 days [1–60]) compared with those without hypertension (24 days [1–156]) or a QRISK2 score of less than 10% (26 days [1–159]). The time interval between pneumonia and an acute cardiovascular event was shorter than for all-cause acute respiratory infection, whereas the time interval for influenza or influenza-like illness was longer (appendix p 19).

After adjustment for confounders, hypertension (hazard ratio [HR] 1·98 [95% CI 1·83–2·15]) and a QRISK2 score of 10% or higher (3·65 [3·42–3·89]) remained associated with acute cardiovascular events after acute respiratory infection (table 3). When acute cardiovascular events were analysed separately, the HRs were largest for acute limb ischaemia, even though there were only a small number of events (hypertension *vs* no hypertension HR 4·63 [95% CI 2·68–7·99]; QRISK2 score of ≥10% *vs* <10% 6·93 [4·43–10·83]), acute coronary syndrome (2·13 [1·86–2·44]; and 4·37 [3·93–4·86]), and cardiovascular death (2·15 [1·69–2·73]; and 4·81 [3·99–5·81]).

In the analysis stratified by prescription of antihypertensive, statin, or antiplatelet drugs, there was no association between hypertension status and acute cardiovascular events after acute respiratory infection among individuals with a prescription (antihypertensives HR 1·00 [95% CI 0·84–1·18], statins 1·09 [0·91–1·30], antiplatelets 0·86 [0·65–1·13]), but the association remained for those without a prescription (no antihypertensives 2·62 [2·37–2·89], no statins 2·14 [1·98–2·32], no antiplatelets 2·05 [1·91–2·21]). A QRISK2 score of 10% or higher remained associated with acute cardiovascular events after acute respiratory infection in patients with and without a prescription of antihypertensives, statins, or antiplatelets, but was higher among individuals without a prescription (antihypertensives 2·45 [2·11–2·85] *vs* no antihypertensives 4·04 [3·76–4·34], statins 2·22 [1·83–2·69] *vs* no statins 3·93 [3·67–4·21], antiplatelets 1·82 [1·36–2·44] *vs* no antiplatelets 3·68 [3·45–3·93]; figure 2; appendix p 11). The interactions between cardiovascular risk and

	Number of events	Incidence per 1000 person-years (95% CI)	Crude HR (95% CI)	Age and sex-adjusted HR (95% CI)	Fully adjusted* HR (95% CI)
Any cardiovascular event					
Hypertension	985	11·2 (10·5–12·0)	2·08 (1·93–2·23)	1·97 (1·84–2·12)	1·98 (1·83–2·15)
No hypertension	3184	5·8 (5·6–6·0)	1 (ref)	1 (ref)	1 (ref)
QRISK2 ≥10%	1526	17·5 (16·7–18·5)	3·74 (3·51–3·98)	NA	3·65 (3·42–3·89)
QRISK2 <10%	2643	4·8 (4·6–5·0)	1 (ref)	NA	1 (ref)
Acute coronary syndrome†					
Hypertension	372	4·2 (3·8–4·7)	2·19 (1·95–2·46)	2·06 (1·83–2·32)	2·13 (1·86–2·44)
No hypertension	1140	2·1 (1·9–2·2)	1 (ref)	1 (ref)	1 (ref)
QRISK2 ≥10%	613	7·0 (6·5–7·6)	4·42 (3·98–4·89)	NA	4·37 (3·93–4·86)
QRISK2 <10%	899	1·6 (1·5–1·7)	1 (ref)	NA	1 (ref)
Heart failure					
Hypertension	290	3·3 (2·9–3·7)	2·04 (1·79–2·32)	1·92 (1·69–2·19)	2·08 (1·79–2·42)
No hypertension	961	1·7 (1·6–1·9)	1 (ref)	1 (ref)	1 (ref)
QRISK2 ≥10%	478	5·5 (5·0–6·0)	4·00 (3·57–4·49)	NA	3·85 (3·42–4·34)
QRISK2 <10%	773	1·4 (1·3–1·5)	1 (ref)	NA	1 (ref)
Acute limb ischaemia					
Hypertension	25	0·3 (0·2–0·4)	2·98 (1·85–4·78)	2·82 (1·74–4·55)	4·63 (2·68–7·99)
No hypertension	55	0·1 (0·1–0·1)	1 (ref)	1 (ref)	1 (ref)
QRISK2 ≥10%	42	0·5 (0·4–0·7)	7·55 (4·62–11·07)	NA	6·93 (4·43–10·83)
QRISK2 <10%	38	0·1 (0·1–0·1)	1 (ref)	NA	1 (ref)
Stroke or transient ischaemic attack‡					
Hypertension	360	4·1 (3·7–4·6)	2·15 (1·91–2·42)	2·08 (1·84–2·34)	2·01 (1·75–2·29)
No hypertension	1120	2·0 (1·9–2·1)	1 (ref)	1 (ref)	1 (ref)
QRISK2 ≥10%	468	5·4 (4·9–5·9)	2·99 (2·68–3·34)	NA	2·93 (2·62–3·28)
QRISK2 <10%	1012	1·8 (1·7–1·9)	1 (ref)	NA	1 (ref)
Cardiovascular-related death					
Hypertension	129	1·5 (1·2–1·8)	2·11 (1·73–2·58)	1·99 (1·63–2·43)	2·15 (1·69–2·73)
No hypertension	413	0·7 (0·7–0·8)	1 (ref)	1 (ref)	1 (ref)
QRISK2 ≥10%	230	2·6 (2·3–3·0)	4·77 (4·03–5·66)	NA	4·81 (3·99–5·81)
QRISK2 <10%	312	0·6 (0·5–0·6)	1 (ref)	NA	1 (ref)

Total person-years per 1000 years of follow-up available was 87·9 for hypertension, 553·6 for no hypertension, 87·0 for a QRISK2 score of 10% or higher, and 554·4 for a QRISK2 score of less than 10%. Likelihood ratio test p values for all comparisons were less than 0·0001. HR=hazard ratio. NA=not applicable. *Hypertension models were adjusted for age, sex, race or ethnicity, socioeconomic status, body-mass index, alcohol consumption, and smoking status; QRISK2 models were adjusted for alcohol consumption. †Fully adjusted HR for myocardial infarction alone was 2·22 (95% CI 1·91–2·59) in the hypertension model and 4·89 (4·35–5·51) in the QRISK2 model; fully adjusted HR for angina alone was 2·03 (1·51–2·72) in the hypertension model and 3·06 (2·38–3·93) in the QRISK2 model. ‡Fully adjusted HR for stroke was 2·10 (95% CI 1·81–2·43) in the hypertension model and 2·90 (2·56–3·30) in the QRISK2 model; and fully adjusted HR for transient ischaemic stroke alone was 1·77 (1·34–2·33) in the hypertension model and 3·10 (2·46–3·90) in the QRISK2 model.

Table 3: **Incidence and risk of acute cardiovascular events after acute respiratory infection by cardiovascular risk group**

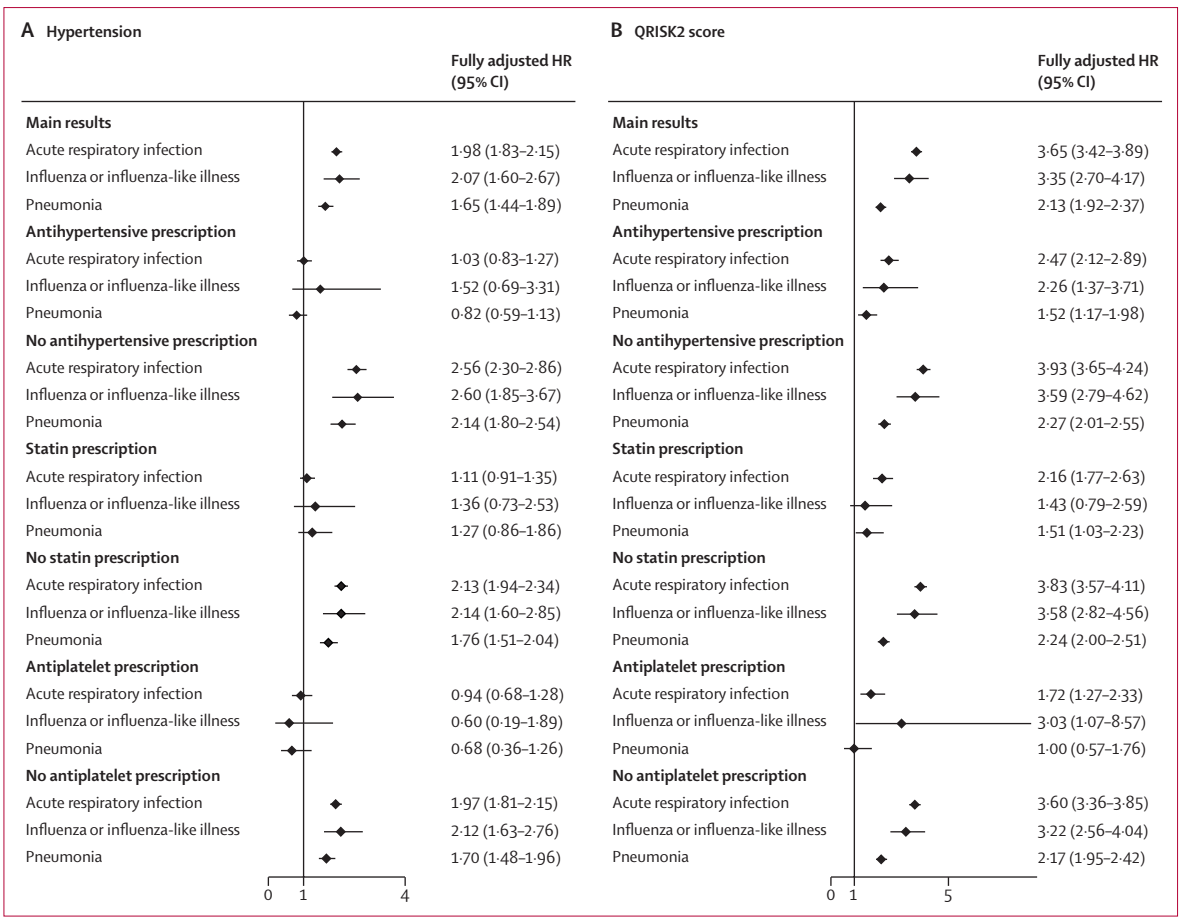


Figure 2: **Risk of an acute cardiovascular event after an acute respiratory infection in the hypertension model (A) and in the QRISK2 score model (B) by stratifying factor**

HRs were adjusted for age, sex, race or ethnicity, socioeconomic status, body-mass index, alcohol intake, and smoking status in the hypertension model, and for alcohol intake in the QRISK2 score model. HR=hazard ratio.

prescription of antihypertensive, statin and antiplatelet drugs were significant (likelihood ratio test p<0·0001).

The sensitivity analyses, in which we analysed the CPRD GOLD and Aurum datasets separately (appendix pp 12–13), redefined the study population (appendix p 14), compared recorded and calculated QRISK2 scores (appendix p 15), redefined the outcome to include only major adverse cardiovascular events (appendix p 16), and excluded individuals who received influenza or pneumococcal vaccines during the follow-up period (appendix p 17), showed similar results to those of the primary analysis.

Discussion

In our study population of more than 4·2 million adults aged 40–64 years who did not have chronic health conditions and were ineligible for influenza vaccination, the incidence of both acute respiratory infection and acute cardiovascular events after acute respiratory infection was increased in individuals with elevated cardiovascular risk, measured by diagnosed hypertension or a QRISK2 score of 10% or higher. The incidence of acute respiratory infection was 1·4–1·5 times higher among individuals with increased cardiovascular risk than in those without. After adjustment for confounders, this increase in incidence was marginal when risk was defined by diagnosed hypertension and more substantial when defined by QRISK2 scores. We observed a substantially larger increase in risk of an acute cardiovascular event after respiratory infection when individuals were stratified by QRISK2 score (3·7 times) than by diagnosed hypertension (2·0 times). Associations were similar for all-cause acute respiratory infection, influenza-like illness, and pneumonia. Half of the acute cardiovascular events occurred within 30 days of acute respiratory infection.

Our finding that the incidence of acute respiratory infection was elevated among individuals with increased cardiovascular risk is consistent with the results of a recent UK Biobank study, which examined the association between blood pressure and the risk of different respiratory infections.[14] Participants with prevalent

hypertension had an increased risk of pneumonia (HR 1·36 [95% CI 1·29–1·43]), influenza or viral pneumonia (1·12 [1·01–1·23]), and other lower respiratory infections (1·15 [1·11–1·19]). In our study, the increased incidence of acute respiratory infection among individuals with hypertension, measured by recorded diagnoses, was smaller than observed in the UK Biobank study. Instead, we found more substantial increases in the incidence of acute respiratory infection in individuals with a QRISK2 score of 10% or higher, which includes systolic blood pressure reading and use of antihypertensive drugs. Only 39% (239 194 of 607 087) of individuals with a QRISK2 score of 10% or higher had diagnosed hypertension. Other UK cohort studies have reported increases in the incidence of influenza-like illness to be associated with non-White ethnicity and social deprivation,[22,23] both of which are cardiovascular risk factors included in the QRISK2 algorithm and can result in higher QRISK2 scores assigned to individuals with these characteristics.

Previous studies have reported transient increases in the risk of myocardial infarction and stroke after acute respiratory infection.[1–4] In the few published observational studies that consider the effect of cardiovascular risk on this association, the results were mixed. In two studies, hypertension was associated with an increase in cardiovascular complications after pneumonia.[15,16] In another study, there was no difference in the likelihood of first myocardial infarction or stroke occurring after acute respiratory infection when stratified by the presence of cardiovascular risk factors.[17] Studies done during the COVID-19 pandemic have generally shown a high prevalence of cardiovascular disease risk factors in people hospitalised or those who died due to COVID-19. Pooled results suggest that the risk of severe COVID-19 outcomes is 2–3 times higher among hospitalised individuals with hypertension or diabetes than those without hypertension or diabetes.[24] The degree to which age drives these associations, especially for hypertension, is unclear, with age-adjusted analyses showing varying results.[24]

Several mechanisms could explain the associations between cardiovascular risk factors, acute respiratory infection, and cardiovascular disease. Hypertension could promote immune dysregulation, leading to infection, or the endothelial dysfunction caused by hypertension, hyperlipidaemia, and diabetes could promote infection.[25,26] Infectious agents, such as SARS-CoV-2, the influenza virus, and *S pneumoniae*, could also exacerbate atherosclerotic processes. The infectious agent could have direct effects on vascular cells, or the infection could induce haemodynamic, inflammatory, and procoagulant processes. The release of proinflammatory cytokines in response to an infection can mediate atherosclerosis or directly affect plaque rupture.[26] Endothelial dysfunction, caused by a range of cardiovascular risk factors, is a key early stage of atherosclerosis.[26]

A strength of our study is that we used large, population-based, linked datasets generalisable to the UK population. We also compared results across two measures of cardiovascular risk. The marked increase in the incidence of acute cardiovascular events after acute respiratory infection in individuals with a QRISK2 score of 10% of higher compared with those with diagnosed hypertension is consistent with the multiple cardiovascular risk factors accounted for in the QRISK2 score.

Our selected study population should have reduced confounding. We included only individuals without chronic health conditions, who were not thought to be at high risk of acute respiratory infection or complications related to acute respiratory infection, and who were not recommended for influenza vaccination in England, according to current guidance. However, the selective study population prevented comparison of the incidence of acute respiratory infection and acute cardiovascular events after acute respiratory infection among individuals with established cardiovascular disease.

In England, diagnoses of acute respiratory infection are based primarily on clinical judgement, with most cases not laboratory confirmed. We included only codes that were most likely to be representative of systemic infection, which could plausibly induce atherosclerotic processes. However, clinically diagnosed influenza is poorly defined.[27] An under-estimation of the incidence of some acute respiratory infections, particularly influenza-like illness, is likely to occur because it is a short-lived illness. Conversely, due to its severity, pneumonia will usually result in health-care attendance. These differences in presentation and EHR data capture could account for the differing incidence of influenza-like illness and pneumonia among individuals with increased cardiovascular risk. From 1995 to 2013, recording of influenza-like illness decreased while recording of cough or fever symptoms increased in UK primary care,[27] and recording of community-acquired pneumonia increased in primary and secondary care.[28] Recording of symptoms, rather than diagnosis, in individuals with a QRISK2 score of 10% or higher, who were likely to attend primary care for comorbid conditions, could account for the lower incidence of influenza-like illness in these individuals than in those who had a QRISK2 score of less than 10%.

We might have misclassified QRISK2 score due to missing data for variables used in its calculation. However, the similar results from primary care recorded and captured in CPRD and our calculated QRISK2 scores suggest minimal misclassification. Furthermore, we adjusted for consultation frequency, as individuals who infrequently attend primary care services are unlikely to have biometric measures such as BMI or blood pressure recorded. We used a pragmatic definition for hypertension based on coded diagnoses only; this coded diagnosis corresponds to how, if eligible, individuals would be identified for influenza and

pneumococcal vaccination in primary care. We are likely to have captured both controlled and uncontrolled hypertension. In the stratified analysis, 38% (25 843 of 68 731) of individuals with diagnosed hypertension had not been prescribed antihypertensive drugs in the year before the acute respiratory infection. Conversely, only 3% (14 912 of 458 069) of individuals without a diagnosis of hypertension were prescribed antihypertensive drugs. Possible misclassification could have biased the results towards the null, which could explain the small association observed between hypertension and acute respiratory infection compared with the larger association for QRISK2 scores in our study and blood pressure reading-defined hypertension in the UK Biobank study.[14]

We assumed that acute cardiovascular events occurring after acute respiratory infection were due to infection. However, some events could have been unrelated, particularly those that occurred several months after infection. The possibility of unrelated events is consistent with our finding of a longer median time between acute respiratory infection and acute cardiovascular events in individuals with increased cardiovascular risk compared with those without an increased risk, as those with an increased cardiovascular risk are more likely to have a cardiovascular event regardless of acute respiratory infection.

Our findings emphasise the importance of improved cardiovascular risk management, which could reduce the incidence of acute respiratory infection and its cardiovascular consequences. The COVID-19 pandemic has highlighted this need. Preventive cardiovascular treatments, particularly among individuals with a QRISK2 score of 10% or higher, could reduce acute respiratory infection-related complications. A meta-analysis of nine observational studies showed a 41% reduction in the odds of mortality within 30 days of having pneumonia among statin users.[29] Targeted interventions for individuals with acute respiratory infection and a QRISK2 score of 10% or higher should be considered. A CPRD-based study published in 2020 found that aspirin use after hospitalisation for pneumonia resulted in a 54% reduction in the incidence of myocardial infarction and a 30% reduction in the incidence of stroke.[30]

Typically, individuals with increased cardiovascular risk do not receive seasonal influenza or pneumococcal vaccines. Previous RCTs have shown that the influenza vaccine provides secondary prevention of cardiovascular disease.[7] However, no RCTs or observational studies have investigated the use of vaccines for primary prevention of cardiovascular complications. Extending influenza vaccination to individuals with a QRISK2 score of 10% or higher would, on the basis of the most recent year in our dataset (2017–18), include approximately 150 000 individuals in England. Vaccinating these individuals could reduce cardiovascular complications and presumably decrease the incidence of influenza and associated complications, such as hospitalisation and mortality. Future studies using laboratory data to ascertain whether specific organisms lead to increases in cardiovascular events in people with increased cardiovascular risk would inform potential expansion of vaccine prioritisation.

In conclusion, the results of our study suggest that the incidence of acute respiratory infection, particularly pneumonia, is increased in people with high cardiovascular risk, and that the incidence of acute cardiovascular events after acute respiratory infection is also increased in these individuals. The QRISK2 score provides a better measure of the risk of a first cardiovascular event following acute respiratory infection than a diagnosis of hypertension alone. Therefore, the QRISK2 score could be used not only to identify individuals who require cardiovascular risk management, but also to target prevention and treatment of acute respiratory infection to individuals at increased risk of cardiovascular events following infection.

Contributors

CW-G conceived the study. JAD led the design of the study with supervision from CW-G, LS, and AB, and further methodological contributions from HIM, DG, EH, and RP. JAD led the development and collation of code lists used to define the variables used in the study, with assistance from HIM and HF and supervision from CW-G, LS, and AB. EH led the development of the QRISK2 coding used in the study. JAD analysed the data with statistical input from DG. JAD wrote the original manuscript draft. JAD, LS, HIM, DG, EH, HF, and CW-G had access to the raw data via the institutional CPRD licence held by the London School of Hygiene & Tropical Medicine; as AB and RP are external to the London School of Hygiene & Tropical Medicine, they were unable to access the securely held raw data. All authors reviewed aggregate data presented in the tables and figures. JAD and DG accessed and verified all the study data. All authors reviewed and commented on the manuscript and had final responsibility for the decision to submit for publication.

Declaration of interests

AB reports receiving grants from AstraZeneca, UK Research and Innovation, and the National Institute for Health Research (NIHR). CW-G reports receiving grants from the Wellcome Trust and the British Heart Foundation; receiving speaker fees from Sanofi Pasteur; and participating in a data safety monitoring board for an investigator-led trial of the effect of influenza vaccination after heart attack on future cardiovascular prognosis (NCT02831608) from January, 2019, to April, 2020. HIM is funded by the NIHR Health Protection Research Unit in Immunisation. JAD is funded by the British Heart Foundation through the grant received by CW-G. All other authors declare no competing interests.

Data sharing

The data used for this study were obtained from the CPRD. All CPRD data are available via an application to the Independent Scientific Advisory Committee (see https://www.cprd.com/Data-access). Data acquisition is associated with a fee and data protection requirements. This study is supported by code lists used to define each health condition, which have been made openly available at https://doi.org/10.17037/DATA.00002240. Our data management and analysis computer code is available via GitHub at https://github.com/jenAdavidson/cvrisk_mace_ari. All code is shared without investigator support.

Acknowledgments

This work was funded in part by the Wellcome Trust (201440/Z/16/Z), which funds an Intermediate Clinical Fellowship for CW-G. This work was also funded by the British Heart Foundation (FS/18/71/33938), which funds a non-clinical PhD studentship for JAD. CW-G is the principal investigator of this PhD studentship. HIM is funded by the

NIHR Health Protection Research Unit in Immunisation at the London School of Hygiene & Tropical Medicine in partnership with Public Health England. The views expressed are those of the authors and do not necessarily reflect those of the NHS, the NIHR, the Department of Health and Social Care, or Public Health England. This study is based on anonymised data from the CPRD obtained under licence from the UK Medicines and Healthcare products Regulatory Agency. The data are provided by patients and collected by the NHS as part of their care and support.

References

1 Smeeth L, Thomas SL, Hall AJ, Hubbard R, Farrington P, Vallance P. Risk of myocardial infarction and stroke after acute infection or vaccination. *N Engl J Med* 2004; **351:** 2611–18.

2 Warren-Gash C, Hayward AC, Hemingway H, et al. Influenza infection and risk of acute myocardial infarction in England and Wales: a CALIBER self-controlled case series study. *J Infect Dis* 2012; **206:** 1652–59.

3 Warren-Gash C, Blackburn R, Whitaker H, McMenamin J, Hayward AC. Laboratory-confirmed respiratory infections as triggers for acute myocardial infarction and stroke: a self-controlled case series analysis of national linked datasets from Scotland. *Eur Respir J* 2018; **51:** 1701794.

4 Ohland J, Warren-Gash C, Blackburn R, et al. Acute myocardial infarctions and stroke triggered by laboratory-confirmed respiratory infections in Denmark, 2010 to 2016. *Euro Surveill* 2020; **25:** 1900199.

5 Ren S, Newby D, Li SC, et al. Effect of the adult pneumococcal polysaccharide vaccine on cardiovascular disease: a systematic review and meta-analysis. *Open Heart* 2015; **2:** e000247.

6 Kwok CS, Aslam S, Kontopantelis E, et al. Influenza, influenza-like symptoms and their association with cardiovascular risks: a systematic review and meta-analysis of observational studies. *Int J Clin Pract* 2015; **69:** 928–37.

7 Clar C, Oseni Z, Flowers N, Keshtkar-Jahromi M, Rees K. Influenza vaccines for preventing cardiovascular disease. *Cochrane Database Syst Rev* 2015; **5:** CD005050.

8 European Centre for Disease Prevention and Control. Seasonal influenza vaccination and antiviral use in EU/EEA member states. 2018. https://ecdc.europa.eu/en/publications-data/seasonal-influenza-vaccination-antiviral-use-eu-eea-member-states (accessed July 16, 2019).

9 Public Health England. Influenza: the green book, chapter 19. https://www.gov.uk/government/publications/influenza-the-green-book-chapter-19 (accessed Jan 10, 2021).

10 Public Health England. Pneumococcal: the green book, chapter 25. 2013. https://www.gov.uk/government/publications/pneumococcal-the-green-book-chapter-25 (accessed Jan 10, 2021).

11 Rapsomaniki E, Timmis A, George J, et al. Blood pressure and incidence of twelve cardiovascular diseases: lifetime risks, healthy life-years lost, and age-specific associations in 1·25 million people. *Lancet* 2014; **383:** 1899–911.

12 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.

13 National Institute for Health and Care Excellence. CVD risk assessment and management. 2019. https://cks.nice.org.uk/cvd-risk-assessment-and-management (accessed Aug 14, 2019).

14 Zekavat SM, Honigberg M, Pirruccello JP, et al. Elevated blood pressure increases pneumonia risk: epidemiological association and mendelian randomization in the UK Biobank. *Med (NY)* 2021; **2:** 137–48.

15 Chen L, Han XD, Li YL, Zhang CX, Xing XQ. Incidence and risk factors for cardiovascular events in patients hospitalized with community-acquired pneumonia. *Zhonghua Xin Xue Guan Bing Za Zhi* 2020; **48:** 228–35.

16 Corrales-Medina VF, Musher DM, Wells GA, Chirinos JA, Chen L, Fine MJ. Cardiac complications in patients with community-acquired pneumonia incidence, timing, risk factors, and association with short-term mortality. *Circulation* 2012; **125:** 773–81.

17 Clayton TC, Thompson M, Meade TW. Recent respiratory infection and risk of cardiovascular disease: case-control study through a general practice database. *Eur Heart J* 2007; **29:** 96–103.

18 Herrett E, Gallagher AM, Bhaskaran K, et al. Data resource profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol* 2015; **44:** 827–36.

19 Wolf A, Dedman D, Campbell J, et al. Data resource profile: Clinical Practice Research Datalink (CPRD) Aurum. *Int J Epidemiol* 2019; **48:** 1740–40g.

20 Herbert A, Wijlaars L, Zylbersztejn A, Cromwell D, Hardelid P. Data resource profile: Hospital episode Statistics Admitted Patient Care (HES APC). *Int J Epidemiol* 2017; **46:** 1093–93i.

21 Davidson J, Banerjee A, Muzambi R, Smeeth L, Warren-Gash C. Validity of acute cardiovascular outcome diagnoses recorded in European electronic health records: a systematic review. *Clin Epidemiol* 2020; **12:** 1095–111.

22 Saberian S, Warren-Gash C, Hayward A, Fragaszy E. Effect of socio-economic status on influenza risk in English households. Options X for the Control of Influenza; Singapore; Aug 28–Sept 1, 2019 (abstr 11280).

23 Davidson J, Banerjee A, Mathur R, et al. Ethnic differences in the incidence of clinically diagnosed influenza: an England population-based cohort study 2008–2018. *Wellcome Open Res* 2021; **6:** 49.

24 Bae SA, Kim SR, Kim MN, Shim WJ, Park SM. Impact of cardiovascular disease and risk factors on fatal outcomes in patients with COVID-19 according to age: a systematic review and meta-analysis. *Heart* 2020; **0:** 1–8.

25 Singh MV, Chapleau MW, Harwani SC, Abboud FM. The immune system and hypertension. *Immunol Res* 2014; **59:** 243–53.

26 Keller TT, Mairuhu ATA, De Kruif MD, et al. Infections and endothelial cells. *Cardiovasc Res* 2003; **60:** 40–48.

27 Hardelid P, Rait G, Gilbert R, Petersen I. Recording of influenza-like illness in UK primary care 1995–2013: cohort study. *PLoS One* 2015; **10:** e0138659.

28 Smith S, Morbey R, de Lusignan S, Pebody RG, Smith GE, Elliot AJ. Investigating regional variation of respiratory infections in a general practice syndromic surveillance system. *J Public Health (Oxf)* 2021; **43:** e153-60.

29 Cheng H-H, Tang T-T, He Q, et al. Beneficial effects of statins on outcomes in pneumonia: a systematic review and meta-analysis. *Eur Rev Med Pharmacol Sci* 2014; **18:** 2294–305.

30 Hamilton F, Arnold D, Henley W, Payne RA. Aspirin reduces cardiovascular events in patients with pneumonia: a prior event rate ratio analysis in a large primary care database. *Eur Respir J* 2020; **57:** 2002795.

The published supplementary material is presented in **Chapter 11 Appendix 3**, which includes the study protocol.

6.3. Further methodological decisions

6.3.1. Detailed explanation of inclusion and exclusion criteria

After I identified patients who met the study inclusion criteria, I excluded those with established CVD. The reasons for this were two-fold: (i) QRISK2 is only for estimating future CVD risk among individuals who do not have pre-existing CVD, and (ii) the first cardiovascular event was the outcome of interest. To identify patients with established CVD, I used CPRD recorded clinical diagnoses of vascular disease or a major intervention specific to heart disease, for example coronary artery bypass graft.

The study population for the main analysis also excluded patients with an influenza vaccine eligible health condition who would not benefit from any further expansion of the current vaccine recommendations. Influenza vaccine recommendations are generally broader than those for the pneumococcal vaccine. Therefore, the study population was redefined in a sensitivity analysis to only exclude patients with a health condition eligible for both influenza and pneumococcal vaccine who have the greatest likelihood of severe disease following ARI. The published supplementary materials summarise the health conditions captured in the main and sensitivity analyses study population exclusions. **Table *6.1*** lists all influenza and pneumococcal vaccine eligible conditions and whether these apply to the study population. Pregnant women are eligible for influenza vaccination, but I did not exclude them from the study population. Identifying pregnancies in CPRD requires linkage to the pregnancy registry, which currently only exists for CPRD GOLD. As CPRD GOLD was the smaller of the two datasets (GOLD and Aurum) used in analysis, it would have been inadequate for identifying pregnancy within the study population. Additionally, the youngest patients included in the study were aged 40 years; therefore, a small proportion of women in the study population would have been pregnant and eligible for one-off influenza vaccination. Patients with a cochlear implant, a cerebrospinal fluid, or

an occupational risk (such as welding) are eligible for pneumococcal vaccination but not excluded from the study population. Again, the number of affected patients would be small. Occupational risk is also challenging to identify in CPRD. Instead, I excluded patients with previous pneumococcal vaccination.

Table 6.1 Influenza and pneumococcal vaccine indications applied to main and sensitivity analyses study populations

Vaccine indication	Influenza vaccine eligibility	Pneumococcal vaccine eligibility	Main study population exclusion	Sensitivity analysis study population exclusion
Heart disease	✓	✓	✓ but excluded based on CVD criterion	✓ but excluded based on CVD criterion
Chronic liver disease	✓	✓	✓	✓
Chronic respiratory disease	✓ asthma included	✓ asthma excluded	✓	✓ not asthma
CKD	✓ stages 3-5	✓ stages 4-5	✓ stages 3-5	✓ stages 4-5
Asplenia/spleen dysfunction	✓	✓	✓	✓
Diabetes	✓	✓ does not include diet-controlled asthma	✓	✓ but only those identified as treated diabetes
Immunosuppression	✓	✓	✓	✓
Neurological disease (including stroke/TIA)	✓	×	✓ but stroke/TIA excluded based on CVD criterion	×
Morbid obesity	✓	×	✓	×
Pregnant	✓	×	×	×
Cochlear implant	×	✓	×	×
Cerebrospinal fluid leak	×	✓	×	×
Occupational groups i.e. welders	×	✓	×	×

6.3.2. Covariate selection: acute respiratory infection outcome

Figure *6.1* shows the directed acyclic graph (DAG) I created and used to inform confounder selection for study objective i in hypertension analysis. I adjusted for age and sex as standard. Non-white ethnicity and socio-economic deprivation are associated with increased risk of hypertension [255–257] and ARI [258,259]. Smoking, excess alcohol consumption and obesity are also associated with hypertension [260–263] and ARI [264–267]. I additionally adjusted for consultation frequency in the year before baseline as hypertension and ARI diagnoses within primary care are more frequent in patients who regularly attend their GP. I did not adjust for any comorbid health conditions as I had excluded most comorbid health conditions from the study population on the basis that these individuals were eligible for influenza vaccination, and this was an analysis of individuals who were not currently eligible to receive influenza vaccine.

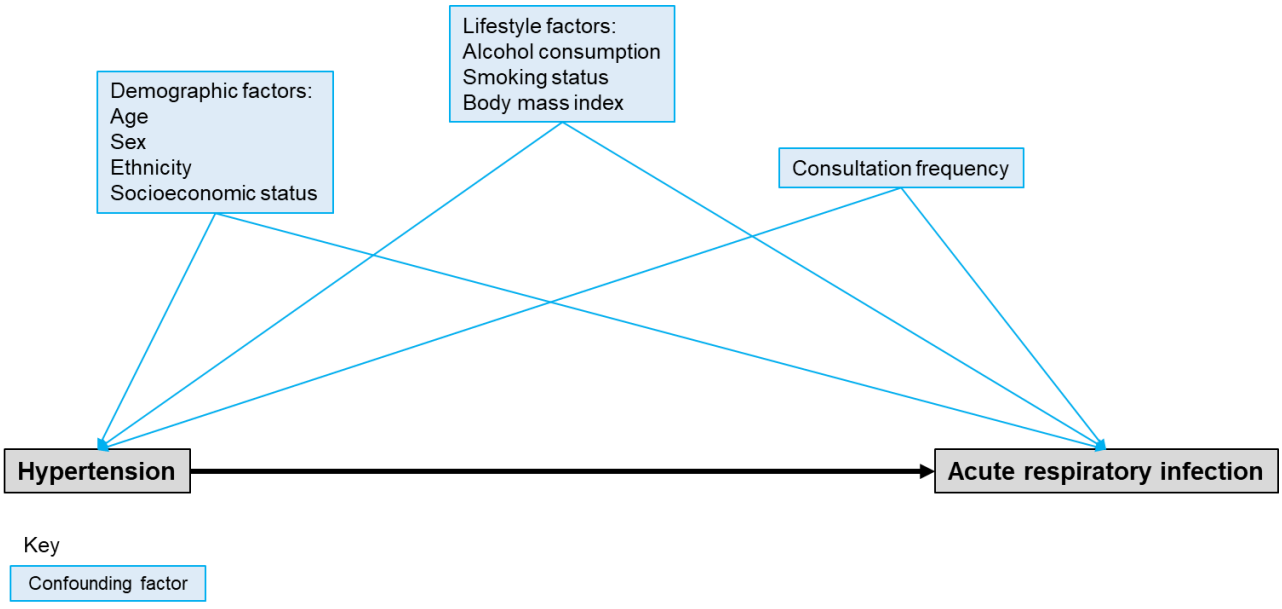


Figure 6.1 Directed acyclic graphic depicting confounders in the association between hypertension and acute respiratory infections

The QRISK2 score algorithm already takes into account age, sex, ethnicity, socio-economic status and smoking status. Therefore, I only adjusted for alcohol consumption and consultation frequency in the QRISK2 analysis.

6.3.3. Covariate selection: acute respiratory infection-related cardiovascular event outcome

Figure *6.2* shows the directed acyclic graph (DAG) I created and used to inform confounder selection for study objective iii in hypertension analysis. In addition to being risk factors for hypertension itself, non-white ethnicity, deprivation, smoking, excess alcohol consumption and obesity are all known risk factors for acute cardiovascular events [268]. I did not include atrial fibrillation, one of the only comorbid health conditions not excluded from the study population, as a covariate due to the likelihood that it lies on the causal pathway between hypertension and cardiovascular complications [269]. I considered antihypertensives and statins as potential effect modifiers and conducted stratified analysis. This is because individuals with diagnosed hypertension are likely to be prescribed antihypertensive treatments [231]. Individuals prescribed antihypertensives may have better-controlled hypertension and be less likely to experience a cardiovascular complication. Similarly, given individuals with hypertension have a high risk of CVD, cholesterol-lowering statins may be prescribed. I additionally considered the impact of antiviral treatment given at the time of infection diagnosis, influenza or pneumococcal vaccines received during follow-up, and baseline antiplatelet treatment, with stratified analysis conducted. Antivirals should be prescribed within 48 hours of symptom onset, so only prescriptions with 48 hours of ARI diagnosis were considered [141].

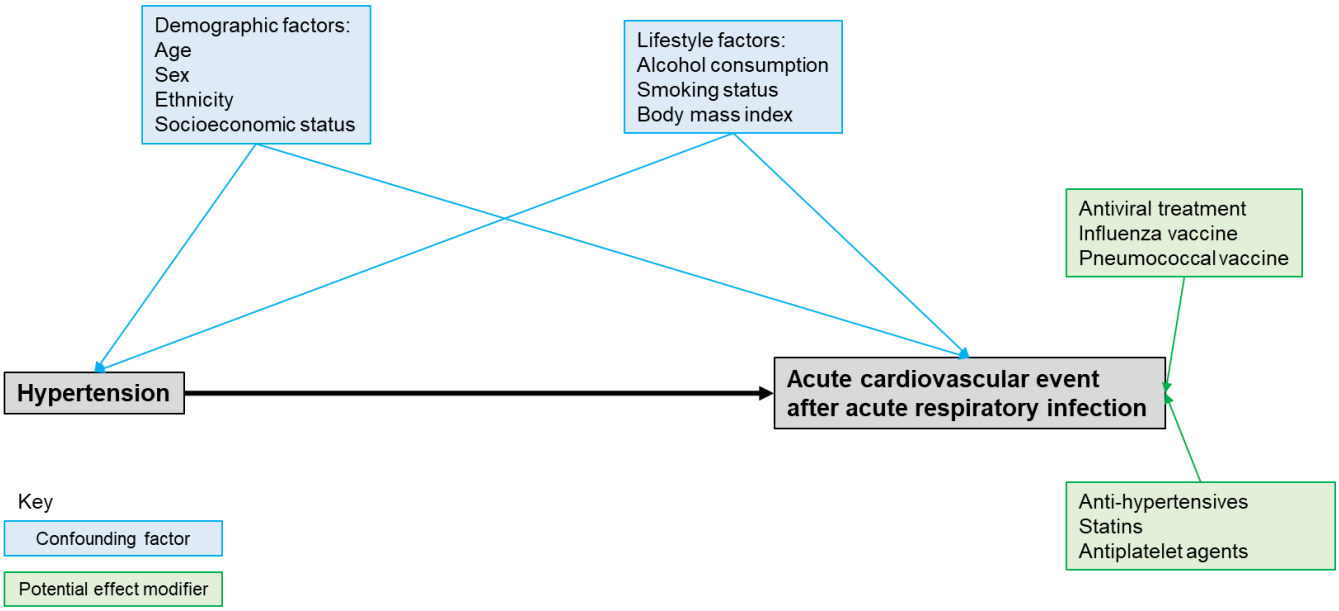


Figure 6.2 Directed acyclic graphic depicting confounders and effect modifiers in the association between hypertension and acute cardiovascular events after acute respiratory infection

6.4. The effect of raised cardiovascular risk on cardiovascular complications independent of acute respiratory infection

6.4.1. Background

In addition to estimating the effect of raised cardiovascular risk on acute cardiovascular events after ARI, I investigated the effect of raised cardiovascular risk on acute cardiovascular events independent of ARI diagnosis study objective ii. ARI recording in EHR data may be variable (see **Chapter 4.2**), so I used the effect of cardiovascular risk on acute cardiovascular events by season to determine if the pattern of events followed the seasonal circulation of ARIs.

6.4.2. Methods

In addition to the methods outlined in the published article, I carried out the following analyses to investigate objective iii. I examined the seasonal and annual trends in acute cardiovascular events using weekly counts by cardiovascular risk. I used Poisson regression models with time-updated age and cardiovascular risk to obtain IRRs. Initially, we adjusted for age and sex, then additional covariates (**Figure *6.3***). I also conducted a further analysis stratified by season, in which I defined season as influenza (1 September to 31 March) or non-influenza (1 April to 31 August).

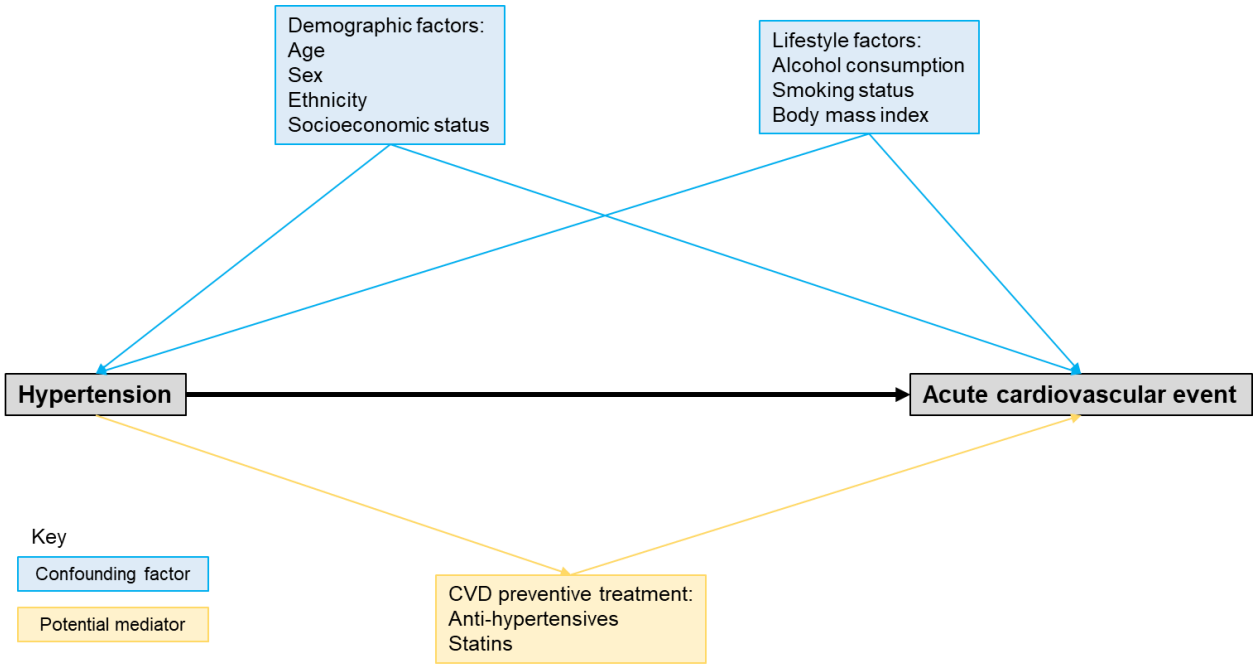


Figure 6.3 Directed acyclic graphic depicting confounders and mediators in the association between hypertension and major adverse cardiovascular events

I also repeated sensitivity analyses 1-4 outlined in the published article. In a further sensitivity analysis, I redefined the season as autumn (September-November), winter (December-February), spring (March-May), and summer (June-August).

In addition to the QRISK2 score algorithm covariates mentioned above (see ARI outcome section), antihypertensive treatment is also included in the algorithm. To analyse cardiovascular risk defined by QRISK2 score, I adjusted for alcohol consumption and stratified results by statin treatment.

6.4.3. Results

Among the 4,212,930 individuals included in the study population, 56,600 had an acute cardiovascular event during follow-up. Weekly counts of acute cardiovascular events suggest high numbers and rates occurred between January and March and lower numbers in the summer months, although weekly fluctuations were present (***Figure 6.4***). There was an increase in the weekly rate of acute cardiovascular events over time among individuals with raised cardiovascular risk, particularly when risk was defined by QRISK2 score.

The incidence of acute cardiovascular events was higher among individuals with hypertension (6.6/1,000 person-years) and QRISK2 ≥10% (10.5/1,000 person-years), compared to those without hypertension (2.5/1,000 person-years) and QRISK2 <10% (2.1/1,000 person-years). Comparing the incidence by method of defining cardiovascular risk, the incidence was substantially higher when raised cardiovascular risk was defined by QRISK2 score compared with diagnosed hypertension (**Table *6.2***).

After adjustment for confounders, the association between cardiovascular risk and acute cardiovascular event outcome remained for both hypertension (aIRR 1.78 [1.74-1.82]) and QRISK2 ≥10% (4.83 [4.74-4.92]), although substantially higher for QRISK2 score (**Table *6.2***). When stratifying the results by season, there was no difference in the incidence of acute cardiovascular events between influenza and non-influenza seasons (**Table *6.3***).

Sensitivity analyses results from published article methods showed:

(i) acute cardiovascular event IRRs differed slightly between CPRD GOLD and Aurum (**Table *6.4***), but meta-analysis results were like those of the combined database analysis.

(ii) redefining the study population to only exclude individuals with a health condition eligible for both pneumococcal and influenza vaccine (**Table *6.5***) produced similar results to the main population.

(iii) CPRD recorded QRISK2 scores (**Table *6.6***) obtained similar results to scores calculated used the LSHTM algorithm.

(iv) redefining the outcome to only include major adverse cardiovascular events (MACE) slightly reduced the estimates obtained when cardiovascular risk was defined by hypertension (all events: 1.78 [1.74-1.82] and MACE: 1.73 [1.69-1.77]) and increased the estimates for QRISK2 score (all events: 4.83 [4.74-4.92] and MACE: 4.94 [4.84-5.04]) (**Table *6.7***).

Redefining season into autumn, winter, spring and summer had little impact on the IRRs (**Table *6.8***).

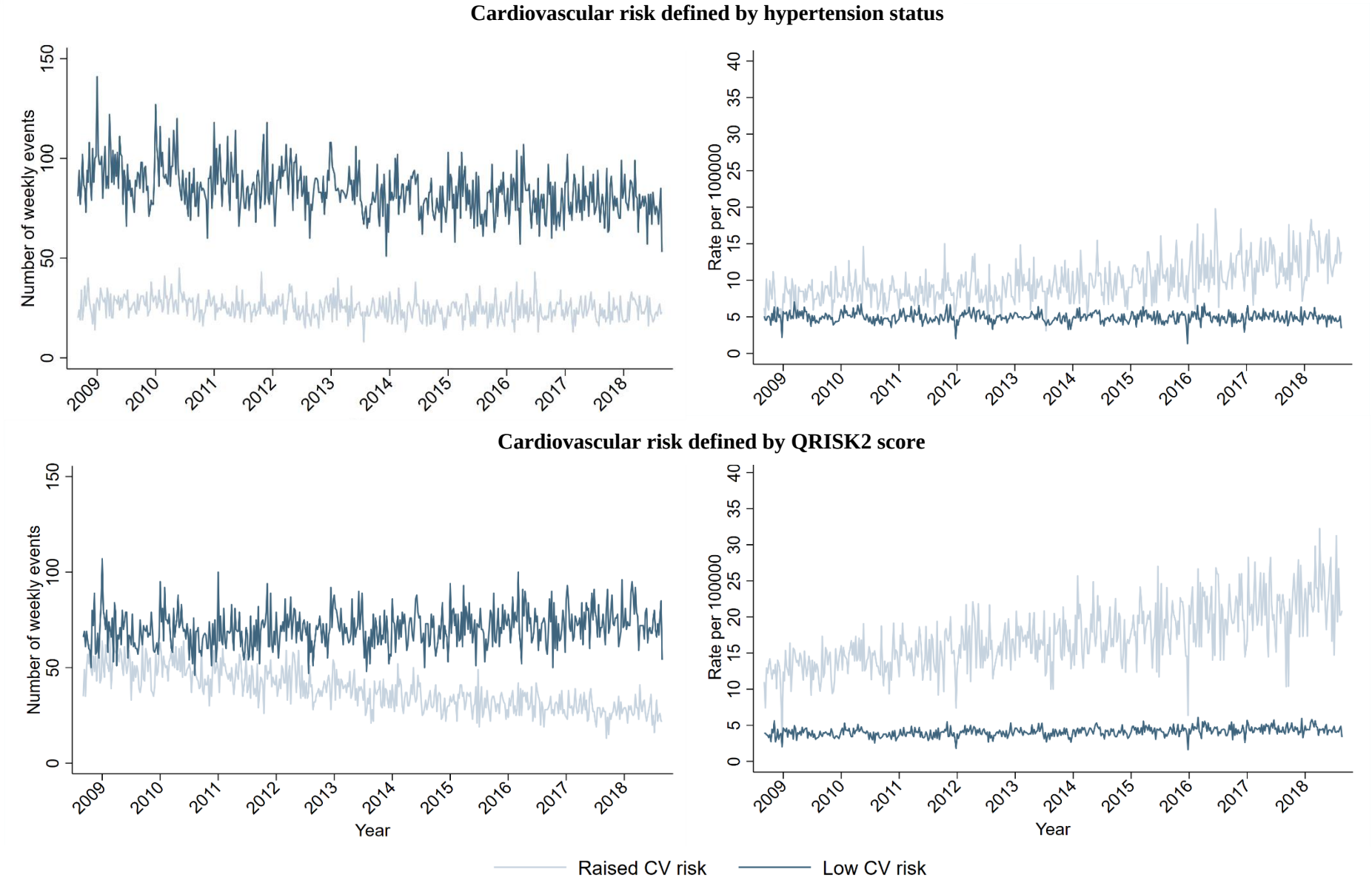


Figure 6.4 Weekly number and rate of acute cardiovascular events, September 2008–August 2018

Table 6.2 Association between cardiovascular risk and acute cardiovascular event

Outcome	Cardiovascular risk	No. of events	Rate per 1,000 person-years (95% CI)	Crude IRR (95% CI)	Age and sex-adjusted IRR (95% CI)	Fully adjusted* IRR (95% CI)
Any cardiovascular event	Hypertension	12,972	6.8 (6.6–6.9)	2.70 (2.65–2.75)	1.96 (1.92–2.00)	1.78 (1.74–1.82)
	No hypertension	43,628	2.5 (2.5–2.5)	1 (ref)	1 (ref)	1 (ref)
	QRISK2 ≥10%	19,876	10.7 (10.5–10.8)	5.08 (5.00–5.17)	NA	4.83 (4.74–4.92)
	QRISK2 <10%	36,724	2.1 (2.1–2.1)	1 (ref)	NA	1 (ref)
Acute coronary syndrome†	Hypertension	5,117	2.7 (2.6–2.7)	2.47 (2.40–2.55)	1.81 (1.75–1.87)	1.58 (1.53–1.64)
	No hypertension	18,733	1.1 (1.1–1.1)	1 (ref)	1 (ref)	1 (ref)
	QRISK2 ≥10%	8,853	4.8 (4.7–4.9)	5.53 (5.39–5.68)	NA	5.26 (5.12–5.40)
	QRISK2 <10%	14,997	0.9 (0.8–0.9)	1 (ref)	NA	1 (ref)
Heart failure	Hypertension	1,659	0.9 (0.8–0.9)	2.90 (2.75–3.07)	2.02 (1.91–2.14)	1.80 (1.69–1.92)
	No hypertension	5,188	0.3 (0.3–0.3)	1 (ref)	1 (ref)	1 (ref)
	QRISK2 ≥10%	3,213	1.5 (1.5–1.6)	6.02 (5.73–6.31)	NA	5.67 (5.39–5.96)
	QRISK2 <10%	4,723	0.3 (0.2–0.3)	1 (ref)	NA	1 (ref)
Acute limb ischaemia	Hypertension	403	0.2 (0.2–0.2)	4.18 (3.72–4.70)	2.68 (2.37–3.02)	2.94 (2.56–3.39)
	No hypertension	869	0.0 (0.0–0.1)	1 (ref)	1 (ref)	1 (ref)
	QRISK2 ≥10%	715	0.4 (0.4–0.4)	11.95 (10.70–13.35)	NA	10.77 (9.60–12.08)
	QRISK2 <10%	557	0.0 (0.0–0.0)	1 (ref)	NA	1 (ref)
Stroke or transient ischaemic attack‡	Hypertension	5240	2.7 (2.7–2.8)	2.82 (2.73–2.90)	2.08 (2.01–2.14)	1.90 (1.83–1.97)
	No hypertension	16,945	1.0 (1.0–1.0)	1 (ref)	1 (ref)	1 (ref)
	QRISK2 ≥10%	6,730	3.6 (3.5–3.7)	4.09 (3.98–4.21)	NA	3.88 (3.77–4.00)
	QRISK2 <10%	15,455	0.9 (0.9–0.9)	1 (ref)	NA	1 (ref)
Cardiovascular-related death	Hypertension	1,412	0.7 (0.7–0.8)	2.54 (2.39–2.69)	1.75 (1.65–1.85)	1.89 (1.75–2.04)
	No hypertension	5,101	0.3 (0.3–0.3)	1 (ref)	1 (ref)	1 (ref)
	QRISK2 ≥10%	2,521	1.4 (1.3–1.4)	5.95 (5.66–6.25)	NA	6.02 (5.70–6.35)
	QRISK2 <10%	3,992	0.2 (0.2–0.2)	1 (ref)	NA	1 (ref)

Total person-years per 1,000 years of follow-up available was 1,921.5 for hypertension, 17,453.5 for no hypertension, 1,859.9 for a QRISK2 score of 10% or higher, and 17,515.1 for a QRISK2 score of less than 10%. LRT p values for all comparisons were <0.0001. *Hypertension models were adjusted for age, sex, ethnicity, socioeconomic status, body-mass index, alcohol consumption and smoking status; and QRISK2 models were adjusted for alcohol consumption. †Fully adjusted IRR for myocardial infarction alone was 1.53 (1.47–1.59) in the hypertension model and 5.60 (5.43–5.77) in the QRISK2 model; fully adjusted IRR for unstable angina alone was 1.89 (1.75–2.05) in the hypertension model and 4.72 (4.43–5.03) in the QRISK2 model. ‡ Fully adjusted IRR for stroke alone was 1.88 (1.80–1.97) in the hypertension model and 3.80 (3.67–3.94) in the QRISK2 model; fully adjusted IRR for transient ischaemic attack was 1.93 (1.82–2.06) in the hypertension model and 4.12 (3.92–4.34) in the QRISK2 model.

Table 6.3 Association between cardiovascular risk and acute cardiovascular event by season

Season	Cardiovascular risk	No. of events	Total person-years per 1,000	Rate per 1,000 person-years (95% CI)	Crude IRR (95% CI)	Age and sex-adjusted IRR (95% CI)	Fully adjusted* IRR (95% CI)
Influenza season	Hypertension	7,538	1,118.5	6.7 (6.6–6.9)	2.69 (2.62–2.76)	1.96 (1.91–2.01)	1.78 (1.73–1.83)
	No hypertension	25,532	10,194.1	2.5 (2.5–2.5)	1 (ref)	1 (ref)	1 (ref)
	QRISK2 ≥10%	11,622	1,094.2	10.6 (10.4–10.8)	5.06 (4.95–5.18)	NA	4.80 (4.69–4.92)
	QRISK2 <10%	21,448	7,259.4	2.1 (2.1–2.1)	1 (ref)	NA	1 (ref)
Non-influenza season	Hypertension	5,434	803.0	6.8 (6.6–6.9)	2.71 (2.63–2.80)	1.97 (1.91–2.03)	1.77 (1.71–1.84)
	No hypertension	18,096	7,259.4	2.5 (2.5–2.5)	1 (ref)	1 (ref)	1 (ref)
	QRISK2 ≥10%	8,254	765.7	10.8 (10.6–11.0)	5.15 (5.01–5.29)	NA	4.90 (4.76–5.03)
	QRISK2 <10%	15,276	7,296.7	2.1 (2.1–2.1)	1 (ref)	NA	1 (ref)

LRT p values for all comparisons were <0.0001. *Hypertension models were adjusted for age, sex, ethnicity, socioeconomic status, body-mass index, alcohol consumption and smoking status; and QRISK2 models were adjusted for alcohol consumption.

Table 6.4 Database comparison of acute cardiovascular events incidence rates and incidence rate ratios

Outcome	Cardiovascular risk	Database	Rate per 1,000 person-years (95% CI)		Crude IRR (95% CI)	Age and sex-adjusted IRR (95% CI)	Fully adjusted* IRR (95% CI)
			High risk	Low risk			
Any cardiovascular event	Hypertension	GOLD	6.5 (6.2–6.8)	2.6 (2.5–2.7)	2.50 (2.38–2.63)	1.81 (1.72–1.91)	1.69 (1.59–1.79)
		Aurum	6.8 (6.7–6.9)	2.5 (2.5–2.5)	2.74 (2.68–2.80)	1.99 (1.95–2.04)	1.79 (1.75–1.84)
		GOLD and Aurum	6.8 (6.6–6.9)	2.5 (2.5–2.5)	2.70 (2.65–2.75)	1.96 (1.92–2.00)	1.78 (1.74–1.82)
		Meta-analysis of GOLD and Aurum	NA	NA	2.63 (2.39–2.86)	1.91 (1.73–2.08)	1.75 (1.66–1.85)
	QRISK2	GOLD	11.1 (10.7–11.5)	2.2 (2.1–2.2)	5.14 (4.91–5.37)	NA	4.91 (4.69–5.14)
		Aurum	10.6 (10.5–10.8)	2.1 (2.1–2.1)	5.09 (5.00–5.19)	NA	4.83 (4.74–4.93)
		GOLD and Aurum	10.7 (10.5–10.8)	2.1 (2.1–2.1)	5.08 (5.00–5.17)	NA	4.83 (4.74–4.92)
		Meta-analysis of GOLD and Aurum	NA	NA	5.10 (5.01–5.19)	NA	4.84 (4.76–4.93)
Acute coronary syndrome	Hypertension	GOLD	2.7 (2.5–2.9)	2.7 (2.5–2.9)	2.25 (2.08–2.44)	1.65 (1.52–1.78)	1.48 (1.36–1.62)
		Aurum	2.7 (2.6–2.7)	1.1 (1.0–1.1)	2.53 (2.44–2.61)	1.85 (1.79–1.91)	1.60 (1.54–1.67)
		GOLD and Aurum	2.7 (2.6–2.7)	1.1 (1.1–1.1)	2.47 (2.40–2.55)	1.81 (1.75–1.87)	1.58 (1.53–1.64)
		Meta-analysis of GOLD and Aurum	NA	NA	2.40 (2.13–2.68)	1.80 (1.56–1.95)	1.55 (1.48–1.67)
	QRISK2	GOLD	5.2 (4.9–5.4)	1.0 (0.9–1.0)	5.44 (5.10–5.81)	NA	5.20 (4.86–5.57)
		Aurum	4.7 (4.6–4.8)	0.8 (0.8–0.9)	5.59 (5.43–5.75)	NA	5.31 (5.15–5.46)
		GOLD and Aurum	4.8 (4.7–4.9)	0.9 (0.8–0.9)	5.53 (5.39–5.68)	NA	5.26 (5.12–5.40)
		Meta-analysis of GOLD and Aurum	NA	NA	5.57 (5.42–5.71)	NA	5.29 (5.15–5.43)
Heart failure	Hypertension	GOLD	0.9 (0.8–1.0)	0.3 (0.3–0.3)	2.67 (2.32–3.07)	1.86 (1.61–2.15)	1.81 (1.54–2.13)
		Aurum	0.9 (0.8–0.9)	0.3 (0.3–0.3)	2.95 (2.78–3.13)	2.06 (1.94–2.19)	1.81 (1.68–1.94)
		GOLD and Aurum	0.9 (0.8–0.9)	0.3 (0.3–0.3)	2.90 (2.75–3.07)	2.02 (1.91–2.14)	1.80 (1.69–1.92)
		Meta-analysis of GOLD and Aurum	NA	NA	2.86 (2.61–3.12)	2.00 (1.82–2.18)	1.81 (1.69–1.93)
	QRISK2	GOLD	1.5 (1.4–1.7)	0.3 (0.2–0.3)	5.86 (5.18–6.63)	NA	5.46 (4.80–6.20)
		Aurum	1.4 (1.4–1.5)	0.2 (0.2–0.2)	5.31 (5.15–5.46)	NA	5.31 (5.15–5.46)
		GOLD and Aurum	1.5 (1.5–1.6)	0.3 (0.2–0.3)	6.02 (5.73–6.31)	NA	5.67 (5.39–5.96)
		Meta-analysis of GOLD and Aurum	NA	NA	5.47 (4.98–5.95)	NA	5.32 (5.17–5.47)

Outcome	Cardiovascular risk	Database	Rate per 1,000 person-years (95% CI)		Crude IRR (95% CI)	Age and sex-adjusted IRR (95% CI)	Fully adjusted* IRR (95% CI)
			High risk	Low risk			
Acute limb ischaemia	Hypertension	GOLD	0.2 (0.1–0.2)	0.0 (0.0–0.1)	3.85 (2.81–5.29)	2.49 (1.80–3.44)	2.53 (1.74–3.68)
		Aurum	0.2 (0.2–0.2)	0.1 (0.0–0.1)	4.28 (3.76–4.86)	2.73 (2.40–3.12)	3.04 (2.61–3.53)
		GOLD and Aurum	0.2 (0.2–0.2)	0.0 (0.0–0.1)	4.18 (3.72–4.70)	2.68 (2.37–3.02)	2.94 (2.56–3.39)
		Meta-analysis of GOLD and Aurum	NA	NA	4.21 (3.71–4.71)	2.69 (2.36–3.02)	2.95 (2.53–3.36)
	QRISK2	GOLD	0.4 (0.3–0.5)	0.0 (0.0–0.0)	14.86 (11.05–19.99)	NA	13.72 (10.04–18.74)
		Aurum	0.4 (0.4–0.4)	0.0 (0.0–0.0)	11.68 (10.36–13.16)	NA	10.48 (9.26–11.87)
		GOLD and Aurum	0.4 (0.4–0.4)	0.0 (0.0–0.0)	11.95 (10.70–13.35)	NA	10.77 (9.60–12.08)
		Meta-analysis of GOLD and Aurum	NA	NA	12.50 (9.78–15.29)	NA	11.41 (8.54–14.28)
Stroke or transient ischaemic attack	Hypertension	GOLD	2.5 (2.3–2.7)	1.0 (0.9–1.0)	2.63 (2.42–2.86)	1.96 (1.80–2.13)	1.84 (1.67–2.02)
		Aurum	2.8 (2.7–2.8)	1.0 (1.0–1.0)	2.84 (2.75–2.94)	2.10 (2.03–2.17)	1.91 (1.83–1.98)
		GOLD and Aurum	2.7 (2.7–2.8)	1.0 (1.0–1.0)	2.82 (2.73–2.90)	2.08 (2.01–2.14)	1.90 (1.83–1.97)
		Meta-analysis of GOLD and Aurum	NA	NA	2.76 (2.56–2.96)	2.05 (1.92–2.18)	1.90 (1.83–1.97)
	QRISK2	GOLD	3.6 (3.4–3.9)	0.9 (0.8–0.9)	4.24 (3.95–4.55)	NA	4.09 (3.81–4.40)
		Aurum	3.6 (3.5–3.7)	0.9 (0.9–0.9)	4.08 (3.95–4.21)	NA	3.86 (3.74–3.98)
		GOLD and Aurum	3.6 (3.5–3.7)	0.9 (0.9–0.9)	4.09 (3.98–4.21)	NA	3.88 (3.77–4.00)
		Meta-analysis of GOLD and Aurum	NA	NA	4.11 (3.99–4.23)	NA	3.93 (3.72–4.14)
Cardiovascular-related death	Hypertension	GOLD	0.8 (0.7–0.9)	0.3 (0.3–0.3)	2.52 (2.18–2.93)	1.72 (1.48–2.00)	1.95 (1.60–2.37)
		Aurum	0.7 (0.7–0.8)	0.3 (0.3–0.3)	2.51 (2.36–2.68)	1.74 (1.62–1.85)	1.86 (1.71–2.02)
		GOLD and Aurum	0.7 (0.7–0.8)	0.3 (0.3–0.3)	2.54 (2.39–2.69)	1.75 (1.65–1.85)	1.89 (1.75–2.04)
		Meta-analysis of GOLD and Aurum	NA	NA	2.51 (2.36–2.66)	1.74 (1.63–1.84)	1.87 (1.73–2.02)
	QRISK2	GOLD	1.5 (1.4–1.7)	0.2 (0.2–0.3)	6.26 (5.58–7.03)	NA	6.19 (5.46–7.03)
		Aurum	1.3 (1.3–1.4)	0.2 (0.2–0.2)	5.88 (5.57–6.21)	NA	5.97 (5.62–6.34)
		GOLD and Aurum	1.4 (1.3–1.4)	0.2 (0.2–0.2)	5.95 (5.66–6.25)	NA	6.02 (5.70–6.35)
		Meta-analysis of GOLD and Aurum	NA	NA	5.94 (5.65–6.24)	NA	6.01 (5.68–6.34)

Table 6.5 Association between cardiovascular risk and acute cardiovascular event in the sensitivity analysis study population*

Outcome	Cardiovascular risk	No. of events	Rate per 1,000 person-years (95% CI)	Crude IRR (95% CI)	Age and sex-adjusted IRR (95% CI)	Fully adjusted† IRR (95% CI)
Any cardiovascular event	Hypertension	15,583	6.9 (6.8–7.0)	2.71 (2.66–2.76)	1.99 (1.96–2.03)	1.78 (1.74–1.82)
	No hypertension	47,069	2.6 (2.5–2.6)	1 (ref)	1 (ref)	1 (ref)
	QRISK2 ≥10%	22,743	10.9 (10.7–11.0)	5.04 (4.96–5.12)	NA	4.79 (4.71–4.87)
	QRISK2 <10%	39,909	2.2 (2.1–2.2)	1 (ref)	NA	1 (ref)
Acute coronary syndrome	Hypertension	5,932	2.6 (2.6–2.7)	2.43 (2.36–2.50)	1.81 (1.75–1.86)	1.58 (1.53–1.63)
	No hypertension	19,961	1.1 (1.1–1.1)	1 (ref)	1 (ref)	1 (ref)
	QRISK2 ≥10%	9,870	4.7 (4.6–4.8)	5.45 (5.31–5.58)	NA	5.18 (5.05–5.32)
	QRISK2 <10%	16,023	0.9 (0.9–0.9)	1 (ref)	NA	1 (ref)
Heart failure	Hypertension	2,159	1.0 (0.9–1.0)	3.05 (2.91–3.21)	2.18 (2.07–2.29)	1.84 (1.73–1.95)
	No hypertension	5,777	0.3 (0.3–0.3)	1 (ref)	1 (ref)	1 (ref)
	QRISK2 ≥10%	3,216	1.5 (1.5–1.6)	6.01 (5.75–6.29)	NA	5.67 (5.42–5.94)
	QRISK2 <10%	4,727	0.3 (0.2–0.3)	1 (ref)	NA	1 (ref)
Acute limb ischaemia	Hypertension	464	0.2 (0.2–0.2)	4.12 (3.68–4.60)	2.66 (2.37–2.98)	2.94 (2.57–3.35)
	No hypertension	921	0.1 (0.0–0.1)	1 (ref)	1 (ref)	1 (ref)
	QRISK2 ≥10%	795	0.4 (0.4–0.4)	11.91 (10.71–13.25)	NA	10.81 (9.67–12.08)
	QRISK2 <10%	590	0.0 (0.0–0.0)	1 (ref)	NA	1 (ref)
Stroke or transient ischaemic attack	Hypertension	6,244	2.8 (2.7–2.9)	2.80 (2.72–2.88)	2.08 (2.02–2.15)	1.90 (1.84–1.97)
	No hypertension	18,231	1.0 (1.0–1.0)	1 (ref)	1 (ref)	1 (ref)
	QRISK2 ≥10%	7,708	3.7 (3.6–3.8)	4.06 (3.96–4.18)	NA	3.86 (3.75–3.96)
	QRISK2 <10%	16,767	0.9 (0.9–0.9)	1 (ref)	NA	1 (ref)
Cardiovascular-related death	Hypertension	1,797	0.8 (0.8–0.8)	2.61 (2.48–2.75)	1.84 (1.74–1.94)	1.87 (1.74–2.00)
	No hypertension	5,628	0.3 (0.3–0.3)	1 (ref)	1 (ref)	1 (ref)
	QRISK2 ≥10%	2,967	1.4 (1.4–1.5)	5.88 (5.62–6.16)	NA	5.94 (5.65–6.25)
	QRISK2 <10%	4,458	0.2 (0.2–0.2)	1 (ref)	NA	1 (ref)

Total person-years per 1,000 years of follow-up available was 2,245.6 for hypertension, 18,356.3 for no hypertension, 2,093.4 for a QRISK2 score of 10% or higher, and 18,508.5 for a QRISK2 score of less than 10%. LRT p values for all comparisons were <0.0001. *The sensitivity analysis study population only excluded individuals eligible for both pneumococcal and influenza vaccination, which is fully defined in the supplementary method for the published paper included in this chapter. †Hypertension models were adjusted for age, sex, ethnicity, socioeconomic status, body-mass index, alcohol consumption and smoking status; and QRISK2 models were adjusted for alcohol consumption.

Table 6.6 Association between QRISK2 score and acute cardiovascular events, by QRISK2 score identification method*

Cardiovascular risk method	No. of events	Rate per 1,000 person-years	Crude IRR (95% CI)	Alcohol intake adjusted IRR (95% CI)
Recorded QRISK2 ≥10%	6,018	27.6 (26.8–28.4)	1.41 (1.37–1.46)	1.42 (1.38–1.47)
Recorded QRISK2 <10%	22,561	19.7 (19.5–20.0)	1 (ref)	1 (ref)
Calculated QRISK2 ≥10%	4,296	29.6 (28.6–30.6)	1.48 (1.43–1.54)	1.40 (1.35–1.46)
Calculated QRISK2 <10%	24,283	20.0 (19.7–20.3)	1 (ref)	1 (ref)

*To validate the QRISK2 algorithm created and used in the main analyses, analyses were repeated with inclusion restricted to individuals with a QRISK2 score recorded in CPRD data (i.e., recorded by the patient's GP) from 2015-2017. A full explanation of the method is explained in the supplementary method for the published paper included in this chapter.

Table 6.7 Association between cardiovascular risk and major adverse cardiovascular event*

Cardiovascular risk	No. of events	Rate per 1,000 person-years (95% CI)	Crude HR (95% CI)	Age and sex-adjusted HR (95% CI)	Fully adjusted† HR (95% CI)
Hypertension	10,068	5.2 (5.1–5.3)	2.62 (2.56–2.68)	1.90 (1.86–1.94)	1.73 (1.69–1.77)
No hypertension	34,946	2.0 (2.0–2.0)	1 (ref)	1 (ref)	1 (ref)
QRISK2 ≥10%	15,995	8.6 (8.5–8.7)	5.19 (5.09–5.29)	NA	4.94 (4.84–5.04)
QRISK2 <10%	29,019	1.7 (1.6–1.7)	1 (ref)	NA	1 (ref)

*Major adverse cardiovascular event was defined as myocardial infarction, stroke, left ventricular heart failure and cardiovascular death (a subset of the broader acute cardiovascular event definition used in the main analyses. †Hypertension models were adjusted for age, sex, ethnicity, socioeconomic status, body-mass index, alcohol consumption and smoking status; and QRISK2 models were adjusted for alcohol consumption.

Table 6.8 Association between cardiovascular risk and acute cardiovascular event by redefined season

	Cardiovascular risk	No. of events	Total person-years per 1,000	Rate per 1,000 person-years (95% CI)	Crude IRR (95% CI)	Age and sex-adjusted IRR (95% CI)	Fully adjusted* IRR (95% CI)
Autumn	Hypertension	3,204	482.86	6.6 (6.4–6.9)	2.77 (2.66–2.88)	2.02 (1.94–2.10)	1.83 (1.74–1.91)
	No hypertension	10,644	4,438.66	2.4 (2.4–2.4)	1 (ref)	1 (ref)	1 (ref)
	QRISK2 ≥10%	4,840	472.28	10.2 (10.0–10.5)	5.06 (4.89–5.24)	NA	4.81 (4.64–4.99)
	QRISK2 <10%	9,008	4,449.25	2.0 (2.0–2.1)	1 (ref)	NA	1 (ref)
Winter	Hypertension	3,218	473.23	6.8 (6.6–7.0)	2.65 (2.55–2.75)	1.93 (1.85–2.01)	1.76 (1.68–1.85)
	No hypertension	11,045	4,300.56	2.6 (2.5–2.6)	1 (ref)	1 (ref)	1 (ref)
	QRISK2 ≥10%	5,011	462.75	10.8 (10.5–11.1)	5.05 (4.88–5.22)	NA	4.77 (4.60–4.94)
	QRISK2 <10%	9,252	4,311.04	2.2 (2.1–2.2)	1 (ref)	NA	1 (ref)
Spring	Hypertension	3,321	473.23	6.8 (6.6–7.1)	2.58 (2.48–2.68)	1.87 (1.79–1.94)	1.68 (1.60–1.76)
	No hypertension	11,471	4327.31	2.7 (2.6–2.7)	1 (ref)	1 (ref)	1 (ref)
	QRISK2 ≥10%	5,299	476.57	11.1 (10.8–11.4)	5.08 (4.91–5.25)	NA	4.86 (4.70–5.04)
	QRISK2 <10%	9,493	4,337.09	2.2 (2.1–2.2)	1 (ref)	NA	1 (ref)
Summer	Hypertension	3,229	479.02	6.7 (6.5–7.0)	2.82 (2.72–2.94)	2.05 (1.97–2.13)	1.85 (1.76–1.93)
	No hypertension	10,468	4386.96	2.4 (2.3–2.4)	1 (ref)	1 (ref)	1 (ref)
	QRISK2 ≥10%	4,726	448.30	10.5 (10.2–10.8)	5.19 (5.01–5.38)	NA	4.91 (4.74–5.10)
	QRISK2 <10%	8,971	4,417.68	2.0 (2.0–2.1)	1 (ref)	NA	1 (ref)

* Hypertension models were adjusted for age, sex, ethnicity, socioeconomic status, body-mass index, alcohol consumption and smoking status; and QRISK2 models were adjusted for alcohol consumption.

6.5. Chapter summary

- In this population-based cohort study I used CPRD GOLD and Aurum data to estimate the risk of raised cardiovascular risk on ARI-related acute cardiovascular events. The study additionally investigated the effect of raised cardiovascular risk on ARI and acute cardiovascular event incidence separately to understand where cardiovascular risk acts in the pathogenesis of cardiovascular complications after an ARI, i.e. whether raised cardiovascular risk increases the likelihood of ARI or only the likelihood of acute cardiovascular events.
- Among the 4,212,930 individuals without established CVD or a health condition included in influenza vaccine recommendation criteria, 12·5% had diagnosed hypertension and 14·4% had a QRISK2 score ≥10%.
- There was only a marginal increased incidence of ARI among individuals with hypertension (adjusted IRR 1.04, 95% CI 1.03-1.05) and a more substantial increased incidence in those with a QRISK2 score ≥10% (adjusted IRR 1.39, 1.37-1.40). There was also an increased incidence of pneumonia among individuals with hypertension (adjusted IRR 1.12, 1.07-1.16), but again this was much more pronounced in those with a QRISK2 score ≥10% (adjusted IRR 2·32, 2·25-2·40). There was a reduced incidence of influenza/ILI among individuals with hypertension (adjusted IRR 0·98, 0·96-1·00) and a QRISK2 score ≥10% (adjusted IRR 0·88, 0·86-0·90).
- The incidence rate ratio for acute cardiovascular events independent of ARI was substantially higher in those with raised cardiovascular risk, particularly when risk was defined by QRISK2 score (adjusted IRR adjusted IRR 4.83, 4.74-4.92) rather than diagnosed hypertension (1.78, 1.74-1.82).
- There was a significant association between raised cardiovascular risk and ARI-related acute cardiovascular event, which was greater when risk was defined by QRISK2 score ≥10% (adjusted HR 3.65, 3.42-3.89) than hypertension (adjusted HR 1.98, 1.83-2.15). Similar results were obtained for pneumonia (QRISK score ≥10%: adjusted HR 2.13, 1.92-2.37 and hypertension:

adjusted HR 1.65, 1.44-1.89) and influenza/ILI (QRISK score ≥10%: adjusted HR 3.35, 2.70-4.17 and hypertension: adjusted HR 2.07, 1.60-2.67).

- The study identified the risk of ARI, acute cardiovascular events independent of ARI and ARI-related cardiovascular complications were all higher in individuals with raised cardiovascular risk defined by hypertension and QRISK2 score, with overall markedly increased risk among individuals with raised QRISK2 scores. The consistent findings show that hypertension and, particularly, QRISK2 predict acute cardiovascular events.
- The results generated by the study emphasise the importance of improved cardiovascular risk management, which could lessen ARI, particularly pneumonia, incidence and its cardiovascular consequences. QRISK2 score provided a better measure for risk of a first cardiovascular event following ARI than hypertension diagnosis alone. Therefore, QRISK2 score could be used to not only identify individuals who require cardiovascular risk management but also targeted ARI prevention and treatment.

Chapter 7 The effectiveness of influenza vaccine in preventing cardiovascular events in people with raised cardiovascular risk

7.1. Chapter overview

In this chapter, I address thesis objective 3; to investigate whether influenza vaccine reduces the risk of acute cardiovascular events and whether any reduction differs between individuals with raised and low cardiovascular risk. The chapter begins with a detailed overview of the different study methodologies I considered to address the research objective. The chapter then shows the research paper submitted to the European Heart Journal, followed by additional methods and results not included in the peer review submission. The study protocol is shown in Chapter 11 Appendix 4 and supplementary material to accompany the research paper in Chapter 11 Appendix 5.

Briefly to address thesis objective 3, I used CPRD Aurum data from 2008 to 2019 to identify individuals aged 40-84 years who received at least one influenza vaccine. For these individuals I then used CPRD Aurum and HES APC to limit the study population to those who had their first acute cardiovascular event in the same year as an influenza vaccine. Using SCCS analyses, I investigated the association between acute cardiovascular events and influenza vaccine, stratifying by cardiovascular risk.

7.2. Study design considerations

Vaccine effectiveness studies using observational data are susceptible to selection bias due to non-randomisation of individuals to vaccinated and unvaccinated groups [270,271]. In an influenza vaccine effectiveness study of adults, healthy vaccinee bias may impact the study as healthier adults, without underlying health conditions or with well managed conditions, are more likely to take up the offer of annual influenza vaccination. Frailty bias occurs when the frailest individuals are preferentially not vaccinated, perhaps because they are nearing the end of life, but have the worse outcomes. Without the

necessary statistical adjustment for frailty and health seeking behaviour, overestimation of vaccines effectiveness in more frequently healthy individuals who have low baseline risk of the outcome of interest (i.e. acute cardiovascular events) can occur. Conversely, confounding by indication occurs as those with underlying health conditions (i.e. less healthy) are more likely to be vaccinated than healthy individuals without any health conditions associated with severe infection or adverse outcomes. Inadequate statistical adjustment for risk factors and comorbidities may, therefore, lead to an underestimation of effectiveness in less healthy individuals who are at higher risk of any adverse outcomes, such as an acute cardiovascular event, regardless of preventing influenza infection [270].

To ensure baseline balance between vaccinated and unvaccinated groups, I initially planned to conduct a cohort study with propensity score matching (**Table *7.1***). A propensity score is the probability of intervention assignment (here vaccination) conditional on observed baseline characteristics [272]. Therefore, the propensity score permits an observational study design and analysis to mimic (to some degree) an RCT by balancing the distribution of baseline characteristics which predict the intervention [273]. Confounding should be reduced as only vaccinated and unvaccinated individuals with similar scores are compared. For a binary exposure, the propensity score is the probability of being exposed, given the measured confounders [274]. Scores are estimated by fitting a logistic regression model for the exposure with potential confounding factors included as explanatory variables. The propensity scores can then be used to obtain estimates of the exposure effect through matching, stratification, covariate-adjustment or inverse-probability weighting [274].

Table 7.1 Possible study design advantages and limitations

Study design	Overview	Advantages	Limitations	Assumptions
Cohort with propensity score matching	Include all eligible individuals regardless of vaccination status. By cardiovascular risk level, use baseline covariates to assign propensity score to match vaccinated and unvaccinated individuals who have similar scores.	- Directly estimate incidence rates in vaccinated and unvaccinated populations. - Can perform ratio of ratios analysis using influenza season and summer data to address residual confounding. - Simple to analyse, present, and interpret	- Difficult to define a suitable control group. Some individuals end up not matched and hence excluded from the analysis, resulting in a loss of both precision and generalizability. Due to high vaccine uptake in ≥65 years, many would be excluded. To avoid exclusion could use matching with replacement but this leads to complex analysis methods. - Unmeasured confounding remains an issue. Can use of negative control exposure and outcome will aid identification of residual confounding but cannot correct for it.	- No unmeasured confounding assumption not met – there is likely differences between patients who are and are not vaccinated which cannot be accounted for in analysis. - Every patient has a non-null probability of receiving the vaccine.
Cohort with clone-censor	Include all eligible individuals regardless of vaccination status. Use two analysis methods and compare results. 1. Analyse with time-varying exposure. 2. Use clone-censor method with inverse probability weighting to account for	- Directly estimate incidence rates in vaccinated and unvaccinated populations. - Can perform ratio of ratios analysis using influenza season and summer data to address residual confounding. - Clone-censor method accounts for measured	- Need to account for baseline being before vaccination status determined (hence time-varying exposure and clone-censor methods to be used). - Unmeasured confounding remains an issue. Can use of negative control exposure and outcome will aid identification	

	small proportion of patients aged ≥65 years who are unvaccinated.	confounding at baseline and immortal-time bias.	of residual confounding but cannot correct for it. - Difficult to define a suitable control group. - Much larger sample size, which can be very computationally intensive and requires bootstrapping.	
Self-controlled case series	Only include individuals who are vaccinated and experience outcome of interest. Thereby patients act as their own controls with exposed period compared with unexposed periods.	- Overcomes limitation of smaller proportion of unvaccinated individuals aged ≥65 years, which makes finding suitable comparators difficult in a cohort study. - Removes issue of residual confounding, thereby overcomes healthy vaccinee effect.	- Only relative incidence is estimated. - May be lack of unexposed time if majority of individuals are vaccinated early in season. - Acute cardiovascular event incidence known to be higher in winter months which is likely after vaccine given so may make interpretation of results difficult. - Need to pre-define period for which vaccine would be protective. - For non-recurrent events, works only when the event risk is small over the observation period.	- Events must occur independently assumption can be met – as cardiovascular recurrences are not independent, only first event is included. - Subsequent exposures should not be affected by previous events assumption should be met – again by only including first cardiovascular event. - Event does not increase the probability of death assumption can be met – sensitivity analysis of only non-fatal events

			- Low vaccine uptake in <65s so smaller (and highly select) study population in this age group	
Case-control	Include all eligible patients with and without outcome of interest. Determine vaccination status for the season in question.	- Simpler analysis methods than cohort or SCCS. - Sample easier to define (as compared with a cohort study).	- Appropriate control group selection, to ensure that the distribution of vaccination is the same, is difficult due to high vaccine uptake. - Need to account for timing of cardiovascular event, as those with event early in influenza season would have less chance of vaccination before event.	- No unmeasured confounding assumption not met – there is likely differences between patients who are and are not vaccinated which cannot be accounted for in analysis. - No informative censoring assumption

Cohort studies with propensity score matching has previously be used to investigate the association between influenza vaccine and secondary cardiovascular events or all-cause mortality [275–277]. For example, Wu *et al* did so for secondary cardiovascular events among individuals aged ≥65 years hospitalised after an MI [277]. The analysis assigned the index date as the 181st day after hospital discharge. Individuals were classified as vaccinated if they received an influenza vaccine between the date of hospital discharge and the index date. Age, sex, socioeconomic status, pre-existing comorbidities, and concomitant medications were used to predict the propensity score. The propensity score was then used to identify a unique matched counterpart from the original unvaccinated cohort for each vaccinated patient (1:1 matching). Individuals were all followed for 12 months after the index date or until loss to follow-up.

However, in the UK, influenza vaccine uptake is high among individuals aged ≥65 years [131]. This would result in an inadequate unvaccinated group to allow propensity score matching when using primary care EHR data. To account for this matching with replacement (i.e., using the same unvaccinated individual to match to multiple vaccinated individuals) could be used, but there would likely still be insufficient similarities between the characteristics of the vaccinated and unvaccinated individuals. Therefore, I instead considered using a clone and censor method to emulate a trial design (***Table 7.1***) [278]. A clone and censor design controls for confounding and immortal-time bias. Follow up would be set as a predefined data, such as 1 September when influenza vaccine eligibility assessment should be started for the season with exposure occurring during follow up [113]. Cloning creates two copies of each patient record: one is allocated to the exposure group and the other to the unexposed group. A clone is censored from follow-up when the exposure status is no longer compatible with the group entered [279]. For example, a clone would be censored from the unexposed group when the patient is vaccinated and censored from the exposure group if influenza vaccine is not received within the predefined vaccination period. **Figure *7.1*** illustrates the design. Inverse probability-weights, similar to propensity scores, would be used in logistic regression to control for variables predictive of influenza vaccination. However,

despite this study design and any additional use of negative exposure or negative outcome analyses to identify residual confounding, such confounding would not be controlled for in the analysis. Similar issues would occur in a case-control study design.

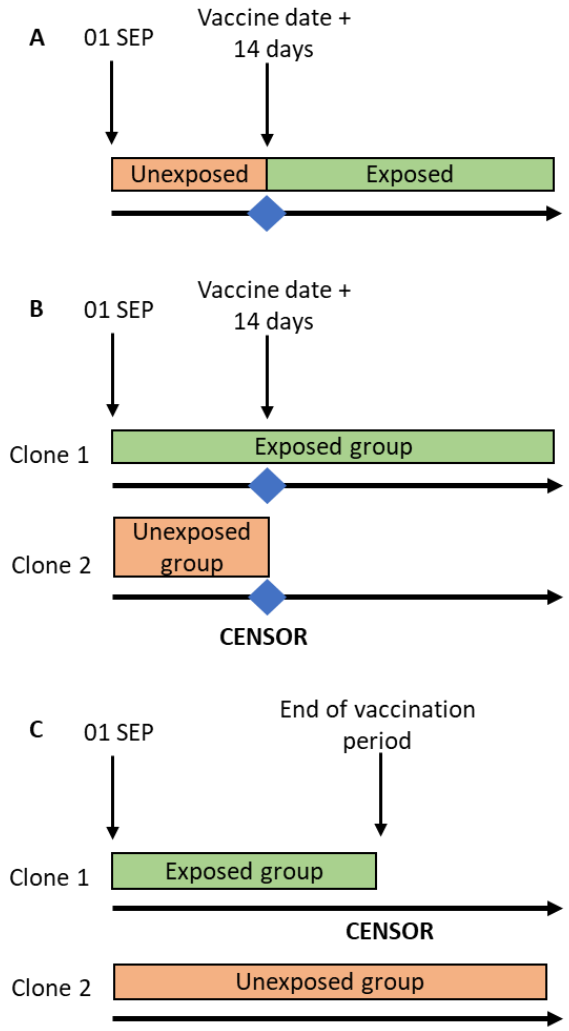


Figure 7.1 Time-updating exposure study design compared to clone and censor study design
In a time-updating exposure design, individual start follow-up as unexposed and move to being exposed after vaccination (A). In clone and censor design individuals are classified as both exposed and unexposed at the start of follow-up and are censored from the unexposed group if the vaccine is received within a predefined vaccination time period (B) or if the vaccine is not received within the predefined vaccination time period, then censored from the exposed group (C). Figure adapted from [278]; which is published by Oxford Academic under a CC NC 4.0 license.

The SCCS method is more robust than case-control or cohort study methods at accounting for the healthy vaccinee effect as it controls for implicitly controls for fixed confounding effects and you can still adjust for time-varying confounders in the models. Smeeth *et al*, who conducted one of the early influenza vaccine SCCS studies, compared the incidence of MI and stroke during exposed and unexposed periods following influenza vaccine using CPRD (then GPRD) data from 1987 to 2001, with the assumption at the time that the vaccine may increase the likelihood of cardiovascular event [163]. Follow up started at the date of first influenza vaccine within the observation period as CVD is an indication for the vaccine, thereby making the probability of vaccine receipt associated with the risk of a cardiovascular events and should have ensured minimal variation in vaccine opportunity during the observation period. However, multiple years were included in the analysis during which some individuals may not have received subsequent vaccination during subsequent influenza seasons and in which time health state may also have changed resulting in increased likelihood of an MI or stroke. Two further CPRD-based SCCS studies investigating the association between influenza vaccine and MI or stroke used a similar methodology [190,191].

After considering cohort, case-control and SCCS study designs, I determined SCCS analysis would be the most suitable methodological approach to address thesis objective 3. Such design minimized the selection biases outlined above. In my SCCS design, I set follow-up to start at 1 September and end at 31 August of the following year to ensure each individual was only included for the year in which they received influenza vaccine and had their first acute cardiovascular event and providing follow-up time before and after vaccination.

7.3. Submitted paper

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	230019	**Title**	Ms
First Name(s)	Jennifer		
Surname/Family Name	Davidson		
Thesis Title	Acute respiratory infections, cardiovascular complications, and prevention among people with raised cardiovascular risk		
Primary Supervisor	Dr Charlotte Warren-Gash		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	European Heart Journal
Please list the paper's authors in the intended authorship order:	Jennifer A Davidson, Prof Amitava Banerjee, Prof Ian Douglas, Clémence Leyrat, Richard Pebody, Helen I McDonald, Emily Herrett, Harriet Forbes, Prof Liam Smeeth, Charlotte Warren-Gash
Stage of publication	**Undergoing revision**

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I am the first author of this research paper. With support of my co-authors, I designed the study, conducted the analysis, drafted and revised the manuscript.

SECTION E

Student Signature	C Davidson
Date	25/07/2022

Supervisor Signature	Charlotte Warren-Gash
Date	25-07-2022

Title: Association between influenza vaccine and first acute cardiovascular events: a self-controlled case-series study, England, 2008-2019

Authors' names: Jennifer A Davidson, MSc[a], Prof Amitava Banerjee, DPhil[b], Prof Ian Douglas, PhD[a], Clémence Leyrat, PhD[a,c], Richard Pebody, PhD[d], Helen I McDonald, PhD[e,f], Emily Herrett, PhD[a], Harriet Forbes, PhD[g], Prof Liam Smeeth, FRCGP[a], Charlotte Warren-Gash, PhD[a]

Total word count: 4,222

Affiliations: [a]Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK; [b]Institute of Health Informatics, University College London, London, UK; [c]Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK; [d]Institute of Epidemiology and Health Care, University College London, London, UK; [e]National Institute for Health Research Health Protection Research Unit in Immunisation, London School of Hygiene and Tropical Medicine in partnership with the UK Health Security Agency, London, UK; [f]Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK; [g]Population Health Sciences, Bristol Medical School, University of Bristol, UK

Address for correspondence: Jennifer Davidson, jennifer.davidson@lshtm.ac.uk, Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, United Kingdom, +44 (0)20 7636 8636

ABSTRACT

Aims Previous studies show a reduced incidence of first myocardial infarction and stroke one to three months after influenza vaccination, but it is unclear how underlying cardiovascular risk impacts the association.

Methods and results We used linked Clinical Practice Research Datalink, Hospital Episode Statistics Admitted Patient Care and Office for National Statistics mortality data from England between 01/09/2008 and 31/08/2019. We included individuals aged 40-84 years with a first acute cardiovascular event and influenza vaccination occurring within 12 months of each September. Using a self-controlled case series analysis, we generated season-adjusted incidence ratios (IRs) for cardiovascular events after vaccination compared with baseline time before and >120 days after vaccination. We stratified by cardiovascular risk using diagnosed hypertension and overall predicted risk using QRISK2 score. We included 193,900 individuals with a first acute cardiovascular event and influenza vaccine. 105,539 had hypertension and 172,050 had a QRISK2 score ≥10%. In our main analysis, acute cardiovascular event risk was reduced in the 15-28 days after vaccination (IR 0.72 [95% CI 0.70-0.74]) and, while the effect size tapered, remained reduced to 91-120 days after vaccination (0.83 [0.81-0.88]). Reduced cardiovascular events were seen after vaccination among individuals of all age groups and with raised and low cardiovascular risk.

Conclusions Influenza vaccine may offer cardiovascular benefit among individuals at varying cardiovascular risk. Further studies are needed to characterize the populations who could derive the most cardiovascular benefits from vaccination.

Keywords influenza vaccine; cardiovascular complications; hypertension; QRISK

INTRODUCTION

Annual trends in acute cardiovascular events, such as myocardial infarction (MI) and cardiovascular mortality mirror influenza seasonality. A population-level association between influenza circulation and cardiovascular events exists after controlling for incidence trends, seasonality, and environmental factors.[1,2] Previous studies show MI and stroke incidences are up to six-times higher in early time periods after clinically diagnosed influenza-like illness or laboratory-confirmed influenza virus infection.[3,4] Several underlying mechanisms may explain influenza-triggered cardiovascular events: influenza may directly affect the vascular cells or induce hemodynamic, inflammatory, and pro-coagulant processes.[5]

Influenza-related complications and mortality are common among older individuals and those with underlying health conditions, such as established cardiovascular disease (CVD).[6] Before the COVID-19 pandemic in England, like many high-income countries, long-standing policy recommended influenza vaccination for everyone aged ≥65 years and adults aged <65 years with an underlying health condition ('clinical risk group').[7,8] Since the winter of 2020/21, vaccine policy recommendations in England have been extended to include all adults aged ≥50 years. Widespread uptake of influenza vaccine aims to protect individuals at risk of severe illness and reduce health service pressure during winter months by lessening influenza morbidity. Reducing winter health service pressure has been critically important during the COVID-19 pandemic, which has caused substantial health system burden as well as significant morbidity and mortality. In England, influenza vaccine uptake in individuals aged ≥65 years is routinely high at nearly 75%[9] but low in adults <65 years in a clinical risk group,[9,10] and was also only 35% and 46% among the newly recommended group of people aged 50-64 and not in a clinical risk group in 2020/21 and 2021/22, respectively.[9] The vaccine uptake seen in England is far higher than that in many other European countries.[8]

Influenza as a trigger of cardiovascular complications provides a potential target for CVD prevention by vaccination. Two meta-analyses of secondary prevention randomized controlled trials (RCTs) among people with chronic heart disease found a significant reduction in cardiovascular mortality (55%)[11] and

cardiovascular complications (36%)[12] after influenza vaccination. Recent RCTs continue to evaluate the cardiovascular benefits of vaccination among individuals with CVD.[13–16] Trial results suggest the cardiovascular protection provided by influenza vaccine is comparable to other secondary CVD prevention strategies.[17] However, there are no RCTs which have examined use of influenza vaccine for primary CVD prevention. Results from observational studies are mixed but suggest a reduction in the relative incidence of first MI and stroke one to three months after vaccination.[18–21] Therefore, influenza vaccine may also have a role in primary CVD prevention.[22] People with raised cardiovascular risk e.g., due to hypertension, but without established CVD, are not specifically recommended to receive the influenza vaccine in England. Recent analyses found an increased incidence of cardiovascular complications after acute respiratory infections, including pneumonia, influenza, and COVID-19, among adults with raised cardiovascular risk.[23,24]

Our current study aimed to investigate the association between influenza vaccination and acute cardiovascular events, considering individual cardiovascular risk, using the self-controlled case series (SCCS) method[25] with linked electronic health records from England.

METHODS

Data sources

We used linked anonymized data from the Clinical Practice Research Datalink (CPRD) Aurum build June 2021,[26] Hospital Episodes Statistics Admitted Patient Care (HES APC), and deaths recorded by the Office of National Statistics (ONS). CPRD Aurum contains longitudinal primary care records, currently comprising >40 million individuals. The dataset includes demographic and lifestyle factors, consultation records with symptoms, diagnoses, prescriptions, immunizations, tests, and referrals.[26] Data are coded using the Systemized Nomenclature of Medicine (SNOMED), Read and local codes. HES APC contains diagnoses and procedures from National Health Service hospital inpatients in England.[27] HES APC and

ONS deaths data, containing death date and cause, are coded using International Classification of Diseases 10th version.

The CPRD Independent Scientific Advisory Committee (application 21_000428) and the London School of Hygiene and Tropical Medicine Ethics Committee (application 26191) approved the study. CPRD provided relevant linked data for the study population.

Study design

SCCS uses within-person comparisons, i.e., individuals act as their own controls during different time periods with only individuals with the exposure (in our study influenza vaccine) and outcome (acute cardiovascular events) of interest included.[25] SCCS analyses investigate the effect of a time-varying exposure on the outcome using conditional Poisson regression models to derive incidence ratios (IRs) by comparing the incidence of events during risk time with the incidence during baseline time.[28] As only cases are sampled, the likelihood is conditional on an event having occurred during the observation period.

The main advantage of the SCCS design is the removal of confounding due to fixed characteristics, recorded or not, that vary between individuals.[25] In observational vaccine effectiveness studies, it is vital to remove confounding: vaccinated and unvaccinated individuals may have health, lifestyle and behavioral differences that are difficult to ascertain in routinely collected data.[29]

For the SCCS method to produce unbiased effect estimates of the association between an exposure and event, some key assumptions are required.[30,31] First, event recurrences must be independent i.e., an event must not increase the probability of a further event. Second, an event should not impact subsequent exposure. Third, an event must not influence the end of the period of observation, but the assumption if often violated when the event increases the likelihood of mortality.

Study population and follow-up

The source population included CPRD Aurum recorded adults aged 40-84 years with ≥12 months current post-registration time from 01/09/2009-31/08/2019. We ended our study in 2019 to prevent the introduction of bias due to COVID-19 circulation from the start of 2020 onwards. We identified those with a first acute cardiovascular event in the same 12-month period (01/09-31/08) as influenza vaccination. We identified influenza vaccination records in CPRD data.[32] In England, the influenza vaccination programme begins annually in September, ahead of the influenza season that usually occurs between December and March.[33]

We defined our outcome of any acute cardiovascular event as; MI, unstable angina, acute left ventricular heart failure, stroke, transient ischemic attack or acute limb ischemia. Our secondary outcomes were each of the cardiovascular conditions separated out (except for acute limb ischemia due to small numbers). We included diagnoses coded in CPRD or HES APC.[32] To ensure we only included first acute cardiovascular events (to meet the first SCCS assumption outlined in our study design section), we excluded individuals with a previous diagnosis (from CPRD and HES APC data), major intervention for or clinical review specific to CVD (as recorded in CPRD) before the start of follow-up. We defined CVD as heart disease (congenital or otherwise), heart failure, stroke or transient ischemic attack.[32]

We stratified the study population by cardiovascular risk. In separate analyses, we defined cardiovascular risk by hypertension and QRISK2 score, the latter being the cardiovascular risk score used in primary care practice in England during our study period. We only included persistent and diagnosed hypertension, defined by coded CPRD diagnoses.[32] QRISK2 uses many risk factors to estimate an individual's absolute ten-year risk of CVD.[34] The risk factors considered are age, sex, ethnicity, deprivation score for area of residence, family history of coronary heart disease in a first degree relative <60 years, diabetes, atrial fibrillation, chronic kidney disease stage 4 or 5, rheumatoid arthritis, ratio of total serum cholesterol to high density lipoprotein cholesterol, systolic blood pressure, treated hypertension, body-mass index, and smoking status. We calculated QRISK2 scores using the published

definitions and weights assigned to each risk factors.[34] As part of the standard approach to calculating QRISK2 scores, we considered the absence of a code for comorbid conditions to equate to absence of the condition (i.e., if the individual had no diabetes code recorded it was determined that the individual did not have diabetes) and imputed missing lifestyle and anthropometric measures (such as body-mass index) with population average values. Our full method is published online.[35] All individuals aged ≥85 years are classed as having a QRISK2 score ≥10% due to age alone,[36] so we limited our study population to those aged <85 years. We classified individuals as having raised cardiovascular risk (hypertension or QRISK2 score ≥10%) or not (no hypertension or QRISK2 score <10%) at baseline (1 September).

We excluded individuals who had their first acute cardiovascular event on the same day as influenza vaccination, as the two events were likely retrospectively recorded. Follow-up started on 1 September each year and ended at the earliest of; date of death, loss to follow-up (date of leaving the practice or the last data collection from the practice), or 31 August of the following year (**Figure *7.2***).

Statistical analysis

We conducted all analyses in Stata (version 16).

We described the baseline characteristics overall and stratified by cardiovascular risk including age group, sex, events associated with a hospital stay, events resulting in death, deaths during follow-up, and loss to follow-up.

We compared the incidence of acute cardiovascular events during risk periods following influenza vaccination with all baseline periods for each person (**Figure *7.2***). Our risk period was the 120 days after vaccination date, subdivided into the stratum of 15-28, 29-59, 60-90, and 91-120 days. We choose a 120-day risk window after influenza vaccination to cover the main period of influenza virus circulation. We excluded the 14 days before and the 14 days after vaccination from risk and baseline time.[19] The 14 days before vaccination were excluded as acute cardiovascular events during this period likely affect the subsequent likelihood of receiving an influenza vaccine, a violation of a SCCS assumption (the second

assumption outlined in our study design section). The 14 days after vaccination were excluded and presented separately as it can take up to 14 days for the vaccine to become effective.[37]

We calculated IRs using conditional Poisson regression for acute cardiovascular events occurring within each risk period stratum compared with baseline. We adjusted for season using the binary classification of warm months (April-September) and cool months (October-March).[18]

We stratified results by age group (40-64, 65-74 and 75-84). Adults aged 40-64 years are selectively offered influenza vaccine based on specific underlying health conditions, so the individuals included in our study from this age group are not representative of the overall age group. Additional stratifying factors were sex (male and female) and the timing of vaccination (≤15 November or >15 November). Late vaccination, after mid-November, has previously been associated with reduced vaccine efficacy.[38] Hypotheses for the difference between early and late vaccine response include an insufficient time for late recipients to develop an immune response before exposure to circulating virus, depletion of susceptibles, or differences in the characteristics and motivations for vaccination, such as late recipients being vaccinated in response to influenza epidemic levels.[39]

We performed three pre-specified sensitivity analyses. Firstly, we repeated our initial analysis excluding fatal acute cardiovascular events. Acute cardiovascular events can result in death, violating the SCCS assumption that observation periods should end independently of event timing (the third assumption outlined in our study design section)[31]. We classified fatal events as those for which the individual's death date was ≤30 days after the event. We also further stratified QRISK2 scores of ≥10% into 10-19% and ≥20% to consider finer definitions of cardiovascular risk.

To assess any violation of the assumption that an event should not influence subsequent exposure (the second SCCS assumption outlined in our study design section), we first assessed, using histograms, the difference in the number of days between vaccination and acute cardiovascular event by age group.[30] We then used a sensitivity analysis to redefine our study population with follow-up from influenza

vaccination date. We used a fixed follow-up until 31 August, regardless of survival, given we only had one exposure and the event, by definition, could only be after the exposure (Supplementary Figure 1). Therefore, all baseline time was from 121 days after vaccination until 31 August. Cardiovascular risk level was defined at the date of vaccination in this sensitivity analysis.

RESULTS

Description of the study population

We included 193,900 individuals aged 40-84 years who had a first acute cardiovascular event in the same year as an influenza vaccine (**Figure *7.3***). 19,868 (10.2%) of individuals died and 9,201 (4.7%) were lost during follow-up. Overall, 90,959 (46.9%) individuals were women, 149,663 (77.2%) were aged 65-84 years, 105,539 (54.4%) had diagnosed hypertension and 172,050 (88.7%) had a QRISK2 score of ≥10% (**Table *7.2***). Individuals with hypertension were older than those without hypertension (40-64 years: 17.0% [17,969] vs 29.7% [26,268]). Individuals with a QRISK2 score ≥10% were much older than those with a QRISK2 score <10% (40-64 years: 14.5% [24,898] vs 88.5% [19,339]) and a higher proportion of individuals with a QRISK2 score ≥10% died compared to those with a QRISK2 score <10% (10.8% [18,641] vs 5.6% [1,227]).

Association between influenza vaccine and first acute cardiovascular event

A significant reduction in the season-adjusted incidence of first acute cardiovascular event was observed throughout the 120-day risk period after influenza vaccination. There was a tapering in the risk reduction over time; with a 28% (IR 0.72 [95% CI 0.70-0.74]) reduction 15-28 days post-vaccination and 16% (0.84 [0.82-0.85]) 91-120 days post-vaccination. When stratified by cardiovascular risk, there was a larger reduction for individuals without hypertension (15-28 days 0.66 [0.64-0.69]) than for those with hypertension (15-28 days 0.76 [0.74-0.79]). Results were similar when raised cardiovascular risk was defined by QRISK2 score ≥10% (15-28 days 0.76 [0.74-0.78]), but there was a more substantial reduction for individuals with a QRISK2 score <10% (15-28 days 0.48 [0.44-0.52]). The full results are in **Table**

7.3. Analysis of the secondary outcomes showed the reduction in relative incidence following influenza vaccination was more substantial for MI (15-28 days 0.60 [0.57-0.64]) than other cardiovascular events (Figure 7.4). Secondary outcomes by cardiovascular risk are presented in Supplementary Figure 2-5.

Results were markedly different between age groups (**Table *7.3***) with a much larger reduction in the relative incidence of first acute cardiovascular event in risk periods for individuals aged 40-64 years (15-28 days 0.54 [0.51-0.57]), compared with 65-74 and 75-84 years (15-28 days 0.80 [0.77-0.84] and 0.80 [0.77-0.83], respectively) (p-value for interaction <0.0001). The pattern was similar across all cardiovascular risk groups, although no one aged 75-84 years had a QRISK2 score <10%.

The relative incidence by sex is in Supplementary Table 1. The IR reduction was larger in men than women, for example IRs for 15-28 days post-vaccination were 0.69 (0.67-0.72) and 0.76 (0.73-0.79), respectively (p-value for interaction <0.0001).

There was a slight difference in the relative incidence of first acute cardiovascular events for individuals vaccinated on or before 15 November and after 15 November (15-28 days: 0.73 [0.71-0.75] vs 0.69 [0.66-0.73], respectively) (p-value for interaction <0.0001) (Supplementary Table 2).

Annual breakdowns did not reveal any substantial differences in reduced relative incidence following vaccination (Supplementary Table 3). Results for the first 14 days after vaccination are presented in Supplementary Table 4.

Sensitivity analysis removing people who had a fatal acute cardiovascular event

After exclusion of fatal acute cardiovascular events (13,193), the IRs during risk periods remained broadly similar to the main analysis overall (15-28 days 0.76 [0.74-0.78] and 91-120 days 0.81 [0.79-0.82]) and across all cardiovascular risk groups (Supplementary Table 5).

Sensitivity analysis with more refined QRISK2 score stratification

When raised cardiovascular risk defined by QRISK2 score was separated into 10-19% and ≥20%, the reduction in relative incidence among individuals aged 65-74 years was greater in those with a risk score ≥20% (15-28 days 0.79 [0.74-0.83]) than a risk score of 10-19% (0.82 [0.77-0.88]) but still broadly similar. Most individuals aged 75-84 years had a QRISK2 score ≥20%. Among those aged 40-64 years, the reduction was greater in those with a QRISK2 score of 10-19% (Supplementary Table 6).

Sensitivity analysis study design

Our investigation of the timing of the event centered to vaccination showed that a high number of events in individuals aged 40-64 years occurred prior to vaccination (Supplementary Figure 6). Overall the baseline characteristics of the sensitivity analysis study population were similar to those of the main study population (Supplementary Table 7), but they were slightly older (40-64 years: 18.6% [29,927] vs 22.8% [44,237]) with a higher proportion of individuals having a QRISK2 score of ≥10% (91.4% [147,023] vs 88.7% [172,050]). Compared to the main study design and population, there was a smaller reduction in the relative incidence of a first acute cardiovascular event during early risk periods after vaccination (15-28 days 0.94 [0.91-0.96]) and no reduction by 91-120 days (1.00 [0.98-1.02]) (Supplementary Table 8). Among individuals aged 40-64 years there was no difference in the relative incidence during risk periods compared to baseline (Supplementary Table 8).

DISCUSSION

Summary

Using English primary and secondary data electronic health records from 2008-2019, we found individuals with both raised and low cardiovascular risk had a reduced incidence of a first acute cardiovascular event after influenza vaccination after adjusting for season (we used a binary classification of warmer and cooler months but when season was adjusted for using four season [results not shown] the change in association was the same). The reduced incidence was largest in in the 15-28 days after

vaccination but persisted to 120 days. The effect size varied from 6-28% across different analyses of study population groups. The protective effect was evident across all age groups in the main analyses but was confined to those ≥65 years in the final sensitivity analysis with follow-up from vaccination date.

Comparison with existing literature

Our main finding for the whole study population was consistent with the results generated by previous SCCS studies using CPRD data. Analysis of data from 1987 to 2001 found a 12% (IR 0.88 [0.80-0.97]) and 13% (IR 0.87 [0.79-0.96]) reduction in the relative incidence of first stroke and MI, respectively, in the 15-28 days after influenza vaccination after which time there was no significant reduction.[18] Two other SCCS studies with CPRD data from 2001 to 2009 found incidence ratios of 0.75 (0.66-0.86) and 0.76 (0.70-0.84) for first MI and stroke, respectively, in the 15-28 days post-vaccination.[19,20] Although we used a composite acute cardiovascular event outcome, when we looked at individual cardiovascular outcomes, the greatest reduction in relative incidence was for MI.

Previous studies have shown that individuals with raised cardiovascular risk have more acute cardiovascular complications following respiratory infection.[23,24] Sen *et al* used Norwegian electronic health record data from 2009/10 to investigate the impact of underlying cardiovascular risk on the association between the H1N1 influenza vaccine and cardiovascular events.[21] The study identified conflicting results, with a reduced relative incidence of MI (15-28 days post-vaccination: IR 0.70 [0.57-0.85]) in those with raised cardiovascular risk and an increase (15-28 days post-vaccination: IR 3.17 [1.99-5.07]) among people at low cardiovascular risk. The study defined cardiovascular risk using cardiovascular prevention prescriptions at the time of vaccination, after follow-up had started, which likely biased results when stratified by cardiovascular risk so comparisons to our results is difficult.

We showed similar protective associations between influenza vaccination and acute cardiovascular events regardless of cardiovascular risk level in people aged ≥65 years. However, among those aged 40-64 years, there was an apparently greater protective association in those at low underlying cardiovascular risk in

main analysis (though not in our sensitivity analysis study design). Individuals who have low cardiovascular risk or who are younger have a lower baseline risk of cardiovascular complications, whereas individuals with raised cardiovascular risk or who are older have a high risk all year round. In higher risk older people influenza-associated cardiovascular complications may explain a lower proportion of cardiovascular events.

Strengths and limitations

We used a large study population from primary and secondary care linked data sources generalizable to the English population. The large study population allowed us to stratify simultaneously by cardiovascular risk, age, and a third factor such as sex. Thereby, allowing us to unpick the findings of our initial analysis in more detail than previous SCCS studies. We compared results across two measures of cardiovascular risk; QRISK2 score and diagnosed hypertension. As our study population was predominantly older, most individuals had a QRISK2 score ≥10%, limiting our ability to conclude any added benefit influenza vaccine may have in younger people with raised cardiovascular risk.

Observational studies, particularly those involving secondary analysis of routinely collected data, of vaccine effects are highly vulnerable to confounding as vaccinated individuals tend to have health, lifestyle, and behavioral differences to those who are not vaccinated.[29] The SCCS design largely overcomes confounding by such fixed individual characteristic by using within-individual comparisons. The design does not control for time-varying confounders within individuals. However, we believe time-varying confounding is likely to be minimised in our study due to the maximum one-year follow-up, with adjustment for season. Another bias not controlled for in the SCCS design is healthcare contact bias. When an individual receives their influenza vaccine this may trigger cardiovascular management for the patient, leading to a reduced incidence of cardiovascular complications including the time periods immediately following vaccination. It is not possible for such healthcare contact bias to be quantified, but our finding of the strongest association immediately after vaccination with a tapering over time could be swift immune response followed antibody waning.[40] Further research could utilize a negative outcome

control, though would need to carefully select a suitable acute event which would not be associated with influenza vaccine. The representativeness of the individuals included aged <65 and ≥65 years differs. In England before the COVID-19 pandemic, universal influenza vaccination only included individuals aged ≥65 years. Before this age, the influenza vaccine was only offered (free of charge) to those with an underlying health condition. One trigger for offering influenza vaccination would be a recent acute cardiovascular event. A high proportion of events in individuals aged <65 years (26%) occurred before vaccination, compared with low proportions in those aged 65-74 years (11%) and 75-84 years (8%). This difference suggests that events led to vaccination in some individuals aged <65 years and may explain why our sensitivity analysis study design which began follow-up at vaccination, showed no protective association in those aged <65 years. Conversely, a higher proportion of older individuals died after their event, which resulted in a short baseline interval in our main study design again potential causing bias, although results from analysis of only non-fatal events suggest this bias was small.

When using secondary data for research it is important to consider its validity. CPRD and HES is widely validated,[41] including cardiovascular events such as MI, heart failure and stroke.[42] Influenza vaccine recording in CPRD has not been specifically validated. Most patients in England will receive their influenza vaccine at their primary care practice, but some vaccines will be administrated by pharmacies or occupational health services. Vaccines received outside of primary care practice are still expected to be recorded within primary care records.[43] Both the overall influenza vaccine uptake and the patterns of regional variation are consistent with national surveillance.[9]

Clinical and public health implications

Measuring the burden and impact of seasonal influenza is difficult, but World Health Organization estimates before the COVID-19 pandemic suggested that influenza infected approximately 20% of people in Europe, depending on the circulating strains.[44] This poses a significant winter healthcare pressure and associated mortality of tens of thousands of deaths in Europe, with an estimated 400,000 respiratory deaths globally.[45] In the United States, during the 2018/19 influenza season there was an estimated 380,000 respiratory hospitalizations and 28,000 respiratory deaths.[46] The mechanisms by which influenza vaccine exerts cardiovascular benefit are uncertain. Here, we have assumed the protective effect is due to prevention of influenza which can trigger a cardiovascular event. However, there is also the possibility of pleiotropic effects between virus and the antigens of atherosclerotic plaque as well as unspecific immunomodulatory effect which in turn prevents cardiovascular complications unrelated to influenza virus circulation and infection.[47] Consideration of the different mechanistic and long-term effects should be explored in future research.

Among adults aged <65 years, some with high cardiovascular risk are already eligible to receive influenza vaccine in many European countries, including those with chronic kidney disease, severe obesity, or diabetes. However, uptake among clinical risk groups is currently moderate in England and low in other European countries.[8,9] This may be due to individual or physician perceived risk. Age eligible vaccination is operationally easier to manage. During the COVID-19 pandemic, influenza vaccine recommendations in England have been extended to all individuals ≥50 years regardless of underlying health conditions.[9] While further studies would help to fully characterize those who would derive the most cardiovascular benefit from influenza vaccine, improving uptake remains a public health priority, both to protect individuals from influenza and complications, including cardiovascular events. On summarizing the evidence generated from RCTs and observations studies to date, a recent editorial emphasized the need for cardiologists, and other physicians, to consider the cardiovascular benefits of influenza vaccine and ensure their patients receive the vaccine in the same way they would advocate the use of statins.[17] Such

promotion of influenza vaccine would be a step towards a "syndemic" approach to healthcare, acknowledging the interaction of infectious diseases and non-communicable diseases, such as CVD.[48] Ultimately, in-hospital vaccination of those hospitalized due to, or at high-risk of, cardiovascular complication is likely one of the most efficient ways to increase vaccine uptake.

CONCLUSIONS

We have shown that influenza vaccine is associated with reduced risk of cardiovascular events, regardless of underlying cardiovascular risk. Improved vaccine uptake could help reduce the risk of first acute cardiovascular events among those already eligible to receive the seasonal influenza vaccine. Furthermore, with continued widespread COVID-19 transmission, minimizing influenza impact is crucial. COVID-19 vaccine boosters are currently being rolled out and offer the opportunity to increase and prioritize influenza vaccine uptake.[49]

REFERENCES

1. Blackburn R, Zhao H, Pebody R, Hayward A, Warren-Gash C. Laboratory-Confirmed Respiratory Infections as Predictors of Hospital Admission for Myocardial Infarction and Stroke: Time-Series Analysis of English Data for 2004–2015. *Clinical Infectious Diseases* 2018;**67**:8–17.
2. Imai C, Barnett A, Hashizume M, Honda Y, Imai C, Barnett AG, Hashizume M, Honda Y. The Role of Influenza in the Delay between Low Temperature and Ischemic Heart Disease: Evidence from Simulation and Mortality Data from Japan. *International Journal of Environmental Research and Public Health* 2016;**13**:454.
3. Warren-Gash C, Hayward AC, Hemingway H, Denaxas S, Thomas SL, Timmis AD, Whitaker H, Smeeth L. Influenza Infection and Risk of Acute Myocardial Infarction in England and Wales: A CALIBER Self-Controlled Case Series Study. *The Journal of Infectious Diseases* 2012;**206**:1652–1659.
4. Warren-Gash C, Blackburn R, Whitaker H, McMenamin J, Hayward AC. Laboratory-confirmed respiratory infections as triggers for acute myocardial infarction and stroke: a self-controlled case series analysis of national linked datasets from Scotland. *European Respiratory Journal* 2018;**51**:1701794.
5. Bazaz R, Marriott HM, Francis SE, Dockrell DH. Mechanistic links between acute respiratory tract infections and acute coronary syndromes. *J Infect* 2013;**66**:1–17.

6. Mauskopf J, Klesse M, Lee S, Herrera-Taracena G. The burden of influenza complications in different high-risk groups: a targeted literature review. *J Med Econ* 2013;**16**:264–277.

7. Public Health England. Influenza: the green book, chapter 19https://www.gov.uk/government/publications/influenza-the-green-book-chapter-19 (December 3, 2020)

8. European Centre for Disease Prevention and Control. Seasonal influenza vaccination and antiviral use in EU/EEA Member Stateshttps://ecdc.europa.eu/en/publications-data/seasonal-influenza-vaccination-antiviral-use-eu-eea-member-states (July 16, 2019)

9. UK Health Security Agency. Seasonal influenza vaccine uptake in GP patients: winter season 2021 to 2022https://www.gov.uk/government/statistics/seasonal-influenza-vaccine-uptake-in-gp-patients-winter-season-2021-to-2022 (July 24, 2022)

10. Oakley S, Bouchet J, Costello P, Parker J. Influenza vaccine uptake among at-risk adults (aged 16–64 years) in the UK: a retrospective database analysis. *BMC Public Health* 2021;**21**:1–11.

11. Clar C, Oseni Z, Flowers N, Keshtkar-Jahromi M, Rees K. Influenza vaccines for preventing cardiovascular disease. *Cochrane Database of Systematic Reviews* 2015:CD005050.

12. Udell JA, Zawi R, Bhatt DL, Keshtkar-Jahromi M, Gaughran F, Phrommintikul A, Ciszewski A, Vakili H, Hoffman EB, Farkouh ME, Cannon CP. Association Between Influenza Vaccination and Cardiovascular Outcomes in High-Risk Patients. *JAMA* 2013;**310**:1711.

13. Frøbert O, Götberg M, Erlinge D, Akhtar Z, Christiansen EH, MacIntyre CR, Oldroyd KG, Motovska Z, Erglis A, Moer R, Hlinomaz O, Jakobsen L, Engstrøm T, Jensen LO, Fallesen CO, Jensen SE, Angerås O, Calais F, Kåregren A, Lauermann J, Mokhtari A, Nilsson J, Persson J, Stalby P, Islam AKMM, Rahman A, Malik F, Choudhury S, Collier T, Pocock SJ, Pernow J. Influenza Vaccination after Myocardial Infarction: A Randomized, Double-Blind, Placebo-Controlled, Multicenter Trial. *Circulation* 2021;**144**:1476–1484.

14. Vardeny O, Kim K, Udell JA, Joseph J, Desai AS, Farkouh ME, Hegde SM, Hernandez AF, McGeer A, Talbot HK, Anand I, Bhatt DL, Cannon CP, DeMets D, Gaziano JM, Goodman SG, Nichol K, Tattersall MC, Temte JL, Wittes J, Yancy C, Claggett B, Chen Y, Mao L, Havighurst TC, Cooper LS, Solomon SD, Investigators IC and. Effect of High-Dose Trivalent vs Standard-Dose Quadrivalent Influenza Vaccine on Mortality or Cardiopulmonary Hospitalization in Patients With High-risk Cardiovascular Disease: A Randomized Clinical Trial. *JAMA* 2021;**325**:39–49.

15. Loeb M, Dokainish H, Dans A, Palileo-Villanueva L, Roy A, Karaye K, Zhu J, Liang Y, Goma F, Damasceno A, KF A, G Y, C M, W A, A AM, S Y. Randomized controlled trial of influenza vaccine in patients with heart failure to reduce adverse vascular events (IVVE): Rationale and design. *Am Heart J* 2019;**212**:36–44.

16. Hollingsworth R, Palmu A, Pepin S, Dupuy M, Shrestha A, Jokinen J, Syrjänen R, Nealon J, Samson S, Bruijn I de. Effectiveness of the quadrivalent high-dose influenza vaccine for prevention of cardiovascular and respiratory events in people aged 65 years and above:

Rationale and design of a real-world pragmatic randomized clinical trial. *American Heart Journal* 2021;**237**:54–61.

17. Michos ED, Udell JA. Am I Getting the Influenza Shot Too?: Influenza Vaccination as Post–Myocardial Infarction Care for the Prevention of Cardiovascular Events and Death. *Circulation* 2021;**144**:1485–1488.

18. Smeeth L, Thomas SL, Hall AJ, Hubbard R, Farrington P, Vallance P. Risk of myocardial infarction and stroke after acute infection or vaccination. *N Engl J Med* 2004;**351**:2611–2618.

19. Gwini SM, Coupland CAC, Siriwardena AN. The effect of influenza vaccination on risk of acute myocardial infarction: Self-controlled case-series study. *Vaccine* 2011;**29**:1145–1149.

20. Asghar Z, Coupland C, Siriwardena N. Influenza vaccination and risk of stroke: Self-controlled case-series study. *Vaccine* 2015;**33**:5458–5463.

21. Sen A, Bakken IJ, Govatsmark RES, Varmdal T, Bønaa KH, Mukamal KJ, Håberg SE, Janszky I. Influenza vaccination and risk for cardiovascular events: a nationwide self-controlled case series study. *BMC Cardiovascular Disorders* 2021;**21**:31.

22. Behrouzi B, Udell JA. Universal flu vaccines: a shot at lifelong cardioprotection? *Nature Reviews Cardiology 2022 19:3* 2022;**19**:145–146.

23. Matsushita K, Ding N, Kou M, Hu X, Chen M, Gao Y, Honda Y, Zhao D, Dowdy D, Mok Y, Ishigami J, Appel LJ. The relationship of COVID-19 severity with cardiovascular disease and its traditional risk factors: A systematic review and meta-analysis. *Global Heart* 2020;**15**:64.

24. Davidson JA, Banerjee A, Smeeth L, McDonald HI, Grint D, Herrett E, Forbes H, Pebody R, Warren-Gash C. Risk of acute respiratory infection and acute cardiovascular events following acute respiratory infection among adults with increased cardiovascular risk in England between 2008 and 2018: a retrospective, population-based cohort study. *The Lancet Digital Health* 2021;**3**:e773–e783.

25. Whitaker HJ, Farrington CP, Spiessens B, Musonda P. Tutorial in biostatistics: The self-controlled case series method. *Statistics in Medicine* 2006;**25**:1768–1797.

26. Clinical Practice Research Datalink. CPRD Aurum June 2021 datasethttps://cprd.com/cprd-aurum-june-2021 (October 4, 2021)

27. Herbert A, Wijlaars L, Zylbersztejn A, Cromwell D, Hardelid P. Data Resource Profile: Hospital Episode Statistics Admitted Patient Care (HES APC). *International Journal of Epidemiology* 2017;**46**:1093–1093i.

28. Whitaker HJ, Hocine MN, Farrington CP. The methodology of self-controlled case series studies. *Statistical Methods in Medical Research* 2009;**18**:7–26.

29. Remschmidt C, Wichmann O, Harder T. Frequency and impact of confounding by indication and healthy vaccinee bias in observational studies assessing influenza vaccine effectiveness: A systematic review. *BMC Infectious Diseases* 2015;**15**:429.

30. Whitaker HJ, Ghebremichael-Weldeselassie Y, Douglas IJ, Smeeth L, Farrington CP. Investigating the assumptions of the self-controlled case series method. *Stat Med* 2018;**37**:643–658.

31. Petersen I, Douglas I, Whitaker H. Self controlled case series methods: an alternative to standard epidemiological study designs. *BMJ* 2016:i4515.

32. Daivdson J, Warren-Gash C, Banerjee A, McDonald H, Smeeth L. London School of Hygiene and Tropical Medicine Data Compass. *Codelists*. https://datacompass.lshtm.ac.uk/id/eprint/2675/ (January 14, 2022)

33. Public Health England. Surveillance of influenza and other respiratory viruses in the UK: Winter 2020 to 2021

34. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, Minhas R, Sheikh A, Brindle P. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. *BMJ* 2008;**336**:1475–1482.

35. Davidson J, Strongman H, Herrett E, Gadd S. emilyherrett/qrisk_cprd_aurum: QRISK Aurum bundle version 2.0. 2022.

36. National Institute for Health and Care Excellence. Cardiovascular disease: risk assessment and reduction, including lipid modificationhttps://www.nice.org.uk/guidance/cg181 (October 26, 2021)

37. Rastogi S, Gross PA, Bonelli J, Dran S, Levandowski RA, Russo C, Weksler ME, Kaye D, levison M, Abrutyn E, Murasko D, Mossey J, Deichmiller S, Hilsen R. Time to peak serum antibody response to influenza vaccine. *Clinical and Diagnostic Laboratory Immunology* 1995;**2**:120–121.

38. Chodick G, Heymann AD, Green MS, Kokia E, Shalev V. Late influenza vaccination is associated with reduced effectiveness. *Preventive Medicine* 2006;**43**:71–76.

39. Yoo BK, Frick K. Determinants of influenza vaccination timing. *Health Economics* 2005;**14**:777–791.

40. Pebody RG, Andrews N, McMenamin J, Durnall H, Ellis J, Thompson CI, Robertson C, Cottrell S, Smyth B, Zambon M, Moore C, Fleming DM, Watson JM. Vaccine effectiveness of 2011/12 trivalent seasonal influenza vaccine in preventing laboratory-confirmed influenza in primary care in the United Kingdom: Evidence of waning intra-seasonal protection. *Eurosurveillance* 2013;**18**:20389.

41. Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. *British Journal of Clinical Pharmacology* 2010;**69**:4–14.

42. Davidson J, Banerjee A, Muzambi R, Smeeth L, Warren-Gash C. Validity of acute cardiovascular outcome diagnoses recorded in european electronic health records: A systematic review. *Clinical Epidemiology* 2020;**12**.

43. NHS England. Enhanced services specification: Seasonal influenza and pneumococcal immunisation enhanced servicehttp://www.england.nhs.uk/ourwork/commissioning/gp-contract/ (July 24, 2022)

44. World Health Organization Europe. Burden of influenzahttps://www.euro.who.int/en/health-topics/communicable-diseases/influenza/seasonal-influenza/burden-of-influenza (November 26, 2021)

45. Paget J, Spreeuwenberg P, Charu V, Taylor RJ, Iuliano AD, Bresee J, Simonsen L, Viboud C. Global mortality associated with seasonal influenza epidemics: New burden estimates and predictors from the GLaMOR Project. *Journal of Global Health* 2019;**9**.

46. Centers for Disease Control and Prevention. Estimated Flu-Related Illnesses, Medical visits, Hospitalizations, and Deaths in the United States — 2018–2019 Flu Seasonhttps://www.cdc.gov/flu/about/burden/2018-2019.html (December 17, 2021)

47. Ciszewski A. Cardioprotective effect of influenza and pneumococcal vaccination in patients with cardiovascular diseases. *Vaccine* 2018;**36**:202–206.

48. Horton R. Offline: COVID-19 is not a pandemic. *The Lancet* 2020;**396**:874.

49. Privor-Dumm LA, Poland GA, Barratt J, Durrheim DN, Deloria Knoll M, Vasudevan P, Jit M, Bonvehí PE, Bonanni P. A global agenda for older adult immunization in the COVID-19 era: A roadmap for action. *Vaccine* 2021;**39**:5240–5250.

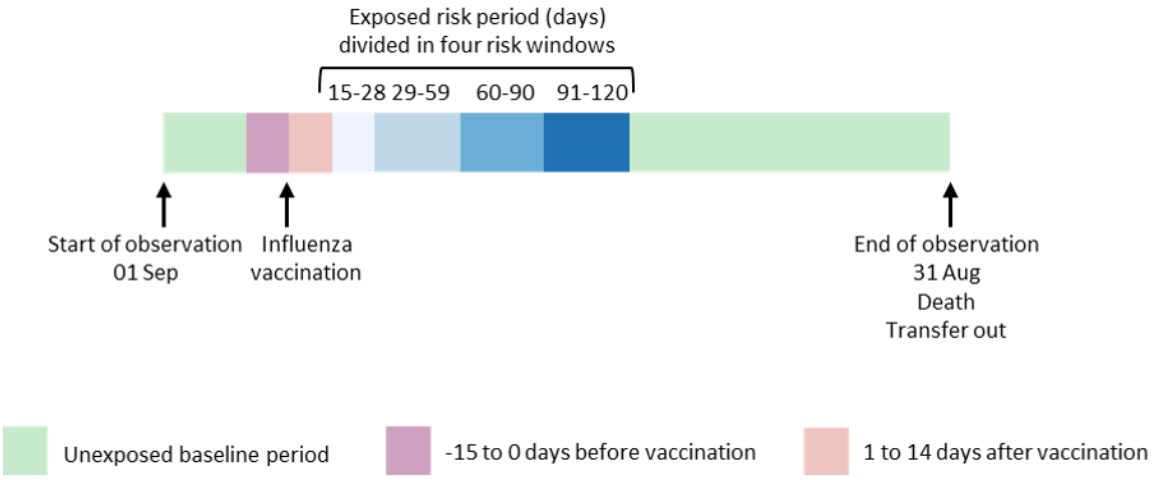


Figure 7.2 Overview of study design

Illustration of baseline and risk contributing follow-up time in relation to start of follow-up (1 September), influenza vaccine receipt and end of follow-up

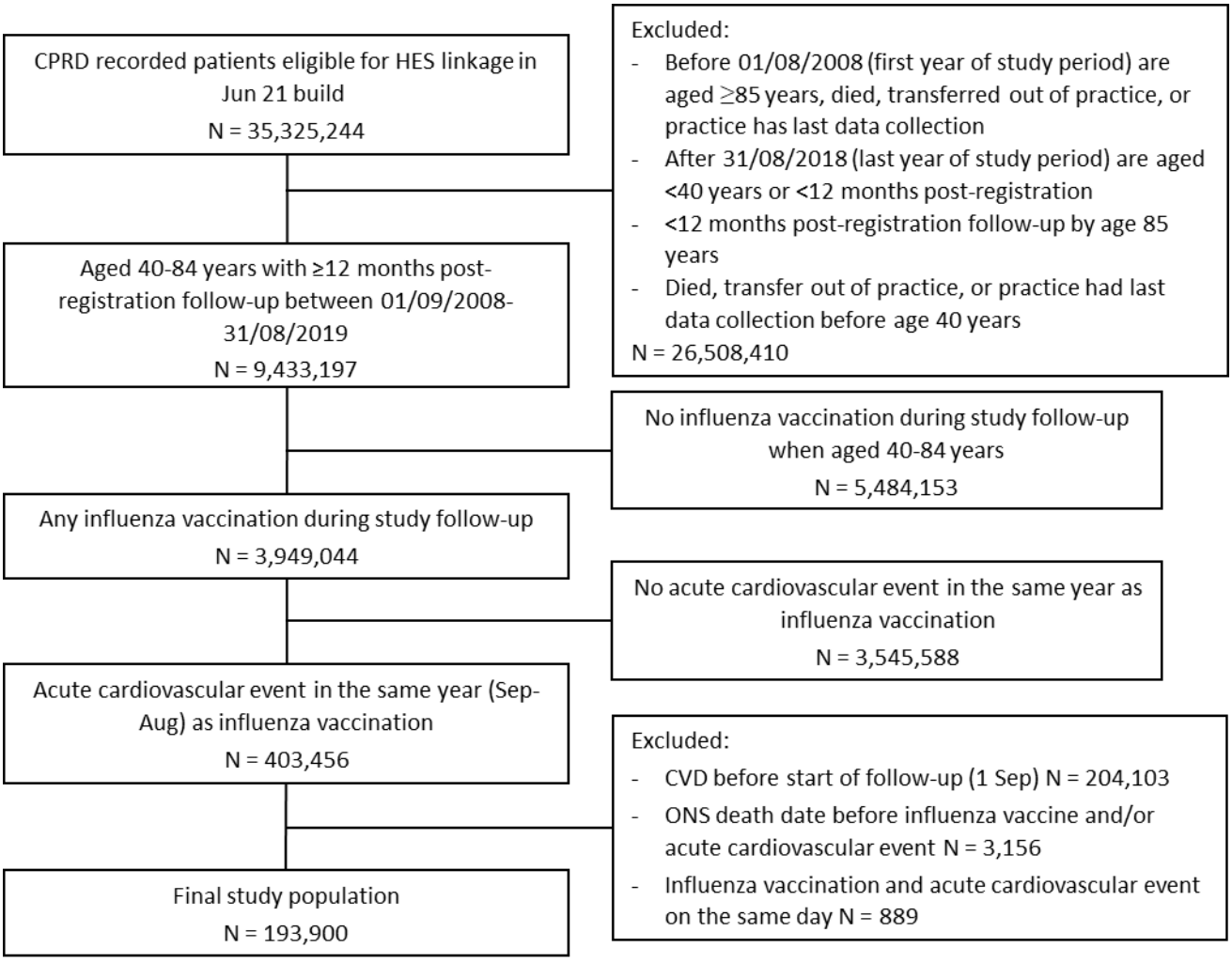


Figure 7.3 Study population flow chart

Overview of study population numbers based on inclusion and exclusion criteria. CPRD = Clinical Practice Research Datalink; CVD = cardiovascular disease; ONS = Office for National Statistics

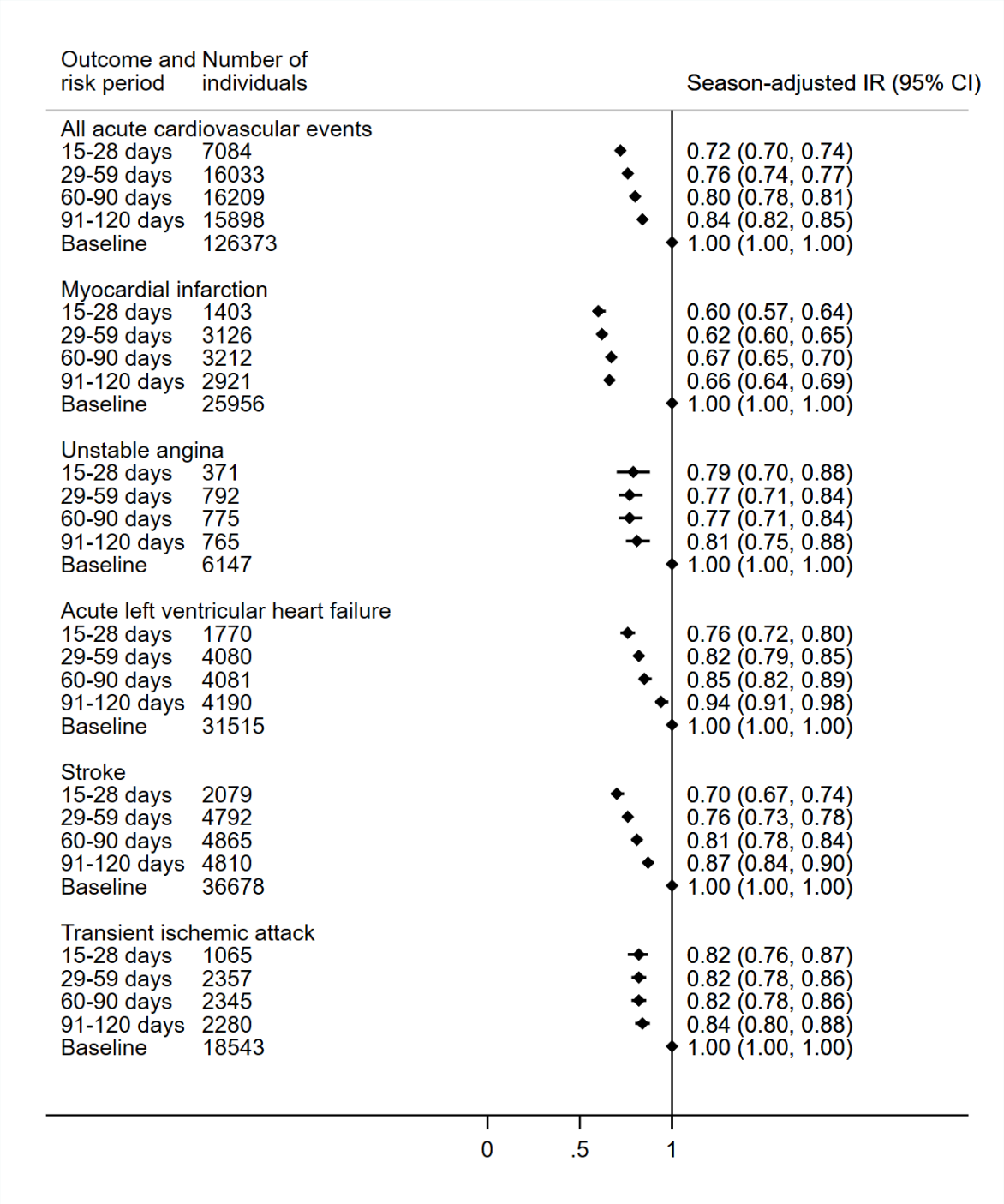


Figure 7.4 Incidence ratios for first acute cardiovascular events in risk periods following influenza vaccination by cardiovascular event type

Forest plot visualization of season-adjusted incidence ratios for primary and secondary outcomes broken down by risk periods of 15-28, 29-59, 60-90 and 91-120 days. CI = confidence interval; IR = incidence ratio

Table 7.2 Baseline characteristics of study population

	All	QRISK2		Hypertension	
		Raised risk	Low risk	Raised risk	Low risk
	n=193,900	n=172,050	n=21,850	n=105,539	n=88,361
Sex					
Female	90,959 (46.9%)	77,453 (45.0%)	13,506 (61.8%)	52,484 (49.7%)	38,475 (43.5%)
Male	102,941 (53.1%)	94,597 (55.0%)	8,344 (38.2%)	53,055 (51.3%)	49,886 (56.5%)
Age group (years)					
40-64	44,237 (22.8%)	24,898 (14.5%)	19,339 (88.5%)	17,969 (17.0%)	26,268 (29.7%)
65-74	68,742 (35.5%)	66,231 (38.5%)	2,511 (11.5%)	36,243 (34.3%)	32,499 (36.8%)
75-84	80,921 (41.7%)	80,921 (47.0%)	0 (0.0%)	51,327 (48.6%)	29,594 (33.5%)
Ethnicity					
White	146,318 (75.5%)	130,207 (75.7%)	16,111 (73.7%)	79,025 (74.9%)	67,293 (76.2%)
Black	6,345 (3.3%)	5,645 (3.3%)	700 (3.2%)	3,840 (3.6%)	2,505 (2.8%)
South Asian	1,206 (0.6%)	810 (0.5%)	396 (1.8%)	857 (0.8%)	349 (0.4%)
Other	9,121 (4.7%)	7,542 (4.4%)	1,579 (7.2%)	5,781 (5.5%)	3,340 (3.8%)
Unknown	30,910 (15.9%)	27,846 (16.2%)	3,064 (14.0%)	16,036 (15.2%)	14,874 (16.8%)
Body mass index					
Underweight (<18·5 kg/m^2)	3,169 (1.6%)	2,781 (1.6%)	388 (1.8%)	1,352 (1.3%)	1,817 (2.1%)
Normal (18·5-24·9 kg/m^2)	42,421 (21.9%)	38,053 (22.1%)	4,368 (20.0%)	20,928 (19.8%)	21,493 (24.3%)
Overweight (25·0-29·9 kg/m^2)	58,499 (30.2%)	53,005 (30.8%)	5,494 (25.1%)	33,858 (32.1%)	24,641 (27.9%)
Obese (30·0-39·9 kg/m^2)	46,425 (23.9%)	41,334 (24.0%)	5,091 (23.3%)	31,208 (29.6%)	15,217 (17.2%)
Severely obese (≥40.0 kg/m^2)	6,731 (3.5%)	5,665 (3.3%)	1,066 (4.9%)	4,871 (4.6%)	1,860 (2.1%)
Unknown	36,655 (18.9%)	31,212 (18.1%)	5,443 (24.9%)	13,322 (12.6%)	23,333 (26.4%)
Smoking status					
Current	83,692 (43.2%)	73,156 (42.5%)	10,536 (48.2%)	49,860 (47.2%)	33,832 (38.3%)
Previous	66,618 (34.4%)	61,509 (35.8%)	5,109 (23.4%)	38,041 (36.0%)	28,577 (32.3%)
Never	31,521 (16.3%)	27,449 (16.0%)	4,072 (18.6%)	13,735 (13.0%)	17,786 (20.1%)
Unknown	12,069 (6.2%)	9,936 (5.8%)	2,133 (9.8%)	3,903 (3.7%)	8,166 (9.2%)
Diabetes	34,257 (17.7%)	33,261 (19.3%)	996 (4.6%)	24,569 (23.3%)	9,688 (11.0%)
Cholesterol:HDL					

	All	QRISK2		Hypertension	
		Raised risk	Low risk	Raised risk	Low risk
Mean (SD)	3.7 (1.2)	3.7 (1.2)	3.7 (1.2)	3.6 (1.2)	3.9 (1.3)
Unknown	59,330 (30.6%)	50,257 (29.2%)	9,073 (41.5%)	23,032 (21.8%)	36,298 (41.1%)
Associated hospital stay					
Yes	136,426 (70.4%)	121,036 (70.3%)	15,390 (70.4%)	74,318 (70.4%)	62,108 (70.3%)
Median (IQR) stay	4.0 (2.0-10.0)	4.0 (2.0-10.0)	3.0 (1.0-7.0)	4.0 (2.0-11.0)	4.0 (2.0-9.0)
Died ≤30 days after event	13,193 (6.8%)	12,338 (7.2%)	855 (3.9%)	7,604 (7.2%)	5,589 (6.3%)
Died in study period	19,868 (10.2%)	18,641 (10.8%)	1,227 (5.6%)	11,487 (10.9%)	8,381 (9.5%)
Loss to follow-up	9,201 (4.7%)	8,756 (5.1%)	445 (2.0%)	5,683 (5.4%)	3,518 (4.0%)

Table 7.3 Incidence ratios for events in risk periods following influenza vaccination by cardiovascular risk and age

	All		QRISK2[a]				Hypertension			
			Raised risk		Low risk		Raised risk		Low risk	
Risk period	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)
All ages										
15-28 days	7,084	0.72 (0.70-0.74)	6,427	0.77 (0.75-0.79)	657	0.48 (0.44-0.52)	3,948	0.77 (0.75-0.80)	3,136	0.67 (0.64-0.69)
29-59 days	16,033	0.76 (0.74-0.77)	14,567	0.81 (0.79-0.82)	1,466	0.49 (0.46-0.52)	8,906	0.81 (0.79-0.83)	7,127	0.70 (0.68-0.72)
60-90 days	16,209	0.80 (0.78-0.81)	14,778	0.85 (0.84-0.87)	1,431	0.50 (0.47-0.53)	9,089	0.86 (0.84-0.88)	7,120	0.73 (0.71-0.75)
91-120 days	15,898	0.84 (0.82-0.85)	14,465	0.89 (0.88-0.91)	1,433	0.55 (0.52-0.58)	8,839	0.90 (0.87-0.92)	7,059	0.78 (0.76-0.80)
Baseline[b]	126,373	ref	111,291	ref	15,082	Ref	68,458	ref	57,915	ref
40-64 years[c]										
15-28 days	1,402	0.54 (0.51-0.57)	837	0.62 (0.58-0.67)	565	0.45 (0.41-0.49)	592	0.63 (0.58-0.69)	810	0.49 (0.45-0.53)
29-59 days	3,121	0.55 (0.53-0.58)	1,845	0.63 (0.60-0.67)	1,276	0.47 (0.44-0.50)	1,375	0.67 (0.63-0.72)	1,746	0.49 (0.46-0.51)
60-90 days	3,057	0.56 (0.54-0.59)	1,810	0.64 (0.61-0.68)	1,247	0.48 (0.45-0.51)	1,339	0.68 (0.64-0.72)	1,718	0.50 (0.48-0.53)
91-120 days	3,034	0.61 (0.58-0.63)	1,790	0.68 (0.65-0.72)	1,244	0.52 (0.49-0.56)	1,340	0.72 (0.68-0.77)	1,694	0.54 (0.51-0.57)
Baseline[b]	30,022	ref	16,628	ref	13,394	Ref	11,970	ref	18,052	ref
65-74 years[c]										
15-28 days	2,517	0.80 (0.77-0.84)	2,425	0.80 (0.77-0.84)	92	0.80 (0.64-1.00)	1,346	0.82 (0.78-0.87)	1,171	0.78 (0.73-0.83)
29-59 days	5,699	0.84 (0.81-0.86)	5,509	0.84 (0.81-0.87)	190	0.76 (0.64-0.90)	2,961	0.83 (0.80-0.87)	2,738	0.84 (0.80-0.88)
60-90 days	5,865	0.89 (0.86-0.92)	5,681	0.90 (0.87-0.92)	184	0.75 (0.63-0.89)	3,107	0.90 (0.87-0.94)	2,758	0.87 (0.84-0.91)
91-120 days	5,769	0.93 (0.90-0.96)	5,580	0.94 (0.91-0.97)	189	0.82 (0.69-0.96)	3,014	0.93 (0.89-0.97)	2,755	0.93 (0.89-0.97)
Baseline[b]	44,708	ref	43,020	ref	1,688	Ref	23,667	ref	21,041	ref
75-84 years[c]										
15-28 days	3,165	0.80 (0.77-0.83)					2,010	0.80 (0.77-0.84)	1,155	0.80 (0.75-0.85)
29-59 days	7,213	0.86 (0.83-0.88)					4,570	0.86 (0.83-0.89)	2,643	0.86 (0.82-0.90)
60-90 days	7,287	0.91 (0.88-0.93)					4,643	0.91 (0.88-0.95)	2,644	0.90 (0.86-0.94)
91-120 days	7,095	0.96 (0.93-0.98)					4,485	0.95 (0.92-0.99)	2,610	0.96 (0.92-1.00)
Baseline[b]	51,643	ref					32,821	ref	18,822	ref

[a]QRISK2 score results are not included for those aged 75-84 years as all individuals were high risk.

[b]Baseline events which occurred before the vaccine risk periods were 25,730 in all ages, 11,605 in 40-64 year-olds, 7,264 in 65-74 year-olds and 6,861 in 75-84 year-olds and baseline events which occurred after the vaccine risk periods were 100,643 in all ages, 18,417 in 40-64year-olds, 37,444 in 65-74 year-olds and 44,782 in 75-84 year-olds.

[c]P-values for age interaction were <0.0001 in models with all age groups i.e. with all individuals and by cardiovascular risk stratification, other than QRISK2 score <10% for which it was 0.0004.

7.3. Additional methods

A priori, in addition to the stratifications presented in the paper, I planned and conducted two further stratifications to investigate the effect of ARI in the 28 days before acute cardiovascular event (binary ARI or no ARI) and vaccine match for the given season strains (binary matched or not matched). I used the ARI definition from my first PhD study (defined in Chapter 4 and results presented in Chapter 6), which included any diagnosis in CPRD Aurum and HES APC of; pneumonia, acute bronchitis, influenza, ILI, or other acute infections suggestive of lower respiratory tract involvement. Determining vaccine match is difficult, particularly as the influenza vaccine contains four strains (two A and two B) and various strains also circulate in a given season.[280] I defined vaccine match to the circulating influenza strains based on the UK Health Security Agency (formerly Public Health England) influenza vaccine effectiveness statistics.[113] All years included in the analysis were a match other than 2009/10 and 2014/15.

A further *a priori* sensitivity analysis not shown in the paper was the reclassification of season from warmer and cooler into three-month blocks of September-November, December-February, March-May, June-August.

To further understand the difference between the main and sensitivity analysis study design results obtained in the paper, I repeated the main study design analysis with the baseline split into pre- and post-vaccine, with the study population limited to those with baseline time before and after vaccination.

7.4. Additional results

There was a larger reduction in the IR among individual who have an ARI within the 28 days before acute cardiovascular event compared to those without an infection (days 15-28: 0.58 [0.53-0.63] vs 0.74 [0.72-

0.76], respectively) (**Table *7.4***). When stratified by age, the aforementioned pattern was limited to older age groups (≥65 years), with the inverse found in those aged <65 years.

Those who experienced their first acute cardiovascular event in a year when the influenza vaccine was not matched to the circulating strains, had a greater reduction in IR compared to those who had their event in a year which was a match, although the number of events was small (**Table *7.5***). Among individuals aged ≥65 years there was, however, no difference in the relative incidence.

When season was redefined into three-month blocks, the relative incidence of first acute cardiovascular event remained reduced for the 120-day risk period (**Table *7.6***). However, the pattern of a large reduction in the 15-28 days post-vaccination followed by a tapering towards unity as the time since vaccination increased not present.

When baseline was split into pre- and -post-vaccination time, there was an overall reduction in the incidence of first acute cardiovascular events during the pre-vaccine baseline compared to post-vaccine baseline (**Table *7.7***). However, when split by age group there was an increase in the incidence ratio during pre-vaccine baseline (compared to post-vaccine baseline) in those aged 40-64 years.

Table 7.4 Incidence ratios for first acute cardiovascular event in risk periods following influenza vaccination by cardiovascular risk, age group and prior acute respiratory infection

Risk period (in days)	All		QRISK2				Hypertension			
			Raised risk		Low risk		Raised risk		Low risk	
	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)
All ages		p <0.0001		p <0.0001		p <0.0001		p <0.0001		p <0.0001
No ARI										
15-28	6,400	0.74 (0.72-0.76)	5,800	0.79 (0.77-0.81)	600	0.47 (0.44-0.52)	3,551	0.80 (0.77-0.83)	2,849	0.68 (0.65-0.71)
29-59	14,352	0.76 (0.75-0.78)	13,017	0.82 (0.80-0.84)	1,335	0.48 (0.46-0.51)	7,948	0.82 (0.80-0.85)	6,404	0.70 (0.68-0.72)
60-90	14,177	0.78 (0.76-0.79)	12,907	0.84 (0.82-0.86)	1,270	0.48 (0.45-0.51)	7,887	0.84 (0.82-0.87)	6,290	0.71 (0.69-0.73)
91-120	13,863	0.81 (0.80-0.83)	12,589	0.87 (0.85-0.89)	1,274	0.52 (0.49-0.55)	7,643	0.87 (0.85-0.89)	6,220	0.75 (0.73-0.78)
Baseline	114,860	ref	100,768	ref	14,092	ref	61,750	ref	53,110	ref
ARI										
15-28	684	0.58 (0.53-0.63)	627	0.59 (0.54-0.64)	57	0.54 (0.41-0.72)	397	0.59 (0.53-0.66)	287	0.57 (0.50-0.64)
29-59	1,681	0.69 (0.65-0.73)	1,550	0.70 (0.66-0.75)	131	0.60 (0.49-0.73)	958	0.69 (0.64-0.75)	723	0.69 (0.63-0.76)
60-90	2,032	0.93 (0.88-0.99)	1,871	0.95 (0.90-1.00)	161	0.82 (0.69-0.99)	1,202	0.97 (0.90-1.04)	830	0.89 (0.82-0.97)
91-120	2,035	1.07 (1.02-1.13)	1,876	1.09 (1.03-1.15)	159	0.92 (0.77-1.10)	1,196	1.11 (1.03-1.19)	839	1.02 (0.94-1.11)
Baseline	11,513	ref	10,523	ref	990	ref	6,708	ref	4,805	ref
40-64 years		p <0.0001		p <0.0001		p <0.0001		p <0.0001		p <0.0001
No ARI										
15-28	1,267	0.53 (0.50-0.56)	756	0.61 (0.57-0.66)	511	0.44 (0.40-0.48)	534	0.62 (0.57-0.68)	733	0.48 (0.45-0.52)
29-59	2,832	0.54 (0.52-0.57)	1,676	0.62 (0.59-0.66)	1,156	0.46 (0.43-0.49)	1,239	0.66 (0.62-0.70)	1,593	0.48 (0.45-0.51)
60-90	2,724	0.54 (0.52-0.57)	1,618	0.62 (0.59-0.66)	1,106	0.46 (0.43-0.49)	1,190	0.65 (0.61-0.69)	1,534	0.48 (0.46-0.51)
91-120	2,736	0.59 (0.56-0.61)	1,638	0.67 (0.64-0.71)	1,098	0.50 (0.47-0.53)	1,217	0.71 (0.66-0.75)	1,519	0.52 (0.49-0.55)
Baseline	27,993	ref	15,483	ref	12,510	ref	11,126	ref	16,867	ref
ARI										
15-28	135	0.66 (0.55-0.79)	81	0.74 (0.58-0.94)	54	0.57 (0.43-0.76)	58	0.76 (0.57-1.01)	77	0.60 (0.47-0.77)
29-59	289	0.68 (0.59-0.78)	169	0.73 (0.61-0.88)	120	0.61 (0.50-0.75)	136	0.85 (0.69-1.05)	153	0.57 (0.48-0.69)
60-90	333	0.86 (0.75-0.97)	192	0.90 (0.76-1.07)	141	0.80 (0.66-0.98)	149	1.02 (0.83-1.24)	184	0.76 (0.64-0.90)
91-120	298	0.87 (0.76-0.99)	152	0.80 (0.67-0.96)	146	0.94 (0.78-1.14)	123	0.95 (0.77-1.16)	175	0.82 (0.69-0.97)
Baseline	2,029	ref	1,145	ref	884	ref	844	ref	1,185	Ref

Risk period (in days)	All		QRISK2				Hypertension			
			Raised risk		Low risk		Raised risk		Low risk	
	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)
65-74 years		p <0.0001		p <0.0001		p = 0.01		p <0.0001		p <0.0001
No ARI										
15-28	2,322	0.83 (0.79-0.87)	2,233	0.83 (0.79-0.87)	89	0.84 (0.67-1.06)	1,233	0.85 (0.80-0.90)	1,089	0.81 (0.76-0.86)
29-59	5,190	0.85 (0.82-0.88)	5,011	0.85 (0.83-0.88)	179	0.78 (0.66-0.93)	2,706	0.85 (0.82-0.89)	2,484	0.85 (0.81-0.89)
60-90	5,203	0.87 (0.85-0.90)	5,039	0.88 (0.85-0.91)	164	0.73 (0.61-0.87)	2,739	0.89 (0.85-0.93)	2,464	0.86 (0.82-0.90)
91-120	5,132	0.91 (0.88-0.94)	4,956	0.92 (0.89-0.95)	176	0.82 (0.69-0.98)	2,655	0.91 (0.87-0.95)	2,477	0.92 (0.87-0.96)
Baseline	41,055	ref	39,473	ref	1,582	ref	21,625	ref	19,430	ref
ARI										
15-28	195	0.56 (0.48-0.65)	192	0.57 (0.49-0.66)	3	0.30 (0.09-0.99)	113	0.61 (0.50-0.75)	82	0.50 (0.40-0.64)
29-59	509	0.71 (0.64-0.79)	498	0.71 (0.64-0.79)	11	0.51 (0.25-1.03)	255	0.66 (0.57-0.77)	254	0.76 (0.65-0.89)
60-90	662	1.02 (0.93-1.13)	642	1.02 (0.93-1.13)	20	1.01 (0.58-1.78)	368	1.06 (0.93-1.21)	294	0.98 (0.85-1.14)
91-120	637	1.12 (1.02-1.23)	624	1.13 (1.03-1.25)	13	0.74 (0.39-1.40)	359	1.18 (1.03-1.34)	278	1.06 (0.92-1.22)
Baseline	3,653	ref	3,547	ref	106	ref	2,042	ref	1,611	ref
75-84 years		p <0.0001						p <0.0001		p <0.0001
No ARI										
15-28	2,811	0.84 (0.81-0.88)					1,784	0.85 (0.81-0.89)	1,027	0.83 (0.77-0.88)
29-59	6,330	0.88 (0.85-0.91)					4,003	0.88 (0.85-0.92)	2,327	0.87 (0.83-0.92)
60-90	6,250	0.90 (0.88-0.93)					3,958	0.91 (0.88-0.95)	2,292	0.89 (0.85-0.94)
91-120	5,995	0.93 (0.90-0.96)					3,771	0.93 (0.89-0.97)	2,224	0.93 (0.88-0.98)
Baseline	45,812	ref					28,999	ref	16,813	ref
ARI										
15-28	354	0.57 (0.51-0.64)					226	0.55 (0.48-0.64)	128	0.61 (0.50-0.74)
29-59	883	0.70 (0.64-0.76)					567	0.68 (0.61-0.75)	316	0.74 (0.64-0.85)
60-90	1,037	0.92 (0.85-0.99)					685	0.92 (0.83-1.01)	352	0.93 (0.81-1.06)
91-120	1,100	1.13 (1.05-1.22)					714	1.11 (1.01-1.22)	386	1.17 (1.03-1.33)
Baseline	5,831	Ref					3,822	ref	2,009	ref

Footnotes: acute respiratory infection in the 28 days before first acute cardiovascular event. P-values in table are for acute respiratory infection interaction

Table 7.5 Incidence ratios for first acute cardiovascular event in risk periods following influenza vaccination by cardiovascular risk, age group vaccine match

Risk period	All		QRISK2				Hypertension			
			Raised risk		Low risk		Raised risk		Low risk	
	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)
All ages		p = 0.0681		p = 0.0494		p = 0.8458		p = 0.2105		p = 0.4535
Not a vaccine match										
15-28 days	1,240	0.72 (0.67-0.76)	1,123	0.76 (0.71-0.81)	117	0.49 (0.40-0.59)	678	0.75 (0.69-0.82)	562	0.68 (0.62-0.74)
29-59 days	2,895	0.78 (0.74-0.81)	2,644	0.83 (0.79-0.87)	251	0.49 (0.42-0.56)	1,591	0.82 (0.77-0.87)	1,304	0.73 (0.68-0.78)
60-90 days	2,836	0.79 (0.76-0.83)	2,586	0.85 (0.81-0.89)	250	0.50 (0.44-0.58)	1,608	0.86 (0.81-0.92)	1,228	0.72 (0.67-0.76)
91-120 days	2,772	0.83 (0.80-0.87)	2,509	0.88 (0.84-0.93)	263	0.57 (0.50-0.65)	1,529	0.88 (0.83-0.94)	1,243	0.78 (0.73-0.83)
Baseline	22,090	Ref	19,459	ref	2,631	ref	11,902	ref	10,188	ref
Vaccine match										
15-28 days	5,844	0.72 (0.70-0.74)	5,304	0.77 (0.75-0.79)	540	0.48 (0.44-0.52)	3,270	0.78 (0.75-0.81)	2,574	0.66 (0.64-0.69)
29-59 days	13,138	0.75 (0.74-0.77)	11,923	0.80 (0.79-0.82)	1,215	0.49 (0.46-0.52)	7,315	0.81 (0.78-0.83)	5,823	0.70 (0.68-0.72)
60-90 days	13,373	0.80 (0.78-0.81)	12,192	0.85 (0.83-0.87)	1,181	0.50 (0.47-0.53)	7,481	0.86 (0.83-0.88)	5,892	0.73 (0.71-0.76)
91-120 days	13,126	0.84 (0.82-0.86)	11,956	0.90 (0.88-0.92)	1,170	0.54 (0.51-0.58)	7,310	0.90 (0.87-0.92)	5,816	0.78 (0.76-0.80)
Baseline	104,283	Ref	91,832	ref	12,451	ref	56,556	ref	47,727	ref
40-64 years		p = 0.1232		p = 0.0176		p = 0.5179		p = 0.5932		p = 0.2352
Not a vaccine match										
15-28 days	225	0.48 (0.42-0.55)	124	0.50 (0.41-0.60)	101	0.47 (0.38-0.57)	103	0.59 (0.48-0.72)	122	0.42 (0.35-0.51)
29-59 days	536	0.53 (0.48-0.58)	323	0.60 (0.53-0.68)	213	0.45 (0.39-0.53)	241	0.63 (0.54-0.73)	295	0.47 (0.42-0.54)
60-90 days	513	0.53 (0.48-0.58)	298	0.57 (0.50-0.65)	215	0.48 (0.42-0.56)	235	0.63 (0.55-0.73)	278	0.46 (0.41-0.53)
91-120 days	544	0.61 (0.55-0.66)	303	0.62 (0.55-0.71)	241	0.59 (0.51-0.67)	241	0.70 (0.61-0.81)	303	0.55 (0.49-0.62)
Baseline	5,285	Ref	2,964	ref	2,321	ref	2,116	ref	3,169	ref
Vaccine match										
15-28 days	1,177	0.55 (0.52-0.59)	713	0.65 (0.60-0.70)	464	0.45 (0.41-0.49)	489	0.64 (0.58-0.70)	688	0.50 (0.47-0.55)
29-59 days	2,585	0.56 (0.53-0.58)	1,522	0.64 (0.60-0.68)	1,063	0.47 (0.44-0.50)	1,134	0.68 (0.64-0.73)	1,451	0.49 (0.46-0.52)
60-90 days	2,544	0.57 (0.55-0.60)	1,512	0.66 (0.62-0.70)	1,032	0.48 (0.45-0.51)	1,104	0.68 (0.64-0.73)	1,440	0.51 (0.48-0.54)
91-120 days	2,490	0.61 (0.58-0.63)	1,487	0.69 (0.66-0.73)	1,003	0.51 (0.48-0.55)	1,099	0.73 (0.68-0.78)	1,391	0.54 (0.51-0.57)
Baseline	24,737	ref	13,664	ref	11,073	ref	9,854	ref	14,883	ref

Risk period	All		QRISK2				Hypertension			
			Raised risk		Low risk		Raised risk		Low risk	
	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)
65-74 years		p = 0.2824		p = 0.3570		p = 0.2685		p = 0.4482		p = 0.7168
Not a vaccine match										
15-28 days	439	0.80 (0.72-0.89)	423	0.80 (0.72-0.89)	16	0.74 (0.43-1.26)	229	0.80 (0.70-0.93)	210	0.80 (0.69-0.93)
29-59 days	1,000	0.85 (0.79-0.91)	962	0.85 (0.79-0.92)	38	0.81 (0.55-1.19)	503	0.82 (0.74-0.91)	497	0.87 (0.79-0.97)
60-90 days	1,006	0.88 (0.81-0.94)	971	0.88 (0.82-0.95)	35	0.75 (0.51-1.12)	533	0.90 (0.81-1.00)	473	0.85 (0.77-0.95)
91-120 days	994	0.92 (0.86-0.99)	972	0.94 (0.87-1.01)	22	0.50 (0.31-0.79)	515	0.92 (0.83-1.02)	479	0.92 (0.83-1.02)
Baseline	7,819	ref	7,509	ref	310	ref	4,118	ref	3,701	ref
Vaccine match										
15-28 days	2,078	0.80 (0.76-0.84)	2,002	0.80 (0.76-0.84)	76	0.82 (0.64-1.04)	1,117	0.83 (0.77-0.88)	961	0.77 (0.72-0.83)
29-59 days	4,699	0.84 (0.81-0.87)	4,547	0.84 (0.81-0.87)	152	0.75 (0.62-0.90)	2,458	0.84 (0.80-0.88)	2,241	0.83 (0.79-0.88)
60-90 days	4,859	0.89 (0.86-0.92)	4,710	0.90 (0.87-0.93)	149	0.75 (0.62-0.91)	2,574	0.91 (0.86-0.95)	2,285	0.88 (0.84-0.92)
91-120 days	4,775	0.93 (0.90-0.97)	4,608	0.94 (0.90-0.97)	167	0.89 (0.74-1.07)	2,499	0.94 (0.89-0.98)	2,276	0.93 (0.89-0.98)
Baseline	36,889	ref	35,511	ref	1,378	ref	19,549	ref	17,340	ref
75-84 years		p = 0.3406						p = 0.6105		p = 0.4453
Not a vaccine match										
15-28 days	576	0.83 (0.76-0.91)					346	0.80 (0.71-0.90)	230	0.89 (0.77-1.03)
29-59 days	1,359	0.92 (0.86-0.99)					847	0.92 (0.84-1.00)	512	0.93 (0.83-1.04)
60-90 days	1,317	0.94 (0.88-1.01)					840	0.96 (0.88-1.04)	477	0.92 (0.82-1.02)
91-120 days	1,234	0.96 (0.90-1.03)					773	0.96 (0.88-1.04)	461	0.96 (0.86-1.07)
Baseline	8,986	ref					5,668	ref	3,318	ref
Vaccine match										
15-28 days	2,589	0.80 (0.76-0.83)					1,664	0.81 (0.76-0.85)	925	0.78 (0.73-0.84)
29-59 days	5,854	0.84 (0.82-0.87)					3,723	0.84 (0.81-0.88)	2,131	0.84 (0.80-0.88)
60-90 days	5,970	0.90 (0.87-0.93)					3,803	0.90 (0.87-0.94)	2,167	0.89 (0.85-0.94)
91-120 days	5,861	0.96 (0.93-0.99)					3,712	0.95 (0.92-0.99)	2,149	0.96 (0.91-1.01)
Baseline	42,657	ref					27,153	ref	15,504	ref

Footnotes: vaccine match or not is determined by compatibility of influenza strains included in the vaccine with those which are circulating. P-values in table are for vaccine match interaction

Table 7.6 Incidence ratios for first acute cardiovascular event in risk periods following influenza vaccination by cardiovascular risk with redefined season

Risk period	All		QRISK2				Hypertension			
			Raised risk		Low risk		Raised risk		Low risk	
	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)
15-28 days	7,058	0.78 (0.76-0.80)	6,403	0.87 (0.85-0.89)	655	0.44 (0.41-0.48)	3,936	0.89 (0.86-0.92)	3,122	0.69 (0.66-0.71)
29-59 days	15,820	0.80 (0.78-0.81)	14,367	0.88 (0.86-0.90)	1,453	0.46 (0.44-0.49)	8,788	0.89 (0.87-0.91)	7,032	0.71 (0.69-0.73)
60-90 days	16,159	0.81 (0.79-0.83)	14,731	0.88 (0.86-0.90)	1,428	0.49 (0.46-0.52)	9,069	0.90 (0.87-0.92)	7,090	0.73 (0.71-0.75)
91-120 days	15,825	0.85 (0.83-0.87)	14,402	0.92 (0.90-0.94)	1,423	0.56 (0.53-0.60)	8,792	0.92 (0.90-0.95)	7,033	0.79 (0.76-0.81)
Baseline	126,702	ref	111,598	ref	15,104	ref	68,649	ref	58,053	ref

Table 7.7 Incidence ratios for first acute cardiovascular event in risk periods following influenza vaccination by cardiovascular risk and age group with baseline split into pre- and post-vaccination time

	All		QRISK2				Hypertension			
			Raised risk		Low risk		Raised risk		Low risk	
Risk period	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)
All ages										
Pre-vaccine baseline	25,730	0.96 (0.95-0.98)	19,439	0.83 (0.81-0.84)	6.291	2.12 (2.05-2.20)	11,535	0.81 (0.79-0.82)	14,195	1.15 (1.13-1.18)
15-28 days	7,054	0.69 (0.67-0.70)	6,332	0.69 (0.67-0.71)	722	0.75 (0.69-0.81)	3,889	0.69 (0.67-0.72)	3,165	0.69 (0.66-0.72)
29-59 days	15,143	0.69 (0.68-0.70)	13,671	0.70 (0.69-0.71)	1,472	0.70 (0.66-0.75)	8,364	0.70 (0.68-0.72)	6,779	0.69 (0.67-0.71)
60-90 days	15,207	0.72 (0.71-0.73)	13,771	0.73 (0.72-0.75)	1,436	0.71 (0.67-0.75)	8,478	0.74 (0.72-0.76)	6,729	0.71 (0.69-0.73)
91-120 days	14,953	0.76 (0.75-0.78)	13,528	0.78 (0.76-0.79)	1,425	0.75 (0.71-0.80)	8,259	0.77 (0.75-0.79)	6,694	0.76 (0.74-0.78)
Post-vaccine baseline	102,750	ref	93,803	ref	8,947	ref	58,101	ref	44,649	ref
40-64 years										
Pre-vaccine baseline	11,603	1.91 (1.86-1.96)	5,588	1.62 (1.56-1.68)	6,015	2.32 (2.23-2.41)	3,675	1.46 (1.40-1.53)	7,928	2.25 (2.18-2.33)
15-28 days	1,512	0.78 (0.74-0.82)	877	0.79 (0.74-0.85)	635	0.77 (0.71-0.84)	627	0.76 (0.70-0.83)	885	0.80 (0.74-0.86)
29-59 days	3,106	0.74 (0.71-0.77)	1,818	0.76 (0.72-0.80)	1,288	0.72 (0.68-0.77)	1,351	0.76 (0.71-0.81)	1,755	0.73 (0.69-0.77)
60-90 days	3,039	0.74 (0.71-0.77)	1,776	0.76 (0.72-0.80)	1,263	0.73 (0.68-0.78)	1,333	0.77 (0.72-0.82)	1,706	0.73 (0.69-0.78)
91-120 days	3,018	0.79 (0.76-0.82)	1,767	0.81 (0.77-0.85)	1,251	0.77 (0.73-0.83)	1,322	0.81 (0.76-0.86)	1,696	0.78 (0.74-0.83)
Post-vaccine baseline	18,681	ref	11,149	Ref	7,532	ref	8,395	ref	10,286	ref
65-74 years										
Pre-vaccine baseline	7,268	0.82 (0.80-0.84)	6,992	0.82 (0.80-0.84)	276	0.86 (0.75-0.98)	3,619	0.80 (0.77-0.83)	3,649	0.85 (0.82-0.88)
15-28 days	2,490	0.73 (0.70-0.76)	2,403	0.73 (0.70-0.76)	87	0.71 (0.57-0.90)	1,330	0.75 (0.70-0.79)	1,160	0.71 (0.67-0.76)
29-59 days	5,415	0.74 (0.71-0.76)	5,231	0.74 (0.72-0.76)	184	0.69 (0.58-	2,811	0.73 (0.70-0.77)	2,604	0.74 (0.71-0.78)
60-90 days	5,513	0.78 (0.75-0.80)	5,340	0.78 (0.76-0.81)	173	0.67 (0.56-0.80)	2,904	0.79 (0.75-0.82)	2,609	0.77 (0.73-0.80)
91-120 days	5,467	0.82 (0.79-0.85)	5,293	0.82 (0.80-0.85)	174	0.71 (0.60-0.84)	2,858	0.82 (0.79-0.86)	2,609	0.82 (0.78-0.86)
Post-vaccine baseline	37,945	ref	36,530	Ref	1,415	ref	20,279	ref	17,666	ref

	All		QRISK2				Hypertension			
			Raised risk		Low risk		Raised risk		Low risk	
Risk period	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)
75-84 years										
Pre-vaccine baseline	6,859	0.60 (0.59-0.62)					4,241	0.59 (0.57-0.61)	2,618	0.62 (0.60-0.65)
15-28 days	3,052	0.65 (0.63-0.68)					1,932	0.65 (0.62-0.69)	1,120	0.66 (0.62-0.70)
29-59 days	6,622	0.68 (0.66-0.70)					4,202	0.68 (0.65-0.70)	2,420	0.68 (0.65-0.71)
60-90 days	6,655	0.72 (0.70-0.74)					4,241	0.72 (0.69-0.75)	2,414	0.71 (0.68-0.75)
91-120 days	6,468	0.76 (0.74-0.78)					4,079	0.75 (0.73-0.78)	2,389	0.77 (0.73-0.80)
Post-vaccine baseline	46,124	ref					29,427	ref	16,697	ref

7.5. Chapter summary

- To minimize confounding by indication and healthy vaccinee bias, I used a SCCS study design with CPRD Aurum and HES APC linked data to investigate the effect of influenza vaccine on first acute cardiovascular events, with stratification by cardiovascular risk.
- I included individuals who had an influenza vaccine and first acute cardiovascular event in the same 365-day period starting from the 1 September.
- There were 193,900 individuals aged 40-84 years who had a first acute cardiovascular event in the same year as an influenza vaccine, 105,539 (54.4%) had diagnosed hypertension and 172,050 (88.7%) had a QRISK2 score of ≥10%.
- I observed a significant reduction in the season-adjusted incidence of first acute cardiovascular event for a pre-defined 120-day risk period after influenza vaccination. There was a tapering in the risk reduction over time; with a 28% reduction 15-28 days post-vaccination and 16% 91-120 days post-vaccination.
- When stratified by cardiovascular risk, there was a larger reduction in the incidence of first acute cardiovascular event after influenza vaccine for individuals with low cardiovascular risk. This result was likely due to a lower all year round risk of a cardiovascular risk in those at low cardiovascular risk.
- Similarly, results were markedly different between age groups with a much larger reduction in the relative incidence of first acute cardiovascular event in risk periods for individuals aged 40-64 years compared with ≥65 years.
- A high number of events in individuals aged 40-64 years occurred prior to vaccination. In a sensitivity analysis study design, with follow up started at the date of vaccination, there was a smaller reduction in the relative incidence of a first acute cardiovascular event during early risk periods after vaccination compared to the main study design. There was no difference in the

relative incidence during risk periods compared to baseline among individuals aged 40-64 years while a difference remained for those aged ≥65 years.

- While adults aged <65 years are already targeted for influenza vaccine when an underlying chronic health condition is present, uptake is low. The results of my study emphasise the importance of influenza vaccine in the prevention of cardiovascular events.

Chapter 8 Investigating the effect of cardiovascular risk on the severe COVID-19 outcomes

8.1. Chapter overview

In this chapter, I present the research conducted to address thesis objective 4. For the study I used CPRD Aurum data linked to HES APC, death registrations, CHESS and SGSS data to investigate the association between cardiovascular risk and severe COVID-19 outcomes, including acute cardiovascular events. The study concept and protocol (Chapter 11 Appendix 6) were developed by my primary supervisor, Dr Warren-Gash. I conducted all analyses and made detailed methodological decisions during the analysis process. Dr Warren-Gash led the drafting of the manuscript, and I contributed by writing the methods and results sections.

The chapter begins with the cohort study research paper which we plan to submit to the Lancet. Supplementary materials which accompany the research paper are presented in Chapter 11 Appendix 7. The chapter then concludes with further methods and results which were not included in the research paper.

8.2. Drafted manuscript

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	230019	**Title**	Ms
First Name(s)	Jennifer		
Surname/Family Name	Davidson		
Thesis Title	Acute respiratory infections, cardiovascular complications, and prevention among people with raised cardiovascular risk		
Primary Supervisor	Dr Charlotte Warren-Gash		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	The Lancet
Please list the paper's authors in the intended authorship order:	Charlotte Warren-Gash, Jennifer A Davidson, Helen Strongman, Emily Herrett, Harriet Forbes, Prof Liam Smeeth, Prof Judith Breuer, Prof Amitava Banerjee
Stage of publication	**Not yet submitted**

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I am the second author of this research paper. Charlotte Warren-Gash conceptualised and lead the design of the study, I conducted the analysis and drafted the methods and results sections of the research paper Charlotte Warren-Gash drafted all remaining elements of the research paper.

SECTION E

Student Signature	[redacted]
Date	25/07/2022

Supervisor Signature	[redacted]
Date	25-07-2022

Severe COVID-19 outcomes among individuals with differing cardiovascular risk profiles in England in 2020: a population-based cohort study

Authors: Charlotte Warren-Gash, PhD[1], Jennifer A Davidson, MSc[1], Helen Strongman, PhD[1], Emily Herrett, PhD[1], Harriet Forbes, PhD[3], Prof Liam Smeeth, FRCGP[1], Judith Breuer, FMedSci[4], Amitava Banerjee, DPhil[2]

Affiliations:

[1] Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK

[2] Institute of Health Informatics, University College London, London, UK

[3] Population Health Sciences, Bristol Medical School, University of Bristol, UK

[4] Institute of Infection and Immunity, University College London, London, UK

Corresponding author: Dr Charlotte Warren-Gash

Email: charlotte.warren-gash1@lshtm.ac.uk Tel: 020 7927 2796

Word count: 3,110

Abstract word count: 296

ABSTRACT

Background While cardiovascular disease is a risk factor for severe COVID-19, the association between underlying cardiovascular risk profile and severe COVID-19 among people without diagnosed cardiovascular disease is unclear.

Methods We carried out historical, population-based incidence and cohort studies among adults aged 40-84 years in England using linked data from the Clinical Practice Research Datalink. Individuals were categorized into: existing cardiovascular disease (CVD), raised cardiovascular risk (QRISK3 score ≥10% or hypertension) and low risk at 12 March 2020. We described incidence and severe outcomes of COVID-19 (deaths, intensive care unit [ICU] admissions, hospitalisations, acute cardiovascular events) for each group. Among those with a COVID-19 record between 12 March 2020 and 31 December 2020, we re-classified cardiovascular risk at infection and assessed the risk of severe outcomes using multivariable Cox proportional hazards regression.

Findings Among 6,059,055 individuals, 741,913 (12.2%) had established CVD, 1,929,627 (31.8%) had a QRISK3 score ≥10% and 3,387,515 (55.9%) had a QRISK3 score <10%. The incidence of COVID-19 death was 7.4 per 1,000 among patients with established CVD, 2.2 per 1,000 for those at raised cardiovascular risk and 0.2 per 1,000 for low cardiovascular risk, with similar gradients for other outcomes. Among those with COVID-19 (N=146,760), there was a strong association between higher QRISK3 score and death (adjusted hazard ratio [aHR] 8.77 (7.62-10.10). Risks of other outcomes were also higher among those at raised cardiovascular risk: aHR 3.66 (3.18-4.21) for ICU admissions, 3.38 (3.22-3.56) for hospitalisations, 5.43 (4.44-6.64) for acute cardiovascular events. When raised cardiovascular risk was classified by hypertension status, only acute cardiovascular events remained associated: aHR 1.49 (1.20-1.85).

Interpretation Individuals without pre-existing CVD but with a raised QRISK3 score were more likely to experience severe COVID-19 outcomes and should be prioritised for prevention and treatment. Addressing cardiovascular risk factors could improve COVID-19 outcomes.

INTRODUCTION

The COVID-19 pandemic due to SARS-CoV-2 infection has resulted in more than 6.3 million deaths to date worldwide[1]. Severe COVID-19 outcomes including admissions to hospitals, intensive care units (ICUs), major complications and deaths are more frequent among older individuals, those with a range of underlying health conditions and in socioeconomically deprived and minority ethnic populations[2-4]. Cardiovascular disease (CVD) and risk factors such as hypertension and diabetes have been associated with severe outcomes in many studies, which have been largely small and mainly hospital-based[5-7], which include selected patient groups and do not capture the full spectrum of either cardiovascular risk or severe COVID-19 outcomes.

An early elevation in acute cardiovascular events such as myocardial infarction (MI) and stroke following COVID-19 has also been shown in population-based self-controlled case series studies from Scotland[8], Sweden[9] and Denmark[10]. Similar transient elevations in the risks of MI and stroke occur following other laboratory-confirmed respiratory infections including influenza and *Streptococcus pneumoniae*[11]. Although evidence from before the COVID-19 pandemic showed that such complications are more frequent among individuals at raised cardiovascular risk[12], this has not been comprehensively investigated for COVID-19. Individuals at raised cardiovascular risk but without existing CVD, were therefore not considered 'clinically vulnerable' in England during the COVID-19 pandemic[13].

Large, population-based studies with detailed cardiovascular risk assessments are needed to assess the burden of severe outcomes of COVID-19, including cardiovascular complications, among individuals with differing levels of underlying cardiovascular risk to guide accurate stratified prevention and management. Here we aimed to quantify the incidence and severe outcomes of SARS-CoV-2 infections and to assess the risk of severe COVID-19 outcomes following infection by underlying cardiovascular risk profile among adults in England.

METHODS

Data sources

We used the Clinical Practice Research Datalink (CPRD) Aurum[14] January 2022 dataset, with linked data from Hospital Episode Statistics Admitted Patient Care (HES APC), Office of National Statistics (ONS) deaths, Second Generation Surveillance System (SGSS) SARS-CoV-2, and COVID-19 Hospitalisations in England Surveillance System (CHESS)[15].

The CPRD Independent Scientific Advisory Committee (application 20_000135) and the London School of Hygiene and Tropical Medicine (LSHTM) Ethics Committee (application 22717) approved the study. CPRD provided relevant HES APC, ONS, SGSS and CHESS data for the study population. All code lists are published on LSHTM Data Compass[16].

Study population and follow-up

Individuals aged 40-84 years with at least one year of follow up post-registration in their primary care practice who are eligible for linkage to HES APC were eligible for inclusion in our incidence study. Follow-up of individuals started at the latest of; age 40 years, 12 months post-registration, or 12 March 2020, and ended at the earliest of; date of death or outcome of interest, administrative censor (date of leaving the practice or date of last data collection from the practice), or 31 December 2020 (**Figure *8.1***). We started follow-up from 12 March 2020 when daily reporting to CHESS was initiated[17].

Our cohort study included individuals with COVID-19. In our main analysis we defined this as laboratory-confirmed SARS-CoV-2, identified using SGSS and CHESS data. All individuals in either of the two datasets were considered to have SARS-CoV-2, with the date of infection taken as the earliest specimen data. In a secondary analysis, we defined COVID as clinically reported COVID-19 (CPRD or HES APC [any diagnostic position] recorded) without laboratory-confirmed SARS-CoV-2. Follow-up in the cohort study started at this date and ended at the earliest of the dates set out in our incidence study (**Figure *8.1***). We stratified the study population further in time based on the UK COVID-19 waves (one;

12 March to 16 August and two; 17 August to 31 December), during which different testing practices were in operation.

Outcome, exposure and covariates

Our primary outcome of interest was death attributable to COVID-19. We defined COVID-19 attributable deaths as those coded as U07.1 or U07.2 in ONS data. In a sensitivity analysis we explored broadening our primary outcome of death attributable to COVID-19 to all-cause death which occurred within 28 days of the individual's diagnosis (based on test result among those with laboratory-confirmed SARS-CoV-2 or consultation date for those with clinically reported COVID-19). Our secondary outcomes were hospitalisation due to COVID-19 (defined by COVID-19 in the primary diagnosis field of any episode recorded in HES APC or presence in CHESS dataset), ICU admission due to COVID-19 (defined by ICU admission recorded in CHESS), need for respiratory support due to COVID-19 (defined by mechanical ventilation recorded in CHESS), or major adverse cardiovascular event (acute cardiovascular events [composite of acute coronary syndrome which included myocardial infarction and unstable angina, ischaemic stroke, acute left ventricular failure, or major ventricular arrhythmia recorded in CPRD or HES APC]).

Our exposure of interest was cardiovascular risk. First, we identified individuals with established CVD (CPRD Aurum or HES APC recorded) diagnosed before baseline. Among individuals without CVD, we then used QRISK3 score and hypertension status to separately identify individuals with and without raised cardiovascular risk. Individuals with established CVD were included in our incidence study but excluded from our cohort study.

QRISK3 is a validated UK ten-year cardiovascular risk prediction score based on a combination of known risk factors[18]. Briefly, the score is calculated using a range of risk factors; age, sex, ethnicity, socioeconomic status, family history of coronary heart disease in a first degree relative aged <60 years, comorbid health conditions (diabetes, treated hypertension, rheumatoid arthritis, systemic lupus

erythematosus [SLE], atrial fibrillation, chronic kidney disease stages 3-5, migraine, severe mental illness, HIV, erectile dysfunction) body mass index (BMI), systolic blood pressure reading and it's variability, total cholesterol to high density lipoprotein cholesterol ratio, smoking status, and corticosteroid treatment. We calculated individual score's using our Stata program[19] with selected codes and measures recorded in patient records prior to baseline for each of the conditions included in the calculator and with the weighted algorithm made available by the QRISK3 developers[20]. We classified hypertension status using coded CPRD diagnoses within the five years before baseline or the most recent to baseline blood pressure (BP) reading with systolic BP of ≥140mmHg or diastolic BP of ≥90mmHg. We classified individuals as having raised cardiovascular risk (hypertension or QRISK3 ≥10%) or low cardiovascular risk (no hypertension or QRISK3 <10%) at baseline. In a secondary analysis, we further stratified QRISK3 scores into <10%, 10-20%, or ≥20%.

In analysis of hypertension, we included covariates of baseline age, sex (male and female), ethnicity (White, south Asian, Black, and mixed or other), socioeconomic status (individual-level Townsend score grouped into quintiles, ranging from least deprived [quintile 1] to most deprived [quintile 5]), BMI (underweight [<18.5 kg/m2], normal [18.5-24.9 kg/m2], overweight [25.0-29.9 kg/m2], obese [30.0-39.9 kg/m2], and severely obese [≥40 kg/m2]), total cholesterol to high-density lipoprotein ratio, alcohol consumption (heavy drinking [defined as either a recorded intake of >42 units per week or a diagnostic code suggestive of alcohol addiction or excessive alcohol consumption] or no known heavy drinking), smoking status (current smoker, never smoker, or former smoker), family history of heart disease, and comorbid conditions or treatments which are potential risk factors for severe COVID-19 outcomes. These comorbidities and treatments were those included in the QRISK3 algorithm (diabetes, atrial fibrillation, migraine, chronic kidney disease stage 3-5, corticosteroid use, severe mental illness or antipsychotic use, and erectile dysfunction) as well as chronic respiratory diseases, asthma, non-haematological and haematological cancer, chronic liver disease, dementia, chronic neurological disease, learning or intellectual disability. Immunosuppression was also included as a covariate grouping elements included in

QRISK3 (rheumatoid arthritis, SLE and HIV) with other elements of any prior solid organ transplant or permanent cellular immunodeficiency, or aplastic anaemia, bone marrow or stem cell transplant recorded within the 24 months before index, or biologic or other immunosuppressant therapy (excluding corticosteroids) or temporary immunodeficiency recorded within the 12 months before index. In analysis by QRISK3, we included covariates which were not part of determining the score (alcohol consumption, treatment with antiplatelets or anticoagulants, diagnosis of chronic liver disease, chronic lung disease, asthma, dementia, chronic neurological disease, learning or intellectual disability, or malignancy, and treatment or diagnosis of a immunosuppressive condition [solid organ transplant or permanent cellular immunodeficiency, or aplastic anaemia, bone marrow or stem cell transplant, biologic or other immunosuppressant therapy, or temporary immunodeficiency]).

Statistical analysis

We described the baseline characteristics, for both the incidence and cohort study populations, using numbers and percentage for categorical variables and mean with standard deviation or median with interquartile range for continuous variables.

For our incidence study population, stratified by cardiovascular risk, we calculated incidence of the primary outcome of COVID-19 death and secondary outcomes of ICU admission, respiratory support, hospitalisation, and acute cardiovascular events, among the whole population, regardless of COVID-19 status. We then calculated the incidence of SARS-CoV-2 infection and clinically reported COVID-19, as well as our primary and secondary outcomes following laboratory-confirmed SARS-CoV-2 or clinically reported COVID-19. We further stratified results by time according to COVID-19 wave.

Among our cohort study population (those with COVID-19), we used Cox proportional hazards regression finely adjusted for calendar time to generate hazard ratios for the association between cardiovascular risk and each outcome, initially adjusting models in hypertension analysis for age and sex, and then in a full model adjusted for all potential confounders. A complete case-analysis approach was used in multivariable analyses. We did not conduct multiple imputation because data in CPRD are

unlikely to be missing at random. We examined non-proportionality using Schoenfeld's residuals. We conducted all analyses in Stata, version 16.

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation or writing of the report.

RESULTS

Description of the study population

The overall study population included 6,059,055 individuals aged 40-84 years of age (**Figure *8.2***), 12.2% (741,913) had established CVD and among those without established CVD, 31.9% (1,929,627) had a QRISK3 score ≥10% and 55.9% (3,387,515) had a QRISK3 score <10%, and 31.1% (1,881,654) had hypertension and 56.7% (3,435,488) had no hypertension. The baseline characteristics of the study population are described in supplementary table 1.

Incidence of COVID-19 and severe outcomes

Among all individuals the incidence of COVID-19 death was 1.7 (95% CI 1.7-1.8) per 1,000 with the highest incidence among those with established CVD (7.4 [7.2-7.7] per 1,000), followed by those with raised cardiovascular risk (QRISK3 ≥10%; 2.2 [2.1-2.2] and hypertension; 1.4 [1.3-1.5] per 1,000), and was lowest among those at low cardiovascular risk (QRISK3 <10%; 0.2 [0.2-0.2] and no hypertension; 0.7 [0.6-0.7] per 1,000). The same gradient by cardiovascular risk level was observed for hospitalisations and acute cardiovascular events, and for the outcomes among individuals with laboratory-confirmed SARS-CoV-2 and clinically reported COVID-19 (**table 8.1**). Results by COVID-19 wave, showed a higher incidence for the outcomes of interest in wave 1 compared to wave 2 (supplementary table 2 and 3). Employing the sensitivity analysis definition of death (all cause within 28 days of diagnosis), resulted

in similar incidence as COVID-19 death among individuals with laboratory-confirmed SARS-CoV-2 and a higher incidence than COVID-19 death among those with clinically reported COVID-19 (**table 8.1**).

Description of cohort study population

After excluding those with established CVD, 146,760 people had laboratory-confirmed SARS-CoV-2 and 56,197 with clinically-reported COVID-19 during our study period (**Figure *8.2***). Among the main study population, with laboratory-confirmed SARS-CoV-2, when cardiovascular risk was classified by QRISK3 score, 26.8% (39,295) had raised risk (a score ≥10%) and 73.2% (107,465) had low risk (a score <10%). When hypertension was used to classify cardiovascular risk, 34.0% (49,955) had raised risk (hypertension) and 66.0% (96,805) had low risk (no hypertension). Individuals with laboratory-confirmed SARS-CoV-2 and raised cardiovascular risk (QRISK3 ≥10% or hypertension) were older and a higher proportion were men. A full summary of the laboratory-confirmed SARS-CoV-2 study population baseline characteristics are shown in **table 8.2** and those for the clinically reported COVID-19 study population are shown in supplementary table 4. Overall, when compared to individuals with laboratory-confirmed SARS-CoV-2, a higher proportion of those with clinically reported COVID-19 were older, women, less affluent, and lived in London.

Risk of severe outcomes after COVID-19

In unadjusted analysis, raised cardiovascular risk resulted in a substantial increase in COVID-19 death, particularly when risk was measured by QRISK3 score (HR 16.33 [14.61-18.24]) (**Figure *8.3*** and supplementary table 5). After adjustment for confounders, the association between QRISK3 score and COVID-19 death attenuated but remained substantial (aHR 8.77 [7.62-10.10]). In comparison, there was no association between hypertension and COVID-19 death. Significant associations were also found for QRISK3 score ≥10% and the outcomes of ICU admission (aHR 3.66 [3.18-4.21]), respiratory support (aHR 3.73 [3.10-4.49]), hospitalisation (aHR 3.38 [3.22-3.56]), and acute cardiovascular events (aHR 5.43 [4.44-6.64]). There was no association between hypertension and ICU admission (aHR 1.15 [0.98-

1.36]), respiratory support (aHR 1.20 [0.97-1.48]), or hospitalisation (aHR 1.05 [0.99-1.11]) but there was an association between hypertension and acute cardiovascular events (aHR 1.49 [1.20-1.85]). Results between wave 1 and wave 2 were broadly similar but for all outcomes (supplementary table 6). Similar results were also obtained for clinically reported COVID-19 (supplementary table 7). Further stratification of the QRISK3 score showed a substantially greater risk of COVID-19 death in individuals with a QRISK3 score of ≥20% (aHR 15.15 [13.05-17.59]) than 10-<20% (aHR 5.32 [4.54-6.23]) when both were compared to those with a score <10% (supplementary table 8). A similar, though less extreme gradient was also observed for the other outcomes.

DISCUSSION

In this large, population-based cohort study of adults aged 40 and 84 years in England in 2020, we found a striking gradient in the occurrence of severe COVID-19 outcomes by underlying cardiovascular risk profile among people without pre-existing cardiovascular disease. The risks of death, ICU admission, hospital admission and acute cardiovascular events were all greater among individuals at raised cardiovascular risk measured by QRISK3 score, compared to those at low risk, despite no increase in recorded infections in this group. When cardiovascular risk was measured by hypertension alone, differences were only evident for acute cardiovascular events outcomes. Analysis by pandemic waves revealed similar patterns, although the incidence of severe outcomes was greatest during the first wave. Our study used linked electronic health record data from primary and secondary, including intensive, care, mortality records and national laboratory surveillance to capture detailed clinical and laboratory data on SARS-CoV-2 infections and outcomes. It is, to our knowledge, the first UK population-based study to assess COVID-19 outcomes using a comprehensive, combined measure of cardiovascular risk, QRISK3, rather than focusing on individual vascular risk factors. Findings from this large, representative cohort should be generalizable to adults over 40 years in England. Our dataset spanned the first and major part of the second wave of the COVID-19 pandemic in England, allowing comparisons of outcomes between

waves. Limiting follow up to the end of December 2020 prevented contamination from the emergence of coronavirus variants or widespread roll out of the COVID-19 vaccination programme in England. Nevertheless, differences in the availability of laboratory PCR testing are likely to have led to differences in the reported incidence of infection between waves: a laboratory-confirmed definition of SARS-CoV-2 lacked sensitivity to identify cases occurring during wave one before mass testing became widely available. It is also possible that some outcomes such as hospitalisation or acute cardiovascular events may have led to in-hospital testing, strengthening the observed association between vascular risk status and severe outcomes in the laboratory-confirmed cohort during the first wave. However, stratifying by pandemic wave to explore the effect of expanded testing and advances in clinical management of COVID-19 in later time periods revealed similar results. When we compared results for individuals with confirmed SARS-CoV-2 infection to those with clinically diagnosed COVID-19, we saw similar patterns. In addition, our descriptive analysis of COVID-19 outcomes alone regardless of recorded infection status supported findings from the cohort analysis.

The magnitude of association between cardiovascular risk status and severe outcomes varied by the method used to classify cardiovascular risk. In general, classification by QRISK3 produced more exaggerated differences between high and low cardiovascular risk groups than classification by hypertension alone. This is perhaps unsurprising as QRISK3 is a more comprehensive measure of cardiovascular risk, which includes additional comorbidities and socio-demographic components of risk. Although misclassification of cardiovascular risk status could have occurred due to the documented reductions in GP visits and healthcare-seeking for non-COVID conditions during the pandemic[21], under-recognition of individuals at raised cardiovascular risk would have led to bias towards the null. In addition, our sensitivity analysis in which QRISK3 status was graded more finely into three strata (<10%, 10-19%, 20%+), confirmed a gradient of increasing risk of severe outcomes with increasing vascular risk level, which suggests that the main results are robust to any minor exposure misclassification.

Residual confounding may also have been present in our study. While we adjusted for a broad range of sociodemographic, lifestyle and clinical confounding factors, some variables are either not measured

(such as genetic risk profiles) or are sub-optimally recorded (such as smoking status and BMI) in EHRs. Nevertheless, population-based self-controlled case series analyses, which use within-person comparisons to control implicitly for the effects of fixed confounding factors[22], support an association between COVID-19 and thrombotic outcomes, and show comparable results to cohort studies[9, 23], suggesting that confounding is unlikely to explain our cohort results.

Our findings extend those from previous smaller studies of individual cardiovascular risk factors and COVID-19 outcomes[24], supporting a strong association between raised cardiovascular risk profile and severe COVID-19 outcomes. Our results are consistent with work on cardiovascular complications of acute respiratory infections in the pre-COVID era, where a gradient in the risk of complications was shown to be aligned with underlying cardiovascular risk status[12]. While a previous Mendelian randomisation study, which by design avoids reverse causation and most confounding, failed to show an association between some genetically-predicted cardiovascular risk factors (blood pressure BMI, type 2 diabetes and coronary artery disease) and COVID-19 hospitalisation[25], estimates had wide confidence intervals and captured full profiles of neither cardiovascular risk nor severe COVID-19.

Potential mechanisms underlying severe outcomes in COVID-19 include pro-inflammatory, pro-thrombotic and vasoconstrictive effects of SARS-CoV-2-mediated imbalances in ACE-2/RAS signalling[26]. It has been suggested that individuals with conditions leading to raised cardiovascular risk are likely to have altered cytokine profiles leading to chronic systemic inflammation, which may have a synergistic effect on disease severity in COVID-19[27]. Further studies are needed into the mechanisms underlying acute severe outcomes of COVID-19 and future emerging respiratory infections among individuals at raised cardiovascular risk. Understanding the natural history of longer-term outcomes including post-COVID-19 syndrome in this group, along with mechanisms underlying long-term health changes should also be a priority for future research[28].

In conclusion, we showed that individuals at raised cardiovascular risk in England were more likely to die or to experience severe outcomes after COVID-19 than those at low cardiovascular risk, despite not

initially being identified as a vulnerable group. Those at raised cardiovascular risk should be considered a priority for targeted prevention and treatment strategies for COVID-19. Addressing cardiovascular risk factors could improve outcomes after COVID-19 and other respiratory infections.

References

1. World Health Organization. Weekly epidemiological update on COVID-19 - 22 June 2022 2022. Available from: https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---22-june-2022 Last accessed 27/06/2022
2. Izcovich A, Ragusa MA, Tortosa F, Lavena Marzio MA, Agnoletti C, Bengolea A, et al. Prognostic factors for severity and mortality in patients infected with COVID-19: A systematic review. PLoS One. 2020;15(11):e0241955
3. Mathur R, Rentsch CT, Morton CE, Hulme WJ, Schultze A, MacKenna B, et al. Ethnic differences in SARS-CoV-2 infection and COVID-19-related hospitalisation, intensive care unit admission, and death in 17 million adults in England: an observational cohort study using the OpenSAFELY platform. Lancet. 2021 May 8;397(10286):1711-24
4. Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, et al. Factors associated with COVID-19-related death using OpenSAFELY. Nature. 2020 Aug;584(7821):430-6
5. Pepera G, Tribali MS, Batalik L, Petrov I, Papathanasiou J. Epidemiology, risk factors and prognosis of cardiovascular disease in the Coronavirus Disease 2019 (COVID-19) pandemic era: a systematic review. Rev Cardiovasc Med. 2022 Jan 17;23(1):28
6. Bae S, Kim SR, Kim MN, Shim WJ, Park SM. Impact of cardiovascular disease and risk factors on fatal outcomes in patients with COVID-19 according to age: a systematic review and meta-analysis. Heart. 2021 Mar;107(5):373-80
7. Silverio A, Di Maio M, Citro R, Esposito L, Iuliano G, Bellino M, et al. Cardiovascular risk factors and mortality in hospitalized patients with COVID-19: systematic review and meta-analysis of 45 studies and 18,300 patients. BMC Cardiovasc Disord. 2021 Jan 7;21(1):23
8. Ho FK, Man KKC, Toshner M, Church C, Celis-Morales C, Wong ICK, et al. Thromboembolic Risk in Hospitalized and Nonhospitalized COVID-19 Patients: A Self-Controlled Case Series Analysis of a Nationwide Cohort. Mayo Clin Proc. 2021 Oct;96(10):2587-97
9. Katsoularis I, Fonseca-Rodríguez O, Farrington P, Lindmark K, Fors Connolly AM. Risk of acute myocardial infarction and ischaemic stroke following COVID-19 in Sweden: a self-controlled case series and matched cohort study. Lancet. 2021 Aug 14;398(10300):599-607
10. Modin D, Claggett B, Sindet-Pedersen C, Lassen MCH, Skaarup KG, Jensen JUS, et al. Acute COVID-19 and the Incidence of Ischemic Stroke and Acute Myocardial Infarction. Circulation. 2020 Nov 24;142(21):2080-2
11. Ohland J, Warren-Gash C, Blackburn R, Mølbak K, Valentiner-Branth P, Nielsen J, et al. Acute myocardial infarctions and stroke triggered by laboratory-confirmed respiratory infections in Denmark, 2010 to 2016. Euro Surveill. 2020 Apr;25(17)
15. Clinical Practice Research Datalink. CPRD Aurum CHESS January 2022 (Version 2022.01.001) 2022. Available 12. Davidson JA, Banerjee A, Smeeth L, McDonald HI, Grint D, Herrett E, et al. Risk of acute respiratory infection and acute cardiovascular events following acute respiratory infection among adults with increased cardiovascular risk in England between 2008 and 2018: a retrospective, population-based cohort study. Lancet Digit Health. 2021 Dec;3(12):e773-e83
13. NHS Digital. Risk criteria (shielded patient list) 2020. Available from: https://digital.nhs.uk/coronavirus/shielded-patient-list/risk-criteria Last accessed 27/06/2022

14. Wolf A, Dedman D, Campbell J, Booth H, Lunn D, Chapman J, et al. Data resource profile: Clinical Practice Research Datalink (CPRD) Aurum. Int J Epidemiol. 2019 Dec 1;48(6):1740-g
from: https://cprd.com/cprd-aurum-chess-january-2022 Last accessed 27/06/2022
16. Davidson JA, McDonald HI, Strongman H, Cadogan SL, Banerjee A, Smeeth L, et al. Codelists for: "Effect of cardiovascular risk profile on severe outcomes of COVID-19 in England in 2020: a population-based cohort study" 2022. Available from: https://datacompass.lshtm.ac.uk/id/eprint/2762/ Last accessed 28/06/2022
17. Public Health England. COVID-19 Hospitalisation in England Surveillance System (CHESS) – daily reporting 2020. Available from: https://www.england.nhs.uk/coronavirus/wp-content/uploads/sites/52/2020/03/phe-letter-to-trusts-re-daily-covid-19-hospital-surveillance-11-march-2020.pdf Last accesse 28/06/2022
18. Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. Bmj. 2017 May 23;357:j2099
19. Davidson JA, Strongman H, Herrett E, Gadd S. qrisk_cprd_aurum: QRISK Aurum bundle version 2.0 2022. Available from: https://zenodo.org/record/5822521#.YrrC6XbMKUl Last accessed 28/06/2022
20. ClinRisk Ltd. QRISK3 algorithm. 2017. Available from: https://qrisk.org/three/src.php Last accessed 28/06/2022
21. Mansfield KE, Mathur R, Tazare J, Henderson AD, Mulick AR, Carreira H, et al. Indirect acute effects of the COVID-19 pandemic on physical and mental health in the UK: a population-based study. Lancet Digit Health. 2021 Apr;3(4):e217-e30
22. Fonseca-Rodríguez O, Fors Connolly AM, Katsoularis I, Lindmark K, Farrington P. Avoiding bias in self-controlled case series studies of coronavirus disease 2019. Stat Med. 2021 Nov 30;40(27):6197-208
23. Katsoularis I, Fonseca-Rodríguez O, Farrington P, Jerndal H, Lundevaller EH, Sund M, et al. Risks of deep vein thrombosis, pulmonary embolism, and bleeding after covid-19: nationwide self-controlled cases series and matched cohort study. Bmj. 2022 Apr 6;377:e069590
24. Krittanawong C, Virk HUH, Narasimhan B, Wang Z, Narasimhan H, Zhang HJ, et al. Coronavirus disease 2019 (COVID-19) and cardiovascular risk: A meta-analysis. Prog Cardiovasc Dis. 2020 Jul-Aug;63(4):527-8
25. Cecelja M, Lewis CM, Shah AM, Chowienczyk P. Cardiovascular health and risk of hospitalization with COVID-19: A Mendelian Randomization study. JRSM Cardiovasc Dis. 2021 Jan-Dec;10:20480040211059374
26. Augustine R, S A, Nayeem A, Salam SA, Augustine P, Dan P, et al. Increased complications of COVID-19 in people with cardiovascular disease: Role of the renin-angiotensin-aldosterone system (RAAS) dysregulation. Chem Biol Interact. 2022 Jan 5;351:109738
27. Srivastava A, Rockman-Greenberg C, Sareen N, Lionetti V, Dhingra S. An insight into the mechanisms of COVID-19, SARS-CoV2 infection severity concerning β-cell survival and cardiovascular conditions in diabetic patients. Mol Cell Biochem. 2022 Jun;477(6):1681-95
28. NICE, SIGN, RCGP. COVID-19 rapid guideline: managing the long-term effects of COVID-19 [NICE Guideline NG188]. 2021. Available from:
https://www.nice.org.uk/guidance/ng188/resources/covid19-rapid-guideline-managing-the-longterm-effects-of-covid19-pdf-51035515742 Last accessed 28/06/2022

Table 8.1 Incidence of laboratory-confirmed SARS-CoV-2 and clinically reported COVID-19 and outcomes of interest

	All		Established CVD		QRISK3 score				Hypertension			
					Raised risk		Low risk		Raised risk		Low risk	
	N	Rate (95% CI) per 1,000	N	Rate (95% CI) per 1,000	N	Rate per 1,000	N	Rate (95% CI) per 1,000	N	Rate per 1,000	N	Rate (95% CI) per 1,000
All individuals	6,059,055		741,913		1,929,627		3,387,515		1,881,654		3,435,488	
COVID-19 death*	7,866	1.7 (1.7-1.8)	4,164	7.4 (7.2-7.7)	3,203	2.2 (2.1-2.2)	499	0.2 (0.2-0.2)	2,014	1.4 (1.3-1.5)	1,688	0.7 (0.6-0.7)
Hospitalisation[$]	28,013	6.1 (6.0-6.2)	10,880	19.4 (19.0-19.8)	10,794	7.3 (7.2-7.4)	6,339	2.5 (2.4-2.6)	8,481	5.9 (5.8-6.0)	8,652	3.4 (3.3-3.4)
Major adverse cardiovascular event	71,035	15.5 (15.4-15.6)	49,318	88.0 (87.2-88.8)	16,604	11.2 (11.0-11.4)	5,113	2.0 (2.0-2.1)	12,633	8.8 (8.6-8.9)	9,084	3.5 (3.4-3.6)
Laboratory-confirmed SARS-CoV-2	174,129	38.0 (37.8-38.2)	24,779	44.2 (43.7-44.8)	41,416	28.0 (27.7-28.2)	107,934	42.5 (42.2-42.7)	50,854	35.3 (35.0-35.6)	98,496	38.2 (37.9-38.4)
COVID-19 death*	6,475	48.9 (47.7-50.1)	3,493	199.9 (193.4-206.6)	2,597	82.9 (79.8-86.2)	385	4.6 (4.2-5.1)	1,664	42.5 (40.5-44.6)	1,318	17.4 (16.5-18.4)
All cause death within 28 days of diagnosis	6,649	50.2 (49.0-51.4)	3,592	205.6 (198.9-212.4)	2,661	85.0 (81.8-88.3)	396	4.7 (4.3-5.2)	1,716	43.9 (41.8-46.0)	1,341	17.7 (16.8-18.7)
ICU admission[†]	2,024	15.3 (14.6-16.0)	499	28.6 (26.2-31.2)	930	29.7 (27.8-31.7)	595	7.1 (6.6-7.7)	811	20.7 (19.4-22.2)	714	9.4 (8.8-10.1)
Respiratory support[‡]	1,084	8.2 (7.7-8.7)	230	13.2 (11.6-15.0)	526	16.8 (15.4-18.3)	328	3.9 (3.5-4.4)	475	12.1 (11.1-13.3)	379	5.0 (4.5-5.5)
Hospitalisation[$]	17,893	135.2 (133.2-137.2)	6,555	375.1 (366.1-384.3)	7,047	225.0 (219.8-230.3)	4,291	51.3 (49.8-52.9)	5,628	143.9 (140.1-147.7)	5,710	75.3 (73.4-77.3)
Major adverse cardiovascular event	2,251	17.0 (16.3-17.7)	1,422	81.4 (77.3-85.7)	616	19.7 (18.2-21.3)	213	2.5 (2.2-2.9)	486	12.4 (11.4-13.6)	343	4.5 (4.1-5.0)
Clinically reported COVID-19	70,700	15.4 (15.3-15.5)	13,668	24.4 (24.0-24.8)	20,240	13.7 (13.5-13.9)	36,792	14.5 (14.3-14.6)	21,374	14.8 (14.6-15.0)	35,658	13.8 (13.7-14.0)
COVID-19 death*	723	13.4 (12.5-14.4)	365	36.2 (32.6-40.1)	318	20.7 (18.5-23.1)	40	1.4 (1.0-1.9)	178	10.8 (9.3-12.5)	180	6.6 (5.7-7.6)
All cause death within 28 days of diagnosis	1,599	29.6 (28.2-31.1)	823	81.5 (76.2-87.3)	657	42.7 (39.5-46.1)	119	4.2 (3.5-5.0)	375	22.8 (20.6-25.2)	401	14.6 (13.3-16.1)
Hospitalisation[$]	3,692	68.4 (66.2-70.6)	1,336	132.4 (125.5-139.7)	1,385	89.9 (85.3-94.8)	971	34.1 (32.0-36.3)	1,124	68.2 (64.3-72.3)	1,232	44.9 (42.5-47.5)
Major adverse cardiovascular event	2,002	37.1 (35.5-38.7)	1,282	127.0 (120.3-134.2)	569	37.0 (34.0-40.1)	151	5.3 (4.5-6.2)	418	25.4 (23.0-27.9)	302	11.0 (9.8-12.3)

*Ascertained from ONS death certificate data in which the COVID related ICD-10 codes U07.1 or U07.2 were present in the record

[$]Ascertained from presence in CHESS dataset or HES APC record coded with primary diagnosis of U07.1 or U07.2

[†]Ascertained from CHESS records coded with ICU/HDU admission, only available for those with laboratory confirmed SARS-CoV-2

[‡]Ascertained from CHESS record coded with use of respiratory support via invasive mechanical ventilation, only available for those with laboratory confirmed SARS-CoV-2

Table 8.2 Baseline characteristics of the laboratory-confirmed SARS-CoV-2 study population by cardiovascular risk

	All	QRISK3 score		Hypertension	
		Raised risk	Low risk	Raised risk	Low risk
	N=146,760	N=39,295	N=107,465	N=49,955	N=96,805
Age (years), Mean (SD)*	54.0 (10.1)	65.3 (9.5)	49.9 (6.6)	57.7 (10.6)	52.2 (9.3)
Age group (years)*					
40-54	84,928 (57.9%)	5,150 (13.1%)	79,778 (74.2%)	21,382 (42.8%)	63,546 (65.6%)
55-64	39,757 (27.1%)	13,967 (35.5%)	25,790 (24.0%)	16,435 (32.9%)	23,322 (24.1%)
65-74	14,782 (10.1%)	12,885 (32.8%)	1,897 (1.8%)	7,819 (15.7%)	6,963 (7.2%)
75-84	7,293 (5.0%)	7,293 (18.6%)	0 (0.0%)	4,319 (8.6%)	2,974 (3.1%)
Sex*					
Women	80,805 (55.1%)	14,316 (36.4%)	66,489 (61.9%)	24,608 (49.3%)	56,197 (58.1%)
Men	65,955 (44.9%)	24,979 (63.6%)	40,976 (38.1%)	25,347 (50.7%)	40,608 (41.9%)
Ethnicity*					
White or not stated	104,902 (71.5%)	28,657 (72.9%)	76,245 (70.9%)	36,280 (72.6%)	68,622 (70.9%)
South Asian	12,340 (8.4%)	4,588 (11.7%)	7,752 (7.2%)	3,893 (7.8%)	8,447 (8.7%)
Black	3,408 (2.3%)	519 (1.3%)	2,889 (2.7%)	1,378 (2.8%)	2,030 (2.1%)
Mixed/Other	11,793 (8.0%)	2,624 (6.7%)	9,169 (8.5%)	4,114 (8.2%)	7,679 (7.9%)
Unknown	14,317 (9.8%)	2,907 (7.4%)	11,410 (10.6%)	4,290 (8.6%)	10,027 (10.4%)
Townsend quintile*					
1 (most affluent)	28,068 (19.1%)	6,224 (15.8%)	21,844 (20.3%)	9,175 (18.4%)	18,893 (19.5%)
2	28,488 (19.4%)	6,968 (17.7%)	21,520 (20.0%)	9,619 (19.3%)	18,869 (19.5%)
3	28,259 (19.3%)	7,281 (18.5%)	20,978 (19.5%)	9,612 (19.2%)	18,647 (19.3%)
4	28,947 (19.7%)	8,099 (20.6%)	20,848 (19.4%)	10,027 (20.1%)	18,920 (19.5%)
5 (least affluent)	32,940 (22.4%)	10,712 (27.3%)	22,228 (20.7%)	11,505 (23.0%)	21,435 (22.1%)
Unknown	58 (0.0%)	11 (0.0%)	47 (0.0%)	17 (0.0%)	41 (0.0%)
Region of residence					
North East	6,207 (4.2%)	1,746 (4.4%)	4,461 (4.2%)	2,276 (4.6%)	3,931 (4.1%)
North West	34,059 (23.2%)	9,696 (24.7%)	24,363 (22.7%)	12,364 (24.8%)	21,695 (22.4%)
Yorkshire and the Humber	4,908 (3.3%)	1,328 (3.4%)	3,580 (3.3%)	1,718 (3.4%)	3,190 (3.3%)
East Midlands	2,508 (1.7%)	690 (1.8%)	1,818 (1.7%)	897 (1.8%)	1,611 (1.7%)
West Midlands	24,071 (16.4%)	6,916 (17.6%)	17,155 (16.0%)	8,972 (18.0%)	15,099 (15.6%)
East of England	5,347 (3.6%)	1,207 (3.1%)	4,140 (3.9%)	1,662 (3.3%)	3,685 (3.8%)

	All	QRISK3 score		Hypertension	
		Raised risk	Low risk	Raised risk	Low risk
South West	32,542 (22.2%)	8,765 (22.3%)	23,777 (22.1%)	10,067 (20.2%)	22,475 (23.2%)
South Central	26,156 (17.8%)	6,050 (15.4%)	20,106 (18.7%)	8,170 (16.4%)	17,986 (18.6%)
London	10,734 (7.3%)	2,812 (7.2%)	7,922 (7.4%)	3,737 (7.5%)	6,997 (7.2%)
Unknown	228 (0.2%)	85 (0.2%)	143 (0.1%)	92 (0.2%)	136 (0.1%)
BMI category*†					
Underweight (<18·5 kg/m^2)	865 (0.6%)	336 (0.9%)	529 (0.5%)	182 (0.4%)	683 (0.7%)
Normal (18·5-24·9 kg/m^2)	25,789 (17.6%)	5,980 (15.2%)	19,809 (18.4%)	5,727 (11.5%)	20,062 (20.7%)
Overweight (25·0-29·9 kg/m^2)	39,501 (26.9%)	12,415 (31.6%)	27,086 (25.2%)	13,618 (27.3%)	25,883 (26.7%)
Obese (30·0-39·9 kg/m^2)	34,394 (23.4%)	12,652 (32.2%)	21,742 (20.2%)	16,361 (32.8%)	18,033 (18.6%)
Severely obese (≥40.0 kg/m^2)	5,558 (3.8%)	1,985 (5.1%)	3,573 (3.3%)	3,153 (6.3%)	2,405 (2.5%)
Unknown	40,653 (27.7%)	5,927 (15.1%)	34,726 (32.3%)	10,914 (21.8%)	29,739 (30.7%)
Cholesterol:HDL, Mean (SD)*†	3.8 (1.2)	4.0 (1.3)	3.7 (1.1)	3.8 (1.2)	3.7 (1.2)
Systolic blood pressure, Mean (SD)*†‡	128.1 (14.6)	134.3 (14.4)	125.5 (13.8)	138.7 (13.5)	122.1 (11.3)
Smoking status*†					
Non-smoker	73,789 (50.3%)	18,643 (47.4%)	55,146 (51.3%)	26,210 (52.5%)	47,579 (49.1%)
Ex-smoker	34,028 (23.2%)	12,543 (31.9%)	21,485 (20.0%)	13,287 (26.6%)	20,741 (21.4%)
Current smoker	12,400 (8.4%)	4,630 (11.8%)	7,770 (7.2%)	3,872 (7.8%)	8,528 (8.8%)
Unknown	26,543 (18.1%)	3,479 (8.9%)	23,064 (21.5%)	6,586 (13.2%)	19,957 (20.6%)
Alcohol consumption†					
No heavy drinking	85,406 (58.2%)	26,746 (68.1%)	58,660 (54.6%)	32,332 (64.7%)	53,074 (54.8%)
Heavy drinking	12,319 (8.4%)	3,760 (9.6%)	8,559 (8.0%)	4,512 (9.0%)	7,807 (8.1%)
Unknown	49,035 (33.4%)	8,789 (22.4%)	40,246 (37.5%)	13,111 (26.2%)	35,924 (37.1%)
Family history of CHD*	13,116 (8.9%)	4,579 (11.7%)	8,537 (7.9%)	4,371 (8.7%)	8,745 (9.0%)
Consultation frequency in prior 12 months, Median (IQR)	3 (1-7)	6 (2-10)	3 (1-6)	4 (1-9)	3 (1-6)
Medication use$					
Regular corticosteroids*	1,545 (1.1%)	1,070 (2.7%)	475 (0.4%)	790 (1.6%)	755 (0.8%)
Antihypertensives*	35,232 (24.0%)	15,938 (40.6%)	19,294 (18.0%)	21,361 (42.8%)	13,871 (14.3%)
Statins	19,723 (13.4%)	13,931 (35.5%)	5,792 (5.4%)	11,230 (22.5%)	8,493 (8.8%)
Antiplatelets	7,415 (5.1%)	4,360 (11.1%)	3,055 (2.8%)	3,830 (7.7%)	3,585 (3.7%)
Anticoagulants	2,720 (1.9%)	1,761 (4.5%)	959 (0.9%)	1,359 (2.7%)	1,361 (1.4%)
Comorbid condition					

	All	QRISK3 score		Hypertension	
		Raised risk	Low risk	Raised risk	Low risk
Atrial fibrillation*	1,420 (1.0%)	1,303 (3.3%)	117 (0.1%)	745 (1.5%)	675 (0.7%)
Migraines*	5,396 (3.7%)	907 (2.3%)	4,489 (4.2%)	1,559 (3.1%)	3,837 (4.0%)
Diabetes*	12,238 (8.3%)	9,811 (25.0%)	2,427 (2.3%)	6,603 (13.2%)	5,635 (5.8%)
CKD stage 3-5*	9,294 (6.3%)	7,013 (17.9%)	2,281 (2.1%)	5,482 (11.0%)	3,812 (3.9%)
Chronic liver disease	1,563 (1.1%)	775 (2.0%)	788 (0.7%)	641 (1.3%)	922 (1.0%)
Chronic respiratory disease (not asthma)	4,880 (3.3%)	3,303 (8.4%)	1,577 (1.5%)	2,267 (4.5%)	2,613 (2.7%)
Asthma with recent OCS use[$]	7,558 (5.1%)	2,597 (6.6%)	4,961 (4.6%)	3,120 (6.2%)	4,438 (4.6%)
Asthma with no recent OCS use	14,861 (10.1%)	3,582 (9.1%)	11,279 (10.5%)	5,056 (10.1%)	9,805 (10.1%)
Severe mental illness / antipsychotic use*	1,700 (1.2%)	956 (2.4%)	744 (0.7%)	595 (1.2%)	1,105 (1.1%)
Dementia	2,407 (1.6%)	2,046 (5.2%)	361 (0.3%)	1,060 (2.1%)	1,347 (1.4%)
Chronic neurological disease	1,932 (1.3%)	1,034 (2.6%)	898 (0.8%)	746 (1.5%)	1,186 (1.2%)
Learning / intellectual disability	1,014 (0.7%)	361 (0.9%)	653 (0.6%)	292 (0.6%)	722 (0.7%)
Non-haematological cancer					
Diagnosed <1 year ago	3,839 (2.6%)	2,275 (5.8%)	1,564 (1.5%)	1,807 (3.6%)	2,032 (2.1%)
Diagnosed 1-4.9 years ago	4,554 (3.1%)	2,085 (5.3%)	2,469 (2.3%)	1,878 (3.8%)	2,676 (2.8%)
Diagnosed ≥5 years ago	8,436 (5.7%)	2,764 (7.0%)	5,672 (5.3%)	2,998 (6.0%)	5,438 (5.6%)
Haematological malignancy					
Diagnosed <1 year ago	558 (0.4%)	373 (0.9%)	185 (0.2%)	246 (0.5%)	312 (0.3%)
Diagnosed 1-4.9 years ago	273 (0.2%)	162 (0.4%)	111 (0.1%)	124 (0.2%)	149 (0.2%)
Diagnosed ≥5 years ago	278 (0.2%)	114 (0.3%)	164 (0.2%)	113 (0.2%)	165 (0.2%)
Rheumatoid arthritis*	1,276 (0.9%)	648 (1.6%)	628 (0.6%)	560 (1.1%)	716 (0.7%)
Systemic lupus erythematosus*	164 (0.1%)	57 (0.1%)	107 (0.1%)	46 (0.1%)	118 (0.1%)
HIV*	234 (0.2%)	56 (0.1%)	178 (0.2%)	100 (0.2%)	134 (0.1%)
Immunosuppression#	1,404 (1.0%)	643 (1.6%)	761 (0.7%)	597 (1.2%)	807 (0.8%)
Erectile dysfunction*	7,183 (10.9%)	5,293 (21.2%)	1,890 (4.6%)	3,556 (14.0%)	3,627 (8.9%)

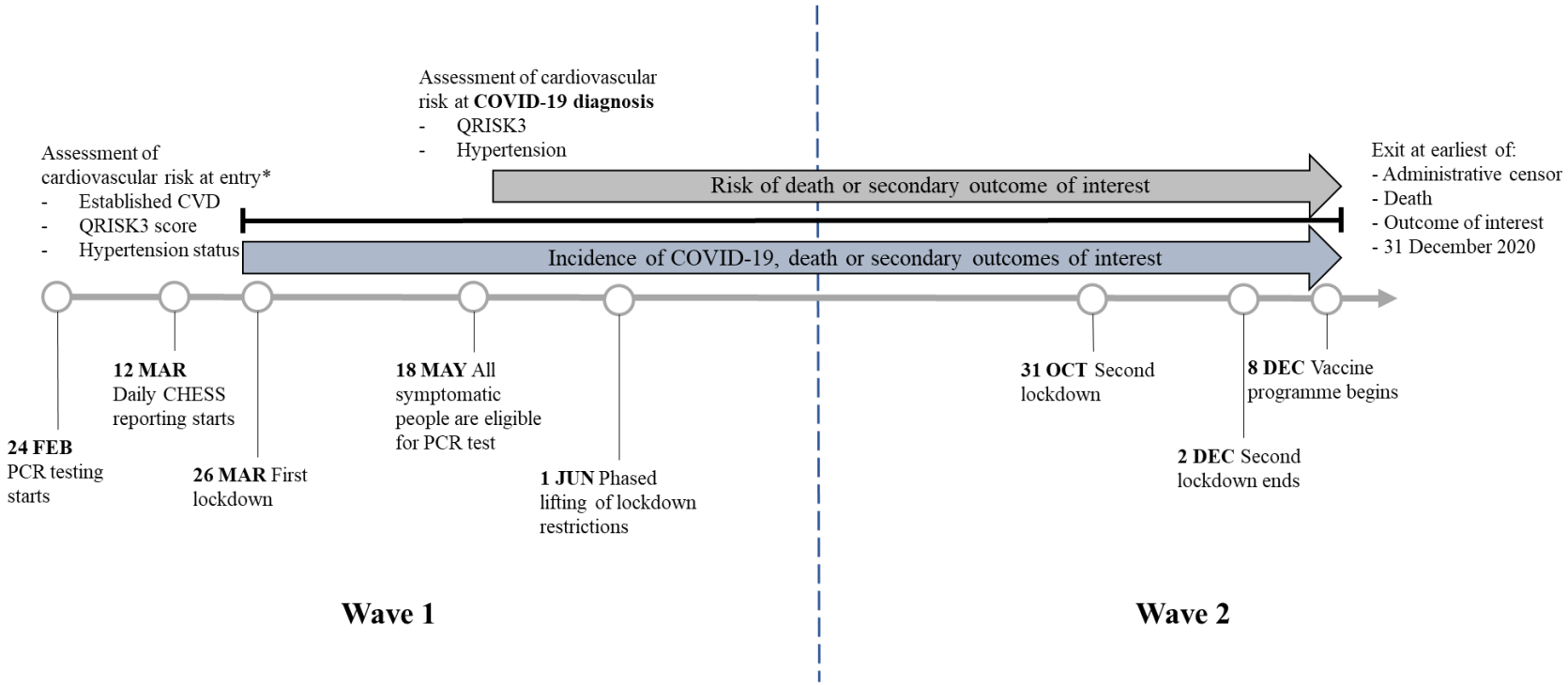
*In QRISK3 algorithm, but non-imputed version included here (for smoking status, cholesterol:HDL ratio, systolic BP and BMI)

†most recent measure before baseline

‡Used on hypertension definition

$ at least 1 prescription in the 12 months before baseline. Other than corticosteroids which was defined as at least 2 prescriptions prior to baseline with the most recent ≤28 days before baseline

#ever history of solid organ transplant or permanent cellular immune deficiency; history in the 24 months before baseline for aplastic anaemia, bone marrow or stem cell transplant; history in the 12 months before baseline for biologics or other immunosuppressant therapy (excluding corticosteroids), other or unspecified cellular immune deficiency



*Entry = latest of 12 March 2020, 40th birthday, 12 months post registration

Figure 8.1 Study design overview with 2020 England COVID-19 timeline

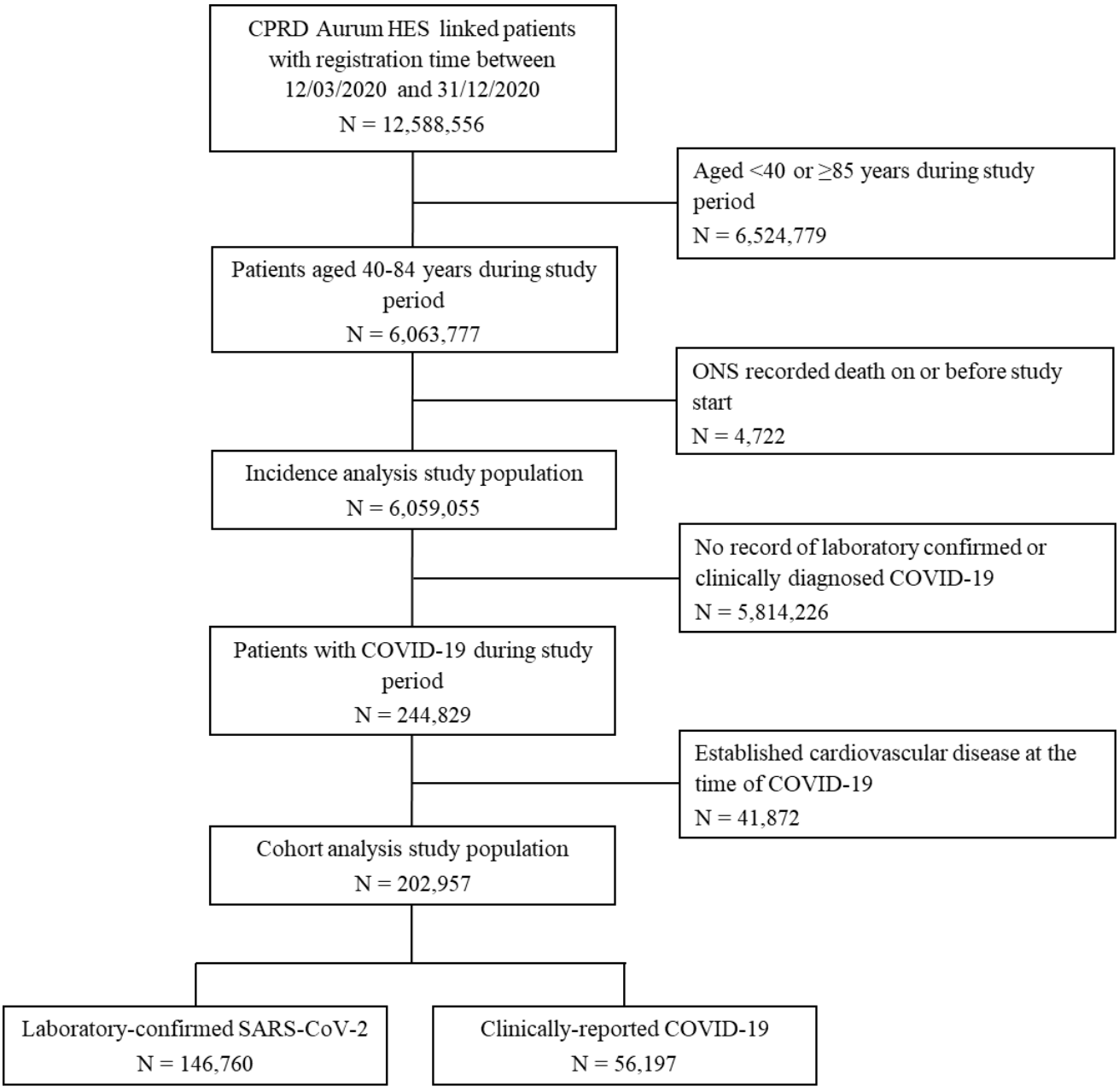


Figure 8.2 Study population flow chart

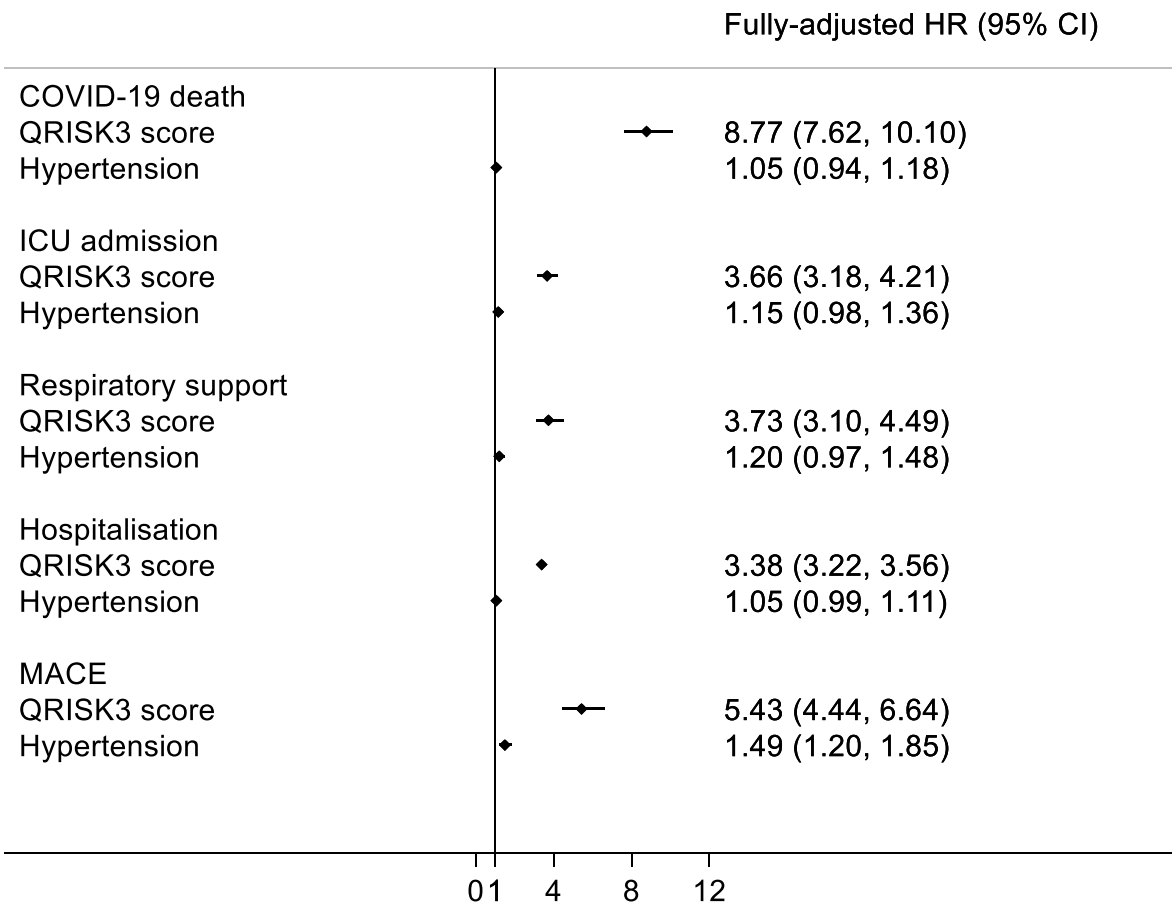


Figure 8.3 Fully-adjusted hazard ratios for raised cardiovascular risk effect on COVID-19 severe outcomes

QRISK3 score hazard ratios are for the effect of a score ≥10% with <10% as the reference. Hypertension hazard ratios are for the effect of having hypertension with not having hypertension as the reference.

8.3. Additional methods

In the research paper, the composite outcome of acute cardiovascular events is presented. I additionally, stratified by individual event type, particularly as acute cardiovascular events were the focus of the SCCS analysis.

Given QRISK3 was new at the time of study design, I also conducted a sensitivity analysis using the QRISK2 algorithm.

Due to the large effect estimates identified for some of the reported outcomes when cardiovascular risk was defined by QRISK3 score, in comparison to hypertension alone, I also conducted a post-hoc analysis with stratification by age group to evaluate the effect of age. Age, like all risk factors included in the calculation of QRISK3 score, had not been adjusted for in analysis as the variable is also considered in the assignment of individual scores.

8.4. Additional results

In adjusted analysis, raised cardiovascular risk resulted in HRs of similar magnitude across all individual events (**Table 8.3**). When risk was measured by QRISK3 score, the HR for highest for acute left ventricular heart failure (5.85 [4.00-8.57]), followed by ischaemic stroke (5.82 [3.55-9.53]), major ventricular arrhythmia (5.43 [3.60-8.18]), and finally ACS (4.99 [3.48-7.16]). In comparison, while there was a significant association between hypertension and ischaemic stroke (2.47 [1.42-4.30]), none of the other events were significantly associated with hypertension. This could be due to a lack of statistical power when individual outcomes were separated.

HRs generated when cardiovascular risk was measured by QRISK2 were similar to those obtained from QRISK3, although the effect estimate for each outcome was slightly lower for raised risk defined using QRISK2 other than hospitalisation which was slightly higher (**Table 8.4**).

There were no individuals aged 75-84 years with a QRISK3 score <10%, so age stratified results were generated for age groups 40-54, 55-64, and 65-74 years. Among all patients aged 40-54 years, 6.1% had a QRISK3 score ≥10% but among those who died from COVID-19, 25.4% had a QRISK3 score ≥10% (**Table 8.5**). Similarly high proportions among those aged 40-54 years were also identified for the outcomes of COVID-19 ICU admission, the need for respiratory support due to COVID-19, COVID-19 hospitalisation, and acute cardiovascular event following COVID-19. For the outcome of COVID-19 death, much lower effect estimates were obtained in age group strata (**Table 8.6**), compared to the overall, suggesting those aged 75-84 years drove the main result obtained (with the age group accounting for 43% of deaths). For COVID-19 death, the effect estimate was larger for the 40-54-year age group (4.32 [3.05-6.12]) than 55-64 (2.72 [2.15-3.43]) or 65-74 (2.90 [1.87-4.50]) years. We found an association between raised cardiovascular risk and severe COVID-19 outcomes in all age strata, even after excluding the stata aged ≥75 (given all had a QRISK3 score ≥10%) who accounted for a large proportion of the deaths. Additionally, the magnitude of effect estimates we obtained were greatest or similar in the youngest age group.

Table 8.3 Hazard ratios for the effect of raised cardiovascular risk on acute cardiovascular events after laboratory-confirmed SARS-CoV-2

	N events	Rate (95% CI) per 1,000 person-years	Crude HR (95% CI)	Age- and sex-adjusted HR (95% CI)	Fully-adjusted* HR (95% CI)
Major adverse cardiovascular event					
QRISK3 ≥10%	570	82.4 (75.9-89.4)	7.51 (6.40-8.81)	NA	5.43 (4.44-6.64)
QRISK3 <10%	204	11.5 (10.0-13.1)	1 (ref)	NA	1 (ref)
Hypertension	450	51.3 (46.8-56.3)	2.63 (2.28-3.04)	1.62 (1.40-1.87)	1.49 (1.20-1.85)
No hypertension	324	20.3 (18.2-22.6)	1 (ref)	1 (ref)	1 (ref)
Acute coronary syndrome					
QRISK3 ≥10%	162	23.2 (19.9-27.0)	7.08 (5.27-9.50)	NA	4.99 (3.48-7.16)
QRISK3 <10%	61	3.4 (2.7-4.4)	1 (ref)	NA	1 (ref)
Hypertension	123	13.9 (11.7-16.6)	2.32 (1.78-3.02)	1.52 (1.16-1.99)	1.12 (0.76-1.64)
No hypertension	100	6.2 (5.1-7.6)	1 (ref)	1 (ref)	1 (ref)
Major ventricular arrhythmia					
QRISK3 ≥10%	105	15.0 (12.3-18.1)	5.58 (4.00-7.80)	NA	5.43 (3.60-8.18)
QRISK3 <10%	51	2.9 (2.2-3.8)	1 (ref)	NA	1 (ref)
Hypertension	91	10.3 (8.4-12.6)	2.68 (1.95-3.68)	1.85 (1.33-2.56)	1.56 (0.94-2.58)
No hypertension	65	4.1 (3.2-5.2)	1 (ref)	1 (ref)	1 (ref)
Acute left ventricular failure					
QRISK3 ≥10%	206	29.5 (25.7-33.8)	9.77 (7.27-13.13)	NA	5.85 (4.00-8.57)
QRISK3 <10%	56	3.1 (2.4-4.1)	1 (ref)	NA	1 (ref)
Hypertension	146	16.5 (14.1-19.5)	2.37 (1.85-3.02)	1.31 (1.02-1.68)	1.44 (0.98-2.11)
No hypertension	116	7.2 (6.0-8.7)	1 (ref)	1 (ref)	1 (ref)
Ischaemic stroke					
QRISK3 ≥10%	97	13.8 (11.3-16.9)	7.24 (4.94-10.61)	NA	5.82 (3.55-9.53)
QRISK3 <10%	36	2.0 (1.5-2.8)	1 (ref)	NA	1 (ref)
Hypertension	90	10.2 (8.3-12.5)	3.97 (2.76-5.71)	2.40 (1.66-3.49)	2.47 (1.42-4.30)
No hypertension	43	2.7 (2.0-3.6)	1 (ref)	1 (ref)	1 (ref)

*Hypertension models were adjusted for age, sex, ethnicity, socioeconomic status, body-mass index, alcohol consumption, smoking status, total cholesterol: high density lipoprotein cholesterol ratio, family history of coronary heart disease, treatment with corticosteroids, antiplatelets, or anticoagulants, diagnosis of atrial fibrillation, migraine, diabetes, chronic kidney disease stage 3-5, chronic liver disease, chronic lung disease, asthma, severe mental illness, dementia, chronic neurological disease, learning disability, or malignancy, and treatment or diagnosis of a immunosuppressive condition; and QRISK3 models were adjusted for alcohol consumption, treatment with antiplatelets or anticoagulants, diagnosis of chronic liver disease, chronic lung disease, asthma, dementia, chronic neurological disease, learning disability, or malignancy, and treatment or diagnosis of a immunosuppressive condition (with are not included in the QRISK3 algorithm

Table 8.4 Hazard ratios for the effect of raised QRISK2 score on severe outcomes after laboratory-confirmed SARS-CoV-2

	N events	Rate (95% CI) per 1,000 person-years	Crude HR (95% CI)	Fully-adjusted[#] HR (95% CI)
COVID-19 death*				
QRISK3 ≥10%	2,178	306.5 (293.9-319.7)	15.82 (14.17-17.67)	9.23 (8.06-10.58)
QRISK3 <10%	370	20.8 (18.8-23.0)	1 (ref)	1 (ref)
ICU admission[†]				
QRISK3 ≥10%	863	116.4 (108.6-124.8)	3.99 (3.58-4.45)	3.39 (2.96-3.89)
QRISK3 <10%	586	31.0 (28.5-33.7)	1 (ref)	1 (ref)
Respiratory support[‡]				
QRISK3 ≥10%	493	65.2 (59.5-71.5)	4.07 (3.52-4.71)	3.52 (2.94-4.22)
QRISK3 <10%	325	17.1 (15.3-19.2)	1 (ref)	1 (ref)
Hospitalisation[$]				
QRISK3 ≥10%	6,510	1186.8 (1158.3-1216.0)	4.27 (4.11-4.44)	3.52 (3.36-3.70)
QRISK3 <10%	4,284	259.8 (252.1-267.7)	1 (ref)	1 (ref)
Acute cardiovascular event				
QRISK3 ≥10%	563	80.5 (74.1-87.5)	7.05 (6.02-8.26)	5.40 (4.45-6.56)
QRISK3 <10%	211	11.9 (10.4-13.6)	1 (ref)	1 (ref)

*Ascertained from ONS death certificate data in which the COVID related ICD-10 codes U07.1 or U07.2 were present in the record

[$]Ascertained from presence in CHESS dataset or HES APC record coded with primary diagnosis of U07.1 or U07.2

[†]Ascertained from CHESS records coded with ICU/HDU admission, only available for those with laboratory confirmed SARS-CoV-2

[‡]Ascertained from CHESS record coded with use of respiratory support via invasive mechanical ventilation, only available for those with laboratory confirmed SARS-CoV-2

[#]Adjusted for alcohol consumption, treatment with corticosteroids, antiplatelets or anticoagulants, diagnosis of migraine, chronic kidney stage 3, chronic liver disease, chronic lung disease, asthma, severe mental illness, dementia, chronic neurological disease, learning disability, or malignancy, and treatment or diagnosis of an immunosuppressive condition (with are not included in the QRISK2 algorithm).

Table 8.5 The number and proportion of patients with outcomes of interest by age group and QRISK3 score level

Age group	QRISK3 score	All study population	COVID-19 death	ICU	Respiratory support	Hospitalisation	Acute cardiovascular event
40-54	<10%	79,778 (93.9%)	170 (74.6%)	346 (78.5%)	196 (76.6%)	2,750 (81.2%)	119 (73.5%)
	≥10%	5,150 (6.1%)	58 (25.4%)	95 (21.5%)	60 (23.4%)	556 (16.8%)	43 (26.5%)
55-64	<10%	25,790 (64.9%)	163 (36.0%)	205 (40.8%)	111 (39.2%)	1,352 (45.9%)	75 (34.1%)
	≥10%	13,967 (35.1%)	290 (64.0%)	298 (59.2%)	172 (60.8%)	1,593 (54.1%)	145 (65.9%)
65-74	<10%	1,897 (12.8%)	32 (4.2%)	22 (5.9%)	13 (5.9%)	145 (6.0%)	10 (4.8%)
	≥10%	12,885 (87.2%)	737 (95.8%)	353 (94.1%)	206 (94.1%)	2,255 (94.0%)	199 (95.2%)
75-84	<10%	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	≥10%	7,293 (100.0%)	1,098 (100.0%)	130 (100.0%)	60 (100.0%)	2,143 (100.0%)	183 (100.0%)

Table 8.6 Hazard ratios for the effect of raised QRISK2 score on severe outcomes after laboratory-confirmed SARS-CoV-2

	Crude HR (95% CI)	Fully-adjusted[#] HR (95% CI)	Fully-adjusted[#] HR (95% CI) ages 40-54 years	Fully-adjusted[#] HR (95% CI) ages 55-64 years	Fully-adjusted[#] HR (95% CI) ages 65-74 years
COVID-19 death*					
QRISK3 ≥10%	15.82 (14.17-17.67)	8.77 (7.62-10.10)	4.32 (3.05-6.12)	2.72 (2.15-3.43)	2.90 (1.87-4.50)
QRISK3 <10%	1 (ref)	1 (ref)	1 (ref)	1 (ref)	1 (ref)
ICU admission[†]					
QRISK3 ≥10%	3.99 (3.58-4.45)	3.66 (3.18-4.21)	3.98 (3.01-5.26)	2.33 (1.88-2.90)	2.31 (1.34-3.96)
QRISK3 <10%	1 (ref)	1 (ref)	1 (ref)	1 (ref)	1 (ref)
Respiratory support[‡]					
QRISK3 ≥10%	4.07 (3.52-4.71)	3.73 (3.10-4.49)	4.01 (2.78-5.79)	2.39 (1.79-3.21)	2.75 (1.29-5.90)
QRISK3 <10%	1 (ref)	1 (ref)	1 (ref)	1 (ref)	1 (ref)
Hospitalisation[$]					
QRISK3 ≥10%	4.27 (4.11-4.44)	3.38 (3.22-3.56)	2.88 (2.59-3.20)	2.18 (2.00-2.37)	2.27 (1.85-2.79)
QRISK3 <10%	1 (ref)	1 (ref)	1 (ref)	1 (ref)	1 (ref)
Acute cardiovascular event					
QRISK3 ≥10%	7.05 (6.02-8.26)	5.43 (4.44-6.64)	4.45 (2.97-6.66)	3.26 (2.34-4.53)	3.23 (1.42-7.32)
QRISK3 <10%	1 (ref)	1 (ref)	1 (ref)	1 (ref)	1 (ref)

*Ascertained from ONS death certificate data in which the COVID related ICD-10 codes U07.1 or U07.2 were present in the record
[$]Ascertained from presence in CHESS dataset or HES APC record coded with primary diagnosis of U07.1 or U07.2
[†]Ascertained from CHESS records coded with ICU/HDU admission, only available for those with laboratory confirmed SARS-CoV-2
[‡]Ascertained from CHESS record coded with use of respiratory support via invasive mechanical ventilation, only available for those with laboratory confirmed SARS-CoV-2
[#]Adjusted for alcohol consumption, treatment with antiplatelets or anticoagulants, diagnosis of chronic liver disease, chronic lung disease, asthma, dementia, chronic neurological disease, learning disability, or malignancy, and treatment or diagnosis of a immunosuppressive condition (with are not included in the QRISK3 algorithm).

8.5. Chapter summary

- In this population-based study I used CPRD Aurum data linked to HES APC, ONS death registrations, CHESS and SGSS data to estimate the impact of raised cardiovascular risk on severe COVID-19 outcomes.
- Among the 6,059,055 individuals aged 40-84 years who were included, 12.2% had established CVD and among those without established CVD, 31.9% had a QRISK3 score ≥10% and 55.9% had a QRISK3 score <10%, and 31.1% had hypertension and 56.7% had no hypertension.
- The incidence of COVID-19 death, ICU admission, hospitalisation, and acute cardiovascular events with the highest incidence among those with established CVD, followed by those with raised cardiovascular risk and was lowest among those at low cardiovascular risk.
- After excluding those with established CVD, 146,760 people had laboratory-confirmed SARS-CoV-2. 26.8% had a QRISK3 score ≥10% and 73.2% had a QRISK3 score <10%, 34.0% had hypertension and 66.0% had no hypertension.
- After adjustment for confounders, risks of COVID-19 death, ICU admission, hospital admission and acute cardiovascular events were all greater among individuals at raised cardiovascular risk measured by QRISK3 score, compared to those at low risk, despite no increase in recorded infections in this group. When cardiovascular risk was measured by hypertension alone, differences were only evident for acute cardiovascular events outcomes.
- Analysis by pandemic waves revealed similar patterns, although the incidence of severe outcomes was greatest during the first wave.
- All individuals aged ≥75 years had a QRISK3 score ≥10%. Post-hoc age stratified results showed an increase in risk of severe COVID-19 outcomes among individuals of all age groups with raised cardiovascular risk. Age has an important associated with the risk of severe COVID-19 outcomes but so do other cardiovascular risk factors.

- Individuals with a raised QRISK3 score should be prioritised for prevention and treatment. Addressing cardiovascular risk factors could improve COVID-19 outcomes.

Chapter 9 Investigating the association between COVID-19 and cardiovascular complications by varying cardiovascular risk

9.1. Chapter overview

In this chapter, I address thesis objective 5. I used CPRD Aurum data linked to HES APC, ONS death registrations, CHESS and SGSS data with a SCCS design to investigate the association between COVID-19 and the relative risk of acute cardiovascular events, including descriptive comparison between individuals with raised and low cardiovascular risk. As with the analysis to address objective 4, the study concept and protocol (Chapter 11 Appendix 6) were developed by my primary supervisor, Charlotte Warren-Gash. I conducted all analyses, made detailed methodological decisions during the analysis process and led the drafting of the manuscript which the author group intend to submit to the New England Journal of Medicine. The chapter begins with presentation of this draft manuscript followed by further analysis scoped *a priori* but which were not included in the manuscript.

9.2. Drafted manuscript

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	230019	**Title**	Ms
First Name(s)	Jennifer		
Surname/Family Name	Davidson		
Thesis Title	Acute respiratory infections, cardiovascular complications, and prevention among people with raised cardiovascular risk		
Primary Supervisor	Dr Charlotte Warren-Gash		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	New England Journal of Medicine
Please list the paper's authors in the intended authorship order:	Jennifer A Davidson, Helen Strongman, Emily Herrett, Harriet Forbes, Prof Liam Smeeth, Prof Judith Breuer, Prof Amitava Banerjee, Charlotte Warren-Gash
Stage of publication	**Not yet submitted**

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I am the first author of this research paper. Charlotte Warren-Gash conceptualised and lead the design of the study. I conducted the analysis and drafted the research paper.

SECTION E

Student Signature	[redacted]
Date	25/07/2022

Supervisor Signature	[redacted]
Date	25-07-2022

Acute cardiovascular events after SARS-CoV-2 infection in England in 2020: a self-controlled case series study

Authors: Jennifer A Davidson, MSc[1], Prof Amitava Banerjee, DPhil[2], Helen Strongman, PhD[1], Emily Herrett, PhD[1], Harriet Forbes, PhD[3], Prof Liam Smeeth, FRCGP[1], Prof Judith Breuer, FMedSci[4], Charlotte Warren-Gash, PhD[1]

Affiliations:

[1] Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK

[2] Institute of Health Informatics, University College London, London, UK

[3] Population Health Sciences, Bristol Medical School, University of Bristol, UK

[4] Institute of Infection and Immunity, University College London, London, UK

Corresponding author: Jennifer Davidson, jennifer.davidson@lshtm.ac.uk, Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, United Kingdom, +44 (0)20 7636 8636

Word count

Abstract: 248

Text: 2,820

Tables/Figures: 2 tables, 3 figures

References: 35

Supplementary material: 4 tables

ABSTRACT

Background: COVID-19 may lead to cardiovascular complications in adults with no pre-existing cardiovascular disease. We aimed to assess risk of incident cardiovascular outcomes after COVID-19 infection by level of cardiovascular risk in wave one and two of the pandemic in England in 2020.

Methods: We conducted a self-controlled case-series study among adults aged 40-84 years with no pre-existing cardiovascular disease using linked data from the Clinical Practice Research Datalink. We generated season-adjusted incidence ratios (IRs) for first acute cardiovascular event after SARS-CoV-2 infection compared with baseline time before and >91 days after infection. We used a composite, and individual, acute cardiovascular event outcomes including myocardial infarction, major ventricular arrhythmia, left ventricular heart failure, and ischaemic stroke. We stratified by cardiovascular risk, using diagnosed hypertension and QRISK3 predicted risk, and by wave one and one of the pandemic.

Results: We included 1,762 individuals, 76.6% had a QRISK3 score ≥10% and 59.4% had hypertension. The risk of any cardiovascular event was elevated in the 1-7 days after infection (IR 7.14 [95% CI 6.06-8.41]) and, while the effect size tapered, risk remained for 15-28 days after infection (1.74 [1.33-2.26]). Risks were similar for individual event type, differing by level of cardiovascular risk, and in wave 1 and 2 of the pandemic.

Conclusions: SARS-CoV-2 infection is associated with early elevations in the risk of first acute cardiovascular event, across cardiovascular risk levels and in both wave one and two of the pandemic. Prevention of COVID-19 is important to avert cardiovascular complications.

INTRODUCTION

COVID-19 has been associated with a number of cardiovascular complications. Multiple self-controlled case series (SCCS) studies have identified elevated incidence of myocardial infarction (MI) and ischaemic stroke. Results from Sweden found more than two-fold increases in the relative incidence of first MI in the two weeks and ischaemic stroke in the month after COVID-19.[1] The matched cohort analysis conducted using the same dataset supported SCCS findings, with three-fold increased odds of MI and ischaemic stroke in the two weeks after infection. A Danish SCCS analysis found three- and six-fold increases in the relative incidence of first MI and stroke, respectively, in the first month after COVID-19.[2] Results from a Scottish SCCS found a five- and seven-fold increase in the relative incidence of MI and ischemic stroke, in the seven days after testing positive for COVID-19.[3] This SCCS study identified even higher risk for pulmonary embolism and deep vein thrombosis. Other studies have also reported increased incidence of myocarditis, acute heart failure and arrhythmias following COVID-19 illness.[4,5]

Biomarkers of cardiac injury, such as troponin, were higher among COVID-19 patients with cardiovascular risk factors, and were associated with a range of severe outcomes including the need for mechanical ventilation and death.[6,7] A sizable proportion of hospitalised individuals who died with COVID-19 had a raised troponin, implicating myocardial ischaemia and injury as a possible mechanism leading to adverse outcomes of COVID-19.[8] Although initial population-based studies conducted using data from the first wave of the pandemic have shown a short-term increased relative incidence of some cardiovascular complications following COVID-19,[1,2] this has not been assessed in a UK population-based study, nor is it known whether risk has evolved over the course of the pandemic. There is also uncertainty on the role which background cardiovascular risk plays in COVID-19-related cardiovascular complications.[9] In England, individuals with raised cardiovascular risk but without existing cardiovascular disease (CVD), were not considered ‘clinically vulnerable’ to COVID-19, therefore, not targeted or prioritised for any prevention measures.[10]

We used a SCCS analysis to investigate the association between laboratory-confirmed SARS-CoV-2 and risk of a range of acute cardiovascular events, with stratification by underlying cardiovascular risk status and wave of the pandemic.

METHODS

Data sources

We used anonymized primary care data from the Clinical Practice Research Datalink (CPRD) Aurum January 2022 dataset,[11] with linked secondary care data from the Hospital Episodes Statistics Admitted Patient Care (HES APC) database, COVID-19 Hospitalisations in England Surveillance System (CHESS) data, Second Generation Surveillance System (SGSS) SARS-CoV-2 PCR test results, and death registration data from the Office of National Statistics (ONS). CPRD Aurum is a growing dataset of longitudinal records from National Health Service (NHS) primary care general practitioners, currently comprising >40 million individuals in England. The data collated include demographics, diagnoses, prescriptions, and immunizations. HES APC is a reimbursement dataset containing diagnosis and procedures collated from inpatient care at NHS hospitals in England. CHESS was set up at the start of the COVID-19 pandemic to provide timely surveillance of outcomes among hospitalized COVID-19 patients.

The CPRD Independent Scientific Advisory Committee (application 20_000135) and the London School of Hygiene and Tropical Medicine Ethics (LSHTM) Committee (application 22717) approved the study.

Study design

We carried out a SCCS study which makes within-person comparisons. In a SCCS study only individuals with both the exposure and outcome are included. Individuals act as their own controls during periods of non-exposure.[12] The main advantage of this design is the removal of fixed characteristics confounding effects that vary between individuals.

SCCS analyses use conditional Poisson regression to model, and derive incidence ratios (IRs), the effect of a time-varying exposure on the outcome by comparing the incidence of events during exposed time with the incidence during unexposed (baseline) time.[13] The SCCS method relies on several key assumptions.[14,15] First, event recurrences must be independent i.e., an event must not increase the probability of a further event. Second, an event should not impact subsequent exposure. Third, an event must not influence the end of the period of observation, which can be violated when the event increases the likelihood of mortality.

Study population and follow-up

CPRD-recorded patients aged 40-84 years (after which age all individuals are classified as having raised cardiovascular risk) who experienced their first acute cardiovascular event and had COVID-19 between 12 March 2020 (when daily reporting to CHESS was initiated)[16] and 31 December 2020 were eligible for inclusion. Follow-up started at the latest of; age 40 years, 12 months post-registration, or 12 March 2020 and ended at the earliest date of; death, patient transfer out of primary care practice, last data collection from the practice, or 31 December 2020. We excluded anyone with established CVD – defined as heart disease (congenital or otherwise), heart failure, stroke or transient ischaemic attack – recorded in CPRD or HES APC prior to 12 March 2020. Individuals with CVD are more likely to have a further cardiovascular event, thereby their inclusion in the study population would violate the first assumption outlined in the study design section.

Outcome

The primary outcome was acute cardiovascular event which we defined as; acute coronary syndrome (ACs) capturing MI and unstable angina, left ventricular heart failure, major ventricular arrhythmia, or ischaemic stroke. We included diagnoses recorded in CPRD or HES APC. Our secondary outcomes were each of the cardiovascular conditions separately. All code lists are published on LSHTM Data Compass.[17]

Exposure

We defined laboratory-confirmed SARS-CoV-2 using CPRD SGSS and CHESS data. Exposure date was the earliest specimen date within the follow-up period. In a secondary analysis we redefined the exposure as clinically reported COVID-19 (recorded in CPRD or HES APC (ICD-10 codes U07.1 or U07.2 in primary diagnostic position) without laboratory-confirmation, as early in the pandemic, COVID-19 testing was largely limited to individuals who were hospitalized.

Statistical analysis

We conducted all analyses in Stata (version 16).

We described key baseline characteristics of the whole study population and stratified by cardiovascular risk. The characteristics were age group (40-54, 55-64, 65-74 and 75-84), sex, COVID-19 associated hospital stay (defined by hospitalization due to COVID-19 based on HES APC ICD-10 codes U07.1 or U07.2 in primary diagnostic position or record in CHESS dataset), events resulting in death, and all deaths during follow-up.

We stratified our analyses by underlying cardiovascular risk using QRISK3 score and hypertension status. QRISK3 is a UK developed cardiovascular risk calculator used to estimate an individual's ten-year likelihood of CVD. The score is widely used in UK primary care, as was the prior version QRISK2, and has been validated.[18,19] The risk factors used in the score are age, sex, ethnicity, socioeconomic status, family history of coronary heart disease in a first degree relative aged <60 years, comorbid health conditions (diabetes, treated hypertension, rheumatoid arthritis, systemic lupus erythematosus, atrial fibrillation, chronic kidney disease stages 3-5, migraine, severe mental illness, HIV, erectile dysfunction) body mass index, systolic blood pressure reading and it's variability, total cholesterol to high density lipoprotein cholesterol ratio, smoking status, and corticosteroid treatment. We used our published Stata scripts[20] to measure baseline risk factors using codes and measures recorded in patient CPRD records and applied the QRISK3 algorithm, published by the developers[21] to assign a final risk score. We assigned

hypertension status using coded CPRD hypertension diagnoses within the five years before baseline or the most recent to baseline blood pressure (BP) reading with systolic BP of ≥140mmHg or diastolic BP of ≥90mmHg. We classified individuals with raised cardiovascular risk (hypertension or QRISK3 ≥10%) and low cardiovascular risk (no hypertension or QRISK3 <10%) at study entry.

We compared the incidence of acute cardiovascular events in exposed periods following COVID-19 infection with baseline (unexposed) periods (**Figure *9.1***). The exposed period was divided into 1-7, 8-14, 15-28, 29-91 days following the exposure date. We excluded the 14 days before the exposure date from the baseline because the temporal relationship between the infection and acute cardiovascular event cannot be determined, with specimens potentially obtained after hospitalization for the cardiovascular event. We used conditional Poisson regression to calculate incidence ratios (IRs) for acute cardiovascular events occurring within each exposed stratum compared with the baseline period. We adjusted for season split into the warmer months of April to September and cooler months of October to March.

We further stratified results by the first UK COVID-19 wave (12 March to 16 August) or second wave (17 August to 31 December), during which different testing strategies and clinical management of infected individuals were in practice in operation.[22] Additional stratifying factors were age group and sex.

We repeated our initial analysis excluding fatal acute cardiovascular events. Acute cardiovascular events can result in death, violating the SCCS assumption that observation periods should end independent of event timing. We classified fatal events as those for which the individual's death date was ≤30 days after the event.

RESULTS

Description of the study population

We identified 1,762 individuals with laboratory-confirmed SARS-CoV-2 and first acute cardiovascular event in our study period (**Figure *9.2***), of which 569 (32.3%) were new onset of left ventricular heart failure, 565 (32.1%) were acute coronary syndrome, 401 (22.8%) were ischaemic stroke and 227 (12.9%)

were major ventricular arrhythmia. The characteristics of the study population are shown in **Table 9.1**. The majority of individuals had raised cardiovascular risk (QRISK3 ≥10%; 76.6%, n=1,350 and hypertension; 59.4%, n=1,047). Overall, there was a higher proportion of men (60.2%, n=1,060) than women (39.8%, n=702). Nearly forty-percent (39.2%, n=691) of individuals were hospitalized due to their COVID-19. Over one-fifth (20.9%, n=369) of individuals died within 30 days of their cardiovascular event and more than one-quarter (26.0%, n=458) died during follow-up.

Association between COVID-19 and first acute cardiovascular event

The season adjusted relative incidence of a first acute cardiovascular event was markedly raised in the first seven days after SARS-CoV-2 infection with an IR of 7.14 (95% CI 6.06-8.41) and fell sharply to an IR of 1.14 (0.92-1.41) by days 29-91 (**Table 9.2**). The same pattern, with similar effect estimates, was observed in both raised and low cardiovascular risk profiles (days 1-7: QRISK3 ≥10% 6.97 [5.79-8.39], QRISK3 <10% 7.78 [5.48-11.05], hypertension 6.55 [5.28-8.14], and no hypertension 8.04 [6.26-10.33]). When stratified by wave one and two, the relative incidence of a cardiovascular event in the first seven days after SARS-CoV-2 was higher, though not significantly, in wave one (6.01 [4.5-8.11]) than wave two (4.19 [3.35-5.23) (**Table 9.2**).

The relative incidence differed by event type; in the first seven days after SARS-CoV-2, the relative incidence was highest for major ventricular arrhythmia (26.62 [17.25-41.09]), although the number of events was small, followed by left ventricular heart failure (7.86 [5.97-10.33]), ischaemic stroke (4.72 [3.27-6.82]), and finally ACS (4.23 [2.98-6.00]) (**Figure *9.3***).

The incidence ratios were marginally higher in men (days 1-7: 7.90 [6.42-9.72]) than women (days 1-7: 6.10 [4.67-7.97]) and also higher in those aged 40-64 years (days 1-7: 9.61 [7.54-12.26]) than 65-84 years (days 1-7: 5.55 [4.44-6.94]) (**Table S9.1**).

Sensitivity analysis using clinically reported COVID-19

Redefining infection as clinically reported non-laboratory confirmed COVID-19 yielded 932 individuals with COVID-19 and a first cardiovascular event. Compared to individuals with laboratory-confirmed SARS-CoV-2, a higher proportion of those with clinically reported COVID-19 were women (44.0% vs 39.8%), and a lower proportion were hospitalized due to their COVID-19 (16.1% vs 39.2%) or died within 30 days of their cardiovascular event (16.1% vs 20.9%) or at all in the study follow-up (20.0% vs 26.0%) (**Table S9.2**). The IR for first acute cardiovascular event was substantially raised, with a larger magnitude that laboratory-confirmed SARS-CoV-2, in the seven days after diagnosis (10.04 [8.05-12.53]) and although the association tapered over time, the relative incidence remained raised through all risk windows (days 29-91: 1.76 [1.41-2.20]) (**Table S9.3**).

Sensitivity analysis removing people who died

After excluding the 369 fatal events, the relative incidence of a first acute cardiovascular event after SARS-CoV-2 decreased compared to when these events were included, but remained raised (days 1-7: 4.36 [3.52-5.41]) (**Table S9.4**).

DISCUSSION

In this population-based SCCS study among adults aged 40 to 84 years with varying cardiovascular risk, we found an increased risk of first acute cardiovascular events in the four weeks after COVID-19 for both raised and low cardiovascular risk profiles, in those with laboratory-confirmed infection or clinically reported illness across all cardiovascular risk profiles. The highest risk was in the first seven days after COVID-19 identification and the association tapered over time. Among the individual outcome of interest, relative incidence post-COVID-19 was highest for major ventricular arrhythmia, followed by left ventricular heart failure, ischaemic stroke, and finally ACS. As we would expect, the effect estimates were marginally higher in men and although the relative incidence was higher for those aged less than 65 years, there is still a low absolute risk of cardiovascular events in this age group. We identified a higher

incidence of first acute cardiovascular event following laboratory-confirmed COVID-19 in the first wave of the pandemic than the second.

Our results are consistent with those of previous SCCS studies investigating COVID-19 associated MI and ischaemic stroke using population-based registers from Denmark, Sweden and Scotland. Danish national register data, to mid-July 2020, showed that in the 31 days after COVID-19, there were three- and six-fold increases in the relative incidence of first MI and stroke, respectively, although the study was based on a small sample size with only 17 MI and 44 stroke events.[2] Swedish EHR data from February to September 2020 illustrated a more than two-fold increase in the relative incidence of first MI in the two weeks after COVID-19 based on 186 events.[1] However, beyond this time the increase in MI relative incidence was not significant. In comparison, the relative incidence of ischaemic stroke remained high in the month after COVID-19, although was only based on 254 events. A SCCS study of Scottish data with 1,449 individuals, using a study period of March 2018 to October 2020, evaluated the risk of MI, ischaemic stroke, pulmonary embolism and deep vein thrombosis. The study identified a 12-fold increase in relative incidence in the first seven days after COVID-19.[3]

Cardiovascular complications are not uniquely triggered by COVID-19. Other acute respiratory infections including GP-diagnosed acute respiratory infections,[23] influenza-like illnesses,[24] laboratory-confirmed respiratory virus infections and *Streptococcus pneumoniae*[25,26] are also associated with a transient increase in the risk of MI and stroke in SCCS studies. We recently demonstrated that underlying cardiovascular risk impacts the likelihood of first acute cardiovascular event after acute respiratory infection; those with raised cardiovascular risk had a 2 to 3 times higher risk of cardiovascular complications after ARI.[27] Nevertheless, individuals at raised cardiovascular risk but without pre-existing cardiovascular disease are not traditionally considered a high-risk group for severe outcomes after respiratory infections.[28]

SARS-CoV-2, like other ARI causative agents however with greater systemic effect, gives rise to a series of inflammatory, thrombotic and microvascular dysfunction events which can lead to new cardiovascular complications.[29] Angiotensin-converting enzyme 2 is a surface protein that is key to SARS-CoV-2 cell

entry but is also part of the renin-angiotensin-aldosterone system which is important for normal functioning of the cardiovascular system, so likely to play a role in the cardiovascular effects of the virus.[30] Numerous cardiovascular risk factors, including obesity, hypertension and diabetes, have been implicated in the interaction between SARS-CoV-2 and cardiovascular complications[31] and may give rise more severe disease and increase COVID-19 associated cardiovascular complications.[29]

Prevention and management of cardiovascular risk factors, along with COVID-19 itself, should be sustained and prioritisied. The results generated by our study, and those of previous studies, show a spectrum of acute cardiovascular complications associated with COVID-19, with the timing of risk varying for different outcomes. Further understanding of longer-term cardiovascular outcomes,[32,33] including the cardiovascular manifestations and mechanisms of post-COVID-19 syndrome, is an important focus for future research.[31,34]

We included a large study population using primary and secondary care linked data sources generalisable to the population of England in terms of age, sex and deprivation. However, the use of the SCCS method means only individuals with COVID-19 who had an acute cardiovascular event were included, limiting generalisability to those at risk, underlying or from an exposure, of cardiovascular complications. At the start of the pandemic, laboratory testing for SARS-CoV-2 was targeted towards individuals with a clinical need. This will have limited the number of individuals entering our study if their COVID-19 illness was early in the pandemic. We included a study population with clinically reported COVID-19 to account for this data bias. To avoid the impact of major SARS-CoV-2 variants of concern or vaccination, we limited our study to 2020. However, improved treatment options were in use by wave two of the pandemic, which may explain the higher relative incidence identified in wave one compared to wave two.

In conclusion, we showed an elevated risk of a range of acute cardiovascular events in the days and weeks after SARS-CoV-2 infection in both waves one and two of the pandemic. Prevention of COVID-19 is important to avert cardiovascular complications. Prevention and management of cardiovascular risk

factors will also alleviate the burden of COVID-19 associated cardiovascular events. However, cardiovascular management has also been impacted by the pandemic.[35]

REFERENCES

1. Katsoularis I, Fonseca-Rodríguez O, Farrington P, Lindmark K, Connolly A-MF. Risk of acute myocardial infarction and ischaemic stroke following COVID-19 in Sweden: a self-controlled case series and matched cohort study. The Lancet 2021;398(10300):599–607.
2. Modin D, Claggett B, Sindet-Pedersen C, et al. Acute COVID-19 and the Incidence of Ischemic Stroke and Acute Myocardial Infarction. Circulation 2020;2080–2.
3. Ho FK, Man KKC, Toshner M, et al. Thromboembolic Risk in Hospitalized and Nonhospitalized COVID-19 Patients: A Self-Controlled Case Series Analysis of a Nationwide Cohort. Mayo Clinic Proceedings 2021;96(10):2587–97.
4. Long B, Brady WJ, Koyfman A, Gottlieb M. Cardiovascular complications in COVID-19. The American Journal of Emergency Medicine 2020;38(7):1504–7.
5. Thakkar S, Arora S, Kumar A, et al. A Systematic Review of the Cardiovascular Manifestations and Outcomes in the Setting of Coronavirus-19 Disease: https://doi.org/101177/1179546820977196 2020;14.
6. Majure DT, Gruberg L, Saba SG, Kvasnovsky C, Hirsch JS, Jauhar R. Usefulness of Elevated Troponin to Predict Death in Patients With COVID-19 and Myocardial Injury. The American Journal of Cardiology 2021;138:100–6.
7. Lippi G, Lavie CJ, Sanchis-Gomar F. Cardiac troponin I in patients with coronavirus disease 2019 (COVID-19): Evidence from a meta-analysis. Progress in Cardiovascular Diseases 2020;63(3):390–1.
8. Tersalvi G, Vicenzi M, Calabretta D, Biasco L, Pedrazzini G, Winterton D. Elevated Troponin in Patients With Coronavirus Disease 2019: Possible Mechanisms. Journal of Cardiac Failure 2020;26(6):470–5.
9. Sabatino J, De Rosa S, Di Salvo G, Indolfi C. Impact of cardiovascular risk profile on COVID-19 outcome. A meta-analysis. PLOS ONE 2020;15(8):e0237131.
10. NHS Digital. COVID-19 Population Risk Assessment - NHS Digital [Internet]. 2022 [cited 2022 Jul 10];Available from: https://digital.nhs.uk/coronavirus/risk-assessment/population
11. Clinical Practice Research Datalink. CPRD Aurum January 2022 dataset [Internet]. 2022 [cited 2022 Jul 3];Available from: https://doi.org/10.48329/db7t-ay41
12. Whitaker HJ, Farrington CP, Spiessens B, Musonda P. Tutorial in biostatistics: The self-controlled case series method. Statistics in Medicine 2006;25(10):1768–97.

13. Whitaker HJ, Hocine MN, Farrington CP. The methodology of self-controlled case series studies. Statistical Methods in Medical Research 2009;18(1):7–26.

14. Whitaker HJ, Ghebremichael-Weldeselassie Y, Douglas IJ, Smeeth L, Farrington CP. Investigating the assumptions of the self-controlled case series method. Stat Med [Internet] 2018 [cited 2021 Dec 17];37(4):643–58. Available from: https://pubmed.ncbi.nlm.nih.gov/29094391/

15. Petersen I, Douglas I, Whitaker H. Self controlled case series methods: an alternative to standard epidemiological study designs. BMJ [Internet] 2016 [cited 2019 Mar 19];i4515. Available from: http://www.bmj.com/lookup/doi/10.1136/bmj.i4515

16. Public Health England. COVID-19 Hospitalisation in England Surveillance System (CHESS) – daily reporting [Internet]. 2020;Available from: https://www.england.nhs.uk/coronavirus/wp-content/uploads/sites/52/2020/03/phe-letter-to-trusts-re-daily-covid-19-hospital-surveillance-11-march-2020.pdf

17. Daivdson J, Warren-Gash C, Banerjee A, McDonald H, Smeeth L. London School of Hygiene and Tropical Medicine Data Compass [Internet]. Codelists. 2022 [cited 2022 Jan 14];Available from: https://datacompass.lshtm.ac.uk/id/eprint/2675/

18. Collins GS, Altman DG. An independent external validation and evaluation of QRISK cardiovascular risk prediction: a prospective open cohort study. BMJ 2009;339(7713):144–7.

19. Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. BMJ [Internet] 2017 [cited 2019 May 15];357:j2099. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28536104

20. Davidson J, Strongman H, Herrett E, Gadd S. qrisk_cprd_aurum: QRISK Aurum bundle version 2.0. 2022;

21. ClinRisk Ltd. QRISK3 algorithm [Internet]. 2017;Available from: https://qrisk.org/three/src.php

22. Office from National Statistics. Coronavirus (COVID-19) Infection Survey technical article: waves and lags of COVID-19 in England, June 2021 [Internet]. 2021 [cited 2021 Oct 27];Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/articles/coronaviruscovid19infectionsurveytechnicalarticle/wavesandlagsofcovid19inenglandjune2021

23. Smeeth L, Thomas SL, Hall AJ, Hubbard R, Farrington P, Vallance P. Risk of myocardial infarction and stroke after acute infection or vaccination. N Engl J Med [Internet] 2004 [cited 2019 Feb 11];351(25):2611–8. Available from: http://www.nejm.org/doi/abs/10.1056/NEJMoa041747

24. Warren-Gash C, Hayward AC, Hemingway H, et al. Influenza Infection and Risk of Acute Myocardial Infarction in England and Wales: A CALIBER Self-Controlled Case Series Study. The Journal of Infectious Diseases 2012;206(11):1652–9.

25. Warren-Gash C, Blackburn R, Whitaker H, McMenamin J, Hayward AC. Laboratory-confirmed respiratory infections as triggers for acute myocardial infarction and stroke: a self-controlled case series analysis of national linked datasets from Scotland. European Respiratory Journal [Internet]

2018 [cited 2019 Feb 11];51(3):1701794. Available from: http://erj.ersjournals.com/lookup/doi/10.1183/13993003.01794-2017

26. Ohland J, Warren-Gash C, Blackburn R, et al. Acute myocardial infarctions and stroke triggered by laboratory-confirmed respiratory infections in Denmark, 2010 to 2016. Eurosurveillance 2020;25(17):1900199.

27. Davidson JA, Banerjee A, Smeeth L, et al. Risk of acute respiratory infection and acute cardiovascular events following acute respiratory infection among adults with increased cardiovascular risk in England between 2008 and 2018: a retrospective, population-based cohort study. The Lancet Digital Health 2021;3(12):e773–83.

28. Public Health England. Influenza: the green book, chapter 19 [Internet]. 2020 [cited 2020 Dec 3];Available from: https://www.gov.uk/government/publications/influenza-the-green-book-chapter-19

29. Nishiga M, Wang DW, Han Y, Lewis DB, Wu JC. COVID-19 and cardiovascular disease: from basic mechanisms to clinical perspectives. Nature Reviews Cardiology 2020 17:9 2020;17(9):543–58.

30. Poznyak A V., Bezsonov EE, Eid AH, et al. ACE2 Is an Adjacent Element of Atherosclerosis and COVID-19 Pathogenesis. International Journal of Molecular Sciences 2021, Vol 22, Page 4691 2021;22(9):4691.

31. Mohamed MO, Banerjee A. Long COVID and cardiovascular disease: a learning health system approach. Nature Reviews Cardiology 2022 19:5 2022;19(5):287–8.

32. Knight R, Walker V, Ip S, et al. Association of COVID-19 with arterial and venous vascular diseases: a population-wide cohort study of 48 million adults in England and Wales. medRxiv [Internet] 2021 [cited 2022 Jul 24];27:2021.11.22.21266512. Available from: https://www.medrxiv.org/content/10.1101/2021.11.22.21266512v1

33. Xie Y, Xu E, Bowe B, Al-Aly Z. Long-term cardiovascular outcomes of COVID-19. Nature Medicine [Internet] 2022 [cited 2022 Jul 24];28(3):583. Available from: /pmc/articles/PMC8938267/

34. Raman B, Bluemke DA, Lüscher TF, Neubauer S. Long COVID: post-acute sequelae of COVID-19 with a cardiovascular focus. European Heart Journal 2022;43(11):1157–72.

35. Ball S, Banerjee A, Berry C, et al. Monitoring indirect impact of COVID-19 pandemic on services for cardiovascular diseases in the UK. Heart [Internet] 2020 [cited 2022 Jul 24];106(24):1890–7. Available from: https://pubmed.ncbi.nlm.nih.gov/33020224/

Table 9.1 Baseline characteristics of study population

	All	QRISK3		Hypertension	
		Raised risk	Low risk	Raised risk	Low risk
	N=1,762	N=1,350	N=412	N=1,047	N=715
Sex					
Women	702 (39.8%)	504 (37.3%)	198 (48.1%)	399 (38.1%)	303 (42.4%)
Men	1,060 (60.2%)	846 (62.7%)	214 (51.9%)	648 (61.9%)	412 (57.6%)
Age group (years)					
40-54	327 (18.6%)	84 (6.2%)	243 (59.0%)	154 (14.7%)	173 (24.2%)
55-64	436 (24.7%)	293 (21.7%)	143 (34.7%)	253 (24.2%)	183 (25.6%)
65-74	454 (25.8%)	428 (31.7%)	26 (6.3%)	276 (26.4%)	178 (24.9%)
75-84	545 (30.9%)	545 (40.4%)	0 (0.0%)	364 (34.8%)	181 (25.3%)
COVID-19 associated hospital stay	691 (39.2%)	543 (40.2%)	148 (35.9%)	410 (39.2%)	281 (39.3%)
Median (IQR)	10 (4-21)	9 (4-20)	10 (4-26)	9 (4-20)	10 (4-25)
Died ≤30 days after event	369 (20.9%)	301 (22.3%)	68 (16.5%)	208 (19.9%)	161 (22.5%)
Died in study period	458 (26.0%)	383 (28.4%)	75 (18.2%)	258 (24.6%)	200 (28.0%)

Table 9.2 Season-adjusted incidence ratios for acute cardiovascular events occurring in exposed periods after SARS-CoV-2 infection by cardiovascular risk and COVID-19 wave

Risk period	All		QRISK3				Hypertension			
			Raised risk		Low risk		Raised risk		Low risk	
	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)
All										
1-7 days	219	7.14 (6.06-8.41)	171	6.97 (5.79-8.39)	48	7.78 (5.48-11.05)	122	6.55 (5.28-8.14)	97	8.04 (6.26-10.33)
8-14 days	88	3.72 (2.96-4.68)	60	3.22 (2.44-4.24)	28	5.62 (3.68-8.58)	46	3.22 (2.35-4.40)	42	4.50 (3.21-6.31)
15-28 days	64	1.74 (1.33-2.26)	46	1.60 (1.18-2.18)	18	2.22 (1.34-3.68)	42	1.86 (1.34-2.59)	22	1.53 (0.98-2.39)
29-91 days	108	1.14 (0.92-1.41)	76	1.05 (0.81-1.34)	32	1.46 (0.97-2.20)	69	1.17 (0.90-1.53)	39	1.08 (0.76-1.54)
Baseline	714	ref	548	ref	166	ref	433	ref	281	ref
Wave 1										
1-7 days	81	6.01 (4.52-7.99)	66	6.13 (4.46-8.43)	15	5.48 (2.91-10.32)	46	5.39 (3.73-7.78)	35	7.15 (4.56-11.23)
8-14 days	20	1.79 (1.11-2.87)	15	1.70 (0.98-2.93)	5	2.09 (0.81-5.44)	9	1.23 (0.62-2.45)	11	2.83 (1.45-5.49)
15-28 days	26	1.36 (0.89-2.07)	20	1.34 (0.83-2.18)	6	1.40 (0.58-3.39)	16	1.28 (0.75-2.19)	10	1.50 (0.75-2.99)
29-91 days	52	0.72 (0.52-1.00)	36	0.65 (0.44-0.95)	16	0.99 (0.53-1.82)	33	0.69 (0.46-1.04)	19	0.78 (0.46-1.34)
Baseline	165	ref	126	ref	39	ref	102	ref	63	ref
Wave 2										
1-7 days	138	4.19 (3.35-5.23)	105	3.73 (2.90-4.79)	33	6.62 (4.05-10.83)	76	3.61 (2.69-4.84)	62	5.16 (3.66-7.26)
8-14 days	68	2.69 (2.03-3.57)	45	2.10 (1.50-2.95)	23	5.85 (3.39-10.08)	37	2.35 (1.61-3.42)	31	3.26 (2.13-5.00)
15-28 days	38	0.99 (0.69-1.41)	26	0.80 (0.52-1.22)	12	1.98 (1.00-3.92)	26	1.06 (0.68-1.64)	12	0.85 (0.46-1.59)
29-91 days	56	0.76 (0.55-1.05)	40	0.65 (0.45-0.95)	16	1.36 (0.71-2.63)	36	0.75 (0.50-1.12)	20	0.78 (0.46-1.32)
Baseline	549	ref	422	ref	127	ref	331	ref	218	ref

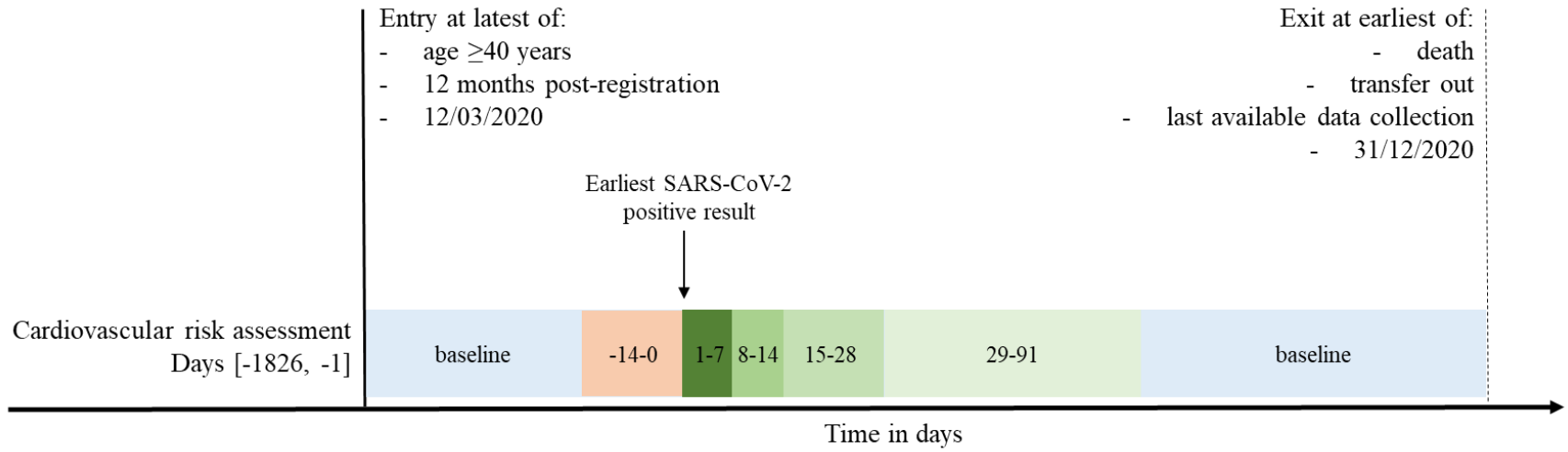


Figure 9.1 Study design with exposure and baseline periods

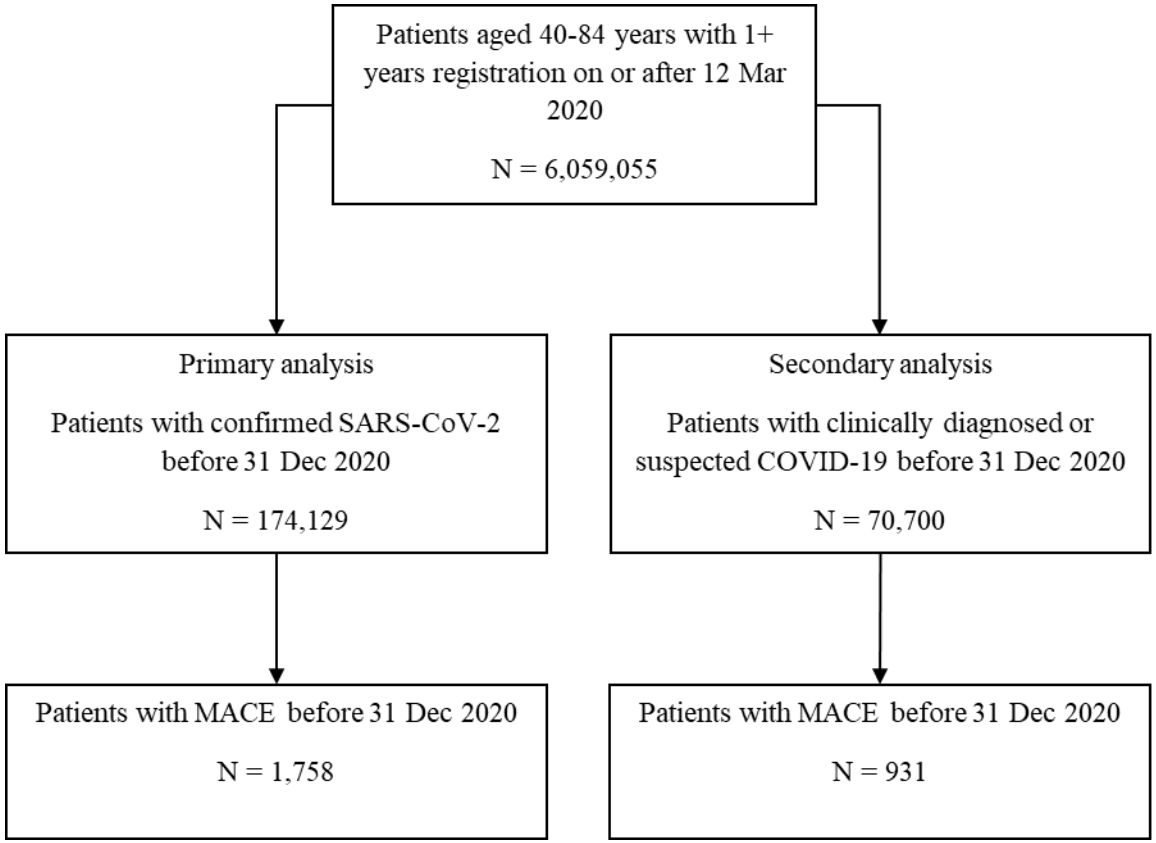


Figure 9.2 Study population flow chart

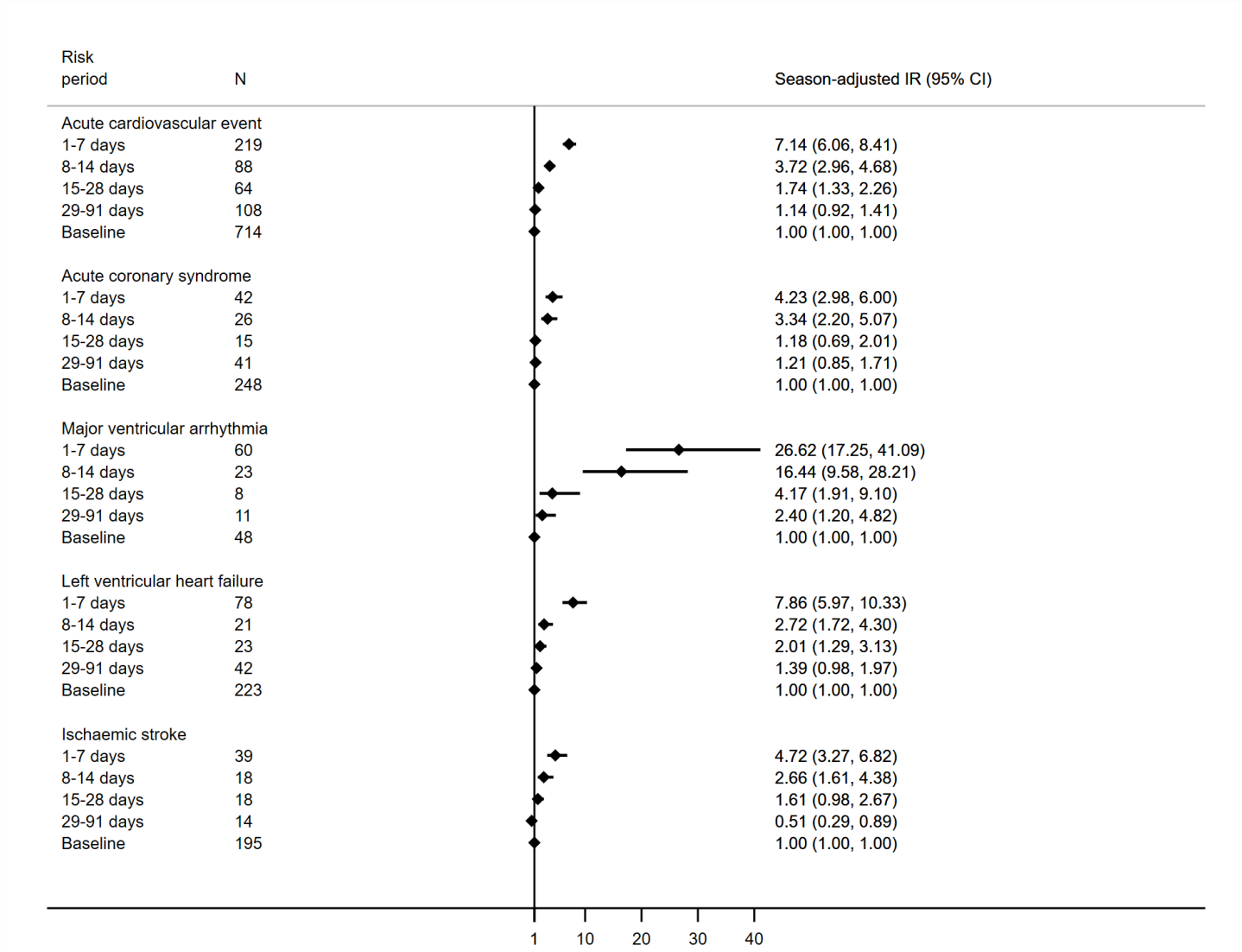


Figure 9.3 Incidence ratios for first acute cardiovascular events in risk periods following SARS-CoV-2 by cardiovascular event type

Table S9.1 Season-adjusted incidence ratios for acute cardiovascular events occurring in exposed periods after SARS-CoV-2 infection by cardiovascular risk, sex and age group

Risk period	All		QRISK3				Hypertension			
			Raised risk		Low risk		Raised risk		Low risk	
	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)
Men										
1-7 days	140	7.90 (6.42-9.72)	114	7.84 (6.23-9.86)	26	8.27 (5.08-13.47)	85	8.12 (6.23-10.58)	55	7.56 (5.41-10.56)
8-14 days	59	4.22 (3.18-5.61)	46	4.08 (2.96-5.62)	13	4.94 (2.66-9.17)	34	4.07 (2.80-5.90)	25	4.44 (2.86-6.90)
15-28 days	44	1.98 (1.43-2.73)	33	1.86 (1.29-2.69)	11	2.50 (1.29-4.85)	28	2.08 (1.39-3.12)	16	1.81 (1.07-3.09)
29-91 days	64	1.10 (0.83-1.45)	51	1.11 (0.82-1.52)	13	1.08 (0.58-2.01)	42	1.18 (0.84-1.67)	22	0.96 (0.60-1.53)
Baseline	429	ref	343	ref	86	ref	266	ref	163	ref
Women										
1-7 days	79	6.10 (4.67-7.97)	57	5.66 (4.13-7.76)	22	7.35 (4.42-12.20)	37	4.45 (3.04-6.53)	42	8.72 (5.97-12.75)
8-14 days	29	3.01 (2.03-4.46)	14	1.89 (1.09-3.29)	15	6.47 (3.62-11.57)	12	2.03 (1.11-3.68)	17	4.57 (2.69-7.75)
15-28 days	20	1.38 (0.86-2.19)	13	1.19 (0.67-2.11)	7	1.92 (0.86-4.25)	14	1.56 (0.89-2.72)	6	1.08 (0.47-2.49)
29-91 days	44	1.21 (0.87-1.69)	25	0.94 (0.61-1.45)	19	1.98 (1.15-3.43)	27	1.16 (0.76-1.79)	17	1.28 (0.75-2.19)
Baseline	285	ref	205	ref	80	ref	167	ref	118	ref
40-64 years										
1-7 days	105	9.61 (7.54-12.26)	59	11.01 (7.92-15.32)	46	8.27 (5.76-11.86)	53	9.56 (6.80-13.45)	52	9.62 (6.79-13.62)
8-14 days	48	5.37 (3.90-7.39)	22	5.05 (3.17-8.06)	26	5.67 (3.65-8.80)	22	4.84 (3.04-7.72)	26	5.90 (3.80-9.16)
15-28 days	30	2.09 (1.41-3.08)	13	1.86 (1.04-3.33)	17	2.30 (1.36-3.88)	19	2.53 (1.54-4.15)	11	1.60 (0.85-3.01)
29-91 days	61	1.51 (1.12-2.03)	31	1.53 (1.01-2.32)	30	1.49 (0.97-2.27)	41	1.88 (1.29-2.73)	20	1.08 (0.66-1.76)
Baseline	300	ref	145	ref	155	ref	155	ref	145	ref
65-84 years										
1-7 days	114	5.55 (4.44-6.94)	112	5.64 (4.50-7.07)	2	2.80 (0.58-13.57)	69	5.04 (3.79-6.69)	45	6.57 (4.57-9.45)
8-14 days	40	2.64 (1.89-3.69)	38	2.58 (1.83-3.64)	2	4.96 (0.98-25.05)	24	2.38 (1.55-3.65)	16	3.18 (1.86-5.44)
15-28 days	34	1.48 (1.03-2.13)	33	1.49 (1.03-2.15)	1	1.26 (0.15-10.66)	23	1.50 (0.97-2.33)	11	1.45 (0.77-2.72)
29-91 days	47	0.87 (0.63-1.19)	45	0.86 (0.62-1.18)	2	1.08 (0.21-5.59)	28	0.76 (0.51-1.14)	19	1.08 (0.66-1.80)
Baseline	414	ref	403	ref	11	ref	278	ref	136	ref

Table S9.2 Baseline characteristics of clinically reported COVID-19 study population

	All	QRISK3		Hypertension	
		Raised risk	Low risk	Raised risk	Low risk
	N=932	N=733	N=199	N=540	N=392
Sex					
Women	410 (44.0%)	297 (40.5%)	113 (56.8%)	221 (40.9%)	189 (48.2%)
Men	522 (56.0%)	436 (59.5%)	86 (43.2%)	319 (59.1%)	203 (51.8%)
Age group (years)					
40-54	168 (18.0%)	46 (6.3%)	122 (61.3%)	79 (14.6%)	89 (22.7%)
55-64	188 (20.2%)	120 (16.4%)	68 (34.2%)	97 (18.0%)	91 (23.2%)
65-74	282 (30.3%)	273 (37.2%)	9 (4.5%)	173 (32.0%)	109 (27.8%)
75-84	294 (31.5%)	294 (40.1%)	0 (0.0%)	191 (35.4%)	103 (26.3%)
COVID-19 associated hospital stay	150 (16.1%)	119 (16.2%)	31 (15.6%)	82 (15.2%)	68 (17.3%)
Median (IQR)	9 (4-16)	9 (4-16)	9 (4-18)	9 (4-16)	8 (4-17)
Died ≤30 days after event					
Died in study period	150 (16.1%)	128 (17.5%)	22 (11.1%)	82 (15.2%)	68 (17.3%)

Table S9.3 Season-adjusted incidence ratios for acute cardiovascular events occurring in exposed periods after clinically-reported COVID-19 by cardiovascular risk and COVID-19 wave

Risk period	All		QRISK3				Hypertension			
			Raised risk		Low risk		Raised risk		Low risk	
	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)
All										
1-7 days	115	10.04 (8.05-12.53)	90	10.10 (7.86-12.99)	25	9.96 (6.23-15.94)	73	12.40 (9.33-16.48)	42	7.42 (5.19-10.61)
8-14 days	49	5.05 (3.72-6.86)	36	4.85 (3.40-6.92)	13	5.78 (3.16-10.58)	31	6.01 (4.07-8.87)	18	3.96 (2.41-6.52)
15-28 days	56	3.24 (2.42-4.33)	45	3.43 (2.48-4.75)	11	2.66 (1.39-5.07)	34	3.67 (2.52-5.35)	22	2.74 (1.74-4.33)
29-91 days	114	1.76 (1.41-2.20)	84	1.70 (1.31-2.19)	30	1.99 (1.27-3.12)	64	1.85 (1.37-2.50)	50	1.65 (1.18-2.31)
Baseline	289	ref	226	ref	63	ref	153	ref	136	ref
Wave 1										
1-7 days	72	7.65 (5.69-10.28)	56	7.57 (5.42-10.59)	16	7.94 (4.22-14.92)	50	9.70 (6.70-14.04)	22	5.11 (3.08-8.46)
8-14 days	35	4.13 (2.82-6.03)	25	3.80 (2.44-5.92)	10	5.23 (2.50-10.92)	24	5.12 (3.20-8.18)	11	2.88 (1.50-5.55)
15-28 days	41	2.63 (1.84-3.76)	33	2.74 (1.84-4.09)	8	2.24 (1.01-4.98)	26	2.93 (1.86-4.62)	15	2.25 (1.26-4.01)
29-91 days	83	1.31 (0.99-1.73)	61	1.25 (0.91-1.72)	22	1.51 (0.86-2.66)	45	1.25 (0.86-1.81)	38	1.41 (0.93-2.14)
Baseline	191	ref	152	ref	39	ref	101	ref	90	ref
Wave 2										
1-7 days	43	11.20 (7.28-17.24)	34	10.69 (6.58-17.36)	9	13.92 (5.41-35.87)	23	10.29 (5.74-18.43)	20	12.38 (6.52-23.51)
8-14 days	14	4.89 (2.65-9.02)	11	4.70 (2.36-9.39)	3	5.88 (1.55-22.32)	7	3.85 (1.64-9.04)	7	6.51 (2.69-15.77)
15-28 days	15	3.13 (1.70-5.78)	12	3.11 (1.56-6.19)	3	3.35 (0.86-12.99)	8	2.75 (1.19-6.31)	7	3.67 (1.48-9.09)
29-91 days	31	2.29 (1.31-4.00)	23	2.00 (1.06-3.77)	8	3.88 (1.20-12.48)	19	2.40 (1.16-4.95)	12	2.13 (0.89-5.11)
Baseline	98	ref	74	ref	24	ref	52	ref	46	ref

Table S9.4 Season-adjusted incidence ratios for non-fatal acute cardiovascular events occurring in exposed periods after SARS-CoV-2 infection by cardiovascular risk

Risk period	All		QRISK3				Hypertension			
			Raised risk		Low risk		Raised risk		Low risk	
	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)
1-7 days	105	4.36 (3.52-5.41)	78	4.10 (3.20-5.26)	27	5.36 (3.48-8.25)	56	3.85 (2.88-5.14)	49	5.17 (3.74-7.13)
8-14 days	61	2.91 (2.22-3.80)	43	2.60 (1.89-3.58)	18	4.04 (2.44-6.71)	32	2.55 (1.76-3.68)	29	3.45 (2.32-5.13)
15-28 days	50	1.37 (1.02-1.84)	36	1.27 (0.90-1.79)	14	1.77 (1.00-3.11)	35	1.59 (1.12-2.27)	15	1.04 (0.61-1.76)
29-91 days	97	0.95 (0.76-1.19)	67	0.86 (0.66-1.12)	30	1.31 (0.86-1.99)	62	0.99 (0.75-1.31)	35	0.89 (0.62-1.29)
Baseline	681	ref	519	ref	162	ref	414	ref	267	ref

9.3. Additional methods

I further stratified results by infection severity. I defined severity by hospitalisation due to COVID-19, which was identified using the CHESS dataset or a primary diagnostic position code for COVID-19 in HES APC. I also conducted an analysis with QRISK3 scores ≥10% further stratified into 10-<20% and ≥20%.

9.4. Additional results

Results stratified by infection severity showed an extreme difference in the magnitude of the association between SARS-CoV-2 and first acute cardiovascular event (**Table 9.3**). Among individuals who were hospitalised, the relative incidence of first acute cardiovascular event was 20.99 (16.60-26.53) in the days 1-7 after SARS-CoV-2 identification compared to 2.63 (2.00-3.47) in those who were not hospitalised. Although it should be noted that there was overlap in the data sources used to identify infection (CHESS and SGSS) as those used to define severity (CHESS and HES APC).

Further stratification of QRISK3 score, showed the highest IR in people with a QRISK3 score 10-<20%, although CIs were overlapping between those with a QRISK3 score ≥20% and <10% (**Table 9.4**). A pattern of decreasing relative incidence as time from infection increased was identified across all QRISK3 groups.

Table 9.3 Season-adjusted incidence ratios for acute cardiovascular events occurring in exposed periods after SARS-CoV-2 infection by cardiovascular risk and infection severity

Risk period	All		QRISK3				Hypertension			
			Raised risk		Low risk		Raised risk		Low risk	
	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)
Severe (hospitalisation required)										
1-7 days	155	20.99 (16.60-26.53)	127	20.59 (15.90-26.67)	28	23.80 (13.58-41.72)	83	18.28 (13.46-24.81)	72	25.69 (17.70-37.30)
8-14 days	59	10.42 (7.64-14.20)	37	7.98 (5.48-11.62)	22	22.27 (12.29-40.33)	32	9.07 (6.02-13.67)	27	12.82 (7.94-20.71)
15-28 days	35	4.02 (2.76-5.85)	27	3.81 (2.49-5.83)	8	5.17 (2.29-11.68)	20	3.71 (2.27-6.07)	15	4.58 (2.54-8.25)
29-91 days	51	2.14 (1.54-2.97)	34	1.74 (1.18-2.58)	17	4.12 (2.17-7.82)	32	2.14 (1.42-3.24)	19	2.17 (1.26-3.73)
Baseline	161	ref	137	ref	24	ref	101	Ref	60	ref
Not severe (hospitalisation not required)										
1-7 days	64	2.63 (2.00-3.47)	44	2.23 (1.61-3.11)	20	4.21 (2.54-7.00)	39	2.64 (1.85-3.76)	25	2.63 (1.70-4.07)
8-14 days	29	1.58 (1.08-2.32)	23	1.57 (1.02-2.41)	6	1.59 (0.69-3.68)	14	1.27 (0.74-2.20)	15	2.05 (1.20-3.51)
15-28 days	29	1.00 (0.68-1.46)	19	0.83 (0.52-1.33)	10	1.60 (0.82-3.12)	22	1.23 (0.79-1.93)	7	0.62 (0.29-1.34)
29-91 days	57	0.79 (0.59-1.06)	42	0.77 (0.55-1.07)	15	0.89 (0.50-1.58)	37	0.83 (0.57-1.18)	20	0.73 (0.45-1.19)
Baseline	553	ref	411	ref	142	ref	332	Ref	221	ref

Table 9.4 Season-adjusted incidence ratios for acute cardiovascular events occurring in exposed periods after SARS-CoV-2 infection by further QRISK3 stratification

Risk period	≥20%		10-<20%		<10%	
	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)
1-7 days	104	5.99 (4.76-7.55)	59	9.67 (7.01-13.33)	46	8.11 (5.70-11.54)
8-14 days	32	2.31 (1.60-3.35)	31	6.22 (4.16-9.32)	26	5.29 (3.42-8.19)
15-28 days	26	1.25 (0.83-1.88)	18	2.29 (1.38-3.80)	18	2.26 (1.36-3.75)
29-91 days	42	0.82 (0.59-1.15)	36	1.69 (1.15-2.48)	32	1.49 (0.99-2.24)
Baseline	393	ref	155	ref	166	ref

9.5. Chapter summary

- In this population-based SCCS, I used CPRD Aurum data linked to HES APC, ONS death registrations, CHESS and SGSS to quantify the association between SARS-CoV-2 infection and first acute cardiovascular event, with stratification by cardiovascular risk.
- I included individuals who had SARS-CoV-2 infection and their first acute cardiovascular event between 12 March 2020 and 31 December 2020, covering wave one of the pandemic as well as the most severe part of wave two.
- My study included 1,762 individuals, the vast majority of whom had raised cardiovascular risk; 76.6% had a QRISK3 score ≥10% and 59.4% had hypertension.
- I observed a significant increase in the season-adjusted incidence of first acute cardiovascular event following SARS-CoV-2 infection; with an IR of 7.14 (6.06-8.41) in the first seven days after infection which fell to 1.14 (0.92-1.41) by days 29-91.
- When stratified by cardiovascular risk, I observed the same pattern with similar effect estimates in both raised and low cardiovascular risk profiles, although the increase in relative incidence was slightly higher in those at low cardiovascular risk. This result was likely due to the higher overall risk of cardiovascular complications in people with raised cardiovascular risk, with the potential for an event at any time and not only following COVID-19.
- The relative incidence of a first acute cardiovascular event after SARS-CoV-2 was higher in wave one, than wave two, I hypothesise this was due the increase in treatment options available by wave two as well as the swift and unprepared for initial onset of the pandemic in March 2020.
- I also identified differences in the increased relative incidence and length of time increased relative incidence was observed by specific cardiovascular event; in the first seven days after SARS-CoV-2, the relative incidence was highest for major ventricular arrhythmia (26.62 [17.25-41.09]), although the number of events was small, followed by left ventricular heart failure (7.86 [5.97-10.33]), ischaemic stroke (4.72 [3.27-6.82]), and finally ACS (4.23 [2.98-6.00]). Increased

relative incidence was observed for the full risk period of 91 days for major ventricular arrhythmia, up to 28 days for left ventricular heart failure, and up to 15 days for ischaemic stroke and ACS.

- The results I generated emphasise the importance of sustaining and prioritizing cardiovascular risk prevention and management, along with that of COVID-19 itself to reduce cardiovascular complications.

Chapter 10 Discussion

10.1. Chapter overview

This thesis explored the association between ARIs, originally including influenza and pneumonia, and then latterly COVID-19, and first acute cardiovascular events. In addition, the effect of vaccination, particularly influenza vaccination, on such cardiovascular complications was investigated using population-based linked EHRs from primary and secondary care, as well as in COVID-19-related analysis laboratory reporting surveillance data. In this concluding chapter, I have summarised the thesis' main findings with a comparison to previous literature. Strengths and limitations, as well as clinical implications, focusing on overarching points that connect the individual research objectives are also captured. I conclude by setting out future research directions and considerations.

10.2. Summary of key research findings

Research objective 1: to assess the validity of acute cardiovascular diagnoses in routinely collected European EHRs

In my systematic review (presented in **Chapter 5**), I identified 81 studies which validated stroke (41 studies), ACS (31 studies including 29 studies specifically validating MI) or heart failure (20 studies) diagnoses in European EHRs. The diagnoses were mainly validated using secondary care EHRs with coded data for ICD-8, -9, and -10. Overall, sensitivity was ≥80% for 91% of MI studies, and ≥70% for 73% of stroke studies, but ≤66% in all but one heart failure study. PPV for all outcomes was ≥80% in the majority of studies; 74% of heart failure studies, 88% of MI studies, and 70% of stroke studies. Few studies reported specificity or NPV. There was substantial heterogeneity between the sensitivity and PPV estimates for all three outcomes. Heterogeneity was likely due to the various ways in which the included studies differed: different reference standards and different study time periods impacted upon the ICD

version used, different ICD codes included, even when the same ICD version was employed, and different study population selection. Having identified variability in definitions and accuracy of coded data, I included a broad definition of acute cardiovascular events in my PhD outcome definition which covered key diagnostic codes for which validation studies were available, with a primary composite outcome and secondary individual event outcomes. In non-COVID-19 analysis, the outcomes included in the composite acute cardiovascular events were ACS (MI and unstable angina), acute left ventricular heart failure, acute cerebrovascular events (stroke and TIA), and acute limb ischaemia. These events are acute and vascular, sharing similar underlying pathologies, due usually to atherosclerosis, and risk factors [281]. In COVID-19 analysis, the outcome included ACS, heart failure, ischaemic stroke, and major ventricular arrhythmia based on initial evidence suggesting these events were linked to COVID-19 [254,282].

Research objective 2: to estimate the effect of cardiovascular risk on systemic ARI, acute cardiovascular events, and acute cardiovascular events after systemic ARI

In **Chapter 6** I presented a cohort study as well as further analysis which investigated the association between cardiovascular risk and ARIs, acute cardiovascular events, and ARI-related cardiovascular events. Using CPRD GOLD or Aurum datasets with linked HES and ONS death registration data, I identified 4,212,930 individuals aged 40-64 years without established CVD or a health condition included in influenza vaccine recommendations. When stratified by cardiovascular risk level (diagnosed hypertension status or QRISK2 score), I found a marginal increased incidence of ARI among individuals with hypertension (aIRR 1.04 [95% CI 1.03-1.05]) and a more substantial increased incidence in those with a QRISK2 score ≥10% (1.39 [1.37-1.40]). Results by coded infection type differed. There was a higher, increase in the incidence of pneumonia among individuals with raised cardiovascular risk compared to ARI overall. This increase was more pronounced when cardiovascular risk was measured by QRISK2 score ≥10% (2.32 [2.25-2.40]) than by hypertension (1.12 [1.07-1.16]). However, there was a

reduced incidence of influenza/ILI among individuals with hypertension (0.98 [0.96-1.00]) and with a QRISK2 score ≥10% (0.88 [0.86-0.90]).

As expected, there was an association between cardiovascular risk and first acute cardiovascular event outcome for both hypertension (aIRR 1.78, [1.74-1.82]) and QRISK2 ≥10% (4.83 [4.74-4.92]), although substantially higher for QRISK2 score.

Among the 442,408 individuals with an ARI, which included 586,147 ARI episodes, I identified a significant association between raised cardiovascular risk and ARI-related cardiovascular complication. As with ARI incidence itself, the association was greater when cardiovascular risk was measured by QRISK2 score (aHR 3.65 [3.42-3.89]) than hypertension (1.98 [1.83-2.15]). Associations followed a similar pattern for pneumonia (QRISK2: 2.13 [1.92-2.37] and hypertension: 1.65 [1.44-1.89]) and influenza/ILI (QRISK2: 3.35 [2.70-4.17] and hypertension: 2·07 [1·60-2·67]). Furthermore, associations between raised cardiovascular risk and ARI-related cardiovascular events were consistent, although slightly higher after excluding individuals who received influenza or pneumococcal vaccines in follow-up (QRISK2: 3.96 [3.70-4.22] and hypertension: 2.07 [1.90-2.24]). Risks of ARI-related cardiovascular complications were also higher among those not on antihypertensive or statin treatment, when compared to those in receipt of treatment.

Research objective 3: to investigate whether influenza vaccine reduces the risk of acute cardiovascular events, and if effect differs between individuals with raised and low cardiovascular risk

I then investigated the association between influenza vaccine and first acute cardiovascular event using a SCCS design containing 193,900 individuals aged 40-84 years (**Chapter 7**). In the analysis, I found a decrease in the season-adjusted relative incidence of first acute cardiovascular events occurring in the days and weeks after influenza vaccination. Acute cardiovascular event risk was reduced in the 15-28 days after vaccination, which is the first time period in which an adequate immune response to the vaccine is likely to be present, (IR 0.72 [0.70-0.74]) and, while the effect size tapered, remained reduced

to 91-120 days after vaccination (0.83 [0.81-0.88]). The protective association was evident across all age groups and cardiovascular risk profiles in the main analyses, with follow-up from 1 September meaning unvaccinated time was included for before and after vaccine risk time. However, the protective effect was confined to those ≥65 years in a sensitivity analysis when follow-up started from vaccination date.

Research objective 4: to investigate the effect of cardiovascular risk on severe outcomes, including acute cardiovascular events after COVID-19

Following analysis of ARI-related cardiovascular complications and the association between influenza vaccine and acute cardiovascular events, my remaining analysis focused on COVID-19. In **Chapter 8**, I presented a cohort study analysis of 6,059,055 adults aged 40-84 years, I showed the incidence of severe outcomes i.e., COVID-19 death, COVID-19 related ICU admission, use of respiratory support, hospitalisation due to COVID-19, and acute cardiovascular event were higher among those with raised cardiovascular risk (measured separately by QRISK3 score or hypertension) with and without prior documented SARS-CoV-2 infection. However, the risk of SARS-CoV-2 itself was not higher among those with raised cardiovascular risk. I then found, among 146,760 individuals without established CVD with SARS-CoV-2, that the risks of COVID-19 death (aHR 8.77 [7.62-10.10]), ICU admission (3.66 [3.18-4.21]), respiratory support (3.73 [3.10-4.49]), hospitalisation (3.38 [3.22-3.56]) and first acute cardiovascular event (5.43 [4.44-6.64]) were greater in those with a QRISK3 score ≥10%, compared to those with a QRISK3 score <10%. When cardiovascular risk was measured by hypertension, difference in outcome risk after infection were only evident for acute cardiovascular event (1.49 [1.20-1.85]). Analysis by pandemic waves revealed similar patterns, although the incidence of severe outcomes was greatest during the first wave.

Research objective 5: to quantify the relative incidence of acute cardiovascular events occurring in periods after COVID-19 to other periods

Finally, to corroborate cohort study results, I conducted a SCCS, presented in **Chapter 9**, investigating the association between SARS-CoV-2 and first acute cardiovascular event. My analysis of 1,762 individuals aged 40-84 years identified that the risk of cardiovascular event was elevated in the month after SARS-CoV-2 infection. The largest increase in risk occurred within the first seven days after infection (IR 7.14 [95% CI 6.06-8.41]) and tapered over time (15-28 days after infection, 1.74 [1.33-2.26]). When I stratified results by cardiovascular risk, I observed the same pattern with similar effect estimates for individuals classified as raised (QRISK3 or hypertension) and low cardiovascular risk. The relative incidence of a first acute cardiovascular event after SARS-CoV-2 was higher in wave one of the pandemic than wave two. I also identified differences in the increased relative incidence and length of time increased relative incidence was observed by specific cardiovascular event; with risk higher and more sustained for in major ventricular arrhythmia and left ventricular heart failure than ischaemic stroke and ACS.

10.3. Explanation of research findings with comparison to previous literature

10.3.1. Raised cardiovascular risk and respiratory infections

While the primary focus of my thesis research was to investigate the association between cardiovascular risk on ARI-related cardiovascular complications, I first wanted to consider the impact of cardiovascular risk on the likelihood of developing infections. Identifying whether cardiovascular risk increased the likelihood of infection, or the likelihood of complications after infection, or both could lead to different intervention strategies. For instances, some interventions such as vaccines address both pathways whereas improved infection treatment would only deal with infection-related complications. This scope was set prior to the onset of the COVID-19 pandemic, after which the topic became even more relevant.

Based on literature review, few studies have examined the effect of cardiovascular risk on ARI, including COVID-19. Instead, most studies have focused on the outcomes associated with cardiovascular risk factors and ARI. However, a UK Biobank study presented estimates for the effect of blood pressure on multiple ARI risks [283]. Biobank participants with high blood pressure had an increased risk of pneumonia (aHR 1.36 [1.29-1.43]), influenza or viral pneumonia (1.12 [1.01-1.23]), and other lower respiratory infections (1.15 [1.11-1.19]).

In our study, the increased incidence of ARI among individuals with hypertension, measured by recorded diagnoses rather than blood pressure readings, was smaller than observed in the UK Biobank study. Instead, I identified more substantial increases in ARI, and pneumonia, incidence among individuals with a QRISK2 score of 10% or higher. Although I identified a decreased incidence of influenza/ILI among those with a high QRISK2 score. QRISK2 is a composite risk score which considers systolic blood pressure reading and use of antihypertensive drugs, as well as capturing cardiovascular risk beyond hypertension. The inclusion of multiple comorbid conditions in QRISK2 score likely contributes to the greater ARI risk among those with a high QRISK2 score, with an individual's score increased due to the presence of comorbid conditions. A single site study in the USA among adults vaccinated against influenza during seasons 2013/14 and 2014/15, reported an increased risk of influenza/ILI among those classified as obese compared to non-obese (RR 2.01 [1.12-3.60]) despite no difference in seroconversion or seroprotection between the groups [284].

While the mechanism will differ, i.e., sociodemographic and socioeconomic rather than biological, other cardiovascular risk factors such as ethnicity and deprivation have also been associated with increased incidence of influenza/ILI. These sociodemographic and socioeconomic mechanisms likely include household overcrowding, use of public transport, and increased mixing through public-facing jobs. In an analysis I conducted using CPRD GOLD and Aurum data, I identified increased ILI incidence among individuals of non-white ethnicity [258]. While another UK study using the Flu Watch cohort found individuals from the most socially deprived populations had a greater risk of influenza/ILI compared to

more affluent populations [259]. Both ethnicity and deprivation are risk factors included in the QRISK2 algorithm and can result in higher QRISK2 scores assigned to individuals with these characteristics.

Conversely, in my analysis for thesis objective 4, I identified individuals with raised cardiovascular risk, QRISK3 score ≥10% or hypertension, to have a lower incidence of laboratory-confirmed SARS-CoV-2. When I employed a clinically reported definition of COVID-19, as was done for my analysis of ARI, those with raised cardiovascular risk did have a higher infection incidence.

The reasons for differences between ARI, influenza/ILI, pneumonia, and COVID-19 incidence between individuals with raised cardiovascular risk are unclear, particularly for ARI and pneumonia the higher effect estimate for those with QRISK2/3 ≥10% being higher than hypertension. While the recording of risk factors included in the QRISK2/3 score as well as infections are more likely in those with a higher rate of primary care attendance, my analysis of ARI and influenza/ILI adjusted for baseline consultation frequency. Having a raised QRISK2/3 score might be a signal that someone presents more frequently than those with recorded hypertension, although those patients would still regularly collect blood pressure medication.

An under-estimation of the incidence of some ARI, particularly ILI, is likely to occur because the infection is short-lived with mild and self-limiting symptoms which individuals will manage at home with over-the-counter treatment, where needed. Conversely, due to its severity or possible development as an end stage complication of an initial ILI, pneumonia will usually result in healthcare attendance. Furthermore, pneumonia occurs in people with lower physiological reserve due to more comorbidities so it is more plausibly linked to the presence of underlying cardiovascular risk factors, whereas ILI or COVID-19 occur at a greater level in the general population. COVID-19 presentation and healthcare interactions are more difficult to interpret. Severe infection will lead to healthcare attendance, but due to the pandemic many will not have informed or presented to healthcare services when illness occurred.

Prior to COVID-19, few individuals with an ARIs who presented to primary care were laboratory-confirmed, and even those hospitalised due to ARI frequently did not have the causative organism identified. However, ARI testing practices during the COVID-19 pandemic have been unprecedented, enabling the identification of laboratory-confirmed asymptomatic as well as mild, moderate and severe laboratory-confirmed ARIs at scale. The differences in presentation and EHR data capture could account for the differing incidence of each infection among individuals with increased cardiovascular risk.

10.3.2. Raised cardiovascular risk and infection associated cardiovascular complications

Across all analysis I conducted, I identified an increase in infection associated cardiovascular complications among individuals at raised cardiovascular risk. The experience of further cardiovascular complications among individuals with established CVD after a severe infection is well established [160–162]. Preventing these, and other complications, is why influenza and pneumococcal vaccines are recommended for individuals with established CVD.

The onset of the COVID-19 pandemic further added to the evidence base with regards to the likelihood of severe infection-related outcomes among those with established CVD and cardiovascular risk factors [285,286]. Prior to the COVID-19 pandemic, there was limited existing evidence with regards to the role of underlying cardiovascular risk in the onset of first acute cardiovascular events. In two studies, hypertension had been found associated with an increase in cardiovascular complications after pneumonia [287,288]. In another study, there was no difference in the likelihood of first MI or stroke occurring after ARI when stratified by the presence of cardiovascular risk factors [289]. The difference in the result of this latter study may be explained by the study design and definition of cardiovascular risk. This was a case-control study which took a data-driven approach to defining cardiovascular risk based on the risk factors which were significantly associated with MI or stroke in univariable analysis, with the resulting multivariable analysis stratified by the number of risk factors.

The association between ARIs and acute cardiovascular events differs by the cardiovascular event type. The risk of specific events different, for instance my analysis of ARI-related cardiovascular events identified a higher risk of ACS among those with raised QRISK2 score than stroke. Furthermore, previous evidence suggests the MI risk after ARI is shorter than that of stroke risk. A SCCS of Scottish data by Warren-Gash *et al* found the relative incidence of stroke after ARI was higher than that of MI and also the association with stroke remained for up to one month after ARI compared to 7-14 days for MI, depending on the causative agent [165]. While the results of my COVID-19 SCCS study were suggestive of a similar pattern, the number of events were small numbers, so until further analysis is conducted the results generated by from my study are only speculative.

Several mechanisms likely contribute to the ARI-triggered acute cardiovascular events. Infections can lead to systemic inflammation and a range of haemodynamic and haemostatic effects. Such processes along with pro-inflammatory cytokine release in response to an infection can mediate atherosclerosis or facilitate plaque rupture. During severe infection, organisms such as the influenza virus and *S. pneumoniae* can invade the myocardium inducing cardiac injury and scarring [170,171]. SARS-CoV-2 can enter host cells by binding to, and then downregulate, the ACE2 receptor. The downregulation may lead to vasoconstriction and inflammation in the renin-angiotensin system, which is central to controlling blood pressure [102,290]. But the differences in risk association by event indicates variation in the specific mechanism responsible for the risk of stroke after ARI compared to the risk of MI. Previous research suggests the pathophysiology and epidemiology differs between cardiac and cerebrovascular events [281].

The COVID-19 pandemic has also accelerated evidence generation with regards to the role of cardiovascular risk factors, in the absence of established CVD, and severe outcomes post-infection [291]. Firstly, studies have generally shown a high prevalence of cardiovascular risk factors in people who were hospitalised or who died due to COVID-19 [292] Multiple studies have demonstrated an association between hypertension, as well as other cardiovascular risk factors such as obesity, diabetes and CKD, and

severe COVID-19 outcomes including mortality, respiratory support, and hospitalisation [293]. While other studies have reported no association [294][294]. The majority of the evidence base was generated from rapid single hospital-based analyses with limited confounder control conducted at the start of the pandemic. However, where large population-based studies were conducted an increased risk of severe outcomes has been identified, such as death among diabetics [295]. Pooled results, from all possible studies, suggest that the risk of severe COVID-19 outcomes is 2-3 times higher among hospitalised individuals with hypertension or diabetes than those without the conditions [285]. However, age is widely hypothesized to drive these associations, especially for hypertension, variable results obtained from age-adjusted analyses [285,294,296]. Analysis of 17 million patients using OpenSafely data showed hypertension was associated with a small increase in in-hospital COVID-19 death after age and sex adjustment, but not after full adjustment [286].

While the relationship between individual risk factors, including cardiovascular risk factors as outlined above, has been well researched, the effect of combined risk factors has been less of a focus. QCOVID, has been developed to predict the risks of death and hospital admission due to COVID-19 [297]. The algorithm includes numerous risk factors, such as CVD and some cardiovascular risk factors. However, no studies appear to have assessed the specific role of underlying cardiovascular risk in severe COVID-19 outcomes, measured using an algorithmic approach such as QRISK3, which captures the full spectrum of conditions associated with raised cardiovascular risk rather than analysis of individual risk factors.

10.3.3. Vaccines and first acute cardiovascular events

The existence of ARI-related cardiovascular complications suggests that reducing ARIs through prevention methods such as vaccines could be used for primary CVD prevention. There is a wealth of evidence examining the effect of influenza vaccine on secondary cardiovascular complications [172,173,298]. Meta-analyses of secondary prevention RCTs among people with IHD found significant

reductions in cardiovascular mortality (55%) [173] and cardiovascular complications (36%) [172] following influenza vaccination. Recent RCTs continue to evaluate the cardiovascular benefits of influenza vaccination among individuals with CVD [175,177,299,300]. Trial results suggest the cardiovascular protection provided by influenza vaccine is comparable to, and in some instances greater, than findings generated for other secondary CVD prevention strategies, including percutaneous coronary intervention or treatment with β-blockers, statins, or ACE inhibitors [301]. However, there are no RCTs which have examined the role of influenza, or other respiratory, vaccination as a primary CVD prevention strategy.

The findings of my main SCCS analysis of the association between influenza vaccine and first acute cardiovascular event were consistent with the results generated by previous SCCS studies. Analysis of data from 1987 to 2001 found a 12% and 13% reduction in the relative incidence of first stroke and MI, respectively, in the 15-28 days after influenza vaccination after which time there was no significant reduction [163]. Two other SCCS studies using data from 2001 to 2009 found 25% and 24% reductions for first MI and stroke, respectively, in the 15-28 days post-vaccination [190,191]. Although we used a composite acute cardiovascular event outcome, when we looked at individual cardiovascular outcomes, the largest reduction in relative incidence was for MI.

Another study has previously considered the impact of underlying cardiovascular risk on the association between influenza vaccine, in this case the H1N1 strain from 2009/10, and cardiovascular events [192]. The study identified conflicting results, with a reduced relative incidence of MI (15-28 days post-vaccination: IR 0.70 [0.57-0.85]) in those with raised cardiovascular risk and an increase (15-28 days post-vaccination: 3.17 [1.99-5.07]) among people at low cardiovascular risk. The study defined cardiovascular risk using cardiovascular prevention prescriptions at the time of vaccination, after follow-up had started, which likely biased results when stratified by cardiovascular risk so comparisons to our results is difficult.

With the rapid development and roll-out of COVID-19 vaccines, RCTs focused on the effect the vaccine had on COVID-19 hospitalisation and death [139]. Future RCTs with cardiovascular or chronic endpoints would be beneficial, particularly in the context of long COVID. To date, observational studies investigating the association between COVID-19 vaccine and cardiovascular complications have tested the hypothesis of increased risk associated with the vaccine, following initial concerns of thrombosis [302]. Multiple SCCS analysis have been conducted to further test these concerns. One large UK-based EHR SCCS found increased risk venous thromboembolism (1.10 [1.02-1.18]), cerebral venous sinus thrombosis (4.01 [2.08-7.71]) after Oxford-AstraZeneca vaccination and arterial thromboembolism (1.06 [1.01 to 1.10] at 15-21 days), cerebral venous sinus thrombosis (3.58 [1.39-9.27]), ischaemic stroke (1.12 [1.04-1.20]) after Pfizer-BioNTech vaccination [303]. The increased risks identified after vaccination were less than those after SARS-CoV-2 infection, with risk of stroke and MI several fold higher after infection. For instance, the risk of venous thromboembolism was 12-fold higher after infection than vaccination. A French EHR SCCS found no difference in the relative incidence of MI, stroke, or pulmonary embolism after COVID-19 vaccination compared to unvaccinated time [304].

10.4. Research strengths and limitations

10.4.1. Data sources

The CPRD datasets collect a breadth of information with regards to demographics, diagnoses, prescriptions, tests, and immunisations, thereby allowing for multiple confounders and effect modifiers to be accounted for my analysis. CPRD datasets, Aurum in particular, provide large study populations to allow adequate statistical power and the derivation of precise effect estimates. The datasets are representative of the national population, in terms of age, sex and ethnicity, and therefore results generated are generalisable. Over time CPRD GOLD has reduced in size due to the growth of CPRD Aurum. At the time of initiating my thesis analysis, CPRD Aurum was new and had not been widely used

in research. Therefore, to substantiate the use of CPRD Aurum in research I conducted my first thesis analysis using both CPRD GOLD, widely used in research for many years, and Aurum datasets and compared results. As the results I obtained were consistent between the two databases, I then proceeded to complete my remaining analysis using only CPRD Aurum.

An additional advantage of CPRD, is the ability to routinely link to other data sources such as HES APC and ONS death registrations. Furthermore, as part of a COVID-19 rapid research generation initiation, CPRD data could also be linked to CHESS and SGSS. This allows exposure, outcome and covariate identification from multiple sources. For example, I used CPRD and HES APC to identify ARI, acute cardiovascular events and ethnicity, and I used CPRD, HES APC, CHESS and SGSS to identify COVID-19. In summary, this assisted me in the misclassification of key variables. Furthermore, the use of multiple data sources permitted the use of sensitivity analysis definitions, i.e., laboratory-confirmed SARS-CoV-2 versus clinically reported COVID-19. While using CPRD linked data has many advantages, data were not originally collected for research. This can lead to a number of limitations, those of which are relevant to my thesis analysis are outlined in below sections.

10.4.2. Ascertainment of cardiovascular risk

In analysing the effect of cardiovascular risk, the use of a score-based measure such as QRISK2/3 has the advantage of incorporates multiple dimensions of cardiovascular risk to consider overall risk. QRISK2 was validated after development and subsequently widely used in primary care as well as cardiovascular prevention treatment recommendations, including the preventive use of statins. However, disadvantages of using QRISK2/3 in research include the reliance on accurate diagnostic coding. While the ascertainment of any variables using EHR data has this requirement, when using an algorithm based on the recording of multiple variables, such as QRISK2/3, incomplete data can have a large impact. Furthermore, the LSHTM QRISK2/3 algorithms are currently based on emulation of the original

algorithm developed by the QRISK developers, therefore, the code lists used may not have been as expansive as those employed to detect any coding of a condition, with many based on QOF codes. Since QRISK2/3 scores are assigned on the presence of coded diagnoses and measurements, the absence of codes will therefore result in a lower QRISK2/3 score and subsequently an underestimate of any association between raised cardiovascular risk, when measured by QRISK2/3 score, and outcomes of interest, such as ARI-related cardiovascular complications. For example, analysis of the validity of diabetes coding in CPRD suggested that code selection was important in determination of diabetes incidence [305].

In each of the analyses I conducted for this thesis, QRISK2/3 score provided the best marker of cardiovascular risk with more marked results obtained when compared to hypertension. However, results between the two measures, QRISK2/3 and hypertension, were consistent whether in relation to positive or negative associations of cardiovascular risk.

10.4.3. Ascertainment of ARIs

I used both CPRD and HES APC to identify ARIs, and further used laboratory surveillance data in the identification of COVID-19. Misclassification of ARI, influenza/ILI, pneumonia and COVID-19 status may have occurred, with individuals missed from inclusion in cohort and SCCS analyses. There may also have been misclassification across coded infection type due to diagnostic uncertainty and lack of availability of testing in primary care. For instance, whether an ILI is actually pneumonia is hard to tell without a chest x-ray. Identification an organisms based on laboratory-testing was (pre-COVID-19) limited for straightforward ILIs. Individuals with mild infections may have been less likely to seek primary care services, particularly when illness is short-lived. However, mild illness suggests less likelihood of systemic illness and, therefore, severe outcomes. There have been no studies which have validated the recording of ARIs in EHR data.

10.4.4. Ascertainment of influenza vaccine

The high uptake of influenza vaccine identified using primary care records among the population aged ≥65 years further suggests low vaccine provision in other settings [131]. Due to low vaccine uptake in clinical risk groups aged <65 years, it is difficult to determine if other routes such as occupational health and pharmacies are more widely used. However, primary care service providers are required to document vaccines delivered outside of their surgery for patients who are in an influenza risk group. Most patients in England receive their influenza vaccine at their primary care practice, but some vaccines will be administrated by pharmacies or occupational health services. Vaccines received outside of primary care practice are still expected to be recorded within primary care records. Furthermore, both the overall uptake and the patterns of regional variation in primary care vaccine data are consistent with national surveillance [131]. Like ARIs, no studies have formally validated the recording of vaccines in EHR data.

10.4.5. Ascertainment of acute cardiovascular events

The majority of acute cardiovascular events, such as MI and stroke will result in hospital admission and therefore recorded in HES APC. Validation studies, the focus of my systematic review as outlined in **Chapter 5**, have shown accurate recording of MI and stroke in secondary care EHRs, including HES APC. With longitudinal primary care data such as CPRD, serious events such as acute cardiovascular events are likely to also be recorded in primary care records. A validation study of MI recording in CPRD, HES APC, and an MI registry has highlighted this [243]. There is also potential for misclassification in the ascertainment of acute cardiovascular events in EHR data. Onset of new heart failure may be identified and managed through secondary care outpatient care, a dataset not utilized in this thesis, due to its poor employment of coded diagnoses [306]. Similarly unstable angina and TIA may not result in inpatient care, and, additionally, have less clear diagnostic pathways given their clinical presentation so less likely to be accurately identified in EHRs. Although, it should be noted that

secondary care diagnoses subsequently recorded in primary care records may be subject to delay. Validation of MI, heart failure and stroke diagnoses have focused on the recording of events rather than the timeliness of event recording.

10.4.6. Study designs

To investigate the association between cardiovascular risk and ARI-related cardiovascular complications, I used a cohort study which included multiple sensitivity analysis, two measures of cardiovascular risk to allow descriptive comparison of results, and multivariable adjusted analysis. In analysis of raised cardiovascular risk and COVD-19 severe outcomes, two study designs were utilized to triangulate results; a cohort study employing two definitions of COVID-19 with multiple secondary outcomes and a SCCS analysis.

SCCS, used both for COVID-19 and influenza vaccine association with cardiovascular complications, uses within-person comparisons, i.e., individuals act as their own controls during different time periods with only individuals with the exposure and outcome included [307]. The main advantage of the SCCS design is the removal of confounding due to fixed characteristics, recorded or not, that vary between individuals. In observational vaccine effectiveness studies, it is vital to remove confounding: vaccinated and unvaccinated individuals may have health, lifestyle and behavioral differences that are difficult to ascertain in routinely collected data [270]. However, healthcare contact bias cannot be accounted for in SCCS. When an individual receives their influenza vaccine this may trigger cardiovascular management, leading to a reduced incidence of cardiovascular complications including the time periods immediately following vaccination.

10.4.7. Residual confounding

Unmeasured confounding is a limiting factor in all secondary use data analysis, such as those which use EHRs. However, I attempted to limit confounding by using linking data, with multiple data sources used for key variable ascertainment, and robust study designs. Additionally, I used strict study population definitions to minimize confounding. For example, in all studies those with established CVD were excluded and in analysis of ARIs I additionally excluded individuals with an indication for influenza vaccination. There are still potential confounders and biases which I was unable to address in my analysis. Some variables are either not measured (such as genetic risk profiles) or are sub-optimally recorded (such as smoking status and BMI).

There is some debate regarding whether influenza vaccine reduces acute cardiovascular events as much as observational studies suggest. Observational studies, including my own, have shown that influenza vaccination is associated with a remarkably large reduction in one's risk of cardiovascular events. The large degree of protection against these outcomes observed in individual level studies have led to a suspicion that uncontrolled confounding and selection bias have exaggerated the effect of influenza vaccination on acute cardiovascular events.

10.4.8. Selection bias

Due to the nature of a SCCS study, only individuals with both the exposure and outcome contribute to the calculation of the effect estimate. Individuals who experience an acute cardiovascular event are likely have some degree of underlying risk regardless of diagnosed risk factors. Therefore, it is unlikely that the SCCS study included individuals with true low cardiovascular risk. The similar effect estimates obtained for both individuals at raised and low cardiovascular risk, in both influenza vaccine and SARS-CoV-2 studies, suggest this as a possibility. A cohort or case control study may, therefore, be more appropriate study designs to investigate the association between underlying cardiovascular risk but they themselves

present difficulties to investigate vaccine effectiveness when using routinely collected EHR data, as outlined in Chapter 7.

10.5. Clinical and public health implications

Results across all of the analyses in my thesis emphasise the importance of improved cardiovascular risk management, which in turn could reduce the incidence of ARIs and their cardiovascular consequences. The COVID-19 pandemic bought this topic to widespread attention. Preventive cardiovascular treatments, particularly among individuals with a QRISK2/3 score of ≥10%, could reduce ARI-related complications, cardiovascular and otherwise.

Typically, individuals with increased cardiovascular risk do not receive seasonal influenza or pneumococcal vaccines. Extending influenza vaccination to individuals with a QRISK3 score of ≥10% would include approximately 150,000 individuals in England (based on the study population included in my analysis for thesis objective 2, extrapolated to the national population). Vaccinating these individuals could reduce cardiovascular complications and potentially decrease the incidence of influenza and associated complications, such as hospitalisation and mortality. However, among adults aged <65 years, some with high cardiovascular risk are already eligible to receive influenza vaccine, for example those with CKD, obesity, or diabetes. However, uptake among clinical risk groups is currently moderate (<50%) in England, like other European countries, despite high uptake (>75%) in individuals aged ≥65 years [131]. Low uptake in <65 years may be due to individual or physician low perceived risk, as well as poor access to healthcare, or vaccine hesitancy. In particular, individuals with the onset of a chronic condition at a younger age links to social deprivation and access to care. Age eligible vaccination is operationally easier to manage compared to risk group eligibility. During the first two winters of the COVID-19 pandemic, influenza vaccine recommendations in England were temporarily extended to include all individuals ≥50 years regardless of underlying health conditions. Maintenance of such policy

expansion, which is again planned for 2022/23 winter [308], would help capture most individuals at raised cardiovascular risk. While further studies would help to fully characterize those who would derive the most cardiovascular benefit from influenza vaccine, improving uptake remains a public health priority, both to protect individuals from influenza and its complications, including cardiovascular events. It is necessary to emphasise to both patients and clinicians, the role of influenza, and other respiratory, vaccines in promotion of cardiovascular fitness. Such promotion of influenza vaccine would be a step towards a "syndemic" approach to healthcare, acknowledging the interaction of infectious diseases and non-communicable diseases, such as CVD. In particular, it is important that not only influenza vaccine recommendations given consideration to CVD, but that CVD guidance include influenza vaccine as a prevention strategy to migrate further cardiovascular events such as been done in the 2021 European Society of Cardiology [309]. Consideration of vaccine hesitancy is also important. This topic has been amplified during the COVID-19 pandemic. Healthcare professionals should work to alleviate patient concerns.

10.6. Further research directions

More observational studies focused on the role of cardiovascular risk, in the absence of established CVD, would bolster understanding on its effect on ARI-related cardiovascular complications and the potential cardiovascular benefits of influenza and COVID-19 vaccines in people with raised cardiovascular risk. As outlined in Chapter 7, I considered multiple study designs to look at the association between cardiovascular risk and influenza vaccine effect on cardiovascular events. Conducting a cohort study, with for example a propensity score matched or clone and censor design, may suffer from unmeasured confounding but alongside the findings from the SCCS study I present in Chapter 7 could still provide useful evidence for the role of influenza vaccine in preventing cardiovascular events by cardiovascular risk profile. In particular, further cohort studies and SCCS could use active comparators or negative control outcomes. Additionally, studies could expansion consideration of waning, and for COVID-19

changes in infection risk such as during lockdowns or variation in virus circulation over time and location.

The ultimate aim of my thesis was to consider if influenza vaccine recommendations should be extended beyond current risk groups to include those with raised cardiovascular risk. During the COVID-19 pandemic, influenza vaccine eligibility was extended to all adults aged ≥50 years. The 50-64 year olds who were included in the 2020/21 and 2021/22 influenza vaccine programme extension, provide a possible non-selective (i.e., not based on presence of clinical risk factors covered by existing vaccine recommendations) study population among whom the association of influenza vaccine and acute cardiovascular events among individuals without established CVD.

However, given the limitations of using observational data to evaluate the effect of vaccines, further RCTs will provide the best evidence of effectiveness. Unlike existing RCTs, future trials would need to consider influenza vaccination as a primary CVD prevention strategy.

Ultimately, to inform vaccine policy, and whether those at raised cardiovascular risk should receive influenza vaccine, then a direct analysis of cardiovascular risk on influenza outcomes such as pneumonia, hospitalisations and death are needed rather than specifically the effect on cardiovascular outcomes, which are not the main complication. A similar approach would be desirable for COVID-19 vaccine if this is to become a routine, perhaps annual vaccine, with clinical risk group prioritisation comparable to that of the influenza vaccine. Analysis building on the element of my PhD which studied the effect of CV risk on ARI, particularly influenza and pneumonia would be necessary. In particular, where influenza and pneumonia result in hospitalization and death. Furthermore, the impact on healthcare resource use and cost of influenza, or COVID-19, among people with raised, compared to low, cardiovascular risk and in the context of those with established CVD, would help to establish the benefit of vaccinating individuals with raised cardiovascular risk. An analysis of Canadian data suggests that half as many individuals who were diabetic, hypertensive or obese need to be vaccinated against COVID-19 to prevent death compared to individuals without these risk factors [310].

Further evidence generation related to cardiovascular risk factors and the vaccine induced prevention of influenza and COVID-19 should help inform a full cost-effectiveness analysis. The evidence to date suggests such strategies could be cost-effective [311]. When considering cardiovascular complications, establishing the cost of events such as MI and stroke in those with raised cardiovascular risk compared to the cost of vaccinating individuals with raised cardiovascular risk would be beneficial. Such analysis has previously been focused on secondary prevention of CVD [312].

Conducting further analysis of the COVID-19-related cardiovascular complications by underlying cardiovascular risk after 2020 and the widespread roll-out of the COVID-19 vaccines will also establish the individual benefit of vaccination as well as the public benefit. Furthermore, the effect of immune waning over time and boosting from further vaccine doses could be considered to increase the existing evidence base [313] and consider the effect of underlying cardiovascular risk in the processes.

The mechanisms by which influenza vaccine exerts cardiovascular benefit are uncertain. In my analysis I assumed the protective effect is due to prevention of influenza which can trigger a cardiovascular event. However, there is also the possibility of pleiotropic effects between virus and the antigens of atherosclerotic plaque as well as unspecific immunomodulatory effect which in turn prevents cardiovascular complications unrelated to influenza virus circulation and infection. Consideration of the different mechanistic and long-term effects of vaccines should be explored in future research.

Lastly, my research has focused on infection specific prevention, namely vaccine, however, there may also a role for short-term cardiovascular prevention methods. Antithrombotic and antiplatelet targeted treatment during infections may offer a preventive option. Short-term use at the time of infection identification could reduce the risk of ACS or stroke without increasing the risk of bleeding due to the temporary nature of the treatment [314].

10.7. Conclusions

In conclusion, my thesis analysis extends existing literature and clinical understanding on the cardiovascular complications associated with ARI by considering the role of underlying cardiovascular risk as well as the protective benefit of influenza vaccine in those at raised cardiovascular risk. The findings I have generated suggest those at raised cardiovascular risk are more likely to develop pneumonia, but it is unclear whether these individuals are also at increased risk of other ARIs such as influenza/ILI and COVID-19. However, I did find an association between raised cardiovascular risk and ARI-related cardiovascular complications. This includes influenza, pneumonia, and COVID-19. Raised cardiovascular risk was more strongly associated with ARI-related cardiovascular complications when a composite measure of cardiovascular risk, in this case QRISK2/3 score was used compared to hypertension alone. Addressing cardiovascular risk factors could improve outcomes after ARIs. Improved vaccine uptake could help reduce the risk of first acute cardiovascular events among those already eligible to receive the seasonal influenza vaccine and regular COVID-19 boosters. Overall, it is important to maintain and improve influenza and COVID-19 vaccine provision.

References

1. World Health Organization. Cardiovascular diseases. World Health Organization, 2021. Available at: https://www.who.int/cardiovascular_diseases/en/. Accessed 28 July 2021.
2. Virtual Health Information Network. Measuring new cardiovascular events in the IDI. 2017. Available at: https://vhin.co.nz/guides/measuring-new-cardiovascular-events-in-the-idi/. Accessed 29 July 2021.
3. British Heart Foundation. Coronary heart disease (Ischaemic heart disease) - types, causes & symptoms. Available at: https://www.bhf.org.uk/informationsupport/conditions/coronary-heart-disease. Accessed 28 July 2021.
4. Bentzon JF, Otsuka F, Virmani R, Falk E. Mechanisms of plaque formation and rupture. Circulation Research **2014**; 114:1852–1866. Available at: https://www.ahajournals.org/doi/abs/10.1161/CIRCRESAHA.114.302721. Accessed 9 January 2022.
5. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Blood Vessels and Endothelial Cells. In: Molecular Biology of the Cell. 4th edition. Garland Science, 2002.
6. National Clinical Guideline Centre. Guideline 94: Unstable Angina and NSTEMI: The Early Management of Unstable Angina and Non-ST-Segment-Elevation Myocardial Infarction. Royal College of Physician **2010**; :1–24.
7. Higashi Y. Assessment of Endothelial Function History, Methodological Aspects, and Clinical Perspectives. International Heart Journal **2015**; 56:125–134.
8. World Health Organization. Hypertension and coronary heart disease: classification and criteria for epidemiological studies. 1958.
9. Char DM, Israel E, Ladenson J. Early laboratory indicators of acute myocardial infarction. Emergency Medicine Clinics of North America **1998**; 16:519–539.
10. Bahit MC, Kochar A, Granger CB. Post-Myocardial Infarction Heart Failure. JACC: Heart Failure **2018**; 6:179–186.
11. Gho JMIH, Schmidt AF, Pasea L, et al. An electronic health records cohort study on heart failure following myocardial infarction in England: incidence and predictors. BMJ Open **2018**; 8:e018331.
12. Definition | Background information | Heart failure - chronic | CKS | NICE. Available at: https://cks.nice.org.uk/topics/heart-failure-chronic/background-information/definition/. Accessed 9 January 2022.
13. National Health Service. Heart failure. 2021. Available at: https://www.nhs.uk/conditions/heart-failure/. Accessed 28 July 2021.
14. Caraballo C, Desai NR, Mulder H, et al. Clinical Implications of the New York Heart Association Classification. J Am Heart Assoc **2019**; 8. Available at: https://www.ahajournals.org/doi/abs/10.1161/JAHA.119.014240. Accessed 9 January 2022.

15. American Heart Association. Ejection Fraction Heart Failure Measurement. 2017. Available at: https://www.heart.org/en/health-topics/heart-failure/diagnosing-heart-failure/ejection-fraction-heart-failure-measurement. Accessed 8 October 2021.

16. British Heart Foundation. Arrhythmias - Abnormal heart rhythms. 2019. Available at: https://www.bhf.org.uk/informationsupport/conditions/arrhythmias. Accessed 8 October 2021.

17. John RM, Tedrow UB, Koplan BA, et al. Ventricular arrhythmias and sudden cardiac death. The Lancet **2012**; 380:1520–1529.

18. Roberts-Thomson KC, Lau DH, Sanders P. The diagnosis and management of ventricular arrhythmias. Nature Reviews Cardiology 2011 8:6 **2011**; 8:311–321.

19. Amarenco P, Bogousslavsky J, Caplan LR, Donnan GA, Hennerici MG. Classification of Stroke Subtypes. Cerebrovascular Diseases **2009**; 27:493–501.

20. Markus H. Stroke: causes and clinical features. Medicine **2008**; 36:586–591.

21. Sacco RL, Kasner SE, Broderick JP, et al. An updated definition of stroke for the 21st century: A statement for healthcare professionals from the American heart association/American stroke association. Stroke **2013**; 44:2064–2089.

22. Fischer U, Baumgartner A, Arnold M, et al. What Is a Minor Stroke? Stroke **2010**; 41:661–666.

23. Yew KS, Cheng E. Acute Stroke Diagnosis. Am Fam Physician **2009**; 80:33. Available at: /pmc/articles/PMC2722757/. Accessed 9 January 2022.

24. Feigin VL, Rinkel GJE, Lawes CMM, et al. Risk factors for subarachnoid hemorrhage: An updated systematic review of epidemiological studies. Stroke **2005**; 36:2773–2780. Available at: https://www.ahajournals.org/doi/abs/10.1161/01.str.0000190838.02954.e8. Accessed 11 January 2022.

25. An SJ, Kim TJ, Yoon BW. Epidemiology, Risk Factors, and Clinical Features of Intracerebral Hemorrhage: An Update. Journal of Stroke **2017**; 19:3. Available at: /pmc/articles/PMC5307940/. Accessed 10 January 2022.

26. Coupland AP, Thapar A, Qureshi MI, Jenkins H, Davies AH. The definition of stroke: http://dx.doi.org/101177/0141076816680121 **2017**; 110:9–12.

27. Rapaport E, Bernard R, Corday E. Nomenclature and criteria for diagnosis of ischemic heart disease. Report of the Joint International Society and Federation of Cardiology/World Health Organization task force on standardization of clinical nomenclature. Circulation **1979**; 59:607–609.

28. Tunstall-Pedoe H, Kuulasmaa K, Amouyel P, Arveiler D, Rajakangas - AM, Pajak A. Myocardial infarction and coronary deaths in the World Health Organization MONICA project: Registration procedures, event rates, and case-fatality rates in 38 populations from 21 countries in four continents. Circulation **1994**; 90:583–612.

29. Myocardial infarction redefined--a consensus document of The Joint European Society of Cardiology/American College of Cardiology Committee for the redefinition of myocardial infarction. Eur Heart J **2000**; 21:1502–1513.

30. Thygesen K, Alpert JS, White HD. Universal definition of myocardial infarction. Circulation. 2007; 116:2634–2653.

31. Thygesen K, Alpert JS, Jaffe AS, Simoons ML, Chaitman BR, White HD. Third Universal Definition of Myocardial Infarction. Circulation **2012**; 126:2020–2035.

32. Thygesen K, Alpert JS, Jaffe AS, et al. Fourth universal definition of myocardial infarction (2018). European Heart Journal **2019**; 40:237–269.

33. Luepker R V, Apple FS, Christenson RH, et al. Case definitions for acute coronary heart disease in epidemiology and clinical research studies: a statement from the AHA Council on Epidemiology and Prevention; AHA Statistics Committee; World Heart Federation Council on Epidemiology and Prevention; the European Society of Cardiology Working Group on Epidemiology and Prevention; Centers for Disease Control and Prevention; and the National Heart, Lung, and Blood Institute. Circulation **2003**; 108:2543–9.

34. Collet JP, Thiele H, Barbato E, et al. 2020 ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation. European Heart Journal **2021**; 42:1289–1367.

35. McKee P, Castelli W, McNamara P, Kannel W. The natural history of congestive heart failure: the Framingham study. N Engl J Med **1971**; 285:1441–1446.

36. Carlson KJ, Lee DCS, Goroll AH, Leahy M, Johnson RA. An analysis of physicians' reasons for prescribing long-term digitalis therapy in outpatients. Journal of Chronic Diseases **1985**; 38:733–739.

37. Nieminen M, Böhm M, Cowie MR, et al. Executive summary of the guidelines on the diagnosis and treatment of acute heart failure: the Task Force on Acute Heart Failure of the European Society of Cardiology. Eur Heart J **2005**; 26:384–416.

38. Yancy CW, Jessup M, Bozkurt B, et al. 2013 ACCF/AHA guideline for the management of heart failure: A report of the american college of cardiology foundation/american heart association task force on practice guidelines. Circulation **2013**; 128.

39. Ponikowski P, Voors AA, Anker SD, et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) Developed with the special contribution o. European Heart Journal **2016**; 37:2129–2200.

40. Bozkurt B, Coats AJ, Tsutsui H, et al. Universal Definition and Classification of Heart Failure: A Report of the Heart Failure Society of America, Heart Failure Association of the European Society of Cardiology, Japanese Heart Failure Society and Writing Committee of the Universal Definition of Heart Failure. Journal of Cardiac Failure **2021**; 27:387–413.

41. Aho K, Harmsen P, Hatano S, Marquardsen J, Smirnov VE, Strasser T. Cerebrovascular disease in the community: results of a WHO collaborative study. Bull World Health Organ **1980**; 58:113–130.

42. Albers G, Caplan L, Easton J, et al. Transient ischemic attack--proposal for a new definition. N Engl J Med **2002**; 347:1713–1716.

43. Easton JD, Saver JL, Albers GW, et al. Definition and Evaluation of Transient Ischemic Attack. Stroke **2009**; 40:2276–2293.

44. World Health Organization. The top 10 causes of death. Available at: https://www.who.int/en/news-room/fact-sheets/detail/the-top-10-causes-of-death. Accessed 11 February 2019.

45. Roth GA, Mensah GA, Johnson CO, et al. Global Burden of Cardiovascular Diseases and Risk Factors, 1990–2019: Update From the GBD 2019 Study. J Am Coll Cardiol **2020**; 76:2982–3021.

46. British Heart Foundation. Heart statistics - Heart and Circulatory Diseases in the UK. 2021. Available at: https://www.bhf.org.uk/what-we-do/our-research/heart-statistics. Accessed 28 July 2021.

47. British Heart Foundation. Death rates over time. 2021. Available at: https://www.bhf.org.uk/what-we-do/our-research/heart-and-circulatory-diseases-in-numbers/death-rates-over-time. Accessed 29 July 2021.

48. University of Washington. Global Burden of Disease Compare. 2021. Available at: https://vizhub.healthdata.org/gbd-compare/. Accessed 28 July 2021.

49. Bhatnagar P, Wickramasinghe K, Wilkins E, Townsend N. Trends in the epidemiology of cardiovascular disease in the UK. Heart **2016**; 102:1945–1952. Available at: https://heart.bmj.com/content/102/24/1945. Accessed 28 July 2021.

50. King D, Wittenberg R, Patel A, Quayyum Z, Berdunov V, Knapp M. The future incidence, prevalence and costs of stroke in the UK. Age and Ageing **2020**; 49:277–282.

51. Lee S, Shafe ACE, 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. Available at: https://bmjopen.bmj.com/content/1/2/e000269. Accessed 14 January 2022.

52. Chevreul K, Durand-Zaleski I, Gouépo A, Fery-Lemonnier E, Hommel M, Woimant F. Cost of stroke in France. European Journal of Neurology **2013**; 20:1094–1100.

53. Evers SMAA, Struijs JN, Ament AJHA, van Genugten MLL, Jager J (Hans) C, van den Bos GAM. International Comparison of Stroke Cost Studies. Stroke **2004**; 35:1209–1215.

54. Saka O, McGuire A, Wolfe C. Cost of stroke in the United Kingdom. Age and Ageing **2008**; 38:27–32.

55. Ambrosy AP, Fonarow GC, Butler J, et al. The Global Health and Economic Burden of Hospitalizations for Heart Failure: Lessons Learned From Hospitalized Heart Failure Registries. J Am Coll Cardiol **2014**; 63:1123–1133.

56. Asaria P, Elliott P, Douglass M, et al. Acute myocardial infarction hospital admissions and deaths in England: a national follow-back and follow-forward record-linkage study. The Lancet Public Health **2017**; 2:e191–e201.

57. Castelli WP. Epidemiology of coronary heart disease: The Framingham study. The American Journal of Medicine **1984**; 76:4–12.

58. Rose G, Hamilton PJS, Keen H, Reid DD, McCartney P, Jarrett RJ. Myocardial ischaemia, risk factors and death from coronary heart-disease. The Lancet **1977**; 309:105–109.

59. Millett ERC, Peters SAE, Woodward M. Sex differences in risk factors for myocardial infarction: cohort study of UK Biobank participants. BMJ **2018**; 363.

60. The GBD 2016 Lifetime Risk of Stroke Collaborators. Global, Regional, and Country-Specific Lifetime Risks of Stroke, 1990 and 2016. https://doi.org/101056/NEJMoa1804492 **2018**; 379:2429–2437.

61. Fuchs FD, Whelton PK. High Blood Pressure and Cardiovascular Disease. Hypertension **2020**; :285–292.

62. Zhou B, Danaei G, Stevens GA, et al. Long-term and recent trends in hypertension awareness, treatment, and control in 12 high-income countries: an analysis of 123 nationally representative surveys. The Lancet **2019**; 394:639–651.

63. Hankey GJ. Population Impact of Potentially Modifiable Risk Factors for Stroke. Stroke **2020**; :719–728.

64. 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 **2015**; 3:105–113.

65. Mosenzon O, Alguwaihes A, Leon JLA, et al. CAPTURE: a multinational, cross-sectional study of cardiovascular disease prevalence in adults with type 2 diabetes across 13 countries. Cardiovascular Diabetology 2021 20:1 **2021**; 20:1–13.

66. Lautsch D, Wang T, Yang L, Rajpathak SN. Prevalence of Established Cardiovascular Disease in Patients with Type 2 Diabetes Mellitus in the UK. Diabetes Therapy 2019 10:6 **2019**; 10:2131–2137.

67. Damen JAAG, Hooft L, Schuit E, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. BMJ **2016**; 353:i2416.

68. Anderson KM, Odell PM, Wilson PW, Kannel WB. Cardiovascular disease risk profiles. Am Heart J **1991**; 121:293–8.

69. Conroy R, Pyörälä K, Fitzgerald AP, et al. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. European Heart Journal **2003**; 24:987–1003.

70. 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.

71. Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. BMJ **2017**; 357:j2099. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28536104. Accessed 15 May 2019.

72. Vries TI de, Visseren FLJ. Cardiovascular risk prediction tools made relevant for GPs and patients. Heart **2021**; 107:332–340.

73. Jiang S-Z, Lu W, Zong X-F, Ruan H-Y, Liu Y. Obesity and hypertension. Experimental and Therapeutic Medicine **2016**; 12:2395.

74. Chaturvedi N, Fuller JH. Ethnic differences in mortality from cardiovascular disease in the UK: do they persist in people with diabetes? J Epidemiol Community Health (1978) **1996**; 50:137–139. Available at: https://pubmed.ncbi.nlm.nih.gov/8762376/. Accessed 14 January 2022.

75. Mindell J, Biddulph JP, Hirani V, et al. Cohort Profile: The Health Survey for England. International Journal of Epidemiology **2012**; 41:1585–1593. Available at: https://academic.oup.com/ije/article/41/6/1585/741862. Accessed 14 January 2022.

76. Folsom AR, Yamagishi K, Hozawa A, Chambless LE. Absolute and attributable risks of heart failure incidence in relation to optimal risk factors. Circ Heart Fail **2009**; 2:11–17.

77. Folsom AR, Yatsuya H, Nettleton JA, Lutsey PL, Cushman M, Rosamond WD. Community prevalence of ideal cardiovascular health, by the american heart association definition, and relationship with cardiovascular disease incidence. J Am Coll Cardiol **2011**; 57:1690–1696.

78. Lloyd-Jones DM, Leip EP, Larson MG, et al. Prediction of Lifetime Risk for Cardiovascular Disease by Risk Factor Burden at 50 Years of Age. Circulation **2006**; 113:791–798.

79. Hobbs FDR, Hobbs R. Cardiovascular disease: different strategies for primary and secondary prevention? Heart **2004**; 90:1217–1223. Available at: https://heart.bmj.com/content/90/10/1217. Accessed 10 January 2022.

80. MI - secondary prevention | Health topics A to Z | CKS | NICE. Available at: https://cks.nice.org.uk/topics/mi-secondary-prevention/. Accessed 17 January 2022.

81. Laing BY, Katz MH. Coronary Arteries, Myocardial Infarction, and History. http://dx.doi.org/101056/NEJMc1201171 **2012**; 366:1258–1260.

82. Centers for Disease Control and Prevention. Achievements in Public Health, 1900-1999: Decline in Deaths from Heart Disease and Stroke -- United States, 1900-1999. Morbidity and Mortality Weekly Report (MMWR) **1999**; 48:649–656.

83. National Institute for Health and Care Excellence. Hypertension in adults: diagnosis and management. 2019. Available at: https://www.guidelines.co.uk/cardiovascular/nice-hypertension-guideline/454934.article. Accessed 2 August 2021.

84. Zhou B, Carrillo-Larco RM, Danaei G, et al. Worldwide trends in hypertension prevalence and progress in treatment and control from 1990 to 2019: a pooled analysis of 1201 population-representative studies with 104 million participants. The Lancet **2021**; 398:957–980.

85. O'Keeffe AG, Nazareth I, Petersen I. Time trends in the prescription of statins for the primary prevention of cardiovascular disease in the United Kingdom: a cohort study using The Health Improvement Network primary care data. Clinical Epidemiology **2016**; 8:123.

86. Backer G De. Prevention of cardiovascular disease: recent achievements and remaining challenges. e-Journal of Cardiology Practice **2017**; 15.

87. Brugts JJ, Yetgin T, Hoeks SE, et al. The benefits of statins in people without established cardiovascular disease but with cardiovascular risk factors: meta-analysis of randomised controlled trials. BMJ **2009**; 338:36.

88. Rodriguez F, Maron DJ, Knowles JW, Virani SS, Lin S, Heidenreich PA. Association of Statin Adherence With Mortality in Patients With Atherosclerotic Cardiovascular Disease. JAMA Cardiology **2019**; 4:206–213.

89. Vera MA De, Bhole V, Burns LC, Lacaille D. Impact of statin adherence on cardiovascular disease and mortality outcomes: a systematic review. British Journal of Clinical Pharmacology **2014**; 78:684–698.

90. National Institute for Health and Care Excellence. CVD risk assessment and management. 2019.

91. Zheng SL, Roddick AJ. Association of Aspirin Use for Primary Prevention With Cardiovascular Events and Bleeding Events: A Systematic Review and Meta-analysis. JAMA **2019**; 321:277–287.

92. Hao Q, Tampi M, O'Donnell M, Foroutan F, Siemieniuk RA, Guyatt G. Clopidogrel plus aspirin versus aspirin alone for acute minor ischaemic stroke or high risk transient ischaemic attack: systematic review and meta-analysis. BMJ **2018**; 363:5108.

93. Woodhead M, Blasi F, Ewig S, et al. Guidelines for the management of adult lower respiratory tract infections - Summary. Clinical Microbiology and Infection **2011**; 17:1–24.

94. Torres A, Blasi F, Peetermans WE, Viegi G, Welte T. The aetiology and antibiotic management of community-acquired pneumonia in adults in Europe: a literature review. European Journal of Clinical Microbiology & Infectious Diseases 2014 33:7 **2014**; 33:1065–1079.

95. Hayward AC, Fragaszy EB, Bermingham A, et al. Comparative community burden and severity of seasonal and pandemic influenza: Results of the Flu Watch cohort study. The Lancet Respiratory Medicine **2014**; 2:445–454. Available at: http://www.thelancet.com/article/S2213260014700347/fulltext. Accessed 10 January 2022.

96. Azziz Baumgartner E, Dao CN, Nasreen S, et al. Seasonality, Timing, and Climate Drivers of Influenza Activity Worldwide. The Journal of Infectious Diseases **2012**; 206:838–846.

97. Chow A, Ma S, Ling AE, Chew SK. Influenza-associated Deaths in Tropical Singapore. Emerging Infectious Diseases **2006**; 12:114. Available at: /pmc/articles/PMC3293465/. Accessed 4 August 2021.

98. World Health Organization. Influenza (Seasonal). 2018. Available at: https://www.who.int/en/news-room/fact-sheets/detail/influenza-(seasonal). Accessed 4 August 2021.

99. Taubenberger JK, Morens DM. The Pathology of Influenza Virus Infections. Annu Rev Pathol **2008**; 3:499.

100. Khandaker G, Dierig A, Rashid H, King C, Heron L, Booy R. Systematic review of clinical and epidemiological features of the pandemic influenza A (H1N1) 2009. Influenza and Other Respiratory Viruses **2011**; 5:148–156.

101. Al-Tawfiq JA, Zumla A, Gautret P, et al. Surveillance for emerging respiratory viruses. The Lancet Infectious Diseases **2014**; 14:992–1000.

102. Gupta A, Madhavan M v., Sehgal K, et al. Extrapulmonary manifestations of COVID-19. Nature Medicine **2020**; 26:1017–1032. Available at: https://www.nature.com/articles/s41591-020-0968-3. Accessed 16 September 2021.

103. Parasher A. COVID-19: Current understanding of its Pathophysiology, Clinical presentation and Treatment. Postgraduate Medical Journal **2021**; 97:312–320.

104. Chest infections - adult | Health topics A to Z | CKS | NICE. Available at: https://cks.nice.org.uk/topics/chest-infections-adult/. Accessed 18 January 2022.

105. Zambon M. Influenza surveillance and laboratory diagnosis. Textbook of Influenza **2013**; :229–249. Available at: https://onlinelibrary.wiley.com/doi/full/10.1002/9781118636817.ch15. Accessed 18 January 2022.

106. Vos LM, Bruyndonckx R, Zuithoff NPA, et al. Lower respiratory tract infection in the community: associations between viral aetiology and illness course. Clinical Microbiology and Infection **2021**; 27:96–104.

107. Ieven M, Coenen S, Loens K, et al. Aetiology of lower respiratory tract infection in adults in primary care: a prospective study in 11 European countries. Clinical Microbiology and Infection **2018**; 24:1158–1163.

108. Chalmers J, Campling J, Ellsbury G, Hawkey PM, Madhava H, Slack M. Community-acquired pneumonia in the United Kingdom: a call to action. Pneumonia **2017**; 9.

109. Chalmers JD, Campling J, Dicker A, Woodhead M, Madhava H. A systematic review of the burden of vaccine preventable pneumococcal disease in UK adults. BMC Pulmonary Medicine **2016**; 16.

110. Paget J, Spreeuwenberg P, Charu V, et al. Global mortality associated with seasonal influenza epidemics: New burden estimates and predictors from the GLaMOR Project. Journal of Global Health **2019**; 9. Available at: /pmc/articles/PMC6815659/. Accessed 26 November 2021.

111. Iuliano AD, Roguski KM, Chang HH, et al. Estimates of global seasonal influenza-associated respiratory mortality: a modelling study. The Lancet **2018**; 391:1285–1300.

112. World Health Organization Europe. Burden of influenza. Available at: https://www.euro.who.int/en/health-topics/communicable-diseases/influenza/seasonal-influenza/burden-of-influenza. Accessed 26 November 2021.

113. Public Health England. Surveillance of influenza and other respiratory viruses in the UK: Winter 2020 to 2021. 2020. Available at: https://www.gov.uk/government/statistics/annual-flu-reports. Accessed 10 December 2021.

114. Adlhoch C, Mook P, Lamb F, et al. Very little influenza in the WHO European Region during the 2020/21 season, weeks 40 2020 to 8 2021. Eurosurveillance **2021**; 26:2100221.

115. European Centre for Disease Prevention and Control. COVID-19 situation update worldwide, as of week 1, updated 13 January 2022. 2022. Available at: https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases. Accessed 18 January 2022.

116. Millett ERC, Quint JK, Smeeth L, Daniel RM, Thomas SL. Incidence of Community-Acquired Lower Respiratory Tract Infections and Pneumonia among Older Adults in the United Kingdom: A Population-Based Study. PLOS ONE **2013**; 8:e75131. Available at: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0075131. Accessed 4 August 2021.

117. Quan TP, Fawcett NJ, Wrightson JM, et al. Increasing burden of community-acquired pneumonia leading to hospitalisation, 1998–2014. Thorax **2016**; 71:535–542. Available at: https://thorax.bmj.com/content/71/6/535. Accessed 20 October 2021.

118. de Miguel-Díez J, Jiménez-García R, Hernández-Barrera V, et al. Trends in hospitalizations for community-acquired pneumonia in Spain: 2004 to 2013. European Journal of Internal Medicine **2017**; 40:64–71.

119. Naucler P, Henriques-Normark B, Hedlund J, Galanis I, Granath F, Örtqvist Å. The changing epidemiology of community-acquired pneumonia: nationwide register-based study in Sweden. Journal of Internal Medicine **2019**; 286:689–701. Available at: https://onlinelibrary.wiley.com/doi/full/10.1111/joim.12956. Accessed 20 October 2021.

120. Thompson WW, Shay D, Weintraub E, et al. Mortality associated with influenza and respiratory syncytial virus in the United States. JAMA **2003**; 289:179–186. Available at: https://pubmed.ncbi.nlm.nih.gov/12517228/. Accessed 5 August 2021.

121. Wong C, Chan K, Hedley A, Peiris J. Influenza-associated mortality in Hong Kong. Clin Infect Dis **2004**; 39:1611–1617. Available at: https://pubmed.ncbi.nlm.nih.gov/15578360/. Accessed 5 August 2021.

122. Feng L, Shay D, Jiang Y, et al. Influenza-associated mortality in temperate and subtropical Chinese cities, 2003-2008. Bull World Health Organ **2012**; 90. Available at: https://pubmed.ncbi.nlm.nih.gov/22511824/. Accessed 5 August 2021.

123. Cromer D, van Hoek AJ, Jit M, Edmunds WJ, Fleming D, Miller E. The burden of influenza in England by age and clinical risk group: A statistical analysis to inform vaccine policy. Journal of Infection **2014**; 68:363–371. Available at: https://www.sciencedirect.com/science/article/pii/S0163445313003733. Accessed 14 August 2019.

124. Welte T, Torres A, Nathwani D. Clinical and economic burden of community-acquired pneumonia among adults in Europe. Thorax **2012**; 67:71–79. Available at: https://thorax.bmj.com/content/67/1/71. Accessed 5 August 2021.

125. Torres A, Peetermans W, Viegi G, Blasi F. Risk factors for community-acquired pneumonia in adults in Europe: a literature review. Thorax **2013**; 68:1057–1065. Available at: https://pubmed.ncbi.nlm.nih.gov/24130229/. Accessed 5 August 2021.

126. Torres A, Blasi F, Dartois N, Akova M. Which individuals are at increased risk of pneumococcal disease and why? Impact of COPD, asthma, smoking, diabetes, and/or chronic heart disease on community-acquired pneumonia and invasive pneumococcal disease. Thorax **2015**; 70:984–989. Available at: https://pubmed.ncbi.nlm.nih.gov/26219979/. Accessed 5 August 2021.

127. Jordan RE, Adab P, Cheng KK. Covid-19: risk factors for severe disease and death. BMJ **2020**; 368. Available at: https://www.bmj.com/content/368/bmj.m1198. Accessed 9 August 2021.

128. Wolff D, Nee S, Hickey NS, Marschollek M. Risk factors for Covid-19 severity and fatality: a structured literature review. Infection **2020**; 49:15–28. Available at: https://link.springer.com/article/10.1007/s15010-020-01509-1. Accessed 9 August 2021.

129. Public Health England. Pneumococcal: the green book, chapter 25. 2018. Available at: https://www.gov.uk/government/publications/pneumococcal-the-green-book-chapter-25. Accessed 14 May 2019.

130. Public Health England. Influenza: the green book, chapter 19. 2020. Available at: https://www.gov.uk/government/publications/influenza-the-green-book-chapter-19. Accessed 3 December 2020.

131. Public Health England. Seasonal influenza vaccine uptake in GP patients: winter season 2019 to 2020. 2020.

132. Andrews N, Waight P, George R, Slack M, Miller E. Impact and effectiveness of 23-valent pneumococcal polysaccharide vaccine against invasive pneumococcal disease in the elderly in England and Wales. Vaccine **2012**; 30:6802–6808.

133. Moberley S, Holden J, Tatham D, Andrews R. Vaccines for preventing pneumococcal infection in adults. Cochrane Database Syst Rev **2008**;

134. European Centre for Disease Prevention and Control. Influenza vaccine effectiveness. 2021. Available at: https://www.ecdc.europa.eu/en/seasonal-influenza/prevention-and-control/vaccine-effectiveness. Accessed 18 January 2022.

135. The Lancet Respiratory Medicine. Realising the potential of SARS-CoV-2 vaccines—a long shot? The Lancet Respiratory Medicine **2021**; 9:117.

136. The Lancet Infectious Diseases. The rocky road to universal COVID-19 vaccination. The Lancet Infectious Diseases **2021**; 21:743.

137. NHS England. COVID-19 Vaccinations Statistics. 2021. Available at: https://www.england.nhs.uk/statistics/statistical-work-areas/covid-19-vaccinations/. Accessed 16 September 2021.

138. The Joint Committee on Vaccination and Immunisation. JCVI statement regarding a COVID-19 booster vaccine programme for winter 2021 to 2022. 2021. Available at: https://www.gov.uk/government/publications/jcvi-statement-september-2021-covid-19-booster-vaccine-programme-for-winter-2021-to-2022. Accessed 16 September 2021.

139. Mohammed I, Nauman A, Paul P, et al. The efficacy and effectiveness of the COVID-19 vaccines in reducing infection, severity, hospitalization, and mortality: a systematic review. Human Vaccines and Immunotherapeutics **2022**; 18. Available at: https://www.tandfonline.com/doi/abs/10.1080/21645515.2022.2027160. Accessed 27 July 2022.

140. Kandel R, Hartshorn KL. Prophylaxis and Treatment of Influenza Virus Infection. BioDrugs 2001 15:5 **2012**; 15:303–323.

141. National Institute for Health and Care Excellence. Influenza treatment summary. 2019. Available at: https://bnf.nice.org.uk/treatment-summary/influenza.html. Accessed 19 October 2021.

142. Mahase E. Covid-19: Molnupiravir reduces risk of hospital admission or death by 50% in patients at risk, MSD reports. BMJ **2021**; 375:n2422.

143. Mahase E. Covid-19: Pfizer's paxlovid is 89% effective in patients at risk of serious illness, company reports. BMJ **2021**; 375:n2713.

144. Ornato JP, Peberdy MA, Chandra NC, Bush DE. Seasonal pattern of acute myocardial infarction in the National Registry of Myocardial Infarction. J Am Coll Cardiol **1997**; 28:1684–1688.

145. Marti-Soler H, Gonseth S, Gubelmann C, et al. Seasonal Variation of Overall and Cardiovascular Mortality: A Study in 19 Countries from Different Geographic Locations. PLoS ONE **2014**; 9:e113500.

146. Weerasinghe DP, MacIntyre CR, Rubin GL. Seasonality of coronary artery deaths in New South Wales, Australia. Heart **2002**; 88:30–34.

147. Skajaa N, Horváth-Puhó E, Sundbøll J, Adelborg K, Rothman KJ, Sørensen HT. Forty-year Seasonality Trends in Occurrence of Myocardial Infarction, Ischemic Stroke, and Hemorrhagic Stroke. Epidemiology **2018**; 29:777–783.

148. Jakovljević D, Salomaa V, Sivenius J, et al. Seasonal variation in the occurrence of stroke in a Finnish adult population. The FINMONICA Stroke Register. Finnish Monitoring Trends and Determinants in Cardiovascular Disease. Stroke **1996**; 27:1774–1779.

149. Turin T, Kita Y, Murakami Y, et al. Higher stroke incidence in the spring season regardless of conventional risk factors: Takashima Stroke Registry, Japan, 1988-2001. Stroke **2008**; 39:745–752.

150. Wang Y, Levi C, Attia J, D'Este C, Spratt N, Fisher J. Seasonal variation in stroke in the Hunter Region, Australia: a 5-year hospital-based study, 1995-2000. Stroke **2003**; 34:1144–1150.

151. Rothwell P, Coull A, Giles M, et al. Change in stroke incidence, mortality, case-fatality, severity, and risk factors in Oxfordshire, UK from 1981 to 2004 (Oxford Vascular Study). The Lancet **2004**; 363:1925–1933.

152. Lee H, Hu C, Chen C, Lin H. Seasonal variation in ischemic stroke incidence and association with climate: a six-year population-based study. Chronobiol Int **2008**; 25:938–949.

153. Bhaskaran K, Hajat S, Haines A, Herrett E, Wilkinson P, Smeeth L. Effects of ambient temperature on the incidence of myocardial infarction. Heart **2009**; 95:1760–1769.

154. Bhaskaran K, Hajat S, Haines A, Herrett E, Wilkinson P, Smeeth L. Effects of air pollution on the incidence of myocardial infarction. Heart **2009**; 95:1746–1759.

155. Tillet HE, Smith JWG, Gooch CD. Excess Deaths Attributable to Influenza in England and Wales: Age at Death and Certified Cause. International Journal of Epidemiology **1983**; 12:344–352.

156. Reichert TA, Simonsen L, Sharma A, Pardo SA, Fedson DS, Miller MA. Influenza and the Winter Increase in Mortality in the United States, 1959-1999. American Journal of Epidemiology **2004**; 160:492–502.

157. Blackburn R, Zhao H, Pebody R, Hayward A, Warren-Gash C. Laboratory-Confirmed Respiratory Infections as Predictors of Hospital Admission for Myocardial Infarction and Stroke: Time-Series Analysis of English Data for 2004–2015. Clinical Infectious Diseases **2018**; 67:8–17.

158. Imai C, Barnett A, Hashizume M, et al. The Role of Influenza in the Delay between Low Temperature and Ischemic Heart Disease: Evidence from Simulation and Mortality Data from Japan. International Journal of Environmental Research and Public Health **2016**; 13:454.

159. Nguyen JL, Yang W, Ito K, Matte TD, Shaman J, Kinney PL. Seasonal Influenza Infections and Cardiovascular Disease Mortality. JAMA Cardiology **2016**; 1:274.

160. Barnes M, Heywood AE, Mahimbo A, Rahman B, Newall AT, Macintyre CR. Acute myocardial infarction and influenza: a meta-analysis of case–control studies. Heart **2015**; 101:1738–1747. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26310262. Accessed 11 February 2019.

161. Kwok CS, Aslam S, Kontopantelis E, et al. Influenza, influenza-like symptoms and their association with cardiovascular risks: a systematic review and meta-analysis of observational studies. Int J Clin Pract **2015**; 69:928–37. Available at: http://doi.wiley.com/10.1111/ijcp.12646. Accessed 11 February 2019.

162. Corrales-Medina VF, Suh KN, Rose G, et al. Cardiac Complications in Patients with Community-Acquired Pneumonia: A Systematic Review and Meta-Analysis of Observational Studies. Available at: www.plosmedicine.org. Accessed 10 April 2019.

163. Smeeth L, Thomas SL, Hall AJ, Hubbard R, Farrington P, Vallance P. Risk of myocardial infarction and stroke after acute infection or vaccination. N Engl J Med **2004**; 351:2611–8. Available at: http://www.nejm.org/doi/abs/10.1056/NEJMoa041747. Accessed 11 February 2019.

164. Warren-Gash C, Hayward AC, Hemingway H, et al. Influenza Infection and Risk of Acute Myocardial Infarction in England and Wales: A CALIBER Self-Controlled Case Series Study. The Journal of Infectious Diseases **2012**; 206:1652–1659.

165. Warren-Gash C, Blackburn R, Whitaker H, McMenamin J, Hayward AC. Laboratory-confirmed respiratory infections as triggers for acute myocardial infarction and stroke: a self-controlled case series analysis of national linked datasets from Scotland. European Respiratory Journal **2018**; 51:1701794. Available at: http://erj.ersjournals.com/lookup/doi/10.1183/13993003.01794-2017. Accessed 11 February 2019.

166. Kwong JC, Schwartz KL, Campitelli MA, et al. Acute Myocardial Infarction after Laboratory-Confirmed Influenza Infection. New England Journal of Medicine **2018**; 378:345–353.

167. Katsoularis I, Fonseca-Rodríguez O, Farrington P, Lindmark K, Connolly A-MF. Risk of acute myocardial infarction and ischaemic stroke following COVID-19 in Sweden: a self-controlled case series and matched cohort study. The Lancet **2021**; 398:599–607.

168. Modin D, Claggett B, Sindet-Pedersen C, et al. Acute COVID-19 and the Incidence of Ischemic Stroke and Acute Myocardial Infarction. Circulation **2020**; :2080–2082.

169. Bazaz R, Marriott HM, Francis SE, Dockrell DH. Mechanistic links between acute respiratory tract infections and acute coronary syndromes. J Infect **2013**; 66:1–17. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0163445312002769. Accessed 11 February 2019.

170. Corrales-Medina VF, Madjid M, Musher DM. Role of acute infection in triggering acute coronary syndromes. The Lancet Infectious Diseases **2010**; 10:83–92. Available at: https://www.sciencedirect.com/science/article/pii/S1473309909703317?via%3Dihub. Accessed 15 July 2019.

171. Reyes LF, Restrepo MI, Hinojosa CA, et al. Severe Pneumococcal Pneumonia Causes Acute Cardiac Toxicity and Subsequent Cardiac Remodeling. American Journal of Respiratory and Critical Care Medicine **2017**; 196:609–620. Available at: http://www.atsjournals.org/doi/10.1164/rccm.201701-0104OC. Accessed 11 February 2019.

172. Udell JA, Zawi R, Bhatt DL, et al. Association Between Influenza Vaccination and Cardiovascular Outcomes in High-Risk Patients. JAMA **2013**; 310:1711. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24150467. Accessed 11 February 2019.

173. Clar C, Oseni Z, Flowers N, Keshtkar-Jahromi M, Rees K. Influenza vaccines for preventing cardiovascular disease. Cochrane Database of Systematic Reviews **2015**; :CD005050. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25940444. Accessed 11 February 2019.

174. Yedlapati SH, Khan SU, Talluri S, et al. Effects of Influenza Vaccine on Mortality and Cardiovascular Outcomes in Patients With Cardiovascular Disease: A Systematic Review and Meta-Analysis. J Am Heart Assoc **2021**; 10:19636.

175. Frøbert O, Götberg M, Erlinge D, et al. Influenza Vaccination after Myocardial Infarction: A Randomized, Double-Blind, Placebo-Controlled, Multicenter Trial. Circulation **2021**; 144:1476–1484. Available at: https://www.ahajournals.org/doi/abs/10.1161/CIRCULATIONAHA.121.057042. Accessed 16 September 2021.

176. Vardeny O, Solomon SD. Influenza vaccination: a one-shot deal to reduce cardiovascular events. European Heart Journal **2016**; :ehw560.

177. Loeb M, Dokainish H, Dans A, et al. Randomized controlled trial of influenza vaccine in patients with heart failure to reduce adverse vascular events (IVVE): Rationale and design. Am Heart J **2019**; 212:36–44. Available at: https://pubmed.ncbi.nlm.nih.gov/30933856/. Accessed 16 September 2021.

178. Shrank WH, Patrick AR, Alan Brookhart M. Healthy User and Related Biases in Observational Studies of Preventive Interventions: A Primer for Physicians. Journal of General Internal Medicine 2010 26:5 **2011**; 26:546–550.

179. Naghavi M, Barlas Z, Siadaty S, Naguib S, Madjid M, Casscells W. Association of influenza vaccination and reduced risk of recurrent myocardial infarction. Circulation **2000**; 102:3039–3045.

180. Nichol KL, Nordin J, Mullooly J, Lask R, Fillbrandt K, Iwane M. Influenza Vaccination and Reduction in Hospitalizations for Cardiac Disease and Stroke among the Elderly. N Engl J Med **2009**; 348:1322–1332.

181. Wang CS, Wang ST, Lai C Te, Lin LJ, Chou P. Impact of influenza vaccination on major cause-specific mortality. Vaccine **2007**; 25:1196–1203.

182. Siriwardena AN, Gwini SM, Coupland CAC. Influenza vaccination, pneumococcal vaccination and risk of acute myocardial infarction: matched case–control study. CMAJ : Canadian Medical Association Journal **2010**; 182:1617.

183. Grau A, Fischer B, Barth C, Ling P, Lichy C, Buggle F. Influenza vaccination is associated with a reduced risk of stroke. Stroke **2005**; 36:1501–1506.

184. Lavallée P, Perchaud V, Gautier-Bertrand M, Grabli D, Amarenco P. Association between influenza vaccination and reduced risk of brain infarction. Stroke **2002**; 33:513–518.

185. Heffelfinger J, Heckbert S, Psaty B, et al. Influenza vaccination and risk of incident myocardial infarction. Hum Vaccin **2006**; 2:161–166.

186. Meyers DG, Beahm DD, Jurisich PD, Milford CJ, Edlavich S. Influenza and Pneumococcal Vaccinations Fail to Prevent Myocardial Infarction. Heart Drug **2004**; 4:96–100.

187. J J, M L, KK T, et al. Influenza vaccination and major adverse vascular events in high-risk patients. Circulation **2012**; 126:278–286.

188. Jackson LA, Group the VSDS, Yu O, et al. Influenza Vaccination Is Not Associated with a Reduction in the Risk of Recurrent Coronary Events. American Journal of Epidemiology **2002**; 156:634–640.

189. Mohseni H, Kiran A, Khorshidi R, Rahimi K. Influenza vaccination and risk of hospitalization in patients with heart failure: a self-controlled case series study. European Heart Journal **2017**; 38:326–333.

190. Gwini SM, Coupland CAC, Siriwardena AN. The effect of influenza vaccination on risk of acute myocardial infarction: Self-controlled case-series study. Vaccine **2011**; 29:1145–1149. Available at: https://www.sciencedirect.com/science/article/pii/S0264410X1001755X?via%3Dihub. Accessed 19 July 2019.

191. Asghar Z, Coupland C, Siriwardena N. Influenza vaccination and risk of stroke: Self-controlled case-series study. Vaccine **2015**; 33:5458–5463.

192. Sen A, Bakken IJ, Govatsmark RES, et al. Influenza vaccination and risk for cardiovascular events: a nationwide self-controlled case series study. BMC Cardiovascular Disorders **2021**; 21:31. Available at: https://bmccardiovascdisord.biomedcentral.com/articles/10.1186/s12872-020-01836-z. Accessed 15 February 2021.

193. Ren S, Newby D, Li SC, et al. Effect of the adult pneumococcal polysaccharide vaccine on cardiovascular disease: a systematic review and meta-analysis. Open Heart **2015**; 2:e000247.

194. Vlachopoulos C V, Terentes-Printzios DG, Aznaouridis KA, Pietri PG, Stefanadis CI. Association between pneumococcal vaccination and cardiovascular outcomes: a systematic review and meta-analysis of cohort studies. European Journal of Preventive Cardiology **2015**; 22:1185–1199.

195. Delamothe T. NHS at 60: Founding principles. British Medical Journal **2008**; 336:1216.

196. Herrett E, Gallagher AM, Bhaskaran K, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). International Journal of Epidemiology **2015**; 44:827–836. Available at: https://academic.oup.com/ije/article-lookup/doi/10.1093/ije/dyv098. Accessed 14 August 2019.

197. Wolf A, Dedman D, Campbell J, et al. Data resource profile: Clinical Practice Research Datalink (CPRD) Aurum. International Journal of Epidemiology **2019**; Available at: https://academic.oup.com/ije/advance-article/doi/10.1093/ije/dyz034/5374844. Accessed 14 August 2019.

198. Herbert A, Wijlaars L, Zylbersztejn A, Cromwell D, Hardelid P. Data Resource Profile: Hospital Episode Statistics Admitted Patient Care (HES APC). International Journal of Epidemiology **2017**; 46:1093–1093i.

199. Office from National Statistics. User guide to mortality statistics. 2021. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/methodologies/userguidetomortalitystatisticsjuly2017. Accessed 25 August 2021.

200. Clinical Practice Research Datalink. CPRD linked data. 2021. Available at: https://www.cprd.com/linked-data#Patient postcode linked measures. Accessed 14 June 2021.

201. Data highlights | CPRD. Available at: https://www.cprd.com/data-highlights. Accessed 11 January 2022.

202. University of Southampton National Centre for Research Methods. Townsend deprivation index. Available at: https://www.restore.ac.uk/geo-refer/36229dtuks00y19810000.php. Accessed 25 August 2021.

203. Benson T. The history of the Read codes: the inaugural James Read Memorial Lecture 2011. Journal of Innovation in Health Informatics **2011**; 19:173–82.

204. Williams R. A Christmas guide to clinical coding. British Medical Journal **2018**; 363.

205. NHS Digital. Hospital Episode Statistics (HES). 2019. Available at: https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics. Accessed 14 August 2019.

206. CLOSER University College London. Understanding Hospital Episode Statistics (HES). 2018.

207. The processing cycle and HES data quality - NHS Digital. Available at: https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics/the-processing-cycle-and-hes-data-quality. Accessed 11 January 2022.

208. Padmanabhan S, Carty L, Cameron E, Ghosh RE, Williams R, Strongman H. Approach to record linkage of primary care data from Clinical Practice Research Datalink to other health-related patient data: overview and implications. European Journal of Epidemiology **2019**; 34:91–99.

209. Hospital Admitted Patient Care Activity 2019-20 - NHS Digital. Available at: https://digital.nhs.uk/data-and-information/publications/statistical/hospital-admitted-patient-care-activity/2019-20#. Accessed 11 January 2022.

210. Hospital Admitted Patient Care Activity 2020-21 - NHS Digital. Available at: https://digital.nhs.uk/data-and-information/publications/statistical/hospital-admitted-patient-care-activity/2020-21. Accessed 11 January 2022.

211. User guide to mortality statistics - Office for National Statistics. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/methodologies/userguidetomortalitystatisticsjuly2017. Accessed 11 January 2022.

212. Mortality statistics in England and Wales QMI - Office for National Statistics. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/methodologies/mortalitystatisticsinenglandandwalesqmi. Accessed 11 January 2022.

213. Public Health England. Notifiable diseases and causative organisms: how to report. 2021. Available at: https://www.gov.uk/guidance/notifiable-diseases-and-causative-organisms-how-to-report. Accessed 25 August 2021.

214. Department of Health & Social Care. COVID-19 testing data: methodology note. 2020. Available at: https://www.gov.uk/government/publications/coronavirus-covid-19-testing-data-methodology/covid-19-testing-data-methodology-note. Accessed 25 August 2021.

215. UK Government. Coronavirus Testing in the UK. 2021. Available at: https://coronavirus.data.gov.uk/details/testing. Accessed 25 August 2021.

216. Public Health England. Sources of COVID-19 surveillance systems. 2021. Available at: https://www.gov.uk/government/publications/national-covid-19-surveillance-reports/sources-of-covid-19-systems. Accessed 25 August 2021.

217. National Health Service. Quality and Outcomes Framework (QOF). 2020. Available at: https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-collections/quality-and-outcomes-framework-qof. Accessed 3 March 2020.

218. Roland M, Guthrie B. Quality and Outcomes Framework: what have we learnt? BMJ **2016**; 354. Available at: https://www.bmj.com/content/354/bmj.i4060. Accessed 11 January 2022.

219. Walker S, Mason AR, Claxton K, et al. Value for money and the Quality and Outcomes Framework in primary care in the UK NHS. British Journal of General Practice **2010**; 60:e213–e220. Available at: https://bjgp.org/content/60/574/e213. Accessed 11 January 2022.

220. Quality and Outcomes Framework Indicators | Standards & Indicators | NICE. Available at: https://www.nice.org.uk/standards-and-indicators/qofindicators?categories=&page=1. Accessed 11 January 2022.

221. Nissen F, Quint JK, Morales DR, Douglas IJ. How to validate a diagnosis recorded in electronic health records. Breathe. 2019; 15:64–68.

222. Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. British Journal of Clinical Pharmacology **2010**; 69:4–14. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20078607. Accessed 7 March 2019.

223. Khan NF, Harrison SE, Rose PW. Validity of diagnostic coding within the General Practice Research Database: a systematic review. British Journal of General Practice **2010**; 60:e128–e136. Available at: https://bjgp.org/content/60/572/e128. Accessed 25 August 2021.

224. Burns EM, Rigby E, Mamidanna R, et al. Systematic review of discharge coding accuracy. Journal of Public Health **2012**; 34:138–148.

225. Davé S, Petersen I. Creating medical and drug code lists to identify cases in primary care databases. Pharmacoepidemiol Drug Saf **2009**; 18:704–707. Available at: https://pubmed.ncbi.nlm.nih.gov/19455565/. Accessed 11 January 2022.

226. Morbey RA, Elliot AJ, Harcourt S, et al. Estimating the burden on general practitioner services in England from increases in respiratory disease associated with seasonal respiratory pathogen activity. Epidemiology and Infection **2018**; 146:1389–1396.

227. EU case definitions. Available at: https://www.ecdc.europa.eu/en/surveillance-and-disease-data/eu-case-definitions. Accessed 11 January 2022.

228. Hardelid P, Rait G, Gilbert R, Petersen I. Recording of influenza-like illness in UK primary care 1995-2013: Cohort study. PLoS ONE **2015**; 10.

229. Troeger C, Forouzanfar M, Rao PC, et al. Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory tract infections in 195 countries: a

systematic analysis for the Global Burden of Disease Study 2015. The Lancet Infectious Diseases **2017**; 17:1133–1161.

230. Greene G, Hood K, Little P, et al. Towards clinical definitions of lower respiratory tract infection (LRTI) for research and primary care practice in Europe: an international consensus study. Primary Care Respiratory Journal 2011 20:3 **2011**; 20:299–306.

231. National Institute for Health and Care Excellence. Hypertension in adults: diagnosis and management. 2019. Available at: https://www.nice.org.uk/guidance/ng136/chapter/Recommendations#diagnosing-hypertension. Accessed 4 February 2022.

232. Peng M, Chen G, Kaplan GG, et al. Methods of defining hypertension in electronic medical records: validation against national survey data. Journal of Public Health **2016**; 38:e392–e399.

233. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, May M, Brindle P. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. BMJ **2007**; 335:136.

234. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, Brindle P. Performance of the QRISK cardiovascular risk prediction algorithm in an independent UK sample of patients from general practice: a validation study. Heart **2008**; 94:34–39.

235. Cooper A, O'Flynn N. Risk assessment and lipid modification for primary and secondary prevention of cardiovascular disease: summary of NICE guidance. BMJ **2008**; 336:1246–1248.

236. Collins GS, Altman DG. An independent external validation and evaluation of QRISK cardiovascular risk prediction: a prospective open cohort study. BMJ **2009**; 339:144–147.

237. Read SH, Diepen M van, Colhoun HM, et al. Performance of Cardiovascular Disease Risk Scores in People Diagnosed With Type 2 Diabetes: External Validation Using Data From the National Scottish Diabetes Register. Diabetes Care **2018**; 41:2010–2018.

238. Li Y, Sperrin M, Belmonte M, Pate A, Ashcroft DM, van Staa TP. Do population-level risk prediction models that use routinely collected health data reliably predict individual risks? Scientific Reports 2019 9:1 **2019**; 9:1–8.

239. National Institute for Health and Care Excellence. Cardiovascular disease: risk assessment and reduction, including lipid modification. 2016. Available at: https://www.nice.org.uk/guidance/cg181. Accessed 26 October 2021.

240. Mathur R, Bhaskaran K, Chaturvedi N, et al. Completeness and usability of ethnicity data in UK-based primary care and hospital databases. Journal of Public Health **2014**; 36:684–692.

241. Yousaf S, Bonsall A. UK Data Service Impact Ambassadors Workshop With The Department for Education Jointly organised by The UK Data Service and The Department for Education (DfE) UK Townsend Deprivation Scores from 2011 census data. **2017**;

242. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. Annals of Internal Medicine **2009**; 150:604–612.

243. 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–f2350. Available at: http://www.bmj.com/cgi/doi/10.1136/bmj.f2350. Accessed 31 January 2019.

244. Pfister R, Michels G, Wilfred J, Luben R, Wareham NJ, Khaw KT. Does ICD-10 hospital discharge code I50 identify people with heart failure? A validation study within the EPIC-Norfolk study. International Journal of Cardiology **2013**; 168:4413–4414.

245. Wright FL, Green J, Canoy D, Cairns BJ, Balkwill A, Beral V. Vascular disease in women: comparison of diagnoses in hospital episode statistics and general practice records in England. BMC Medical Research Methodology **2012**; 12:161.

246. Sansom LT, Ramadan H. Stroke incidence: sensitivity of hospital data coding of acute stroke. Int J Stroke **2015**; 10:E70.

247. Cook M, Baker N, Lanes S, Bullock R, Wentworth C, Michael Arrighi H. Incidence of stroke and seizure in Alzheimer's disease dementia. Age and Ageing **2015**; 44:695–699.

248. Hammad TA, McAdams MA, Feight A, Iyasu S, Dal Pan GJ. Determining the predictive value of Read/OXMIS codes to identify incident acute myocardial infarction in the General Practice Research Database. Pharmacoepidemiology and Drug Safety **2008**; 17:1197–1201.

249. Gaist D, Wallander M-A, González-Pérez A, García-Rodríguez LA. Incidence of hemorrhagic stroke in the general population: validation of data from The Health Improvement Network. Pharmacoepidemiology and Drug Safety **2013**; 22:176–182.

250. Ruigómez A, Martín-Merino E, Rodríguez LAG. Validation of ischemic cerebrovascular diagnoses in the health improvement network (THIN). Pharmacoepidemiology and Drug Safety **2010**; 19:579–585.

251. Van Staa T -P, Abenhaim L. The quality of information recorded on a UK database of primary care records: A study of hospitalizations due to hypoglycemia and other conditions. Pharmacoepidemiology and Drug Safety **1994**; 3:15–21.

252. Zhou EH, Gelperin K, Levenson MS, Rose M, Hsueh YH, Graham DJ. Risk of acute myocardial infarction, stroke, or death in patients initiating olmesartan or other angiotensin receptor blockers - a cohort study using the clinical practice research datalink. Pharmacoepidemiology and Drug Safety **2014**; 23:340–347.

253. Donnan PT, Dougall HT, Sullivan FM. Optimal strategies for identifying patients with myocardial infarction in general practice. Fam Pract **2003**; 20:706–10.

254. Dherange P, Lang J, Qian P, et al. Arrhythmias and COVID-19: A Review. JACC: Clinical Electrophysiology **2020**; 6:1193–1204.

255. Quan H, Chen G, Walker RL, et al. Incidence, cardiovascular complications and mortality of hypertension by sex and ethnicity. Heart **2013**; 99:715–721. Available at: https://heart.bmj.com/content/99/10/715. Accessed 19 October 2021.

256. Deere BP, Ferdinand KC. Hypertension and race/ethnicity. Curr Opin Cardiol **2020**; 35:342–350. Available at: https://journals.lww.com/co-cardiology/Fulltext/2020/07000/Hypertension_and_race_ethnicity.7.aspx. Accessed 19 October 2021.

257. Claudel SE, Adu-Brimpong J, Banks A, et al. Association between neighborhood-level socioeconomic deprivation and incident hypertension: A longitudinal analysis of data from the Dallas heart study. American Heart Journal **2018**; 204:109–118.

258. Davidson J, Banerjee A, Mathur R, et al. Ethnic differences in the incidence of clinically diagnosed influenza: an England population-based cohort study 2008-2018. Wellcome Open Research **2021**; 6:49. Available at: https://wellcomeopenresearch.org/articles/6-49/v3. Accessed 14 June 2021.

259. Saberian S, Warren-Gash C, Hayward A, Fragaszy E. Effect of socio-economic status on influenza risk in English households - Abstract from OPTIONS X for the Control of Influenza. 2019. Available at: https://isirv.org/site/images/conferences/Optionsx/Options X_Abstracts _Oral and Poster.pdf. Accessed 4 February 2021.

260. Fuchs FD, Chambless LE, Whelton PK, Nieto FJ, Heiss G. Alcohol Consumption and the Incidence of Hypertension. Hypertension **2001**; 37:1242–1250. Available at: https://www.ahajournals.org/doi/abs/10.1161/01.HYP.37.5.1242. Accessed 19 October 2021.

261. Sesso HD, Cook NR, Buring JE, Manson JE, Gaziano JM. Alcohol Consumption and the Risk of Hypertension in Women and Men. Hypertension **2008**; 51:1080–1087. Available at: https://www.ahajournals.org/doi/abs/10.1161/HYPERTENSIONAHA.107.104968. Accessed 19 October 2021.

262. A V, C G, MF N, S T, L G. Cigarette smoking and hypertension. Curr Pharm Des **2010**; 16:2518–2525. Available at: https://pubmed.ncbi.nlm.nih.gov/20550499/. Accessed 19 October 2021.

263. Narkiewicz K. Obesity and hypertension—the issue is more complex than we thought. Nephrology Dialysis Transplantation **2006**; 21:264–267. Available at: https://academic.oup.com/ndt/article/21/2/264/1850864. Accessed 19 October 2021.

264. Cocoros NM, Lash TL, DeMaria A, Klompas M. Obesity as a risk factor for severe influenza-like illness. Influenza and Other Respiratory Viruses **2014**; 8:25–32. Available at: https://onlinelibrary.wiley.com/doi/full/10.1111/irv.12156. Accessed 19 October 2021.

265. Kornum JB, Nørgaard M, Dethlefsen C, et al. Obesity and risk of subsequent hospitalisation with pneumonia. European Respiratory Journal **2010**; 36:1330–1336. Available at: https://erj.ersjournals.com/content/36/6/1330. Accessed 19 October 2021.

266. SAMOKHVALOV A v., IRVING HM, REHM J. Alcohol consumption as a risk factor for pneumonia: a systematic review and meta-analysis. Epidemiology & Infection **2010**; 138:1789–1795. Available at: https://www.cambridge.org/core/journals/epidemiology-and-infection/article/alcohol-consumption-as-a-risk-factor-for-pneumonia-a-systematic-review-and-metaanalysis/90ECC91CEBA682AAF42E3EA55417D719. Accessed 19 October 2021.

267. Han L, Ran J, Mak YW, et al. Smoking and Influenza-associated Morbidity and Mortality: A Systematic Review and Meta-analysis. Epidemiology **2019**; 30:405–417. Available at: https://journals.lww.com/epidem/Fulltext/2019/05000/Smoking_and_Influenza_associated_Morbidity_and.15.aspx. Accessed 19 October 2021.

268. Mahmood SS, Levy D, Vasan RS, Wang TJ. The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. Lancet **2014**; 383:999–1008. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24084292. Accessed 21 March 2019.

269. Verdecchia P, Reboldi G, Gattobigio R, et al. Atrial Fibrillation in Hypertension. Hypertension **2003**; 41:218–223. Available at: https://www.ahajournals.org/doi/abs/10.1161/01.HYP.0000052830.02773.E4. Accessed 19 October 2021.

270. Remschmidt C, Wichmann O, Harder T. Frequency and impact of confounding by indication and healthy vaccinee bias in observational studies assessing influenza vaccine effectiveness: A systematic review. BMC Infectious Diseases **2015**; 15:429. Available at: http://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-015-1154-y. Accessed 3 March 2021.

271. Simonsen L, Taylor RJ, Viboud C, Miller MA, Jackson LA. Mortality benefits of influenza vaccination in elderly people: an ongoing controversy. Lancet Infect Dis **2007**; 7:658–66. Available at: https://linkinghub.elsevier.com/retrieve/pii/S1473309907702360. Accessed 11 February 2019.

272. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behavioral Research **2011**; 46:399–424.

273. Williamson E, Morley R, Lucas A, Carpenter J. Propensity scores: From naïve enthusiasm to intuitive understanding. Statistical Methods in Medical Research **2012**; 21:273–293. Available at: https://journals.sagepub.com/doi/10.1177/0962280210394483?url_ver=Z39.88-2003&rfr_id=ori%3Arid%3Acrossref.org&rfr_dat=cr_pub++0pubmed. Accessed 6 February 2022.

274. Williamson EJ, Forbes A. Introduction to propensity scores. Respirology **2014**; 19:625–635. Available at: https://onlinelibrary.wiley.com/doi/full/10.1111/resp.12312. Accessed 6 February 2022.

275. Walzer P, Estève C, Barben J, et al. Impact of Influenza Vaccination on Mortality in the Oldest Old: A Propensity Score-Matched Cohort Study. Vaccines 2020, Vol 8, Page 356 **2020**; 8:356. Available at: https://www.mdpi.com/2076-393X/8/3/356/htm. Accessed 8 February 2022.

276. Lavallée PC, Labreuche J, Fox KM, et al. Influenza vaccination and cardiovascular risk in patients with recent TIA and stroke. Neurology **2014**; 82:1905–1913. Available at: https://n.neurology.org/content/82/21/1905. Accessed 8 February 2022.

277. Wu HH, Chang YY, Kuo SC, Chen YT. Influenza vaccination and secondary prevention of cardiovascular disease among Taiwanese elders—A propensity score-matched follow-up study. PLOS ONE **2019**; 14:e0219172. Available at: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0219172. Accessed 6 February 2022.

278. Maringe C, Benitez Majano S, Exarchakou A, et al. Reflections on modern methods: trial emulation in the presence of immortal-time bias. Assessing the benefit of major surgery for elderly lung cancer patients using observational data. International Journal of Epidemiology **2020**; Available at: https://academic.oup.com/ije/advance-article/doi/10.1093/ije/dyaa057/5835351. Accessed 28 September 2020.

279. Hernán MA. How to estimate the effect of treatment duration on survival outcomes using observational data. BMJ **2018**; 360:k182. Available at: http://dx.doi.org/10.1136/bmj.k182. Accessed 28 September 2020.

280. Ainslie KEC, Haber M, Orenstein WA. Challenges in estimating influenza vaccine effectiveness. Expert Rev Vaccines **2019**; 18:615. Available at: /pmc/articles/PMC6594904/. Accessed 15 February 2022.

281. Rothwell PM, Coull AJ, Silver LE, et al. Population-based study of event-rate, incidence, case fatality, and mortality for all acute vascular events in all arterial territories (Oxford Vascular Study). The Lancet **2005**; 366:1773–1783. Available at: http://www.thelancet.com/article/S0140673605677021/fulltext. Accessed 27 July 2022.

282. Guo T, Fan Y, Chen M, et al. Cardiovascular Implications of Fatal Outcomes of Patients With Coronavirus Disease 2019 (COVID-19). JAMA Cardiology **2020**; 5:811–818. Available at: https://jamanetwork.com/journals/jamacardiology/fullarticle/2763845. Accessed 27 July 2022.

283. Zekavat SM, Honigberg M, Pirruccello J, et al. Influence of blood pressure on pneumonia risk: Epidemiological association and Mendelian randomisation in the UK Biobank. medRxiv. 2020; Fax:617–726. Available at: https://doi.org/10.1101/2020.04.19.20071936. Accessed 1 February 2021.

284. Neidich SD, Green WD, Rebeles J, et al. Increased risk of influenza among vaccinated adults who are obese. International Journal of Obesity 2017 41:9 **2017**; 41:1324–1330. Available at: https://www.nature.com/articles/ijo2017131. Accessed 22 July 2022.

285. Bae SA, Kim SR, Kim MN, Shim WJ, Park SM. Impact of cardiovascular disease and risk factors on fatal outcomes in patients with COVID-19 according to age: A systematic review and meta-analysis. Heart **2020**; 0:1–8. Available at: http://dx.doi.org/10.1136/heartjnl-2020-317901. Accessed 4 February 2021.

286. Williamson EJ, Walker AJ, Bhaskaran K, et al. Factors associated with COVID-19-related death using OpenSAFELY. Nature 2020 584:7821 **2020**; 584:430–436. Available at: https://www.nature.com/articles/s41586-020-2521-4. Accessed 23 July 2022.

287. Chen L, Han XD, Li YL, Zhang CX, Xing XQ. [Incidence and risk factors for cardiovascular events in patients hospitalized with community-acquired pneumonia]. Zhonghua xin xue Guan Bing za zhi **2020**; 48:228–235. Available at: https://europepmc.org/article/med/32234181. Accessed 22 July 2022.

288. Corrales-Medina VF, Musher DM, Wells GA, Chirinos JA, Chen L, Fine MJ. Cardiac Complications in Patients With Community-Acquired Pneumonia. Circulation **2012**; 125:773–781. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22219349. Accessed 19 July 2019.

289. Clayton TC, Thompson M, Meade TW. Recent respiratory infection and risk of cardiovascular disease: case-control study through a general practice database. European Heart Journal **2007**; 29:96–103. Available at: https://academic.oup.com/eurheartj/article-lookup/doi/10.1093/eurheartj/ehm516. Accessed 1 February 2021.

290. Sanchis-Gomar F, Lavie CJ, Perez-Quilis C, Henry BM, Lippi G. Angiotensin-Converting Enzyme 2 and Antihypertensives (Angiotensin Receptor Blockers and Angiotensin-Converting Enzyme Inhibitors) in Coronavirus Disease 2019. Mayo Clinic Proceedings **2020**; 95:1222. Available at: /pmc/articles/PMC7129862/. Accessed 27 July 2022.

291. Matsushita K, Ding N, Kou M, et al. The relationship of COVID-19 severity with cardiovascular disease and its traditional risk factors: A systematic review and meta-analysis. Global Heart **2020**; 15:64. Available at: /pmc/articles/PMC7546112/?report=abstract. Accessed 4 February 2021.

292. Yang J, Zheng Y, Gou X, et al. Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: a systematic review and meta-analysis. International Journal of Infectious Diseases **2020**; 94:91–95.

293. Basu A, Agwu JC, Barlow N, Lee B. Hypertension is the major predictor of poor outcomes among inpatients with COVID-19 infection in the UK: a retrospective cohort study. BMJ Open **2021**; 11:e047561. Available at: https://bmjopen.bmj.com/content/11/6/e047561. Accessed 22 July 2022.

294. McFarlane E, Linschoten M, Asselbergs FW, Lacy PS, Jedrzejewski D, Williams B. The impact of pre-existing hypertension and its treatment on outcomes in patients admitted to hospital with COVID-19. Hypertension Research 2022 45:5 **2022**; 45:834–845. Available at: https://www.nature.com/articles/s41440-022-00893-5. Accessed 22 July 2022.

295. Barron E, Bakhai C, Kar P, et al. Associations of type 1 and type 2 diabetes with COVID-19-related mortality in England: a whole-population study. The Lancet Diabetes and Endocrinology **2020**; 8:813–822. Available at: http://www.thelancet.com/article/S2213858720302722/fulltext. Accessed 22 July 2022.

296. Ho FK, Petermann-Rocha F, Gray SR, et al. Is older age associated with COVID-19 mortality in the absence of other risk factors? General population cohort study of 470,034 participants. PLOS ONE **2020**; 15:e0241824. Available at: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0241824. Accessed 23 July 2022.

297. Clift AK, Coupland CAC, Keogh RH, et al. Living risk prediction algorithm (QCOVID) for risk of hospital admission and mortality from coronavirus 19 in adults: national derivation and validation cohort study. BMJ **2020**; 371. Available at: https://www.bmj.com/content/371/bmj.m3731. Accessed 23 July 2022.

298. Behrouzi B, Bhatt DL, Cannon CP, et al. Association of Influenza Vaccination With Cardiovascular Risk: A Meta-analysis. JAMA Network Open **2022**; 5:e228873–e228873. Available at: https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2791733. Accessed 23 July 2022.

299. Vardeny O, Kim K, Udell JA, et al. Effect of High-Dose Trivalent vs Standard-Dose Quadrivalent Influenza Vaccine on Mortality or Cardiopulmonary Hospitalization in Patients With High-risk Cardiovascular Disease: A Randomized Clinical Trial. JAMA **2021**; 325:39–49. Available at: https://jamanetwork.com/journals/jama/fullarticle/2773989. Accessed 17 September 2021.

300. Hollingsworth R, Palmu A, Pepin S, et al. Effectiveness of the quadrivalent high-dose influenza vaccine for prevention of cardiovascular and respiratory events in people aged 65 years and above: Rationale and design of a real-world pragmatic randomized clinical trial. American Heart Journal **2021**; 237:54–61.

301. Michos ED, Udell JA. Am I Getting the Influenza Shot Too?: Influenza Vaccination as Post–Myocardial Infarction Care for the Prevention of Cardiovascular Events and Death. Circulation **2021**; 144:1485–1488.

302. Østergaard SD, Schmidt M, Horváth-Puhó E, Thomsen RW, Sørensen HT. Thromboembolism and the Oxford–AstraZeneca COVID-19 vaccine: side-effect or coincidence? The Lancet **2021**; 397:1441–1443. Available at: http://www.thelancet.com/article/S0140673621007625/fulltext. Accessed 23 July 2022.

303. Hippisley-Cox J, Patone M, Mei XW, et al. Risk of thrombocytopenia and thromboembolism after covid-19 vaccination and SARS-CoV-2 positive testing: self-controlled case series study. BMJ **2021**; 374. Available at: https://www.bmj.com/content/374/bmj.n1931. Accessed 23 July 2022.

304. Jabagi MJ, Botton J, Bertrand M, et al. Myocardial Infarction, Stroke, and Pulmonary Embolism After BNT162b2 mRNA COVID-19 Vaccine in People Aged 75 Years or Older. JAMA **2022**; 327:80–82. Available at: https://jamanetwork.com/journals/jama/fullarticle/2786667. Accessed 23 July 2022.

305. Tate AR, Dungey S, Glew S, Beloff N, Williams R, Williams T. Quality of recording of diabetes in the UK: how does the GP's method of coding clinical data affect incidence estimates? Cross-sectional study using the CPRD database. BMJ Open **2017**; 7:e012905. Available at: https://bmjopen.bmj.com/content/7/1/e012905. Accessed 23 July 2022.

306. Clinical Practice Research Datalink. Hospital Episode Statistics (HES) Outpatient Care and CPRD primary care data Documentation (set 21). 2021. Available at: https://doi.org/10.48329/cp5e-7790. Accessed 23 July 2022.

307. Whitaker HJ, Farrington CP, Spiessens B, Musonda P. Tutorial in biostatistics: The self-controlled case series method. Statistics in Medicine **2006**; 25:1768–1797.

308. UK Health Security Agency. Statement of amendments to annual flu letter – 21 July 2022. 2022. Available at: https://www.gov.uk/government/publications/national-flu-immunisation-programme-plan/statement-of-amendments-to-annual-flu-letter-21-july-2022. Accessed 23 July 2022.

309. Visseren FLJ, MacH F, Smulders YM, et al. 2021 ESC Guidelines on cardiovascular disease prevention in clinical practiceDeveloped by the Task Force for cardiovascular disease prevention in clinical practice with representatives of the European Society of Cardiology and 12 medical societies With the special contribution of the European Association of Preventive Cardiology (EAPC). European Heart Journal **2021**; 42:3227–3337. Available at: https://academic.oup.com/eurheartj/article/42/34/3227/6358713. Accessed 23 July 2022.

310. Leong DP, Banerjee A, Yusuf S. COVID-19 Vaccination Prioritization on the Basis of Cardiovascular Risk Factors and Number Needed to Vaccinate to Prevent Death. Canadian Journal of Cardiology **2021**; 37:1112–1116. Available at: http://www.onlinecjc.ca/article/S0828282X2100218X/fulltext. Accessed 27 July 2022.

311. Ding H, Huang J, Ngai CH, et al. The cost-effectiveness of starting 23-valent pneumococcal polysaccharide vaccine and influenza vaccination at 50 vs. 65 years: A comparative modelling study. Vaccine **2022**; 40:1282–1288.

312. Peasah SK, Meltzer MI, Vu M, Moulia DL, Bridges CB. Cost-effectiveness of increased influenza vaccination uptake against readmissions of major adverse cardiac events in the US. PLoS ONE **2019**; 14. Available at: /pmc/articles/PMC6488048/. Accessed 27 July 2022.

313. Menni C, May A, Polidori L, et al. COVID-19 vaccine waning and effectiveness and side-effects of boosters: a prospective community study from the ZOE COVID Study. The Lancet Infectious

Diseases **2022**; 22:1002–1010. Available at: http://www.thelancet.com/article/S1473309922001463/fulltext. Accessed 27 July 2022.

314. Lee JJ, Koshiaris C, Hobbs FDR, Sheppard JP. Beyond COVID-19: respiratory infection and cardiovascular events. British Journal of General Practice **2021**; 71:342–343. Available at: https://bjgp.org/content/71/709/342. Accessed 23 July 2022.

Appendices

Appendix 1 Example do file to create code lists

```
/*=====================================================================
DO FILE NAME:           Hypertension_Gold
AUTHOR:                 Jennifer Davidson
DATE:                   17/10/2019
STUDY:                  PhD Study 1 - risk of MACE by CVD risk level
DESCRIPTION:            Identify atrial fibrillation diagnosis in CPRD Gold
DATASETS USED:          Gold Medical Dictionary, Jul 19 version
DATASETS CREATED:       Hypertension_Gold_Jul19
*=====================================================================*/
***SET UP DIRECTORIES***
global dictionary "J:\EHR Share\3 Database guidelines and info\GPRD_Gold\Medical & Product
Browsers\2019_07_Browsers\forpullrecords"
global codelist "J:\EHR-
Working\Jennifer\ARI_CVcomplications\CodelistDevelopment\StataDatasets\Final"
*************************
***ATRIAL FIBRILLATION***
*************************
use "$dictionary\medical", clear
replace readterm=lower(readterm)
**DEFINE SEARCH TERMS**
loc interm " "*atrial fibrillation*" "*atrial flutter*" "
loc interm "`interm'"
**WORD SEARCH ON MEDICAL DICTIONARY**
gen atrialfib=.
foreach word in `interm'{
        replace atrialfib=1 if strmatch(readterm, "`word'")
       }
br if atrialfib==1
**EXCLUDE TERMS TO MAKE SEARCH MORE SPECIFIC**
loc exterm " "*resolved*" "*except*" "*monitoring*" "*written info*" "*exclude*" "*care pathway*"
"*referral*" "
foreach word in `exterm' {
      replace atrialfib=. if strmatch(readterm,"`word'")
       }
drop if atrialfib==.
*save data
sort medcode
keep medcode readcode readterm atrialfib
save "$codelist\AtrialFibrillation_Gold_Jul19.dta", replace
```

Appendix 2 Chapter 5 supplementary materials

Supplemental Digital Content

Validity of acute cardiovascular outcome diagnoses recorded in European electronic health records: a systematic review

Jennifer A. Davidson[1]; Amitava Banerjee[2]; Rutendo Muzambi[1]; Liam Smeeth[1]; Charlotte Warren-Gash[1]

[1] Faculty of Epidemiology & Population Health, London School of Hygiene and Tropical Medicine, London, UK
[2] Faculty of Population Health Sciences, University College London, London, UK

Contents

S1 Appendix. Search strategy 305

S2 Appendix. Data items extracted 314

S3 Appendix. Adaptation of original QUADAS assessment questions to fit this review 315

S4 Appendix. Grade assessment 316

S5 Appendix. Additional validation results – time, age and sex 317

S6 Appendix. Outcome definitions used in included studies 318

S1 Table. Characteristics of included studies 320

S2 Table. Study results 333

S3 Table. QUADAS-2 assessment of studies 378

S4 Table. Assessment of quality of evidence for outcomes 384

S1 Fig. Positive predictive value for ischaemic stroke diagnoses 386

S2 Fig. Positive predictive value for intracerebral haemorrhage diagnoses 387

S3 Fig. Positive predictive value for subarachnoid haemorrhage diagnoses 388

S4 Fig. Sensitivity for ischaemic stroke diagnoses 389

S5 Fig. Sensitivity for intracerebral haemorrhage diagnoses 390

S6 Fig. Sensitivity for subarachnoid haemorrhage diagnoses 391

S1 Appendix. Search strategy

Medline (OVID)

1. (Europe or Albania or Andorra or Armenia or Austria or Azerbaijan or Balkan* or Belgium or Belarus or Byelarus or Belorussia or Bosnia or Herzegovia or Bulgaria or Croatia or Cyprus or Czechoslovakia or Czech Republic or Denmark or Faeroe Islands or Estonia or Finland or France or Germany or Great Britain or GBR or United Kingdom or UK or (Wales not New South Wales) or (England not New England) or Northern Ireland or Scotland or Channel Islands or Isle of Man or Greece or Gibraltar or Hungary or Iceland or Ireland or Eire or Italy or Latvia or Liechtenstein or Lithuania or Luxembourg or Kosovo or Macedonia or Malta or Mediterranean or Moldova or Monaco or Montenegro or Netherlands or Holland or Norway or Poland or Portugal or Romania or Russia or Russian Federation or USSR or "Union of Soviet Socialist Republics" or Soviet Union or San Marino or Scandinavia or Serbia or Slovakia or Slovak Republic or Slovenia or Spain or Balearic Islands or Canary Islands or Sweden or Switzerland or Ukraine).ti,ab.
2. turkey.ti,ab. not animal/
3. exp Europe/
4. or/1-3
5. exp Stroke/ or exp Brain Infarction/ or exp Cerebral Hemorrhage/ or exp Subarachnoid Hemorrhage/
6. stroke.ti,ab.
7. cerebrovascular accident.ti,ab.
8. ((brain* or cerebr* or intracerebral or intracran* or subarachnoid) adj (infarct* or thrombo* or emboli* or h?emorrhage or h?ematoma or bleed*)).ti,ab.
9. or/5-8
10. exp Myocardial Infarction/ or Acute Coronary Syndrome/
11. (myocardial infarct* or MI or AMI or acute coronary syndrome or ACS).ti,ab.
12. ((cardiac or heart) adj (infarct* or attack* or arrest* or event*)).ti,ab.
13. (stemi or st-segment or st segment or st-elevat* or st elevat*).ti,ab.
14. (nstemi or non-st-segment or non-st segment or non st segment or non-st-elevat* or non-st elevat* or non st elevat*).ti,ab.
15. or/10-14
16. exp Heart Failure/
17. (heart failure or cardiac failure or CCF or left ventricular failure).ti,ab.
18. (left ventricular adj (systolic or diastolic) adj (dysfunction or impairment)).ti,ab.
19. or/16-18
20. 9 or 15 or 19
21. Hospital Records/ or exp Medical Records/
22. ((electronic* or digital* or computer* or longitudinal) adj (health* or medical or clinical or patient) adj (record* or data* or regist*)).ti,ab.
23. (EHR or EPR or EMR or EPD).ti,ab.
24. (routine* collected adj2 data).ti,ab.
25. ((primary care or general practice or secondary care or hospital* or health* or administrative or automated) adj2 (record* or data* or regist*)).ti,ab.
26. exp Clinical Coding/ or exp International Classification of Diseases/
27. ((clinical or medical or read or OXMIS) adj cod*).ti,ab.
28. SNOMED.ti,ab.
29. (International Classification of Diseases or ICD10* or ICD-10* or ICD9* or ICD-9* or International Classification of Primary Care or ICPC).ti,ab.
30. (Clinical Practice Research Datalink or CPRD or General Practice Research Database or GPRD or Value Added Medical Products or VAMP or Hospital Episode Statistics or HES or The Health Improvement Network or QResearch or ResearchOne or Danish National Patient Registry or DNPR or Intego or National Patient Register or NPR or patientregistret or National Inpatient Register or IPR or slutenvardsregistret or Finnish Hospital Discharge Register or FHDR or BIFAP or DIRAYA or Dutch National Basic Registration of Hospital Care or Landelijke Basisregistratie Ziekenhuiszorg or LBZ or Dutch Hospital Discharge Register or Landelijke

Medische registratie or LMR or Mondriaan or Netherlands Primary Care Research Database or NPCRD or Integrated Primary Care Information or IPCI or Information Network of General Practice or LINH or French Hospital Discharge Database or FHDDB or Health Longitudinal Patient Database or HSD).ti,ab.
31. or/21-30
32. Validation Studies/ or exp Reproducibility of Results/ or exp "sensitivity and specificity"/
33. (sensitivity or specificity or positive predictive value or PPV or negative predictive value or NPV).ti,ab.
34. (accura* or consisten* or reliab* or reproduc* or complet* or precis* or concordance or variation or variab* or replicat* or verif* or valid* or predict*).ti,ab.
35. (ROC or receiver operating characteristic or kappa).ti,ab.
36. or/32-35
37. 4 and 20 and 31 and 36

Embase

1. (Europe or Albania or Andorra or Armenia or Austria or Azerbaijan or Balkan* or Belgium or Belarus or Byelarus or Belorussia or Bosnia or Herzegovia or Bulgaria or Croatia or Cyprus or Czechoslovakia or Czech Republic or Denmark or Faeroe Islands or Estonia or Finland or France or Germany or Great Britain or GBR or United Kingdom or UK or (Wales not New South Wales) or (England not New England) or Northern Ireland or Scotland or Channel Islands or Isle of Man or Greece or Gibraltar or Hungary or Iceland or Ireland or Eire or Italy or Latvia or Liechtenstein or Lithuania or Luxembourg or Kosovo or Macedonia or Malta or Mediterranean or Moldova or Monaco or Montenegro or Netherlands or Holland or Norway or Poland or Portugal or Romania or Russia or Russian Federation or USSR or "Union of Soviet Socialist Republics" or Soviet Union or San Marino or Scandinavia or Serbia or Slovakia or Slovak Republic or Slovenia or Spain or Balearic Islands or Canary Islands or Sweden or Switzerland or Ukraine).ti,ab.
2. turkey.ti,ab. not animal/
3. exp Europe/
4. or/1-3
5. exp cerebrovascular accident/ or exp brain infarction/ or exp brain hemorrhage/ or exp subarachnoid hemorrhage/ or exp brain hematoma/
6. stroke.ti,ab.
7. cerebrovascular accident.ti,ab.
8. ((brain* or cerebr* or intracerebral or intracran* or subarachnoid) adj (infarct* or thrombo* or emboli* or h?emorrhage or h?ematoma or bleed*)).ti,ab.
9. or/5-8
10. exp heart infarction/ or acute coronary syndrome/
11. (myocardial infarct* or MI or AMI or acute coronary syndrome or ACS).ti,ab.
12. ((cardiac or heart) adj (infarct* or attack* or arrest* or event*)).ti,ab.
13. (stemi or st-segment or st segment or st-elevat* or st elevat*).ti,ab.
14. (nstemi or non-st-segment or non-st segment or non st segment or non-st-elevat* or non-st elevat* or non st elevat*).ti,ab.
15. or/10-14
16. exp heart failure/
17. (heart failure or cardiac failure or CCF or left ventricular failure).ti,ab.
18. (left ventricular adj (systolic or diastolic) adj (dysfunction or impairment)).ti,ab.
19. or/16-18
20. 9 or 15 or 19
21. exp medical record/
22. ((electronic* or digital* or computer* or longitudinal) adj (health* or medical or clinical or patient) adj (record* or data* or regist*)).ti,ab.
23. (EHR or EPR or EMR or EPD).ti,ab.
24. routine* collected adj2 data).ti,ab.

25. ((primary care or general practice or secondary care or hospital* or health* or administrative or automated) adj2 (record* or data* or regist*)).ti,ab.
26. exp Clinical Coding/ or exp International Classification of Diseases/
27. ((clinical or medical or read or OXMIS) adj cod*).ti,ab.
28. SNOMED.ti,ab.
29. (International Classification of Diseases or ICD10* or ICD-10* or ICD9* or ICD-9* or International Classification of Primary Care or ICPC).ti,ab.
30. (Clinical Practice Research Datalink or CPRD or General Practice Research Database or GPRD or Value Added Medical Products or VAMP or Hospital Episode Statistics or HES or The Health Improvement Network or QResearch or ResearchOne or Danish National Patient Registry or DNPR or Intego or National Patient Register or NPR or patientregistret or National Inpatient Register or IPR or slutenvardsregistret or Finnish Hospital Discharge Register or FHDR or BIFAP or DIRAYA or Dutch National Basic Registration of Hospital Care or Landelijke Basisregistratie Ziekenhuiszorg or LBZ or Dutch Hospital Discharge Register or Landelijke Medische registratie or LMR or Mondriaan or Netherlands Primary Care Research Database or NPCRD or Integrated Primary Care Information or IPCI or Information Network of General Practice or LINH or French Hospital Discharge Database or FHDDB or Health Longitudinal Patient Database or HSD).ti,ab.
31. or/21-30
32. exp validation study/ or exp predictive value/ or exp reproducibility/ or exp "sensitivity and specificity"/
33. (sensitivity or specificity or positive predictive value or PPV or negative predictive value or NPV).ti,ab.
34. (accura* or consisten* or reliab* or reproduc* or complet* or precis* or concordance or variation or variab* or replicat* or verif* or valid* or predict*).ti,ab.
35. (ROC or receiver operating characteristic or kappa).ti,ab.
36. or/32-35
37. 4 and 20 and 31 and 36

Scopus

(TITLE-ABS KEY (europe OR albania OR andorra OR armenia OR austria OR azerbaijan OR balkan* OR belgium OR belarus OR byelarus OR belorussia OR bosnia OR herzegovia OR bulgaria OR croatia OR cyprus OR czechoslovakia OR "Czech Republic" OR denmark OR "Faeroe Islands" OR estonia OR finland OR france OR germany OR "Great Britain" OR gbr OR "United Kingdom" OR uk OR "Northern Ireland" OR scotland OR "Channel Islands" OR "Isle of Man" OR greece OR gibraltar OR hungary OR iceland OR ireland OR eire OR italy OR latvia OR liechtenstein OR lithuania OR luxembourg OR kosovo OR macedonia OR malta OR mediterranean OR moldova OR monaco OR montenegro OR netherlands OR holland OR norway OR poland OR portugal OR romania OR russia OR "Russian Federation" OR ussr OR "Union of Soviet Socialist Republics" OR "Soviet Union" OR "San Marino" OR scandinavia OR serbia OR slovakia OR "Slovak Republic" OR slovenia OR spain OR "Balearic Islands" OR "Canary Islands" OR sweden OR switzerland OR turkey OR ukraine)) OR (TITLE-ABS-KEY (wales AND NOT "New South Wales")) OR (TITLE-ABS-KEY (england AND NOT "New England")) AND

(TITLE-ABS-KEY (stroke OR "cerebrovascular accident")) OR (TITLE-ABS-KEY (brain* OR cerebr* OR intracerebral OR intracran* OR subarachnoid PRE/0 infarct* OR thrombo* OR emboli* OR h*emorrhage OR h*ematoma* OR bleed*)) OR (TITLE-ABS-KEY ("myocardial infarct*" OR mi OR ami OR "acute coronary syndrome" OR acs OR stemi OR "st segment" OR "st elevat*" OR nstemi OR "non st segment" OR "non st elevat*")) OR (TITLE-ABS-KEY (cardiac OR heart PRE/0 infarct* OR attack* OR arrest* OR event*)) OR (TITLE-ABS-KEY ("heart failure" OR "cardiac failure" OR ccf OR "left ventricular failure")) OR (TITLE-ABS-KEY ("left ventricular" PRE/0 systolic OR diastolic PRE/0 dysfunction OR impairment)) AND

(TITLE-ABS-KEY (electronic* OR digital* OR computer* OR longitudinal PRE/0 health* OR medical OR clinical OR patient PRE/0 record* OR data* OR regist*)) OR (TITLE-ABS-KEY (ehr OR epr OR emr OR epd)) OR (TITLE-ABS-KEY ("routine* collected" PRE/2 data)) OR (TITLE-ABS-KEY (clinical OR medical OR read OR oxmis PRE/0 cod*)) OR (TITLE-ABS-KEY (snomed OR {International Classification

of Diseases} OR icd10* OR icd-10* OR icd9* OR icd-9* OR {International Classification of Primary Care} OR icpc OR {Clinical Practice Research Datalink} OR cprd OR {General Practice Research Database} OR gprd OR {Value Added Medical Products} OR vamp OR {Hospital Episode Statistics} OR hes OR {The Health Improvement Network} OR qresearch OR researchone OR {Danish National Patient Registry} OR dnpr OR intego OR {National Patient Register} OR npr OR patientregistret OR {National Inpatient Register} OR ipr OR slutenvardsregistret OR {Finnish Hospital Discharge Register} OR fhdr OR bifap OR diraya OR {Dutch National Basic Registration of Hospital Care} OR {Landelijke Basisregistratie Ziekenhuiszorg} OR lbz OR {Dutch Hospital Discharge Register} OR {Landelijke Medische registratie} OR lmr OR mondriaan OR {Netherlands Primary Care Research Database} OR npcrd OR {Integrated Primary Care Information} OR ipci OR {Information Network of General Practice} OR linh OR {French Hospital Discharge Database} OR fhddb OR {Health Longitudinal Patient Database} OR hsd)) AND

(TITLE-ABS-KEY (sensitivity OR specificity OR "positive predictive value" OR ppv OR "negative predictive value" OR npv OR accura* OR consisten* OR reliab* OR reproduc* OR complet* OR precis* OR concordance OR variation OR variab* OR replicat* OR verif* OR valid* OR predict* OR roc OR "receiver operating characteristic" OR kappa))

Web of Science

1. TS=(Europe or Albania or Andorra or Armenia or Austria or Azerbaijan or Balkan* or Belgium or Belarus or Byelarus or Belorussia or Bosnia or Herzegovia or Bulgaria or Croatia or Cyprus or Czechoslovakia or "Czech Republic" or Denmark or "Faeroe Islands" or Estonia or Finland or France or Germany or "Great Britain" or GBR or "United Kingdom" or UK or "Northern Ireland" or Scotland or "Channel Islands" or "Isle of Man" or Greece or Gibraltar or Hungary or Iceland or Ireland or Eire or Italy or Latvia or Liechtenstein or Lithuania or Luxembourg or Kosovo or Macedonia or Malta or Mediterranean or Moldova or Monaco or Montenegro or Netherlands or Holland or Norway or Poland or Portugal or Romania or Russia or "Russian Federation" or USSR or "Union of Soviet Socialist Republics" or "Soviet Union" or "San Marino" or Scandinavia or Serbia or Slovakia or "Slovak Republic" or Slovenia or Spain or "Balearic Islands" or "Canary Islands" or Sweden or Switzerland or Turkey or Ukraine)
2. TS=(Wales NOT "New South Wales")
3. TS=(England NOT "New England")
4. #1 or #2 or #3
5. TS=(stroke or "cerebrovascular accident" or "brain* infarct*" or "brain* thrombo*" or "brain* emboli*" or "brain* h?emorrhage" or "brain* h?ematoma" or "brain* bleed*" or "cerebr* infarct*" or "cerebr* thrombo*" or "cerebr* emboli*" or "cerebr* h?emorrhage" or "cerebr* h?ematoma" or "cerebr* bleed*" or "intracerebral infarct*" or "intracerebral thrombo*" or "intracerebral emboli*" or "intracerebral h?emorrhage" or "intracerebral h?ematoma" or "intracerebral bleed*" or " intracran* infarct*" or " intracran* thrombo*" or " intracran* emboli*" or " intracran* h?emorrhage" or " intracran* h?ematoma" or " intracran* bleed*")
6. TS=("myocardial infarct*" or mi or ami or "cardiac infarct*" or "cardiac attack*" or "cardiac arrest*" or "cardiac event*" or "heart infarct*" or "heart attack*" or "heart arrest*" or "heart event*" or stemi or "st-segment" or "st-elevat*" or nstemi or "non-st-segment" or "non-st-elevat*" or "acute coronary syndrome" or ACS)
7. TS=("heart failure" or "cardiac failure" or CCF or "left ventricular failure" or "left ventricular systolic dysfunction" or "left ventricular systolic impairment" or "left ventricular diastolic dysfunction" or "left ventricular diastolic impairment")
8. #5 or #6 or #7
9. TS=("electronic* health* record*" or "electronic* health* data*" or "electronic* health* regist*" or "digital* health* record*" or "digital* health* data*" or "digital* health* regist*" or "computer* health* record*" or "computer* health* data*" or "computer* health* regist*" or "longitudinal health* record*" or "longitudinal health* data*" or "longitudinal health* regist*" or "electronic* medical record*" or "electronic* medical data*" or "electronic* medical regist*" or "digital* medical record*" or "digital* medical data*" or "digital* medical regist*" or "computer* medical record*" or "computer* medical data*" or "computer* medical regist*" or "longitudinal medical record*" or "longitudinal medical data*" or "longitudinal medical regist*" or "electronic*

clinical record*" or "electronic* clinical data*" or "electronic* clinical regist*" or "digital* clinical record*" or "digital* clinical data*" or "digital* clinical regist*" or "computer* clinical record*" or "computer* clinical data*" or "computer* clinical regist*" or "longitudinal clinical record*" or "longitudinal clinical data*" or "longitudinal clinical regist*" or EHR or EPR or EMR or EPD)

10. TS=("routine* collected" NEAR/2 data or "primary care" NEAR/2 record* or "general practice" NEAR/2 record* or "secondary care" NEAR/2 record* or hospital* NEAR/2 record* or health* NEAR/2 record* or administrative NEAR/2 record* or automated NEAR/2 record* or "primary care" NEAR/2 data* or "general practice" NEAR/2 data* or "secondary care" NEAR/2 data* or hospital* NEAR/2 data* or health* NEAR/2 data* or administrative NEAR/2 data* or automated NEAR/2 data* or "primary care" NEAR/2 regist* or "general practice" NEAR/2 regist* or "secondary care" NEAR/2 regist* or hospital* NEAR/2 regist* or health* NEAR/2 regist* or administrative NEAR/2 regist* or automated NEAR/2 regist*)
11. TS=("clinical cod*" or "medical cod*" or "read cod*" or "oxmis cod*" or SNOMED or "International Classification of Diseases" or ICD-10* or ICD-9* or "International Classification of Primary Care" or ICPC)
12. TS=("Clinical Practice Research Datalink" or CPRD or "General Practice Research Database" or GPRD or "Value Added Medical Products" or VAMP or "Hospital Episode Statistics" or HES or "The Health Improvement Network" or QResearch or ResearchOne or "Danish National Patient Registry" or DNPR or Intego or "National Patient Register" or NPR or patientregistret or "National Inpatient Register" or IPR or slutenvardsregistret or "Finnish Hospital Discharge Register" or FHDR or BIFAP or DIRAYA or "Dutch National Basic Registration of Hospital Care" or "Landelijke Basisregistratie Ziekenhuiszorg" or LBZ or "Dutch Hospital Discharge Register" or "Landelijke Medische registratie" or LMR or Mondriaan or "Netherlands Primary Care Research Database" or NPCRD or "Integrated Primary Care Information" or IPCI or "Information Network of General Practice" or LINH or "French Hospital Discharge Database" or FHDDB or "Health Longitudinal Patient Database" or HSD)
13. #9 or #10 or #11 or #12
14. TS=(sensitivity or specificity or "positive predictive value" or PPV or "negative predictive value" or NPV or accura* or consisten* or reliab* or reproduc* or complet* or precis* or concordance or variation or variab* replicat* or verif* or valid* or predict* or ROC or "receiver operating characteristic" or kappa)
15. #4 and #8 and #13 and #14

Cochrane Library

1. MeSH descriptor: [Europe] explode all trees
2. (Europe or Albania or Andorra or Armenia or Austria or Azerbaijan or Balkan* or Belgium or Belarus or Byelarus or Belorussia or Bosnia or Herzegovia or Bulgaria or Croatia or Cyprus or Czechoslovakia or Czech Republic or Denmark or Faeroe Islands or Estonia or Finland or France or Germany or Great Britain or GBR or United Kingdom or UK or Wales or England or Northern Ireland or Scotland or Channel Islands or Isle of Man or Greece or Gibraltar or Hungary or Iceland or Ireland or Eire or Italy or Latvia or Liechtenstein or Lithuania or Luxembourg or Kosovo or Macedonia or Malta or Mediterranean or Moldova or Monaco or Montenegro or Netherlands or Holland or Norway or Poland or Portugal or Romania or Russia or Russian Federation or USSR or "Union of Soviet Socialist Republics" or Soviet Union or San Marino or Scandinavia or Serbia or Slovakia or Slovak Republic or Slovenia or Spain or Balearic Islands or Canary Islands or Sweden or Switzerland or Turkey or Ukraine):ti,ab,kw
3. #1 or #2
4. MeSH descriptor: [Stroke] explode all trees
5. MeSH descriptor: [Brain Infarction] explode all trees
6. MeSH descriptor: [Cerebral Hemorrhage] explode all trees
7. MeSH descriptor: [Subarachnoid Hemorrhage] explode all trees
8. (stroke or "cerebrovascular accident" or "brain* infarct*" or "brain* thrombo*" or "brain* emboli*" or "brain* h?emorrhage" or "brain* h?ematoma" or "brain* bleed*" or "cerebr* infarct*" or "cerebr* thrombo*" or "cerebr* emboli*" or "cerebr* h?emorrhage" or "cerebr* h?ematoma" or "cerebr* bleed*" or "intracerebral infarct*" or "intracerebral thrombo*" or "intracerebral emboli*" or "intracerebral h?emorrhage" or "intracerebral h?ematoma" or "intracerebral bleed*" or " intracran* infarct*" or " intracran* thrombo*" or " intracran* emboli*" or " intracran* h?emorrhage" or " intracran* h?ematoma" or " intracran* bleed*"):ti,ab,kw
9. MeSH descriptor: [Acute Coronary Syndrome] explode all trees

10. MeSH descriptor: [Myocardial Infarction] explode all trees
11. ("myocardial infarct*" or mi or ami or "cardiac infarct*" or "cardiac attack*" or "cardiac arrest*" or "cardiac event*" or "heart infarct*" or "heart attack*" or "heart arrest*" or "heart event*" or stemi or "st-segment" or "st segment" or "st-elevat*" or "st elevat*" or nstemi or "non-st-segment" or "non-st segment" or "non st segment" or "non-st-elevat*" or "non-st elevat*" or "non st elevat*" or "acute coronary syndrome" or ACS):ti,ab,kw
12. MeSH descriptor: [Heart Failure] explode all trees
13. ("heart failure" or "cardiac failure" or CCF or "left ventricular failure" or "left ventricular systolic dysfunction" or "left ventricular systolic impairment" or "left ventricular diastolic dysfunction" or "left ventricular diastolic impairment"):ti,ab,kw
14. #4 or #5 or #6 or #7 or #8 or #9 or #10 or #11 or #12 or #13
15. MeSH descriptor: [Hospital Records] explode all trees
16. MeSH descriptor: [Medical Records] explode all trees
17. ("electronic* health* record*" or "electronic* health* data*" or "electronic* health* regist*" or "digital* health* record*" or "digital* health* data*" or "digital* health* regist*" or "computer* health* record*" or "computer* health* data*" or "computer* health* regist*" or "longitudinal health* record*" or "longitudinal health* data*" or "longitudinal health* regist*" or "electronic* medical record*" or "electronic* medical data*" or "electronic* medical regist*" or "digital* medical record*" or "digital* medical data*" or "digital* medical regist*" or "computer* medical record*" or "computer* medical data*" or "computer* medical regist*" or "longitudinal medical record*" or "longitudinal medical data*" or "longitudinal medical regist*" or "electronic* clinical record*" or "electronic* clinical data*" or "electronic* clinical regist*" or "digital* clinical record*" or "digital* clinical data*" or "digital* clinical regist*" or "computer* clinical record*" or "computer* clinical data*" or "computer* clinical regist*" or "longitudinal clinical record*" or "longitudinal clinical data*" or "longitudinal clinical regist*" or EHR or EPR or EMR or EPD):ti,ab,kw
18. ("routine* collected" NEAR/2 data or "primary care" NEAR/2 record* or "general practice" NEAR/2 record* or "secondary care" NEAR/2 record* or hospital* NEAR/2 record* or health* NEAR/2 record* or administrative NEAR/2 record* or automated NEAR/2 record* or "primary care" NEAR/2 data* or "general practice" NEAR/2 data* or "secondary care" NEAR/2 data* or hospital* NEAR/2 data* or health* NEAR/2 data* or administrative NEAR/2 data* or automated NEAR/2 data* or "primary care" NEAR/2 regist* or "general practice" NEAR/2 regist* or "secondary care" NEAR/2 regist* or hospital* NEAR/2 regist* or health* NEAR/2 regist* or administrative NEAR/2 regist* or automated NEAR/2 regist*):ti,ab,kw
19. ("clinical cod*" or "medical cod*" or "read cod*" or "oxmis cod*" or SNOMED or "International Classification of Diseases" or ICD-10* or ICD-9* or "International Classification of Primary Care" or ICPC):ti,ab,kw
20. ("Clinical Practice Research Datalink" or CPRD or "General Practice Research Database" or GPRD or "Value Added Medical Products" or VAMP or "Hospital Episode Statistics" or HES or "The Health Improvement Network" or QResearch or ResearchOne or "Danish National Patient Registry" or DNPR or Intego or "National Patient Register" or NPR or patientregistret or "National Inpatient Register" or IPR or slutenvardsregistret or "Finnish Hospital Discharge Register" or FHDR or BIFAP or DIRAYA or "Dutch National Basic Registration of Hospital Care" or "Landelijke Basisregistratie Ziekenhuiszorg" or LBZ or "Dutch Hospital Discharge Register" or "Landelijke Medische registratie" or LMR or Mondriaan or "Netherlands Primary Care Research Database" or NPCRD or "Integrated Primary Care Information" or IPCI or "Information Network of General Practice" or LINH or "French Hospital Discharge Database" or FHDDB or "Health Longitudinal Patient Database" or HSD):ti,ab,kw
21. #15 or #16 or #17 or #18 or #19 or #20
22. MeSHdescriptor [Validation Studies] explode all trees
23. MeSH descriptor [Reproducibility of Results] explode all trees
24. MeSH descriptor [Sensitivity and Specificity] explode all trees
25. (sensitivity or specificity or "positive predictive value" or PPV or "negative predictive value" or NPV or accura* or consisten* or reliab* or reproduc* or complet* or precis* or concordance or variation or variab* replicat* or verif* or valid* or predict* or ROC or "receiver operating characteristic" or kappa):ti,ab,kw
26. #22 or #23 or #24 or #25
27. #3 and #14 and #21 and #26

Open Grey

1. stroke and electronic* NEAR/2 data* and valid*
2. stroke and "routine* collected" NEAR/2 data* and valid*

3. stroke and "primary care" NEAR/2 data* and valid*
4. stroke and "general practice" NEAR/2 data* and valid*
5. stroke and "secondary care" NEAR/2 data* and valid*
6. stroke and hospital* NEAR/2 data* and valid*
7. stroke and administrative NEAR/2 data* and valid*
8. stroke and electronic* NEAR/2 record* and valid*
9. stroke and "primary care" NEAR/2 record* and valid*
10. stroke and "general practice" NEAR/2 record* and valid*
11. stroke and "secondary care" NEAR/2 record* and valid*
12. stroke and hospital* NEAR/2 record* and valid*
13. stroke and administrative* NEAR/2 record* and valid*
14. stroke and electronic* NEAR/2 data* and complet*
15. stroke and "routine* collected" NEAR/2 data* and complet*
16. stroke and "primary care" NEAR/2 data* and complet*
17. stroke and "general practice" NEAR/2 data* and complet*
18. stroke and "secondary care" NEAR/2 data* and complet*
19. stroke and hospital* NEAR/2 data* and complet*
20. stroke and administrative NEAR/2 data* and complet*
21. stroke and electronic* NEAR/2 record* and complet*
22. stroke and "primary care" NEAR/2 record* and complet*
23. stroke and "general practice" NEAR/2 record* and complet*
24. stroke and "secondary care" NEAR/2 record* and complet*
25. stroke and hospital* NEAR/2 record* and complet*
26. stroke and administrative* NEAR/2 record* and complet*
27. "acute coronary syndrome" and electronic* NEAR/2 data* and valid*
28. "acute coronary syndrome" and "routine* collected" NEAR/2 data* and valid*
29. "acute coronary syndrome" and "primary care" NEAR/2 data* and valid*
30. "acute coronary syndrome" and "general practice" NEAR/2 data* and valid*
31. "acute coronary syndrome" and "secondary care" NEAR/2 data* and valid*
32. "acute coronary syndrome" and hospital* NEAR/2 data* and valid*
33. "acute coronary syndrome" and administrative NEAR/2 data* and valid*
34. "acute coronary syndrome" and electronic* NEAR/2 record* and valid*
35. "acute coronary syndrome" and "primary care" NEAR/2 record* and valid*
36. "acute coronary syndrome" and "general practice" NEAR/2 record* and valid*
37. "acute coronary syndrome" and "secondary care" NEAR/2 record* and valid*
38. "acute coronary syndrome" and hospital* NEAR/2 record* and valid*
39. "acute coronary syndrome" and administrative NEAR/2 record* and valid*
40. "acute coronary syndrome" and electronic* NEAR/2 data* and complet*
41. "acute coronary syndrome" and "routine* collected" NEAR/2 data* and complet*
42. "acute coronary syndrome" and "primary care" NEAR/2 data* and complet*
43. "acute coronary syndrome" and "general practice" NEAR/2 data* and complet*
44. "acute coronary syndrome" and "secondary care" NEAR/2 data* and complet*
45. "acute coronary syndrome" and hospital* NEAR/2 data* and complet*
46. "acute coronary syndrome" and administrative NEAR/2 data* and complet*
47. "acute coronary syndrome" and electronic* NEAR/2 record* and complet*
48. "acute coronary syndrome" and "primary care" NEAR/2 record* and complet*
49. "acute coronary syndrome" and "general practice" NEAR/2 record* and complet*
50. "acute coronary syndrome" and "secondary care" NEAR/2 record* and complet*
51. "acute coronary syndrome" and hospital* NEAR/2 record* and complet*
52. "acute coronary syndrome" and administrative NEAR/2 record* and complet*
53. "myocardial infarction" and electronic* NEAR/2 data* and valid*
54. "myocardial infarction" and "routine* collected" NEAR/2 data* and valid*
55. "myocardial infarction" and "primary care" NEAR/2 data* and valid*
56. "myocardial infarction" and "general practice" NEAR/2 data* and valid*
57. "myocardial infarction" and "secondary care" NEAR/2 data* and valid*
58. "myocardial infarction" and hospital* NEAR/2 data* and valid*
59. "myocardial infarction" and administrative NEAR/2 data* and valid*

60. "myocardial infarction" and electronic* NEAR/2 record* and valid*
61. "myocardial infarction" and "primary care" NEAR/2 record* and valid*
62. "myocardial infarction" and "general practice" NEAR/2 record* and valid*
63. "myocardial infarction" and "secondary care" NEAR/2 record* and valid*
64. "myocardial infarction" and hospital* NEAR/2 record* and valid*
65. "myocardial infarction" and administrative NEAR/2 record* and valid*
66. "myocardial infarction" and electronic* NEAR/2 data* and complet*
67. "myocardial infarction" and "routine* collected" NEAR/2 data* and complet*
68. "myocardial infarction" and "primary care" NEAR/2 data* and complet*
69. "myocardial infarction" and "general practice" NEAR/2 data* and complet*
70. "myocardial infarction" and "secondary care" NEAR/2 data* and complet*
71. "myocardial infarction" and hospital* NEAR/2 data* and complet*
72. "myocardial infarction" and administrative NEAR/2 data* and complet*
73. "myocardial infarction" and electronic* NEAR/2 record* and complet*
74. "myocardial infarction" and "primary care" NEAR/2 record* and complet*
75. "myocardial infarction" and "general practice" NEAR/2 record* and complet*
76. "myocardial infarction" and "secondary care" NEAR/2 record* and complet*
77. "myocardial infarction" and hospital* NEAR/2 record* and complet*
78. "myocardial infarction" and administrative NEAR/2 record* and complet*
79. "heart failure" and electronic* NEAR/2 data* and valid*
80. "heart failure" and "routine* collected" NEAR/2 data* and valid*
81. "heart failure" and "primary care" NEAR/2 data* and valid*
82. "heart failure" and "general practice" NEAR/2 data* and valid*
83. "heart failure" and "secondary care" NEAR/2 data* and valid*
84. "heart failure" and hospital* NEAR/2 data* and valid*
85. "heart failure" and administrative NEAR/2 data* and valid*
86. "heart failure" and electronic* NEAR/2 record* and valid*
87. "heart failure" and "primary care" NEAR/2 record* and valid*
88. "heart failure" and "general practice" NEAR/2 record* and valid*
89. "heart failure" and "secondary care" NEAR/2 record* and valid*
90. "heart failure" and hospital* NEAR/2 record* and valid*
91. "heart failure" and administrative* NEAR/2 record* and valid*
92. "heart failure" and electronic* NEAR/2 data* and complet*
93. "heart failure" and "routine* collected" NEAR/2 data* and complet*
94. "heart failure" and "primary care" NEAR/2 data* and complet*
95. "heart failure" and "general practice" NEAR/2 data* and complet*
96. "heart failure" and "secondary care" NEAR/2 data* and complet*
97. "heart failure" and hospital* NEAR/2 data* and complet*
98. "heart failure" and administrative* NEAR/2 data* and complet*
99. "heart failure" and electronic* NEAR/2 record* and complet*
100. "heart failure" and "primary care" NEAR/2 record* and complet*
101. "heart failure" and "general practice" NEAR/2 record* and complet*
102. "heart failure" and "secondary care" NEAR/2 record* and complet*
103. "heart failure" and hospital* NEAR/2 record* and complet*
104. "heart failure" and administrative* NEAR/2 record* and complet*

EThOS

1. stroke and "electronic* health* data*" and valid*
2. stroke and "routine* collected data*" and valid*
3. stroke and "primary care data*" and valid*
4. stroke and "general practice data*" and valid*
5. stroke and "secondary care data*" and valid*
6. stroke and "hospital* data*" and valid*
7. stroke and "administrative health* data*" and valid*

8. stroke and "electronic* health* record*" and valid*
9. stroke and "primary care record*" and valid*
10. stroke and "general practice record*" and valid*
11. stroke and "secondary care record*" and valid*
12. stroke and "hospital* record*" and valid*
13. stroke and "administrative health* record*" and valid*
14. stroke and "electronic* health* data*" and complet*
15. stroke and "routine* collected data*" and complet*
16. stroke and "primary care data*" and complet*
17. stroke and "general practice data*" and complet*
18. stroke and "secondary care data*" and complet*
19. stroke and "hospital* data*" and complet*
20. stroke and "administrative health* data*" and complet*
21. stroke and "electronic* health* record*" and complet*
22. stroke and "primary care record*" and complet*
23. stroke and "general practice record*" and complet*
24. stroke and "secondary care record*" and complet*
25. stroke and "hospital* record*" and complet*
26. stroke and "administrative* health* record*" and complet*
27. "acute coronary syndrome" and "electronic* health* data*" and valid*
28. "acute coronary syndrome" and "routine* collected data*" and valid*
29. "acute coronary syndrome" and "primary care data*" and valid*
30. "acute coronary syndrome" and "general practice data*" and valid*
31. "acute coronary syndrome" and "secondary care data*" and valid*
32. "acute coronary syndrome" and "hospital* data*" and valid*
33. "acute coronary syndrome" and "administrative health* data*" and valid*
34. "acute coronary syndrome" and "electronic* health* record*" and valid*
35. "acute coronary syndrome" and "primary care record*" and valid*
36. "acute coronary syndrome" and "general practice record*" and valid*
37. "acute coronary syndrome" and "secondary care record*" and valid*
38. "acute coronary syndrome" and "hospital* record*" and valid*
39. "acute coronary syndrome" and "administrative* health* record*" and valid*
40. "acute coronary syndrome" and "electronic* health* data*" and complet*
41. "acute coronary syndrome" and "routine* collected data*" and complet*
42. "acute coronary syndrome" and "primary care data*" and complet*
43. "acute coronary syndrome" and "general practice data*" and complet*
44. "acute coronary syndrome" and "secondary care data*" and complet*
45. "acute coronary syndrome" and "hospital* data*" and complet*
46. "acute coronary syndrome" and "administrative health* data*" and complet*
47. "acute coronary syndrome" and "electronic* health* record*" and complet*
48. "acute coronary syndrome" and "primary care record*" and complet*
49. "acute coronary syndrome" and "general practice record*" and complet*
50. "acute coronary syndrome" and "secondary care record*" and complet*
51. "acute coronary syndrome" and "hospital* record*" and complet*
52. "acute coronary syndrome" and "administrative* health* record*" and complet*
53. "myocardial infarction" and "electronic* health* data*" and valid*
54. "myocardial infarction" and "routine* collected data*" and valid*
55. "myocardial infarction" and "primary care data*" and valid*
56. "myocardial infarction" and "general practice data*" and valid*
57. "myocardial infarction" and "secondary care data*" and valid*
58. "myocardial infarction" and "hospital* data*" and valid*
59. "myocardial infarction" and "administrative health* data*" and valid*

60. "myocardial infarction" and "electronic* health* record*" and valid*
61. "myocardial infarction" and "primary care record*" and valid*
62. "myocardial infarction" and "general practice record*" and valid*
63. "myocardial infarction" and "secondary care record*" and valid*
64. "myocardial infarction" and "hospital* record*" and valid*
65. "myocardial infarction" and "administrative* health* record*" and valid*
66. "myocardial infarction" and "electronic* health* data*" and complet*
67. "myocardial infarction" and "routine* collected data*" and complet*
68. "myocardial infarction" and "primary care data*" and complet*
69. "myocardial infarction" and "general practice data*" and complet*
70. "myocardial infarction" and "secondary care data*" and complet*
71. "myocardial infarction" and "hospital* data*" and complet*
72. "myocardial infarction" and "administrative health* data*" and complet*
73. "myocardial infarction" and "electronic* health* record*" and complet*
74. "myocardial infarction" and "primary care record*" and complet*
75. "myocardial infarction" and "general practice record*" and complet*
76. "myocardial infarction" and "secondary care record*" and complet*
77. "myocardial infarction" and "hospital* record*" and complet*
78. "myocardial infarction" and "administrative* health* record*" and complet*
79. "heart failure" and "electronic* health* data*" and valid*
80. "heart failure" and "routine* collected data*" and valid*
81. "heart failure" and "primary care data*" and valid*
82. "heart failure" and "general practice data*" and valid*
83. "heart failure" and "secondary care data*" and valid*
84. "heart failure" and "hospital* data*" and valid*
85. "heart failure" and "administrative health* data*" and valid*
86. "heart failure" and "electronic* health* record*" and valid*
87. "heart failure" and "primary care record*" and valid*
88. "heart failure" and "general practice record*" and valid*
89. "heart failure" and "secondary care record*" and valid*
90. "heart failure" and "hospital* record*" and valid*
91. "heart failure" and "administrative* health* record*" and valid*
92. "heart failure" and "electronic* health* data*" and complet*
93. "heart failure" and "routine* collected data*" and complet*
94. "heart failure" and "primary care data*" and complet*
95. "heart failure" and "general practice data*" and complet*
96. "heart failure" and "secondary care data*" and complet*
97. "heart failure" and "hospital* data*" and complet*
98. "heart failure" and "administrative health* data*" and complet*
99. "heart failure" and "electronic* health* record*" and complet*
100. "heart failure" and "primary care record*" and complet*
101. "heart failure" and "general practice record*" and complet*
102. "heart failure" and "secondary care record*" and complet*
103. "heart failure" and "hospital* record*" and complet*
104. "heart failure" and "administrative* health* record*" and complet*

S2 Appendix. Data items extracted

1. *Population:* general description, age (mean/median), sex, inclusion criteria, exclusion criteria;
2. *Index:* EHR name, setting (primary or secondary care), coding system, coverage (regional or national), diagnoses validated, codes validated;
3. *Comparators:* method of validation, description of method, definition of diagnosis used;

4. *Outcomes:* number of validations planned, number of validations conducted, main measures of validity (sensitivity, specificity, positive predictive value, negative predictive value), main raw data (true positives, false positives, true negatives, false negatives), individual codes validity (and raw data), age and sex validity (and raw data), time period validity (and raw data), diagnostic position validity (and raw data), any other relevant measures of validity;
5. *Study characteristics:* authors, publication year, study period (years), country, study aim, source of participants.

S3 Appendix. Adaptation of original QUADAS assessment questions to fit this review[18]

Patient selection domain

1. Was a consecutive/ random sample of participants enrolled?
2. Did the study avoid inappropriate exclusions?
3. Were the included patients representative of the patients who will receive the diagnosis in practice?
4. Were selection criteria clearly described?

Index test domain

1. Were the index test results interpreted without knowledge of the results using the reference standard?
2. Was the execution of the index test described in sufficient detail to permit replication?

Reference standard domain

1. Is the reference standard likely to correctly classify the target condition?
2. Were the reference standard results interpreted without knowledge of the results of the index test?
3. Was the execution of the reference standard described in sufficient detail to permit replication?
4. Were the same clinical data available when results were interpreted as would be available in practice?

Flow and timing domain

1. Is the interval between index test(s) and reference standard appropriate? i.e. source of reference standard before time point in index test but time period short enough to be reasonably sure that the target condition did not change between the two?
2. Did all patients receive a reference standard? If not were the reasons explained?
3. Did all participants receive the same reference standard (regardless of index test result)?
4. Were all participants included in the analysis? If not were exclusions and reasons explained?

S4 Appendix. Grade assessment

	Scoring
Risk of bias	• Not serious: If >50% the studies have no QUADAS-2 domain at high risk of bias • Serious: If studies fall between not serious and very serious • Very serious: If studies with ≥2 QUADAS-2 domains at high risk of bias represent >50% of the total studies
Inconsistency	Not serious if have 0, serious if have 1, and very serious if have ≥2 of: • Heterogeneity is substantial (50-90%) or considerable (75-100%) • Wide variance in sensitivity or PPV across studies • Minimal/no overlap in sensitivity or PPV confidence intervals
Imprecision	• Not serious: Narrow confidence intervals • Serious: Wide confidence intervals • Very serious: Very wide confidence intervals
Large magnitude of effect	• Upgrade: If >90% of studies have validity estimate of >80%
Gradient of result	• Upgrade: If studies with definite, probable and/or possible diagnostic categories show a suitable gradient in the sensitivity or PPV

S5 Appendix. Additional validation results – time, age and sex

Heart failure

There was no apparent trend in validity estimates over time; in 1986, a study reported a sensitivity of 65% and PPV 87%,[88] while corresponding results from 2014 were 64% and 88%.[77] Where individual studies stratified by time period,[29,43,65,77,83] results were mixed with some increases and decreases in validity estimates.

By age breakdown, studies[29,43,65] obtaining differing results. Merry *et al* found lower sensitivity and PPV in patients aged <50 years (30% and 60%, respectively) than those ≥50 years (44% and 81%, respectively). Pfister *et al* reported no difference in the PPV for patients aged <65 years and ≥65 years (95% and 96%). Sundboll *et al* found PPV a lower PPV with increasing age for first HF diagnosis (<60 years 100%, 60-80 years 74% and >80 years 69%) and a mixed trend for recurrent HF diagnosis (<60 years 50%, 60-80 years 84% and >80 years 75%). By sex, the PPVs were lower in women (63-94%) than men (83-97%) in each of the three studies[29,33,43] which provided estimates.

Myocardial infarction

Similar to HF, there was no change in validity estimates for MI over time. One study[80] used data from 1981 with PPV estimates of 78% and 81% obtained from two Swedish cities, and a 2014 study[32] reported a PPV of 75%. Pajunen *et al*[68] found estimates fluctuated over time; for example, among men aged 35-74 years sensitivity and PPV were 64% and 93%, respectively, then in 1993-1997 sensitivity increased to 78% while PPV decreased to 86%, and in 1998-2002 the sensitivity and PPV were 81% and 90%, respectively.

By age breakdown,[29,34,65,70] sensitivity and PPV were largely similar between younger and older populations. By sex, the sensitivity and PPV were generally lower in women than men,[34,99] although Dalsgaard *et al*[32] found a much higher PPV in women (88%) than men (71%).

Stroke

There were no temporal trends in sensitivity or PPV for stroke diagnoses.

Aboa-Eboule *et al*[101] and Kivimaki *et al*[56] both included breakdowns by age, each study found sensitivity (76-79%) and PPV (72-82%) was higher in older populations, defined and ≥70 and ≥60 years, respectively, compared to younger ages (sensitivity 64-71%, PPV 67-73%). By sex results varied, Aboa-Eboule *et al*[101] and Kivimaki *et al*[56] found sensitivity was higher for diagnoses among women. While Aboa-Eboule *et al* also found PPV was higher, three other studies[25,32,56] identified the PPV was lower for women than men.

S6 Appendix. Outcome definitions used in included studies

Studies included multiple published definitions of HF, ACS and stroke. Below we have summarised the main features of each definition used.

Heart failure			
ESC definition[15]	ACC/AHA definition[105]	Framingham criteria[103]	Boston criteria[104]
Clinical syndrome characterised by symptoms such as breathlessness, persistent coughing or wheezing, ankle swelling and fatigue, that may be accompanied by the following signs: jugular venous pressure, pulmonary crackles, increased heart rate and peripheral oedema. Acute heart failure is defined as rapid onset or worsening of symptoms, which may occur with or without previous cardiac disease.	Describe heart failure as a progressive disease into four stages. Stage A: high risk but without structural heart disease or symptoms. Stage B:structural heart disease but without signs or symptoms – equivalent to NYHA class I (no limitation / ordinary physical activity). Stage C: structural heart disease with prior or current symptoms – equivalent to NYHA class II-III (ordinary / slightly limited physical activity or comfortable at rest but slightly / markedly limited physical activity). Stage D: refractory HF requiring intervention – equivalent to NYHA class IV (unable to carry on any physical activity or symptoms at rest).	Diagnosis requires two major or one major and two minor criteria Major criteria: paroxysmal nocturnal dyspnea or orthopnea; neck vein distension; rales; radiographic cardiomegaly; acute pulmonary edema; S3 gallop; central venous pressure >I6 cm water at the right atrium; circulation time ≥25 seconds; hepatojugular reflux Minor criteria: bilateral ankle edema; nocturnal cough; dyspnoea on ordinary exertion; hepatomegaly; pleural effusion; decrease in vital capacity by 33% from maximal value; tachycardia	Composite score based on points from three categories Category I: history rest dyspnea; orthopnea; paroxysmal nocturnal dyspnea; dyspnea while walking on level area; dyspnea while climbing Category II: physical examination heart rate; jugular venous elevation; lung crackles; wheezing; third heart sound Category III: chest radiography alveolar pulmonary edema; interstitial pulmonary edema; bilateral pleural effusion; cardiothoracic ratio >0.5; upper zone flow redistribution.
Myocardial infarction			
MONICA[106]	AHA/ESC definition[16]	Universal definition[107]	Third universal definition[108]
Definite: a. definite ECG, or b. symptoms with probable ECG and abnormal enzymes, or	Definite: 1. evolving diagnostic ECG, or 2. diagnostic enzymes Probable: 1. positive ECG, cardiac symptoms and missing biomarkers, or	1. typical rise and gradual fall of troponin or more rapid rise and fall of CK-MB biomarkers of myocardial necrosis with at least one of: a. ischemic symptoms	Update of previous definition to account for more sensitive biomarkers

c. symptoms and abnormal enzymes with ischemic or non-codable ECG or ECG not available, or d. fatal case with “naked-eye appearance” of fresh MI and/or recent coronary occlusion found at necropsy Possible: symptoms but lesser or no ECG and enzyme findings to classify as definite	2. positive ECG and equivocal biomarkers Possible: 1. equivocal biomarkers and nonspecific ECG findings, or 2. equivocal biomarkers and cardiac symptoms or signs, or 3. missing biomarkers and positive ECG	b. development of pathologic Q waves on the ECG c. ECG changes indicative of ischemia d. coronary artery intervention (e.g. coronary angioplasty) Or 2. pathologic findings	
Stroke			
WHO[109]		MONICA[110]	
Rapidly developing clinical signs of focal or global cerebral function disturbance lasting ≥24 hours or leading to death with no known cause other than vascular origin		WHO definition with clinical signs and symptoms specified to be suggestive of subarachnoid hemorrhage, intracerebral hemorrhage, or cerebral infarction with the use of CT to aid identification	

ACC, American College of Cardiology; AHA, American Heart Association; CK-MB, creatine kinase myocardial band; CT, computed tomography; ECG, electrocardiogram; ESC, European Society of Cardiology; HF, heart failure; MI, myocardial infarction; MONICA, Monitoring trends and determinants in cardiovascular disease; NYHA, New York Heart Association; WHO, World Health Organization

S1 Table. Characteristics of included studies

Author, year	Data collection period	Validation study	Country	Source population	Population description	Mean (SD) age	N (%) female	EHR description	Coding system	Reference standard
Heart failure										
Bosco-Lévy 2019[77]	2014[a]	Y	France	PMSI & electronic discharge records at Hospital Center University De Bordeaux & Paris Hopital Européen Georges Pompidou	2 random samples of HF patients, 1 from PMSI & 1 from electronic discharge records using free-text word searches	EHR 76.8 (13.4), RS 69.1 (11.5) inclusion age: ≥40 years	EHR 90 (45.0), RS 66 (28.8)	PMSI, national secondary care system	ICD-10	Medical record review
Delekta 2018[33]	2007	Y	Denmark	DNPR at Northern Denmark hospitals	Random sample of HF patients	77 (68-84)[b]	47[c]	DNPR, national secondary care system	ICD-10	Medical record review
Gini 2016[97]	unknown	Y	Italy	HSD at 12 GP practices	HF patients[d]	NR	NR	HSD, national primary care system	unknown	GP questionnaire
Heerdink 1998[59]	1990[e]	N	The Netherlands	PHARMO in 6 cities	Random sample of CHF patients	NR, inclusion age: ≥55 years	NR	PHARMO, national secondary care system	ICD-9	Medical record review
Hjerpe 2010[82]	2002-2003	Y	Sweden	SPCD GP practices	Random sample of patients with CVD prescription from each practice	69 (28-95)[b]	630 (52)	SPCD, regional primary care system	ICD-10	Medical record review
Ingelsson 2005[83]	1976-2001	Y	Sweden	Uppsala Longitudinal Study of Adult Men cohort study participants with	HF patients	NR, inclusion age: ≥50 years	NA, only men included	HDR, national secondary care system	ICD-8, ICD-9, ICD-10	Medical record review

Author, year	Data collection period	Validation study	Country	Source population	Population description	Mean (SD) age	N (%) female	EHR description	Coding system	Reference standard
				linked data in HDR						
Kaspar 2018[85]	2009	Y	Germany	Würzburg University Hospital EHR	Random sample of consecutive patients treated at the Medical Department	NR	NR	Clinical data warehouse, secondary care system	ICD unknown version	Medical record review
Khand 2005[54]	1997-1998	Y	Scotland, UK	Dishcarge diagnoses at Glasgow Royal Infirmary Hospital & affiliated (principally geriatric) units	HF hospitalised patients[f]	HF discharge code 68.2 (11.9)	HF discharge code male 54[c]	Unnamed, secondary care system	ICD-10	Medical record review
Kümler 2008[24]	1998-1999	Y	Denmark	DNPR at Amager Hospital in Copenhagen	Consecutive patients admitted to the hospital	78.0[g] inclusion age: ≥40 years	72 (50.8)[g]	DNPR, national secondary care system	ICD-10	Clinical examination
Mähönen 2013[67]	1969-1997	Y	Finland	FINRISK survey data participants with linked data in HDR, CDR[h], DRR[h] or pharmacy prescription[h]	All participants of the 1997 FINRISK survey	NR	NR	HDR, national secondary care system	ICD-8, ICD-9, ICD-10	Medical record review
Mard 2010[28]	2005-2007	Y	Denmark	DNPR at Herlev University Hospital	Patients referred to the outpatient clinic, HF clinic or admitted to the cardiology ward	75 (65-82, 33-99)[i]	320 (42.2)	DNPR, national secondary care system	ICD-10	Medical record review
Pfister 2013[43]	1997-2009	Y	England, UK	EPIC study participants linked	Random sample of HF patients	64.7 (7.5)	42.2[c]	Unnamed, secondary care system	ICD-10	Medical record review

Author, year	Data collection period	Validation study	Country	Source population	Population description	Mean (SD) age	N (%) female	EHR description	Coding system	Reference standard
				to hospital discharge data		inclusion age: 39-79 years				
Valk 2016[94]	2011	Y	The Netherlands	EHR at 30 Amersfoort GPs	HF patients	77.9 (11.4)	57.8[c]	Unnamed, national primary care system	ICPC	Medical record review
Van Doorn 2017[95]	2013-2014	Y	The Netherlands	CAFe study participants with GP EHR linked data	ECG confirmed AF patients	77 (68-84)[b]	47.7[c]	Unnamed, national primary care system	ICPC	Medical record review
Verdú-Rotellar 2017[96]	2014	Y	Spain	EHR at 2 Barcelona GP practices	HF patients	78 (10) inclusion age: ≥15 years	58[c]	Unnamed, primary care system	ICD-10	Medical record review
Acute coronary syndrome										
Bezin 2015[76]	2011	Y	France	PMSI at Bordeaux teaching hospitals	Random sample of ACS patients	66.5 (11.5)	18 (18.0)	PMSI, national secondary care system	ICD-10	Medical record review
Bork 2007[22]	2007	Y	Denmark	DNPR at Northern Denmark hospitals	Random sample of ACS patients	71.0 (41.4-91.0)[j]	178 (36.0)	DNPR, national secondary care system	ICD-10	Medical record review
Joensen 2009[37]	1994-2003	Y	Denmark	Diet, Cancer, and Health cohort study participants with linked data in DNPR	ACS patients born in Denmark & living in Copenhagen or Aarhus	56.1 (52.4-60.3)[b]	NR	DNPR, national secondary care system	ICD-8, ICD-10	Medical record review
Pajunen 2005[68]	1988-2002	Y	Finland	FINMONICA/ FINAMI Registry, CDR[h] & HDR	Fatal & non-fatal IHD patients	NR	NR	HDR, national secondary care system	ICD-9 ICD-10	AMI registry
Acute coronary syndrome – only myocardial infarction										

Author, year	Data collection period	Validation study	Country	Source population	Population description	Mean (SD) age	N (%) female	EHR description	Coding system	Reference standard
Barchielli 2012[75]	2003	Y	Italy	Tuscany regional Tosc-AMI Registry & hospital EHR	Random sample of AMI & IHD patients	NR	NR	Unnamed, national secondary care system	ICD-9-CM	Medical record review
Coloma 2013[23]	1996-2009	Y	Italy, The Netherlands, Denmark	3 databases from the Exploring & Understanding Adverse Drug Reactions project	Random sample of AMI patients	Italy 68, NL 66, DK 67	Italy 68 (34), NL 154 (38.5), DK 45 (30.4)	Italy HSD & NL IPCI, national primary care systems DK: Aarhus, regional secondary care system	Italy ICD9-CM, NL ICPC, DK ICD-10	Medical record review
Donnan 2003[50]	unknown	Y	Scotland	10 Tayside GP practices	Random sample of AMI patients	NR, inclusion age: ≥35 years	NR	Unnamed, primary care system SMR1, national secondary care system	Read, ICD-9	Medical record review
Egholm 2016[34]	2006-2012	Y	Denmark	Clinical drug-eluting coronary stent studies participants with linked data in the DNPR	PCI treated patients at Aarhus University Hospital cardiology ward	66 (58-73)[b]	1,448 (25.3)	DNPR, national secondary care system	ICD-10	AMI registry & medical record review
Hammad 2008[52]	1997-2004	Y	UK	GPRD participating GP practices	Random sample of AMI patients with NSAID & non-NSAID prescriptions	NR, inclusion age: 40-84 years	NR	GPRD (now CPRD), national primary care system	read	GP questionnaire
Hammar 2001[58]	1987 & 1995[e]	N	Sweden	HDR in Swedish hospitals & CRD[h]	National samples of AMI & IHD[k] patients	NR	NR	HDR, national secondary care system	ICD-9	Medical record review

Author, year	Data collection period	Validation study	Country	Source population	Population description	Mean (SD) age	N (%) female	EHR description	Coding system	Reference standard
Hammar 1994[80]	1981	Y	Sweden	Stockholm & Gavleborg hospitals	Random sample of patients by age group from 11 Stockholm hospitals & random sample of patients at 5 Gavleborg hospitals	NR, inclusion age: 30-94 years	NR	HDR, national secondary care system	ICD-8	Medical record review
Herrett 2013[53]	2003-2009	Y	England, UK	CPRD, HES, MINAP & ONS deaths[h]	AMI patients	CPRD 73 (61-81)[b] HES 73 (61-82)[b]	CPRD 5,810 (36.7) HES 5,072 (36.7)	CPRD, national primary care system HES, national secondary care system	read, ICD-10	ACS registry
Joensuu 1992[84]	1977-1986	Y	Finland	HDR in Finnish hospitals	AMI patients resident in rural districts of Oulu	67 (33-96)[j]	NR	HDR, national secondary care system	ICD-8	Medical record review
Madsen 1990[26]	1979-1980	Y	Denmark	Danish Heart Registry & DNPR	AMI patients from hospitals in the Copenhagen & 3 hospitals in Aarhus	NR	NR	DNPR, national secondary care system	ICD-8	Heart registry & medical record review
Madsen 2003[27]	1982-1991	Y	Denmark	Danish Heart Registry, DNPR & CDR[h]	AMI patients in Copenhagen hospitals	NR, inclusion age: 25-74 years	NR	DNPR, national secondary care system	ICD-8	AMI registry
Mähönen 1997[99]	1983-1990	Y	Finland	FINMONICA AMI Register linked to HDR	IHD & cerebrovascular disease patients[l] in Southwestern Finland	NR, inclusion age: 25-64 years	NR	HDR, national secondary care system	ICD-8, ICD-9	AMI registry
McAlpine 1998[42]	1993-1995	Y	Scotland, UK	SMR1 in Tayside	AMI or possible AMI patients	NR, inclusion age: 16-44 years	NA, only women included	SMR1, national secondary care system	ICD-9	Medical record review

Author, year	Data collection period	Validation study	Country	Source population	Population description	Mean (SD) age	N (%) female	EHR description	Coding system	Reference standard
Palomäki 1994[100]	1987-1990	Y	Finland	HDR at Kuopio University Hospital	Patients with suspected IHD events[l]	NR, inclusion age: 35-64 years	NR	HDR, national secondary care system	ICD-9	AMI registry
Pietilä 1997[69]	1980-1990	Y	Finland	Helsinki Heart Study RCT participants with linked data in HDR & CDR[h]	Coronary primary prevention patients followed up for IHD[l]	NR, inclusion age: 40-55 years	NA, only men included	HDR, national secondary care system	ICD-8, ICD-9	Medical record review
Rapola 1997[70]	1985-1989 & 1991	Y	Finland	ATBC RCT conducted in the Southern & Western Finland participants with HDR or CDR[h] linked data	Patients who smoked ≥5 cigarettes a day	NR, inclusion age: 50-69 years	NA, only men included	HDR, national secondary care system	ICD-8 ICD-9	Medical record review
Stroke										
Aboa-Eboulé 2013[101]	2004-2008	Y	France	Dijon Stroke Registry & FHDDB at Dijon Teaching Hospital	Stroke patients resident in Dijon	EHR 75.4 (15.5), RS 75.4 (14.8)	EHR 504 (55.8), RS 452 (55.7)	FHDDB, national secondary care system	ICD-10	Stroke registry
Appelros 2011[71]	1999-2000	Y	Sweden	Obrebro Stroke Registry, CDR[h] & IPR	Stroke patients resident in Orebro	women 78.3 & men 83.9	208 (55)	IPR, national secondary care system	ICD-10	Stroke registry
Baldereschi 2018[74]	2015	Y	Italy	HDR at 6 hospitals in Florence	Stroke patients recorded in the emergency or any inpatient department	NR	NR	HDR, secondary care system	ICD-9-CM	Medical record review
Barer 1996[40]	unknown	Y	England, UK	Stroke registry & Liverpool Teaching Hospital EHR	Stroke patients	NR	NR	Unnamed, secondary care system	ICD-9	Stroke registry

Author, year	Data collection period	Validation study	Country	Source population	Population description	Mean (SD) age	N (%) female	EHR description	Coding system	Reference standard
Cook 2013[41]	1990-2009	N	UK	THIN participating GP practices	Stroke patients with and without AD	Study cohort 79.8 (8.1) inclusion age: ≥50 years	Study cohort 6,748 (67.8)	THIN, national primary care system	Read	GP questionnaire & medical record review
Davenport 1996[49]	unknown	Y	Scotland, UK	Lothian Stroke Register & SMR1	Stroke patients	NR	NR	SMR1, national secondary care system	ICD-9	Stroke registry & medical record review
Ekker 2019[57]	1998-2017	N	The Netherlands	Whole study: linked national registries (HDR, CDR & population register) Validation cohort: HDR at 3 teaching hospital	Stroke patients	Study cohort women 41.4 (7.0), men 42.3 (6.5) inclusion age: 18-50 years	NR	HDR, national secondary care system	ICD-9, ICD-10	Medical record review
Ellekjaer 1999[102]	1994-1996	Y	Norway	Innherred Nord-Trøndelag County Stroke Registry	Cerebrovascular patients[l]	NR, inclusion age: ≥15 years	NR	HDR, national secondary care system	ICD-9	Stroke registry
Frost 2007[35]	1980-2002	N	Denmark	DNPR at 11 hospitals in Aarhus country	Random sample of stroke patients with AF	NR	NR	DNPR, national secondary care system	ICD-8, ICD-10	Medical record review
Gaist 2000[36]	1977-1995	N	Denmark	Whole study linked DNPR & DCPR, validation cohort included Funen County hospitals	SAH patients	Study cohort 53.5	Study cohort 5,466 (58.4)	DNPR, national secondary care system	ICD-8, ICD-10	Medical record review
Gaist 2013[51]	2000-2008	Y	UK	THIN participating GP practices	Patients followed until stroke, age 90 years, death, or study end	NR, inclusion age: 20-89 years	NR	THIN, national primary care system	read	GP questionnaire
Giroud 2015[78]	2009-2010	Y	France	FHDDB at 31 hospitals	Random sample of 56 stroke &	≥65 years 1,056 (63.3)[n]	786 (47.1)	FHDDB, national	ICD-10	Medical record review

Author, year	Data collection period	Validation study	Country	Source population	Population description	Mean (SD) age	N (%) female	EHR description	Coding system	Reference standard
					TIA[m] hospital stays from each hospital	inclusion age: ≥18 years		secondary care system		
Haesebaert 2013[79]	2006-2007	Y	France	AVC69 cohort study participants with linked data in PMSI	Stroke patients resident in Rhône	TP 72.8 (14.3), FN 77.6 (14.5) inclusion age: ≥18 years	TP 153 (48.9), FN 90 (59.2)	PMSI, national secondary care system	ICD-10	Medical record review
Holmqvist 2012[60]	1997-2009	N	Sweden	Epidemiological Investigation of Rheumatoid Arthritis case-control study participants with linked data in NPR	Stroke patients with & without RA	NR	NR	NPR, national secondary care system	ICD-10	Medical record review
Johnsen 2002[38]	1993-1998/9[o]	Y	Denmark	Diet, Cancer, and Health cohort study participants with linked data in DNPR	Cerebrovascular patients[l] born in Denmark & living in Copenhagen or Aarhus	58 (50-65)[p]	41.4[c]	DNPR, national secondary care system	ICD-10	Medical record review
Kirkman 2009[55]	2002-2007	Y	England, UK	Discharge diagnoses at 4 hospitals in Newcastle upon Tyne	Haemorrhagic stroke patients	ICH 66.4 (11-96)[j], SAH 55.4 (2-91)[j]	ICH 441 (47.0), SAH 713 (63.5)	PAS, secondary care system	ICD-10	Medical record review
Kivimäki 2017[56]	1997-2009	Y	England, UK	UK Whitehall II cohort study participants with linked HES data	London-based nonindustrial government workers followed up for stroke & IHD[l]	Cohort study 56[q]	Cohort study 30[c]	HES, national secondary care system	ICD-9, ICD-10	Medical record review

Author, year	Data collection period	Validation study	Country	Source population	Population description	Mean (SD) age	N (%) female	EHR description	Coding system	Reference standard
Köster 2013[64]	2004	Y	Sweden	Northern Sweden Stroke Registry, CDR[h] & NPR	Stroke patients	NR, inclusion age: ≥20 years	NR	NPR, national secondary care system	ICD-10	Stroke registry
Krarup 2007[39]	1998-1999	Y	Denmark	Copenhagen City Heart Study participants with linked DNPR data	Stroke & TIA[m] patients living in Østerbro & Nørrebro (urban Copenhagen) residents	women 75 & men 71.4	120 (51)	DNPR, national secondary care system	ICD-10	Medical record review
Leone 2004[86]	1998	Y	Italy	SISR at Novara Hospital	Stroke & TIA patients[m]	NR	NR	SISR, regional secondary care system	ICD-9	Medical record review
Leppälä 1999[72]	1985-1992	Y	Finland	ATBC RCT conducted in the Southern & Western Finland participants with HDR or CDR[h] linked data	Cerebrovascular disease patients[l] who smoked ≥5 cigarettes a day	62.9 (51.0-76.6)[p]	NA, only men included	HDR, national secondary care system	ICD-8, ICD-9	Medical record review
Lühdorf 2017[25]	1993-2009	Y	Denmark	Diet, Cancer, and Health cohort study participants with linked data in DNPR	Stroke patients born in Denmark and living in Copenhagen or Aarhus	NR	NR	DNPR, national secondary care system	ICD-8, ICD-10	Medical record review
Nieuwkamp 2014[61]	1997-2008	N	The Netherlands	HDR at University Medical Center Utrecht	SAH patients	58.1 inclusion age: ≥20 years	64.2[c]	HDR, national secondary care system	ICD-9	Medical record review
Øie 2018[89]	2008-2014	Y	Norway	NPR at St Olavs University Hospital	Intracranial haemorrhage[r] & clinically similar diagnosed patients resident	NR, inclusion age: ≥18 years	NR	NPR, national secondary care system	ICD-10	Medical record review

Author, year	Data collection period	Validation study	Country	Source population	Population description	Mean (SD) age	N (%) female	EHR description	Coding system	Reference standard
					in Sør-Trøndelag County					
Rinaldi 2003[90]	1999	Y	Italy	General Hospital of Lugo di Romagna EHR	Clinically diagnosed stroke patients	78 (10) 80 (4-99)[p]	77 (49.0)	Unnamed, secondary care system	ICD-9	Clinical assessment or medical record review
Ruigómez 2010[44]	2000-2004	Y	UK	THIN participating GP practices	Cerebrovascular disease patients[l]	NR, inclusion age: 40-84 years	NR	THIN, national primary care system	read	GP questionnaire & medical record review
Sansom 2015[45]	unknown	Y	UK	Unnamed acute hospital trust EHR	Consecutive acute stroke patients	NR	NR	Unnamed, national secondary care system	ICD-10	Medical record review
Sedova 2015[92]	2011	Y	Czech Republic	NRHOSP from sample of hospitals	10 randomly selected hospitals with randomly selected stroke & TIA[m] patients	NR	NR	NRHOSP, national secondary care system	ICD-10	Medical record review
Spolaore 2005[93]	1999	Y	Italy	HDR at Veneto hospitals	Stroke patients	NR	NR	HDR, national secondary care system	ICD-9	Medical record review
Stegmayr 1992[66]	1985-1989	Y	Sweden	Northern Sweden MONICA study participants linked to hospital EHR	Stroke patients	NR, inclusion age: 25-74 years	NR	Unnamed, secondary care system	ICD-9	Stroke registry
Tolonen 2007[73]	1993-1998	Y	Finland	FINSTROKE registry linked to HDR & CDR[h]	Stroke patients	NR, inclusion age: ≥25 years	NR	HDR, national secondary care system	ICD-9, ICD-10	Stroke registry
Varmdal 2016[63]	2012	Y	Norway	Stroke register & NPR	Confirmed & potential acute stroke patients	NR, inclusion age: ≥18 years	NR	NPR, national secondary care system	ICD-10	Medical record review

Author, year	Data collection period	Validation study	Country	Source population	Population description	Mean (SD) age	N (%) female	EHR description	Coding system	Reference standard
Vila-Corcoles 2014[62]	2008-2011	N	Spain	Cohort study participants with linked MBDS data from Joan XXIII University Hospital or Santa Tecla Hospital	Ischaemic stroke patients	71.7 (8.6) inclusion age: ≥60 years	55[c]	MBDS, national secondary care system	ICD-9	Medical record review
Wildenschild 2014[31]	2009 & 2010[s]	Y	Denmark	Stroke registry & DNPR	Confirmed & potential stroke patients	NR, inclusion age: ≥18 years	NR	DNPR, national secondary care system	ICD-10	Medical record review
Zhou 2014[48]	2003-2011	N	UK	Cohort study participants with linked CPRD data	Patients initiating angiotensin receptor blockers	NR, inclusion age: 40-95 years	NR	CPRD, national primary care system	read	GP questionnaire
Heart failure and acute coronary syndrome										
Merry 2009[65]	1987-2003	Y	The Netherlands	CAREMA cohort study participants linked to the University Hospital Maastricht CIS & HDR	Random sample of Maastricht residents	Study cohort 41.7 (20.1-60.9)[j] inclusion age: 20-59 years	Study cohort 11,175 (52.9)	HDR, national secondary care system	ICD-9	Local cardiology database
Sundbøll 2016[29]	2010-2012	Y	Denmark	DNPR at Aarhus University Hospital & regional hospitals of Randers & Herning in Central Denmark	Patients from cardiology, internal medicine, acute medicine & neurology wards	NR	NR	DNPR, national secondary care system	ICD-10	Medical record review
Heart failure and myocardial infarction										
Nilsson 1994[88]	1986	Y	Sweden	HDR for discharges from internal medicine,	Patients with cerebrovascular disease & IHD[l]	NR	NR	NPR, national secondary care system	ICD-8	Medical record review

Author, year	Data collection period	Validation study	Country	Source population	Population description	Mean (SD) age	N (%) female	EHR description	Coding system	Reference standard
				pediatric, general surgical, orthopaedic, urological or gynecological departments	born on a specific date with an odd year of birth					
Thygesen 2011[30]	1998-2007	Y	Denmark	DNPR at all hospitals in North Jutland Region	Patients with a Charlson comorbidity condition diagnosed	NR	NR	DNPR, national secondary care system	ICD-10	Medical record review
Van Staa 1994[46]	1990-1992	Y	UK	VAMP participating GP practices	Patients with ≥1 prescription of Sulphonyl Urea	NR, inclusion age: ≥20 years	NR	VAMP (now CPRD), national primary care system		Medical record review
Myocardial infarction and stroke										
Bernal 2019[98]	2012	Y	Spain	DIOCLES cross-sectional study participants with linked data in MBDS	DIOCLES participants hospitalized with suspected ACS	67.0 (13.0) inclusion age: ≥18 years	375 (24.4)	MBDS, national secondary care system	ICD-9-CM	ACS registry
Dalsgaard 2019[32]	2001-2014	Y	Denmark	ADDITION cohort study participants with linked data in DNPR	AMI & stroke patients with type 2 diabetes	NR, inclusion age: 40-69 years	NR	DNPR, national secondary care system	ICD-10	Medical record review
Heliövaara 1984[81]	1970-1975	Y	Finland	The Social Insurance Institution's mobile clinic screening participants	Patients screened 1966-1972 who were re-examined 1973-1977	NR, inclusion age: 30-59 years	NR	HDR, national secondary care system	ICD-8	Medical record review

Author, year	Data collection period	Validation study	Country	Source population	Population description	Mean (SD) age	N (%) female	EHR description	Coding system	Reference standard
Lindblad 1993[87]	1977-1987	Y	Sweden	Hypertension register linked to CDR[h] & IPR	AMI & stroke patients with hypertension	NR, inclusion age: 40-69 years	NR	IPR, national secondary care system	ICD-8, ICD-9	Medical record review
Rodrigo-Rincon 2015[91]	2008 & 2010	Y	Spain	Hospital de Navarra CADB	Surgical patients with adverse events during hospitalization	60.8 (18.2)	660 (41.2)	CADB, secondary care system	ICD-9-CM	Medical record review
Wright 2012[47]	1997-2005	Y	England, UK	Million Women Study participants with linked HES data	Random sample of IHD, cerebrovascular disease & venous thrombo-embolism patients[l]	NR	NA, only women included	HES, national secondary care system	ICD-10	GP questionnaire

[a] 2009 included for time comparison in sub-analysis; [b] median age with interquartile range; [c] percentage only; [d] also included diabetes, IHD & hypertension patients to validate diagnoses but results not included here; [e] years for which data were validated from a wider study period; [f] also included atrial fibrillation patients and patients in the local ECG database to identify heart failure patients but results not included here as could not determine results; [g] result for patients with a heart failure diagnosis in the EHR and reference standard, also reported results for patients with; 1) no HF in EHR or RS (mean age 69.7, 848 (40.1%) female), HF in EHR only (mean age 79.5, 8 (26.7%) female), HF in RS only (mean age 76.8, 138 (45.5%) female); [h] results for this source not included here; [i] median with interquartile range and range; [j] mean age with range; [k] IHD patients included to calculate sensitivity of MI diagnoses rather than to validate IHD; [l] only results for sub-diagnoses relevant to this review (HF, ACS and stroke) included here; [m] also included TIA patients to validate diagnosis but results not included here; [n] number with proportion in age group; [o] end year depended on area of residence; [p] median age with range; [q] mean with no SD; [r] also validated subdural haemorrhage but results not included here; [s] two separate validations methods, one in 2009 and one in 2010

AMI, acute myocardial infarction; ATBC, Alpha-Tocopherol, Beta-carotene Cancer prevention study; CADB, Clinical and Administrative DataBase; CAREMA, Cardiovascular registry Maastricht; CDR, Cause of Death Register; CIS, Cardiology Information System; CPRD, Clinical Practice Research Datalink; DCPR, Danish Central Person Registry; DIOCLES, Description of Ischemic Heart Disease in the Spanish Territory; DNPR, Danish National Patient Register; DRR, Drug Reimbursement Register; EHR, Electronic Health Record; EPIC, European Prospective Investigation into Cancer & nutrition study; FHDDB, French Hospital Discharge DataBase; GPRD, General Practice Research Database; HDR, Hospital Discharge Register; HES, Hospital Episode Statistics; HF, heart failure; HSD, Health Search Database; ICPC, International Classification of Primary Care; ICD, International Classification of Diseases; IPCI, Integrated Primary Care Information; IPR, In-Patient Register; MBDS, Minimum Basic Data Set; MINAP, Myocardial Ischaemia National Audit Project; MONICA, MONItoring of trends and determinants in CArdiovascular disease; NPR, National Patient Register; NR, not reported; NRHOSP, National Registry of Hospitalized Patients; ONS, Office for National Statistics; PAS, Patient Adminstration System; PCI, percutaneous coronary intervention; PMSI, Programme de Médicalisation des Systèmes d'Information; RCT, randomized controlled trial; SISR, Sistema Informativo Sanitario Regionale; SMR1, Scottish Morbidity Record; SPCD, Skaraborg Primary Care Database; THIN, The Health Improvement Network; VAMP, Value Added Information Medical Products Ltd

S2 Table. Study results

A. Heart failure

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
Secondary care EHRs								
Bosco-Lévy 2019[77]	ICD-10 I50	200 (EHR), 229 (RS)	-ESC definite & probable -First diagnosis -Any diagnostic position	64.2 (58.0-70.4)		88.0 (83.5-92.5)		High
			-ESC definite -First diagnosis -Any diagnostic position			60.5 (53.7-67.3)		
		73	-ESC definite & probable -First diagnosis -Primary diagnostic position			95.9 (88.5-99.1)		
			-ESC definite -First diagnosis -Primary diagnostic position			69.9 (59.3-80.4)		
		127	-ESC definite & probable -First diagnosis -Secondary diagnostic position			83.5 (77.0-89.9)		
			-ESC definite -First diagnosis -Secondary diagnostic position			55.1 (46.5-63.8)		

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
		200	-ESC definite & probable -First diagnosis -Any diagnostic position -2009 comparison			74.5 (68.5-80.5)		
Delekta 2018[33]	ICD-10 I50	500	-ESC definite & probable -Any diagnostic position			83.6 (80.1-86.7)		High
			-ESC definite -Any diagnostic position			61.0 (56.7-65.2)[b,c]		
		400	-ESC definite & probable -Primary diagnostic position			88.0 (84.4-91.0)		
		100	-ESC definite & probable -Secondary diagnostic position			66.0 (55.8-75.2)		
		267	-ESC definite & probable -Any diagnostic position -Men			85.4 (80.6-89.4)		
		233	- ESC definite & probable -Any diagnostic position -Women			81.5 (76.0-86.3)		
Heerdink 1998[59]	ICD-9 428	138	-Framingham definition			79.7		High
			-Boston definition			81.2		
Ingelsson 2005[83]	ICD-8 427.00, 427.10, 428.99; ICD-9 428;	317	-ESC definite -Any diagnostic position			81.7 (77.1-85.6)[c]		Medium

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
	ICD-10 I11.0, I50	140	-ESC definite & questionable -Any diagnostic position			97.8 (95.5-98.9)[b,c]		
			-ESC definite -Primary diagnostic position			95.0 (90.0-97.6)[c]		
			-ESC definite & questionable -Primary diagnostic position			99.3 (96.1-99.9)[b,c]		
		112	-ESC definite -Secondary diagnostic position			75.9 (67.2-82.9)[c]		
			-ESC definite & questionable -Secondary diagnostic position			97.3 (92.4-99.1)[b,c]		
		65	-ESC definite -Third-sixth diagnostic position			63.1 (50.9-73.8)[c]		
			-ESC definite & questionable -Third-sixth diagnostic position			95.4 (87.3-98.4)[b,c]		
	ICD-8 427.00, 427.10, 428.99; ICD-9 428	108	-ESC definite -Any diagnostic position -1976-1991			88.0 (80.5-92.8)[c]		
			-ESC definite & questionable -Any diagnostic position -1976-1991			97.2 (92.1-99.1)[b,c]		
	ICD-9 428	101	-ESC definite -Any diagnostic position			79.2 (70.3-86.0)[c]		

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
			-1992-1996 -ESC definite & questionable -Any diagnostic position			97.0 (91.6-99.0)[b,c]		
			-1992-1996 -ESC definite -Any diagnostic position			77.7 (69.1-84.6)[c]		
	ICD-10 I11.0, I50	108	-1997-2001 -ESC definite & questionable -Any diagnostic position			99.1 (94.9-99.8)[b,c]		
			-1997-2001					
Kaspar 2018[85]	ICD-10 I11, I13.0, I13.2, I50	1,042		50.0 (43.0-56.1)[c]		94.0 (88.2-97.1)[c]		High
			-ESC definite, probable & possible -Any diagnostic position			86.7 (82.6-89.9)[c]		
Khand 2005[54]	ICD-10 I11.0-1, I25.5, I42.9, I50.1-2	339	-ESC definite & probable -Any diagnostic position			77.3 (72.5-81.5)[c]		High
			-ESC definite -Any diagnostic position			65.5 (60.2-70.4)[c]		
Kümler 2008[24]	ICD-10 I50	3,201	-ESC definition	29.4 (25.3-33.8)[c]	98.9 (98.5-99.2)[c]	80.8 (73.9-86.2)[c]	90.0 (88.9-91.1)[c]	High
Mähönen 2013[67]	ICD-8 427.00, 427.10, 428; ICD-9 402.9B, 414.8, 428;	325	-Study defined probable or possible	10.5 (7.6-14.4)[c]	99.8 (99.7-99.9)[c]	73.3 (59.0-84.0)[c]	96.4 (96.0-96.8)[c]	High

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
Mard 2010[28]	ICD-10 I11.0, I13.0, I13.2, I50							
	ICD-10 I11.0, I13.0, I13.2, I42.0, I42.6-9, I50.0-1, I50.9	758	-ESC definition -First or recurrent diagnosis			84.0 (81.2-86.6)		High
		479	-ESC definition -First diagnosis			77.9 (74.1-81.6)		
Merry 2009[65]	ICD-9 428	84 (EHR), 154 (RS)		43 (35-51)		80 (71-88)		High
		Unknown	-Aged <50 years	30 (2-58)		60 (17-100)		
			-Aged ≥50 years	44 (36-52)		81 (72-90)		
			-1987-1995	45 (28-62)		68 (49-88)		
			-1996-2003	42 (33-51)		84 (74-93)		
Nilsson 1994[88]	ICD-8 427	24	-Primary diagnostic position	65.0 (43.3-81.9)[c,d]		86.7 (62.1-96.3)[b,c]		High
Pfister 2013[43]	ICD-10 I50	396	-ESC definite, probable & possible -Any diagnostic position			95.7 (93.2-97.3)[c]		High
			-ESC definite & probable -Any diagnostic position			88.1 (84.6-91.0)[c]		
		165	-ESC definite -Any diagnostic position			77.8 (73.4-81.6)[c]		
			-ESC definite, probable & possible -Any diagnostic position			95.1 (90.7-97.5)[c]		
			-Aged <65 years					

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
			-ESC definite & probable					
			-Any diagnostic position			88.5 (82.7-92.5)[c]		
			-Aged <65 years					
			-ESC definite					
			-Any diagnostic position			78.8 (71.9-84.3)[c]		
			-Aged <65 years					
			-ESC definite, probable & possible					
			-Any diagnostic position			96.1 (92.8-97.9)[c]		
			-Aged ≥65 years					
			-ESC definite & probable					
		231	-Any diagnostic position			87.9 (83.0-91.5)[c]		
			-Aged ≥65 years					
			-ESC definite					
			-Any diagnostic position			77.1 (71.2-82.0)[c]		
			-Aged ≥65 years					
			-ESC definite, probable & possible					
			-Any diagnostic position			96.9 (93.8-98.5)[c]		
			-Men					
		229	-ESC definite & probable					
			-Any diagnostic position			90.4 (85.9-93.6)[c]		
			-Men					
			-ESC definite			81.2 (75.7-85.8)[c]		

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
			-Any diagnostic position					
			-Men					
			-ESC definite, probable & possible					
			-Any diagnostic position			94.0 (89.3-96.7)[c]		
			-Women					
			-ESC definite & probable					
		167	-Any diagnostic position			85.0 (78.8-89.6)[c]		
			-Women					
			-ESC definite					
			-Any diagnostic position			73.1 (65.9-79.2)[c]		
			- Women					
			-ESC definite, probable & possible					
			-Any diagnostic position			99.1 (94.9-99.8)		
			-Apr 1997-Mar 2003					
			-ESC definite & probable					
		107	-Any diagnostic position			91.6 (84.8-95.5)		
			-Apr 1997-Mar 2003					
			-ESC definite					
			-Any diagnostic position			82.2 (73.9-88.3)		
			-Apr 1997-Mar 2003					
			-ESC definite, probable & possible					
		289	-Any diagnostic position			94.5 (91.2-96.6)		

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
			-Apr 2003-Mar 2009					
			-ESC definite & probable					
			-Any diagnostic position			86.8 (82.5-90.3)		
			-Apr 2003-Mar 2009					
			-ESC definite					
			-Any diagnostic position			76.1 (70.9-80.7)		
			-Apr 2003-Mar 2009					
Sundbøll 2016[29]	ICD-10 I11.0, I13.0, I13.2, I42.0, I42.6-9, I50.0-3, I50.8-9	95	-First diagnosis -Any diagnostic position			76 (66-83)		High
		96	-Recurrent diagnosis -Any diagnostic position			76 (67-83)		
		50	-First diagnosis -Any diagnostic position			83 (72-91)		
		22	-Men -First diagnosis -Any diagnostic position			63 (46-77)		
		45	-Women -Recurrent diagnosis -Any diagnostic position			80 (68-89)		
		28	-Men -Recurrent diagnosis -Any diagnostic position			70 (55-82)		

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
			- Women					
			-First diagnosis					
		13	-Any diagnostic position			100 (77-100)		
			-Aged <60 years					
			-First diagnosis					
		50	-Any diagnostic position			74 (60-84)		
			-Aged 60-80 years					
			-First diagnosis					
		32	-Any diagnostic position			69 (51-82)		
			-Aged >80 years					
			-Recurrent diagnosis					
		14	-Any diagnostic position			50 (27-73)		
			-Aged <60 years					
			-Recurrent diagnosis					
		50	-Any diagnostic position			84 (71-92)		
			-Aged 60-80 years					
			-Recurrent diagnosis					
		32	-Any diagnostic position			75 (58-87)		
			-Aged >80 years					
			-First diagnosis					
		13	-Any diagnostic position			81 (64-91)		
			-2010					

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
			-Recurrent diagnosis					
		14	-Any diagnostic position			75 (57-87)		
			-2010					
			-First diagnosis					
		50	-Any diagnostic position			71 (55-83)		
			-2011					
			-Recurrent diagnosis					
		50	-Any diagnostic position			79 (64-89)		
			-2011					
			-First diagnosis					
		32	-Any diagnostic position			77 (58-89)		
			-2012					
			-Recurrent diagnosis					
		32	-Any diagnostic position			73 (56-86)		
			-2012					
Thygesen 2011[30]	ICD-10 I11.0, I13.0, I13.2, I50	50	-Primary diagnostic position			100 (92.9-100)		Medium
Primary care EHRs								
Gini 2016[97]	unknown	243	-ACC / AHA definition			55 (49-61)		Low
Hjerpe 2010[82]	ICD-10 I50	unknown		66 (58-73)[e]				Medium
Valk 2016[94]	ICPC K77	683	-ESC definite			63.5 (59.9-67.1)		High

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
Van Doorn 2017[95]	ICPC K77	unknown	-Study specific definition	54.5	95.7	83.3	84.3	Medium
Van Staa 1994[46]	unknown	31		100.0		100.0		Medium
Verdú-Rotellar 2017[96]	ICD-10 I50	595	-Study specific definition			53.6 (50.0-57.6)[b,c]		High

B. Acute coronary syndrome

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
Acute coronary syndrome – secondary care EHRs								
Bezin 2015[76]	ICD-10 I20.0, I21-I22, I24	100				84.2 (72.1-92.5)		High
	ICD-10 I20.0, I24		-ESC definition			70.6 (44.0-89.7)		
	ICD-10 I21, I24		-Primary diagnostic position			90.5 (77.4–97.3)		
	ICD-10 I20.0, I21					83.6 (71.2–92.2)		
Bork 2007[22]	ICD-10 I20.0, I21- I22	494	-ESC definite -First diagnosis -Any diagnostic position			86.6 (83.3-89.5)		High
			-ESC definite or possible -First diagnosis			87.8 (84.6-90.6)		

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
		316	-Any diagnostic position			87.6 (83.5-91.1)		
			-ESC definite					
			-First diagnosis					
			-Any diagnostic position					
			-Men					
			-ESC definite or possible			88.9 (84.9-92.2)		
			-First diagnosis					
			-Any diagnostic position					
			-Men					
		178	-ESC definite			84.8 (78.7-89.8)		
			-First diagnosis					
			-Any diagnostic position					
			-Women					
			-ESC definite or possible			86.0 (80.0-90.7)		
			-First diagnosis					
			-Any diagnostic position					
			-Women					
		398	-ESC definite			90.2 (86.8-92.9)		
			-First diagnosis					
			-Primary diagnostic position					
			-ESC definite or possible			91.5 (88.3-94.0)		
			-First diagnosis					

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
		96	-Primary diagnostic position -ESC definite -First diagnosis -Secondary diagnostic position			71.9 (61.8-80.6)		
			-ESC definite or possible -First diagnosis -Secondary diagnostic position			72.9 (62.9-81.5)		
Joensen 2009[37]	ICD-8 410, 427.27; ICD-10 I20.0, I21, I46	1,558	-AHA & ESC definitions -Any diagnostic position			65.5 (63.1-67.9)		Medium
		1,067	-AHA & ESC definitions -Any diagnostic position -Men			72.6 (69.9-75.3)		
		491	-AHA & ESC definitions -Any diagnostic position -Women			50.1 (45.6-54.6)		
		1,425	-AHA & ESC definitions -Primary diagnostic position			67.1 (64.6-69.5)		
		115	-AHA & ESC definitions -Secondary diagnostic position			47.0 (37.6-56.5)		
Pajunen 2005[68]	ICD-9 410, 411.0; ICD-10 I20.0, I21-I22	2,727	-Any diagnostic position -Men -Aged 35-74 years	67		86		High

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
		4,715	-1988-1992 -Any diagnostic position -Men -Aged 35-74 years	83		71		
		3,493	-1993-1997 -Any diagnostic position -Men -Aged 35-74 years	87		76		
		1,020	-1998-2002 -Any diagnostic position -Women -Aged 35-74 years	66		80		
		2,880	-1988-1992 -Any diagnostic position - Women -Aged 35-74 years	79		63		
		1,921	-1993-1997 -Any diagnostic position - Women -Aged 35-74 years	84		69		
		1,451	-1998-2002 -Any diagnostic position	77		72		

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
			-Men -Aged ≥75 years -1993-1997 -Any diagnostic position					
		1,544	-Men -Aged ≥75 years -1998-2002 -Any diagnostic position	82		78		
		3,313	-Women -Aged ≥75 years -1993-1997 -Any diagnostic position	71		69		
		3,156	-Women -Aged ≥75 years -1998-2002	80		76		
Acute myocardial infarction – secondary care EHRs								
Barchielli 2012[75]	ICD-9-CM 410	372	-AHA definite -Primary diagnostic position			86.0 (82.4-89.6)		High
			-AHA definite & probable -Primary diagnostic position			87.3 (84.0-90.8)		
			-AHA definite, probable & possible			94.6 (92.3-96.9)		

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
			-Primary diagnostic position					
			-MONICA definite -Primary diagnostic position			52.7 (47.6-57.8)		
			-MONICA definite & possible -Primary diagnostic position			65.3 (60.5-70.2)		
Bernal 2019[98]	ICD-9-CM 410.*1	12	-Primary diagnostic position	91.9 (90.2-93.5)	92.6 (90.0-95.1)	96.8 (95.7-97.9)	82.2 (78.8-85.7)	High
Bezin 2015[76]	ICD-10 I21	unknown (100 for all included codes)	-ESC definition -Primary diagnostic position			85.0 (70.2-94.3)		High
Coloma 2013[23]	ICD-10 I21	148	-Universal (v1) & AHA definitions			100.0		High
		69	-Study definition -First diagnosis			75 (64-84)		
Dalsgaard 2019[32]	ICD-10 I21-I24	Unknown	-Study definition -First diagnosis -Men			71		High
			-Study definition -First diagnosis -Women			88		
Donnan 2003[50]	ICD-9 410	207	-Adapted MONICA definition	59 (52-66)		95 (91-100)		High
	ICD-10 I21	285	-Universal definition (v3)	95.2 (92.2-97.3)	93.4 (92.7-94.0)	41.7 (37.9-45.7)	99.7 (99.6-99.9)	High

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
			-Any diagnostic position					
			-All inpatients					
			-Universal definition (v3)					
			-Any diagnostic position	93.9 (90.6-96.3)	98.3 (97.9-98.6)	73.4 (68.6-77.8)	99.7 (99.5-99.8)	
			-Acute admissions					
			-Universal definition (v3)					
			-Primary diagnostic position	85.0 (80.5-88.8)	98.2 (97.8-98.5)	70.4 (65.4-75.1)	99.2 (99.0-99.4)	
			-All inpatients					
			-Universal defined (v3)					
			-Primary diagnostic position	82.1 (77.3-86.3)	99.0 (98.7-99.2)	81.0 (76.1-85.2)	99.1 (98.8-99.3)	
Egholm 2016[34]			-Acute admissions					
			-Universal definition (v3)					
		Unknown	-Any diagnostic position	94.7 (91.0-97.1)	98.3 (97.9-98.7)	74.2 (68.7-79.2)	99.7 (99.5-99.9)	
			-Acute admissions					
			-Men					
			-Universal definition (v3)					
			-Any diagnostic position	91.7 (83.6-96.4)	98.0 (97.2-98.7)	71.0 (61.2-79.4)	99.6 (99.1-99.8)	
			-Acute admissions					
			-Women					
			-Universal definition (v3)					
			-Any diagnostic position	95.7 (90.8-98.3)	98.3 (97.7-98.7)	70.7 (63.3-77.4)	99.8 (99.6-99.9)	
			-Acute admissions					

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
			-Aged ≤65 years -Universal definition (v3) -Any diagnostic position -Acute admissions -Aged >65 years	92.6 (87.9-95.9)	98.2 (97.7-98.7)	75.5 (69.2-81.1)	99.6 (99.3-99.8)	
Hammar 1994[80]	ICD-8 unknown codes	512	-Study defined definite or possible			81.2 (75.7-85.7)[b,c]		Low
			-2 separate cities			77.8 (70.6-83.6)[b,c]		
		Unknown	-Study defined definite			77.5 (72.8-81.5)[b,c]		
Hammar 2001[58]	ICD-9 410	750	-Nationally defined definite	94.3 (92.2-95.8)[c]		85.8 (83.1-88.2)[c]		Medium
Heliövaara 1984[81]	ICD-8 410	75	-MONICA definite	84.7 (75.6-90.8)[c]				Medium
Herrett 2013[53]	ICD-10 I21-I23	7,489	-Universal definition (v3)			91.5 (90.8-92.1)		High
Joensen 2009[37]	ICD-8, presume 410; ICD-10 presume I21	1,072	-AHA & ESC definition -First diagnosis			81.9 (79.5-84.2)		Medium
Joensuu 1992[84]	ICD-8 410	671	-MONICA definite & possible -First diagnosis			81.4 (78.2-84.1)[c]		Medium
			-'Normal clinical' definition -First diagnosis			89.1 (86.5-91.3)[c]		
Lindblad 1993[87]	ICD-8 410.00, 410.99; ICD-9 410	413	-Study defined definite			95.7 (93.2-97.2)[c]		High

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
		7,197	-WHO definition -Any diagnostic position -Registry reference standard			80.0 (79.1-80.9)[b,c]		
		692 (EHR), 457 (RS)	-WHO definition -Any diagnostic position -Medical record review reference standard	90.4 (87.3-92.7)[c,d]		59.7 (56.0-63.3)[b,c]		
Madsen 1990[26]	ICD-8 410-414	6,108	-WHO definition -Primary diagnostic position -Registry reference standard			88.7 (87.9-89.5)[b,c]		High
		1,089	-WHO definition -Secondary diagnostic position -Registry reference standard			31.1 (28.4-33.9)[b,c]		
		538	-WHO definition -Primary diagnostic position -Medical record review reference standard			66.7 (62.6-70.6)[b,c]		
		154	-WHO definition -Secondary diagnostic position -Medical record review reference standard			35.1 (28.0-42.9)[b,c]		
Madsen 2003[27]	ICD-8 410, 427.24, 427.27,	5,511	-MONICA definite -Non-fatal	96.6		78.8 (77.6-79.9)[c]		Medium

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
	427.91 & 427.97		-Any diagnostic position					
			-MONICA definite					
		4,956	-Non-fatal	92.9		79.9 (78.7-81.0)[c]		
			-Primary diagnostic position					
			-MONICA definite	84.4 (83.2-86.3)		79.5 (77.8-81.1)		
			-Men					
			-MONICA definite	85.5 (82.3-88.7)		72.5 (68.8-76.2)		
	ICD-8 410		-Women					
			-MONICA definite & possible	72.0 (70.5-73.5)		90.7 (89.6-91.8)		
			-Men					
			-MONICA definite & possible	62.2 (59.2-65.3)		87.5 (85.0-90.0)		
			-Women					
Mähönen 1997[99]		18,288 in total	-MONICA definite & possible	72.9 (71.5-74.2)		85.4 (84.3-86.5)		High
	ICD-8 410, 411		-Men					
			-MONICA definite & possible	66.6 (64.2-69.1)		80.5 (78.3-82.8)		
			-Women					
			-MONICA definite	86.0 (84.4-87.6)		85.9 (84.2-87.5)		
			-Men					
			-MONICA definite	81.3 (77.6-85.0)		80.7 (77.0-84.4)		
	ICD-9 410		-Women					
			-MONICA definite & possible	66.8 (65.1-68.5)		93.3 (92.3-94.4)		
			Men					

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
			-MONICA definite & possible -Women	55.9 (52.6-59.2)		89.6 (87.1-92.2)		
	ICD-9 410, 411		-MONICA definite & possible -Men	79.6 (78.3-80.9)		84.8 (83.6-85.9)		
			-MONICA definite & possible -Women	73.9 (71.4-76.3)		79.0 (76.6-81.3)		
McAlpine 1998[42]	ICD-9 410	204	-Definite & possible	67	100	100		High
	ICD-9 411			5.6	99	50		
Merry 2009[65]	ICD-9 410	656 (EHR), 815 (RS)		84 (81-87)		97 (96-99)		High
		Unknown	-Aged <50 years	86 (80-93)		99 (97-100)		
			-Aged ≥50 years	83 (80-87)		97 (95-99)		
			-1987-1995	82 (76-87)		94 (91-98)		
			-1996-2003	85 (81-89)		99 (97-100)		
Nilsson 1994[88]	ICD-8 410	900	-Primary diagnostic position	92.3 (79.7-97.3)[c,d]		100.0[2]		High
Pajunen 2005[68]	ICD-9 410; ICD-10 I21-I22	2,727	-Any diagnostic position -Men -Aged 35-74 years -1988-1992	64		93		High
		4,715	-Any diagnostic position -Men	78		86		

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
			-Aged 35-74 years -1993-1997 -Any diagnostic position -Men					
		3,493	-Aged 35-74 years -1998-2002 -Any diagnostic position -Women	81		90		
		1,020	-Aged 35-74 years -1988-1992 -Any diagnostic position - Women	61		88		
		2,880	-Aged 35-74 years -1993-1997 -Any diagnostic position -Women	75		79		
		1,921	-Aged 35-74 years -1998-2002 -Any diagnostic position -Men	78		86		
		1,451	-Aged ≥75 years -1993-1997	73		85		

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
		1,544	-Any diagnostic position -Men -Aged ≥75 years -1998-2002	78		90		
		3,313	-Any diagnostic position -Women -Aged ≥75 years -1993-1997	69		80		
		3,156	-Any diagnostic position -Women -Aged ≥75 years -1998-2002	76		87		
Palomäki 1994[100]	ICD-9 4100, 4109	734	-MONICA definite			71.0 (67.6-74.1)[b,c]		High
			-MONICA definite & possible			93.5 (91.4-95.0)[b,c]		
			-FINMONICA definite			71.0 (67.6-74.1)[b,c]		
			-FINMONICA definite & possible			92.0 (89.8-93.7)[b,c]		
	ICD-9 4100	566	-MONICA definite			86.4 (83.3-89.0)[b,c]		
			-FINMONICA definite			86.4 (83.3-89.0)[b,c]		
Pietilä 1997[69]	ICD-8 410,00, 410,99	211	-Definite & possible			96.6 (90.3-98.8)[b,c]		High
			-Definite			93.1 (85.8-96.8)[b,c]		
	ICD-9 4100	178	-Definite & possible			95.0 (86.3-98.2)[b,c]		
			-Definite			88.3 (77.8-94.2)[b,c]		

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
Rapola 1997[70]	ICD-8 410.00, 410.07, 410.97, 410.99; ICD-9 4100, 4109	217	-MONICA definite & possible			93.5 (89.5-96.1)[3]		High
			-MONICA definite			77.9 (71.9-82.9)[3]		
		105	-MONICA definite & possible -Aged <60 years			96.2 (90.1-98.5)		
			-MONICA definite -Aged <60 years			80.0 (71.4-86.5)		
		112	-MONICA definite & possible -Aged ≥60 years			91.1 (84.3-95.1)		
			-MONICA definit -Aged ≥60 years			75.9 (67.2-82.9)		
Sundbøll 2016[29]	ICD-10 I21	99	-First diagnosis -Any diagnostic position			97 (91-99)		High
		100	-Recurrent diagnosis -Any diagnostic position			88 (80-93)		
		63	-First diagnosis -Any diagnostic position -Men			97 (89-99)		
		36	-First diagnosis -Any diagnostic position -Women			97 (86-100)		
		69	-Recurrent diagnosis -Any diagnostic position			88 (79-94)		

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
			-Men					
			-Recurrent diagnosis					
		31	-Any diagnostic position			88 (71-95)		
			-Women					
			-First diagnosis					
		30	-Any diagnostic position			97 (83-99)		
			-Aged <60 years					
			-First diagnosis					
		48	-Any diagnostic position			98 (89-100)		
			-Aged 60-80 years					
			-First diagnosis					
		21	-Any diagnostic position			95 (77-99)		
			-Aged >80 years					
			-Recurrent diagnosis					
		19	-Any diagnostic position			89 (69-97)		
			-Aged <60 years					
			-Recurrent diagnosis					
		57	-Any diagnostic position			89 (79-95)		
			-Aged 60-80 years					
			-Recurrent diagnosis					
		24	-Any diagnostic position			83 (64-93)		
			-Aged >80 years					

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
			-First diagnosis					
		35	-Any diagnostic position			97 (85-99)		
			-2010					
			-First diagnosis					
		29	-Any diagnostic position			97 (83-99)		
			-2011					
			-First diagnosis					
		35	-Any diagnostic position			97 (85-99)		
			-2012					
			-Recurrent diagnosis					
		33	-Any diagnostic position			85 (69-93)		
			-2010					
			-Recurrent diagnosis					
		38	-Any diagnostic position			92 (79-97)		
			-2011					
			-Recurrent diagnosis					
		29	-Any diagnostic position			86 (69-95)		
			-2012					
		89	-First diagnosis -Primary diagnostic position			99 (94-100)		
		10	-First diagnosis -Secondary diagnostic position			80 (49-94)		

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
Thygesen 2011[30]	ICD-10 I21-I23	50	-Primary diagnostic position			98.0 (89.4-99.9)		Medium
Wright 2012[47]	ICD-10 I21-I22	130				89.2 (82.7-93.5)[c]		High
Unstable angina – secondary care EHRs								
Bezin 2015[76]	ICD-10 I20.0	unknown (100 for all included codes)	-ESC definition -Primary diagnostic position			20.0 (4.3-48.1)		High
Joensen 2009[37]	ICD-8 code unclear; ICD-10 presume I20.0	444	-AHA & ESC definition			27.5 (23.4-31.9)		Medium
Merry 2009[65]	ICD-9 411.1, 413.1	656 (EHR), 815 (RS)		53 (48-58)		78 (74-83)		High
		Unknown	-Aged <50 years	54 (41-67)		76 (62–89)		
			-Aged ≥50 years	53 (48-59)		79 (74-84)		
			-1987-1995	53 (44-63)		69 (59-79)		
			-1996-2003	53 (48-59)		83 (77-88)		
Sundbøll 2016[29]	ICD-10 I20.0	96	-Any diagnostic position			88 (79-93)		High
		55	-Any diagnostic position -Men			87 (76-94)		
		41	-Any diagnostic position -Women			88 (74-95)		
		28	-Any diagnostic position			86 (69-94)		

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
		57	-Aged <60 years -Any diagnostic position			89 (79-95)		
		11	-Aged 60-80 years -Any diagnostic position			82 (52-95)		
		29	-Aged >80 years -Any diagnostic position			90 (74-96)		
		36	-2010 -Any diagnostic position			86 (71-94)		
		31	-2011 -Any diagnostic position			87 (71-95)		
			-2012					
		90	-Primary diagnostic position			89 (81-94)		
		6	-Secondary diagnostic position			67 (30-90)		
Cardiac arrest – secondary care EHRs								
Joensen 2009[37]	ICD-8 presume 427.27; ICD-10 presume I46	42	-AHA & ESC definition			50.0 (34.2-65.8)		Medium
Acute myocardial infarction – primary care EHRs								
Coloma 2013[23]	ICPC K75	124	-Universal (v1) & AHA definition			75.0 (67.4-82.6)		High
	ICD9-CM 410	116				96.6 (93.2-99.9)		

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
Donnan 2003[50]	Read	207	-Adapted MONICA definition	83 (71-96)		91 (83-99)		High
Hammad 2008[52]	Read/OMIS	217				93 (90-96)		Medium
Herrett 2013[53]	Read	7,224	-Universal definition (v3)			92.2 (91.6-92.8)		High
Van Staa 1994[46]	Unknown	34		89.3		85.3		Medium

C. Stroke

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
Stroke – secondary care EHRs								
Aboa-Eboulé 2013[101]	ICD-10 I61, I63, I64, G46	903 (EHR), 811 (RS)	-WHO definition -First or recurrent diagnosis -Primary diagnostic position	77.1 (74.2-80.0)		69.2 (66.1-72.2)		High
		230 (EHR), 201 (RS)	-WHO definition -First or recurrent diagnosis -Primary diagnostic position -Aged <70 years	70.6 (64.0-76.5)[c]		61.9 (55.3-67.8)[c]		
		673 (EHR), 610 (RS)	-WHO definition -First or recurrent diagnosis -Primary diagnostic position -Aged ≥70 years	79.2 (75.8-82.2)[c]		71.7 (68.3-75.0)[c]		

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
		399 (EHR), 359 (RS)	-WHO definition -First or recurrent diagnosis -Primary diagnostic position -Men	74.7 (69.9-78.9)[c]		67.2 (62.4-71.6)[c]		
		504 (EHR), 452 (RS)	-WHO definition -First or recurrent diagnosis -Primary diagnostic position -Women	79.0 (75.0-82.5)[c]		70.8 (66.7-74.6)[c]		
		136 (EHR), 107 (RS)	-WHO definition -First or recurrent diagnosis -Primary diagnostic position -2004	80.3 (72.6-86.3)[c]		52.8 (45.8-59.8)[c,i]		
		178 (EHR), 192 (RS)	-WHO definition -First or recurrent diagnosis -Primary diagnostic position -2005	75.5 (68.0-81.8)[c]		62.7 (55.4-69.5)[c,i]		
		232 (EHR), 254 (RS)	-WHO definition -First or recurrent diagnosis -Primary diagnostic position -2006	70.7 (63.4-77.2)[c]		74.7 (67.5-81.0)[c]		
		132 (EHR), 164 (RS)	-WHO definition -First or recurrent diagnosis -Primary diagnostic position -2007	76.0 (69.5-81.5)[c]		75.6 (69.1-81.2)[c,i]		
		177 (EHR), 94 (RS)	-WHO definition -First or recurrent diagnosis -Primary diagnostic position -2008	82.9 (76.7-87.7)[c]		81.2 (74.8-86.1)[c]		

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
Appelros 2011[71]	ICD-10 I61, I63, I64	328 (EHR), 377 (RS)	-WHO definition -First diagnosis	89.1 (85.4-91.9)[c,d]		97.0 (94.5-98.3)[b,c]		Low
Barer 1996[40]	ICD-9 431, 433-4, 436	340 (EHR), 420 (RS)		66.2 (61.5-70.5)[c,d]		81.8 (77.3-85.5)[b,c]		Low
Bernal 2019[98]	ICD-9-CM 094.87, 430-434[6], 436	12 (EHR), 15 (RS)	-Primary diagnostic position	66.7 (39.5-93.9)	99.9 (99.7-100.0)	83.3 (58.1-100.0)	99.7 (99.4-100.0)	High
Dalsgaard 2019[32]	ICD-10 I61–I65	46	-Study definition -First diagnosis			70 (54-80)		High
		Unknown	-Study definition -First diagnosis -Men			75		
			-Study definition -First diagnosis -Women			61		
Davenport 1996[49]	ICD-9 431, 432.9, 433, 434, 436, 438	557 (EHR), 613 (RS)	-Primary diagnostic position	86.3 (83.3-88.8)[c]	99.9 (99.9-99.9)[c]	95.0 (92.8-96.5)[b,c]	99.9 (99.9-99.9)[c,g]	Medium
Ellekjaer 1999[102]	ICD-9 430, 431, 434, 436	508 (EHR), 389 (RS)	-WHO definition	89.2 (85.7-91.9)[c]		68.3 (64.1-72.2)		High
Frost 2007[35]	ICD-8 430-434, 436; ICD-10 I60-I64	164	-WHO definition			97.0 (93.0-98.7)[c]		Medium

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
Heliövaara 1984[81]	ICD-8 430-434	59 (EHR), 78 (RS)	WHO definite	81.7 (70.1-89.4)[c]		83.1 (71.5-90.5)[b,c]		Medium
Holmqvist 2012[60]	ICD-10 I61, I63	76	-MONICA definition -RA patients			92		High
			-MONICA definition -Patients without RA			89		
Johnsen 2002[38]	ICD-10 I60, I61, I63, I64	377	-WHO definition			79.3 (74.9-83.3)		High
Kivimäki 2017[56]	ICD-9 430, 431, 434, 436; ICD-10 I60, I61, I63, I64	107 (EHR), 7,837 (RS)	-MONICA definition -First or recurrent diagnosis -Any diagnostic position	71 (62-79)	100 (100-100)	79 (70-86)	100 (99-100)	High
		44 (EHR), 5,475 (RS)	-MONICA definition -First or recurrent diagnosis -Any diagnostic position	64 (49-77)	100 (100-100)	73 (57-86)	100 (100-100)	
		66 (EHR), 2,362 (RS)	-Aged <60 years -MONICA definition -First or recurrent diagnosis -Any diagnostic position	76 (65-85)	100 (99-100)	82 (70-90)	99 (99-100)	
		76 (EHR), 5,455 (RS)	-Aged ≥60 years -MONICA definition -First or recurrent diagnosis -Any diagnostic position	70 (59-80)	100 (100-100)	80 (70-89)	100 (99-100)	
			-Men					

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
		31 (EHR), 2,382 (RS)	-MONICA definition -First or recurrent diagnosis -Any diagnostic position -Women	74 (55-88)	100 (99-100)	74 (55-88)	100 (99-100)	
Köster 2013[64]	ICD-10 I61, I63, I64	2,032 (EHR), 2,166 (RS)	-MONICA definite & possible -First or recurrent diagnosis	89.3 (87.9-90.7)		80.5 (78.8-82.2)		High
			-MONICA definite, possible & unclassifiable -First or recurrent diagnosis	82.7 (81.1-84.3)		88.1 (86.7-89.5)		
		1,426 (EHR), 1,468 (RS)	-MONICA definite & possible -First diagnosis	90.6 (89.0-92.2)		85.8 (84.0-87.6)		
			-MONICA definite, possible & unclassifiable -First diagnosis	85.9 (84.2-87.6)		94.0 (92.8-95.2)		
Krarup 2007[39]	ICD-10 I60, I61, I63, I64	236	-WHO definition -2 separate medical record reviewers			80.5 (73.6-86.3) & 86.0 (79.7-90.9)		High
Leone 2004[86]	ICD-9 430, 431, 434, 436	411 (EHR), 698 (RS)	-WHO definition -Any diagnostic position	53.2 (49.4-56.8)[c]		90.3 (87.0-92.8)[b,c]		High
Leppälä 1999[72]	ICD-8 430, 431[h], 432, 433, 434, 436; ICD-9 430, 431, 433, 434, 436	326	-MONICA & nationally defined definite & probable			89.9 (86.1-92.7)[c]		High

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
Lindblad 1993[87]	ICD-8 & ICD-9 430-434, 436	251	-Study defined definite			94.0 (90.4-96.3)[c]		High
		3,326	-WHO definition -Any diagnostic position			69.3 (67.8-70.9)		
			-WHO definition -Any diagnostic position -Men			71.1 (69.1-73.1)		
			-WHO definition -Any diagnostic position -Women			66.8 (64.3-69.3)		
Lühdorf 2017[25]	ICD-10 I60, I61, I63, I64	Unknown	-WHO definition -Any diagnostic position -1994-1999			75.1 (71.6-78.7)		High
			-WHO definition -Any diagnostic position -2000-2004			70.2 (67.7-72.7)		
			-WHO definition -Any diagnostic position -2005-2009			66.3 (63.9-68.7)		
Rodrigo-Rincon 2015[91]	ICD-9-CM 997.0X, 431-434.9X, 436, 346.XX, 430	6 (EHR), 10 (RS)		33.3		20.0		Low

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
Sansom 2015[45]	ICD-10 I60-I64	502 (EHR), 664 (RS)		75.6 (72.2-78.8)		96.3 (94.4-97.8)		Low
Sedova 2015[92]	ICD-10 I60, I61, I63, I64	484	-WHO definition			85 (81-88)		High
	ICD-10 I60, I61, I63	Unknown				90 (87-93)		
Stegmayr 1992[66]	ICD-9 presume 430-434, 436	5,101 (EHR), 3,606 (RS)	-WHO definition			68.5 (67.2-69.7)[2,3]		Medium
Tolonen 2007[73]	ICD-9 430-434, 436-438; ICD-10 I60-I69	2,866	-First diagnosis -Primary diagnostic position	85 (84-86)		85 (84-86)		High
Varmdal 2016[63]	ICD-10 I61, I63, I64	5,192	-WHO definition -Any diagnostic position	96.8 (95.7-97.9)	99.6 (99.6-99.7)	79.7 (77.5-82.0)	99.9 (99.9-100.0)	High
			-WHO definition -Primary diagnostic position	86.1 (84.0-88.3)	99.9 (99.9-99.9)	93.5 (92.0-95.1)	99.7 (99.7-99.8)	
		25	-WHO definition			20 (9-39)		
Wildenschild 2014[31]	ICD-10 I61, I63, I64	10,000 (EHR), 10,015 (RS)	-WHO definition -Extrapolated results	79 (73-84)		79 (62-88)		High
		156	-WHO definition -Possible stroke patients	58 (46-69)	96 (90-99)	93 (82-99)	72 (62-80)	
Ischaemic stroke – secondary care EHRs								

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
Aboa-Eboulé 2013[101]	presume ICD-10 I63	178 (EHR), 192 (RS)	-WHO definition -First or recurrent diagnosis -Primary diagnostic position -Cardiac embolism ischaemic stroke	83.3 (77.4-87.9)[c]		89.9 (84.6-93.5)[c,i]		High
		232 (EHR), 254 (RS)	-WHO definition -First or recurrent diagnosis -Primary diagnostic position -Large-artery atherosclerosis	74.0 (68.3-79.0)[c]		81.0 (75.5-85.6)[c,i]		
		132 (EHR), 164 (RS)	-WHO definition -First or recurrent diagnosis -Primary diagnostic position -Large-artery lacunar infarct	73.2 (65.9-79.4)[c]		90.1 (84.8-94.7)[c,i]		
		177 (EHR), 77 (RS)	-WHO definition -First or recurrent diagnosis -Primary diagnostic position -Other ischaemic stroke	67.0 (57.0-75.7)[c]		35.6 (28.9-42.9)[c,i]		
Baldereschi 2018[74]	ICD-9-CM 433*1, 434*1	1,002 (EHR), 1,273 (RS)	-MONICA definition -Primary diagnostic position	70.5 (68.0-73.0)		89.9 (87.6-91.4)		Medium
	ICD-9-CM 433*1	63				85.7 (74.6-93.3)		

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
	ICD-9-CM 434*1	979				89.9 (87.8-91.7)		
Dalsgaard 2019[32]	ICD-10 I63	23	-Study definition -First diagnosis			78.3 (58.1-90.3)[c]		High
Ekker 2019[57]	ICD-9 433, 434; ICD-10 I63	301	-First diagnosis			90.4 (86.5-93.2)[c]		Medium
Ellekjaer 1999[102]	ICD-9 434	313 (EHR), 281 (RS)	-WHO definition	73.3 (67.8-78.1)[c,d]		65.8 (60.3-70.8)[b,c]		High
Giroud 2015[78]	ICD-10 I63	914	-First or recurrent diagnosis -Primary diagnostic position			96.3 (94.8-97.3)[c]		High
Haesebaert 2013[79]	ICD-10 I63	329 (EHR), 465 (RS)		67.3 (63.1-71.5)		95.1 (92.8-97.4)		High
Heliövaara 1984[81]	ICD-8 432-434	39 (EHR), 59 (RS)	-WHO definition	66.1 (53.4-76.9)[c]				Medium
Johnsen 2002[38]	ICD-10 I63	113	-WHO definition			87.6 (80.3-92.5)[b,c]		High
Krarup 2007[39]	presume ICD-10 I63	33	-WHO definition -2 separate medical record reviewers			97.0 (84.2–99.9) 100.0		High
Leone 2004[86]	ICD-9 434	202(EHR), 478 (RS)	-WHO definition -Any diagnostic position	36.8 (32.6-41.2)[c]		87.1 (81.8-91.1)[c]		High
		188 (EHR), 478 (RS)	-WHO definition -Primary diagnostic position	35.4 (31.2-39.7)[c]		89.9 (84.8-93.4)[c]		

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
Leppälä 1999[72]	ICD-8 432, 433, 434; ICD-9 433, 434	252 (EHR), 258 (RS)	-MONICA & nationally defined definite & probable stroke			90.1 (85.8-93.2)[c]		High
Lühdorf 2017[25]	ICD-10 I63	Unknown	-WHO definition -Any diagnostic positions			80.1 (77.9-82.3)		High
			-WHO definition -Primary diagnostic position			81.0 (78.6-83.4)		
Rinaldi 2003[90]	ICD-9 434, 436	180 (EHR), 157 (RS)	-WHO definition -Any diagnostic positions	81.5 (74.7-86.2)[c]		71.1 (64.0-77.2)[c]		High
		157 (EHR), 157 (RS)	-WHO definition -Primary diagnostic position	75.8 (68.5-81.8)[c]		75.8 (68.5-81.8)[c]		
	ICD-9 434	3	-WHO definition			100.0		
	ICD-9 436	177	-Any diagnostic positions			70.6 (63.5-76.8)[b,c]		
Sedova 2015[92]	ICD-10 I63	unknown	-WHO definition			82 (74-89)		High
Spolaore 2005[93]	ICD-9 433	unknown (4,015 for all included codes)	-MONICA definition -Primary diagnostic position			9 (6-13)		High
			-MONICA definition -Secondary diagnostic position			4 (1-7)		
	ICD-9 434		-MONICA definition -Primary diagnostic position			77 (73-82)		
			-MONICA definition			29 (22-36)		

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
			-Secondary diagnostic position					
Tolonen 2007[73]	ICD-9 433, 434, 436; ICD-10 I63, I64	unknown (2,866 for all included codes)	-First diagnosis -Primary diagnostic position	82 (80–84)		84 (83–85)		High
	ICD-9 433, 434; ICD-10 I63			81 (79–83)		83 (81–85)		
Vila-Corcoles 2014[62]	ICD-9 433, 434, 436, 437 but later states 434, 435, 436, 437	406	-MONICA definition			84.5 (80.6-87.7)[b,c]		High
Wright 2012[47]	ICD-10 I63	190				86.3 (80.7-90.5)[c]		High
Haemorrhagic stroke including both intracerebral and subarachnoid haemorrhages – secondary care EHRs								
Dalsgaard 2019[32]	ICD-10 I60–I62	5	-Study definition -First diagnosis			60.0 (23.1-88.2)[c]		High
Intracerebral haemorrhage – secondary care EHRs								
Aboa-Eboulé 2013[101]	ICD-10 presume I61	189 (EHR), 107 (RS)	-WHO definition -First or recurrent diagnosis -Primary diagnostic position	87.9 (80.3-92.8)[b]		67.6 (59.5-74.8)[c,i]		High
Ekker 2019[57]	ICD-9 431; ICD-10 I61	183	-First diagnosis			86.3 (80.6-90.6)[c]		Medium
Ellekjaer 1999[102]	ICD-9 431	56 (EHR), 41 (RS)	-WHO definition	97.6 (87.4-99.6)[c,d]		71.4 (58.5-81.6)[b,c]		High

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
Giroud 2015[78]	ICD-10 I61	188	-First or recurrent diagnosis -Primary diagnostic position			89.4 (84.1-93.0)[c]		High
Heliövaara 1984[81]	ICD-8 431	0 (EHR), 4 (RS)	-WHO definition	0.0				Medium
Johnsen 2002[38]	ICD-10 I61	35	-WHO definition			65.7 (49.2-79.2)[b,c]		High
Kirkman 2009[55]	ICD-10 I61	978	-Primary diagnostic position			95.9 (94.5-97.0)		Medium
		204	-Primary diagnostic position -2002			95.1 (91.2-97.3)		
		179	-Primary diagnostic position -2003			97.2 (93.6-98.8)		
		170	-Primary diagnostic position -2004			95.3 (91.0-97.6)		
		152	-Primary diagnostic position -2005			98.0 (94.4-99.3)		
		151	-Primary diagnostic position -2006			94.7 (90.0-97.3)		
		122	-Primary diagnostic position -2007			95.1 (90.0-97.7)		
Krarup 2007[39]	presume ICD-10 I61	17	-WHO definition -2 separate medical record reviewers			73.9 (51.6–89.8)		High

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
Leone 2004[86]	ICD-9 431	110 (EHR), 138 (RS)	-WHO definition -Any diagnostic positions	59.4 (51.1-67.3)[c]		74.5 (65.7-81.8)[c]		High
		102 (EHR), 138 (RS)	-WHO definition -Primary diagnostic position	56.5 (48.2-64.5)[c]		76.5 (67.4-83.6)[c]		
Leppälä 1999[72]	ICD-8 431[h]; ICD-9 431	28 (EHR), 26 (RS)	-MONICA & nationally defined definite & probable			82.1 (64.4-92.1)[c]		High
Lindblad 1993[87]	presume ICD-8 431; ICD-9 431	20	-Study defined definite			55.0 (34.2-74.2)[b,c]		High
Lühdorf 2017[25]	ICD-10 I61	unknown	-WHO definition -Any diagnostic positions			73.1 (68.1-78.0)		High
			-WHO definition -Primary diagnostic position			76.7 (71.6-81.8)		
Øie 2018[89]	ICD-10 I61	545	-Primary diagnostic position			96.9 (95.1-98.0)[c]		High
		143	-Secondary diagnostic position			83.1 (76.4-88.2)[c]		
Sedova 2015[92]	ICD-10 I61	unknown	-WHO definition			91 (85-96)		High
Spolaore 2005[93]	ICD-9 431	unknown (4,015 for all included codes)	-WHO-MONICA definition -Primary diagnostic position			78 (74-83)		High
			-WHO-MONICA definition -Secondary diagnostic position			40 (31-49)		
Tolonen 2007[73]	ICD-9 431; ICD-10 I61	unknown (2,866 for all	-First diagnosis -Primary diagnostic position	94 (91–97)		84 (80–88)		High

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
		included codes)						
Wright 2012[47]	ICD-10 I61, I62	69				78.3 (67.2-86.4)[3]		High
Subarachnoid haemorrhage – secondary care EHRs								
Ellekjaer 1999[102]	ICD-9 430	13 (EHR), 12 (RS)	-WHO definition	75.0 (46.8-91.1)[3,4]		69.2 (42.4-87.3)[2,3]		High
Gaist 2000[36]	ICD-8 430; ICD-10 I60	191	-First diagnosis -Neurosurgery ward			93 (85-98)		Medium
			-First diagnosis -Neurology ward			75 (60-87)		
			-First diagnosis -Non-neurosurgery or neurology ward			47 (36-59)		
Giroud 2015[78]	ICD-10 I60	72	-First or recurrent diagnosis -Primary diagnostic position			45.8 (34.8-57.3)[2]		High
Heliövaara 1984[81]	ICD-8 430	13 (EHR), 15 (RS)	-WHO definition	73.3 (48.0-89.1)[3]				Medium
Johnsen 2002[38]	ICD-10 I60	29	-WHO definition			48.3 (31.4-65.6)[2,3]		High
Kirkman 2009[55]	ICD-10 I60	1,169	-Primary diagnostic position			96.1 (94.8-97.0)		Medium
		201	-Primary diagnostic position -2002			95.5 (91.7-97.6)		

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
		198	-Primary diagnostic position -2003			95.0 (91.0-97.2)		
		191	-Primary diagnostic position -2004			95.8 (92.0-97.9)		
		180	-Primary diagnostic position -2005			97.8 (94.4-99.1)		
		198	-Primary diagnostic position -2006			95.5 (91.6-97.6)		
		201	-Primary diagnostic position -2007			97.0 (93.6-98.6)		
Krarup 2007[39]	presume ICD-10 I60	3	-WHO definition -2 separate medical record reviewers			66.7 (9.4-99.2)		High
Leone 2004[86]	ICD-9 430	42 (EHR), 51 (RS)	-WHO definition -Any diagnostic position	35.3 (23.6-49.0)[c]		42.9 (29.1-57.8)[c]		High
		37 (EHR), 51 (RS)	-WHO definition -Primary diagnostic position	33.3 (22.0-47.0)[c]		45.9 (31.0-61.6)[c]		
Leppälä 1999[72]	ICD-8 430; ICD-9 430	28 (EHR), 25 (RS)	-MONICA & nationally defined definite & probable			78.6 (60.5-89.8)[3]		High
Lindblad 1993[87]	presume ICD-8 430; ICD-9 430	23	-Study defined definite			78.3 (58.1-90.3)[2,3]		High
Lühdorf 2017[25]	ICD-10 I60	unknown	-WHO definition -Any diagnostic position			60.6 (53.3-67.7)		High

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
			-WHO definition -Primary diagnostic position			62.1 (54.5-69.7)		
Nieuwkamp 2014[61]	ICD-9 430	1,472	-Primary diagnostic position			95.4 (94.2-96.3)[c]		Medium
Øie 2018[89]	ICD-10 I60	171	-Primary diagnostic position			95.3 (91.0-97.6)[c]		High
		34	-Secondary diagnostic position			55.9 (39.5-71.1)[c]		
Sedova 2015[92]	ICD-10 I60	unknown	-WHO definition			91 (85-97)		High
Spolaore 2005[93]	ICD-9 430	unknown (4,015 for all included codes)	-WHO-MONICA definition -Primary diagnostic position			76 (73-79)		High
			-WHO-MONICA definition -Secondary diagnostic position			26 (18-34)		
Tolonen 2007[73]	ICD-9 430; ICD-10 I60	unknown (2,866 for all included codes)	-First diagnosis -Primary diagnostic position	92 (88–96)		81 (75–87)		High
Wright 2012[47]	ICD-10 I60	78				96.1 (89.3-98.7)[c]		High
Acute stroke – primary care EHRs								
Cook 2013[41]	Unknown Read codes	95				90.5 (83.0-94.9)[b,c]		Low
Ruigómez 2010[44]	Read codes	400	-Review of computerized record excluding free text (step 1)			82.8 (70.6-91.4)		High

Author, year	Diagnostic code(s)	Records assessed[a]	Parameters	Sensitivity	Specificity	PPV	NPV	Quality of study
			-Review of free text in computerized record (step 2)			90.2 (78.6-96.7)		
			-Confirmed by GP & medical records (step 3)			77.5 (69.2-84.1)		
Zhou 2014[48]	Unknown Read codes	1,000				89		Low
Haemorrhagic stroke – primary care EHRs								
Gaist 2013[51]	41 different read codes	306				82 (77-86)		High
Intracerebral haemorrhage – primary care EHRs								
Gaist 2013[51]	41 different read codes	150				73 (65-80)		High
Subarachnoid haemorrhage – primary care EHRs								
Gaist 2013[51]	41 different read codes	156				91 (85-95)		High

[a] EHR unless otherwise stated; [b] positive predictive value calculated from raw data: TP/(TP+FP); [c] confidence interval calculated from raw data; [d] sensitivity calculated from raw data: TP/(TP+FN); [e] mean sensitivity calculated from results of each primary care centre; [f] 432 included codes ended in .0, .1 or .9, 433 included codes ended in .01, .11, .21, .31, .81 or .91, and 434 included codes ended in .01, .11 or .91; [g] negative predictive value calculated from raw data: TN/(FN+TN); [h] 431 included codes ended in .00, .08, .09, .98, .99; [i] PPV presented in paper is different but based on numbers presented this would be correct percentage

EHR, electronic health record; RS, reference standard

Main study result(s) highlighted in blue

S3 Table. QUADAS-2 assessment of studies

		Patient selection				Index test		Reference standard					Flow and timing			Overall quality
		1	2	3	4	1	2	1	2	3	4	5	1	2	3	
Aboa-Eboulé 2013[101]	Response	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Low**					**Low**			
Appelros 2011[71]	Response	Y	Y	Y	Y	Y	N	Y	N	N	Y	Y	Y	Y	Y	LOW
	Risk	**Low**				**High**		**High**					**Low**			
Baldereschi 2018[74]	Response	Y	Y	Y	Y	Y	Y	U	U	N	Y	Y	Y	Y	Y	MEDIUM
	Risk	**Low**				**Low**		**High**					**Low**			
Barchielli 2012[75]	Response	Y	Y	Y	Y	Y	Y	Y	U	U	Y	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Unclear**					**Low**			
Barer 1996[40]	Response	U	U	U	N	Y	N	Y	Y	U	N	Y	Y	U	Y	LOW
	Risk	**High**				**High**		**High**					**Low**			
Bernal 2019[98]	Response	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	U	Y	Y	HIGH
	Risk	**Low**				**Low**		**Unclear**					**Low**			
Bezin 2015[76]	Response	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Low**					**Low**			
Bork 2007[22]	Response	Y	Y	Y	Y	Y	Y	U	U	Y	Y	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Unclear**					**Low**			
Bosco-Lévy 2019[77]	Response	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Low**					**Low**			
Coloma 2013[23]	Response	Y	Y	Y	N	Y	Y	Y	U	Y	Y	Y	Y	Y	Y	HIGH
	Risk	**Unclear**				**Low**		**Low**					**Low**			
Cook 2013[41]	Response	U	Y	Y	Y	Y	N	Y	U	U	U	Y	U	U	Y	LOW
	Risk	**Low**				**High**		**Unclear**					**Unclear**			
Dalsgaard 2019[32]	Response	Y	Y	Y[a]	Y	Y	Y	Y	U	Y	Y	Y	U	Y	Y	HIGH
	Risk	**Low**				**Low**		**Low**					**Low**			
Davenport 1996[49]	Response	Y	Y	Y	Y	Y	Y	Y	Y	N	U	U	Y	Y	Y	MEDIUM
	Risk	**Low**				**Low**		**High**					**Low**			
Delekta 2018[33]	Response	Y	Y	Y	Y	Y	Y	U	U	Y	Y	Y	Y	Y	Y	HIGH

		Patient selection				Index test		Reference standard					Flow and timing			Overall quality
		1	2	3	4	1	2	1	2	3	4	5	1	2	3	
	Risk	**Low**				**Low**		**Low**					**Low**			
Donnan 2003[50]	Response	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Low**					**Low**			
Egholm 2016[34]	Response	Y	Y	Y[b]	Y	Y	Y	Y	U	Y	Y	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Low**					**Low**			
Ekker 2019[57]	Response	Y	Y	Y	Y	Y	Y	U	U	U	U	Y	Y	Y	Y	MEDIUM
	Risk	**Low**				**Low**		**High**					**Low**			
Ellekjaer 1999[102]	Response	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Low**					**Low**			
Frost 2007[35]	Response	Y	Y	Y[c]	Y	Y	Y	U	U	Y	N	Y	Y	Y	Y	MEDIUM
	Risk	**Low**				**Low**		**High**					**Low**			
Gaist 2000[36]	Response	U	Y	Y	Y	Y	Y	U	U	Y	U	Y	N	Y	Y	MEDIUM
	Risk	**Low**				**Low**		**Unclear**					**Unclear**			
Gaist 2013[51]	Response	U	Y	Y	Y	Y	Y	Y	U	U	Y	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Unclear**					**Low**			
Gini 2016[97]	Response	Y	U	Y	Y	Y	N	Y	N	U	Y	Y	U	Y	Y	LOW
	Risk	**Low**				**High**		**High**					**Low**			
Giroud 2015[78]	Response	Y	Y	Y	Y	Y	Y	Y	U	Y	U	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Unclear**					**Low**			
Haesebaert 2013[79]	Response	U	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Low**					**Low**			
Hammad 2008[52]	Response	Y	Y	Y	Y	Y	N	Y	U	U	Y	Y	Y	Y	Y	MEDIUM
	Risk	**Low**				**High**		**Unclear**					**Low**			
Hammar 1994[80]	Response	Y	Y	Y	Y	Y	N	U	U	U	U	Y	Y	Y	Y	LOW
	Risk	**Low**				**High**		**High**					**Low**			
Hammar 2001[58]	Response	U	U	Y	Y	Y	Y	U	U	U	Y	Y	U	Y	Y	MEDIUM
	Risk	**Unclear**				**Low**		**Unclear**					**Low**			
Heerdink 1998[59]	Response	Y	Y	Y[d]	Y	Y	Y	U	U	Y	U	Y	Y	U	Y	HIGH

		Patient selection				Index test		Reference standard					Flow and timing			Overall quality
		1	2	3	4	1	2	1	2	3	4	5	1	2	3	
	Risk	**Low**				**Low**		**Unclear**					**Low**			
Heliövaara 1984[81]	Response	Y	Y	Y	N	Y	Y	U	U	Y	U	Y	Y	U	Y	MEDIUM
	Risk	**Unclear**				**Low**		**Unclear**					**Low**			
Herrett 2013[53]	Response	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Low**					**Low**			
Hjerpe 2010[82]	Response	Y	Y	Y[e]	Y	Y	Y	Y	U	N	U	Y	Y	U	Y	MEDIUM
	Risk	**Low**				**Low**		**High**					**Low**			
Holmqvist 2012[60]	Response	Y	U	Y	Y	Y	Y	U	U	Y	U	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Unclear**					**Low**			
Ingelsson 2005[83]	Response	Y	U	Y[f]	Y	Y	Y	Y	U	U	Y	Y	U	Y	Y	MEDIUM
	Risk	**Low**				**Low**		**Unclear**					**Low**			
Joensen 2009[37]	Response	Y	Y	Y	Y	Y	N	U	U	Y	Y	Y	Y	Y	Y	MEDIUM
	Risk	**Low**				**High**		**Unclear**					**Low**			
Joensuu 1992[84]	Response	Y	Y	Y	U	Y	Y	U	U	Y	N	Y	Y	Y	Y	MEDIUM
	Risk	**Low**				**Low**		**High**					**Low**			
Johnsen 2002[38]	Response	Y	Y	Y	Y	Y	Y	Y	U	U	Y	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Unclear**					**Low**			
Kaspar 2018[85]	Response	Y	U	Y	Y	Y	Y	U	U	Y	U	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Unclear**					**Low**			
Khand 2005[54]	Response	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Low**					**Low**			
Kirkman 2009[55]	Response	Y	Y	Y	Y	Y	Y	N	U	Y	Y	Y	Y	Y	Y	MEDIUM
	Risk	**Low**				**Low**		**High**					**Low**			
Kivimäki 2017[56]	Response	Y	Y	Y	Y	Y	Y	Y	U	U	Y	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Unclear**					**Low**			
Köster 2013[64]	Response	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	U	Y	Y	HIGH
	Risk	**Low**				**Low**		**Low**					**Low**			
Krarup 2007[39]	Response	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	Y	Y	Y	Y	HIGH

		Patient selection				Index test		Reference standard					Flow and timing			Overall quality
		1	2	3	4	1	2	1	2	3	4	5	1	2	3	
	Risk	**Low**				**Low**		**Low**					**Low**			
Kümler 2008[24]	Response	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Low**					**Low**			
Leone 2004[86]	Response	Y	Y	Y	Y	Y	Y	U	U	Y	Y	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Unclear**					**Low**			
Leppälä 1999[72]	Response	Y	Y	Y[g]	Y	Y	Y	U	U	Y	Y	Y	U	Y	Y	HIGH
	Risk	**Low**				**Low**		**Unclear**					**Low**			
Lindblad 1993[87]	Response	Y	Y	Y	Y	Y	Y	Y	U	U	Y	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Unclear**					**Low**			
Lühdorf 2017[25]	Response	Y	U	Y	Y	Y	Y	Y	U	Y	Y	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Low**					**Low**			
Madsen 1990[26]	Response	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Low**					**Low**			
Madsen 2003[27]	Response	Y	Y	Y	Y	Y	Y	Y	U	N	Y	Y	Y	Y	Y	MEDIUM
	Risk	**Low**				**Low**		**High**					**Low**			
Mähönen 1997[99]	Response	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	Y	U	Y	HIGH
	Risk	**Low**				**Low**		**Low**					**Low**			
Mähönen 2013[67]	Response	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Low**					**Low**			
Mard 2010[28]	Response	Y	U	Y	Y	Y	Y	Y	U	Y	Y	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Low**					**Low**			
McAlpine 1998[42]	Response	Y	U	Y[h]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Low**					**Low**			
Merry 2009[65]	Response	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Low**					**Low**			
Nieuwkamp 2014[61]	Response	Y	Y	Y	Y	Y	Y	U	U	U	N	Y	Y	Y	Y	MEDIUM
	Risk	**Low**				**Low**		**High**					**Low**			
Nilsson 1994[88]	Response	Y	Y	U	Y	Y	Y	Y	U	Y	U	Y	Y	Y	Y	HIGH

		Patient selection				Index test		Reference standard					Flow and timing			Overall quality
		1	2	3	4	1	2	1	2	3	4	5	1	2	3	
	Risk	**Low**				**Low**		**Unclear**					**Low**			
Øie 2018[89]	Response	Y	Y	Y	Y	Y	Y	Y	U	Y	U	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Unclear**					**Low**			
Pajunen 2005[68]	Response	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	Y	U	Y	HIGH
	Risk	**Low**				**Low**		**Low**					**Low**			
Palomäki 1994[100]	Response	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	U	Y	HIGH
	Risk	**Low**				**Low**		**Unclear**					**Low**			
Pfister 2013[43]	Response	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Low**					**Low**			
Pietilä 1997[69]	Response	Y	Y	Y[i]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Low**					**Low**			
Rapola 1997[70]	Response	Y	Y	Y[g]	Y	Y	Y	U	U	Y	U	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Unclear**					**Low**			
Rinaldi 2003[90]	Response	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Low**					**Low**			
Rodrigo-Rincon 2015[91]	Response	U	U	U	N	Y	Y	Y	U	Y	N	Y	Y	U	Y	LOW
	Risk	**High**				**Low**		**High**					**Low**			
Ruigómez 2010[44]	Response	Y	U	Y	Y	Y	Y	Y	U	U	Y	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Unclear**					**Low**			
Sansom 2015[45]	Response	Y	Y	Y	Y	Y	N	U	U	Y	N	Y	Y	Y	Y	LOW
	Risk	**Low**				**High**		**High**					**Low**			
Sedova 2015[92]	Response	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Unclear**					**Low**			
Spolaore 2005[93]	Response	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Low**					**Low**			
Stegmayr 1992[66]	Response	Y	Y	Y	Y	U	N	Y	Y	U	Y	Y	Y	Y	Y	MEDIUM
	Risk	**Low**				**High**		**Low**					**Low**			
Sundbøll 2016[29]	Response	Y	Y	Y	Y	Y	Y	U	U	Y	U	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Unclear**					**Low**			

		Patient selection				Index test		Reference standard					Flow and timing			Overall quality
		1	2	3	4	1	2	1	2	3	4	5	1	2	3	
Thygesen 2011[30]	Response	Y	Y	Y	Y	Y	Y	Y	N	Y	U	Y	Y	Y	Y	MEDIUM
	Risk	**Low**				**Low**		**High**					**Low**			
Tolonen 2007[73]	Response	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Low**					**Low**			
Valk 2016[94]	Response	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Low**					**Low**			
Van Doorn 2017[95]	Response	Y	U	Y[c]	Y	Y	Y	Y	U	Y	N	Y	Y	Y	Y	MEDIUM
	Risk	**Low**				**Low**		**High**					**Low**			
Van Staa 1994[46]	Response	Y	Y	Y[j]	Y	Y	N	Y	U	Y	Y	Y	Y	Y	Y	MEDIUM
	Risk	**Low**				**High**		**Low**					**Low**			
Varmdal 2016[63]	Response	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Low**					**Low**			
Verdú-Rotellar 2017[96]	Response	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Low**					**Low**			
Vila-Corcoles 2014[62]	Response	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	Y	U	Y	Y	HIGH
	Risk	**Low**				**Low**		**Low**					**Low**			
Wildenschild 2014[31]	Response	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Low**					**Low**			
Wright 2012[47]	Response	Y	Y	Y[k]	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	HIGH
	Risk	**Low**				**Low**		**Unclear**					**Low**			
Zhou 2014[48]	Response	Y	Y	Y	Y	Y	N	U	U	U	N	Y	U	U	Y	LOW
	Risk	**Low**				**High**		**High**					**Unclear**			

[a] generalisable to diabetics; [b] generalisable to patients treated with PCI; [c] generalisable to patients with AF; [d] generalisable to patients ≥55 years treated with diuretics & NSAIDs; [e] generalisable to patients treated with a CVD drug; [f] generalisable to men aged 50 in 1970-74 with HF when aged 52- 90; [g] generalisable to men who smoke 5+ cigarettes a day; [h] generalisable to women without hysterectomy, pre-menopausal and have not had breast or ovarian cancer; [i] generalisable to men; [j] generalisable to patients treated with sulphonylurea medication; [k] generalisable to women

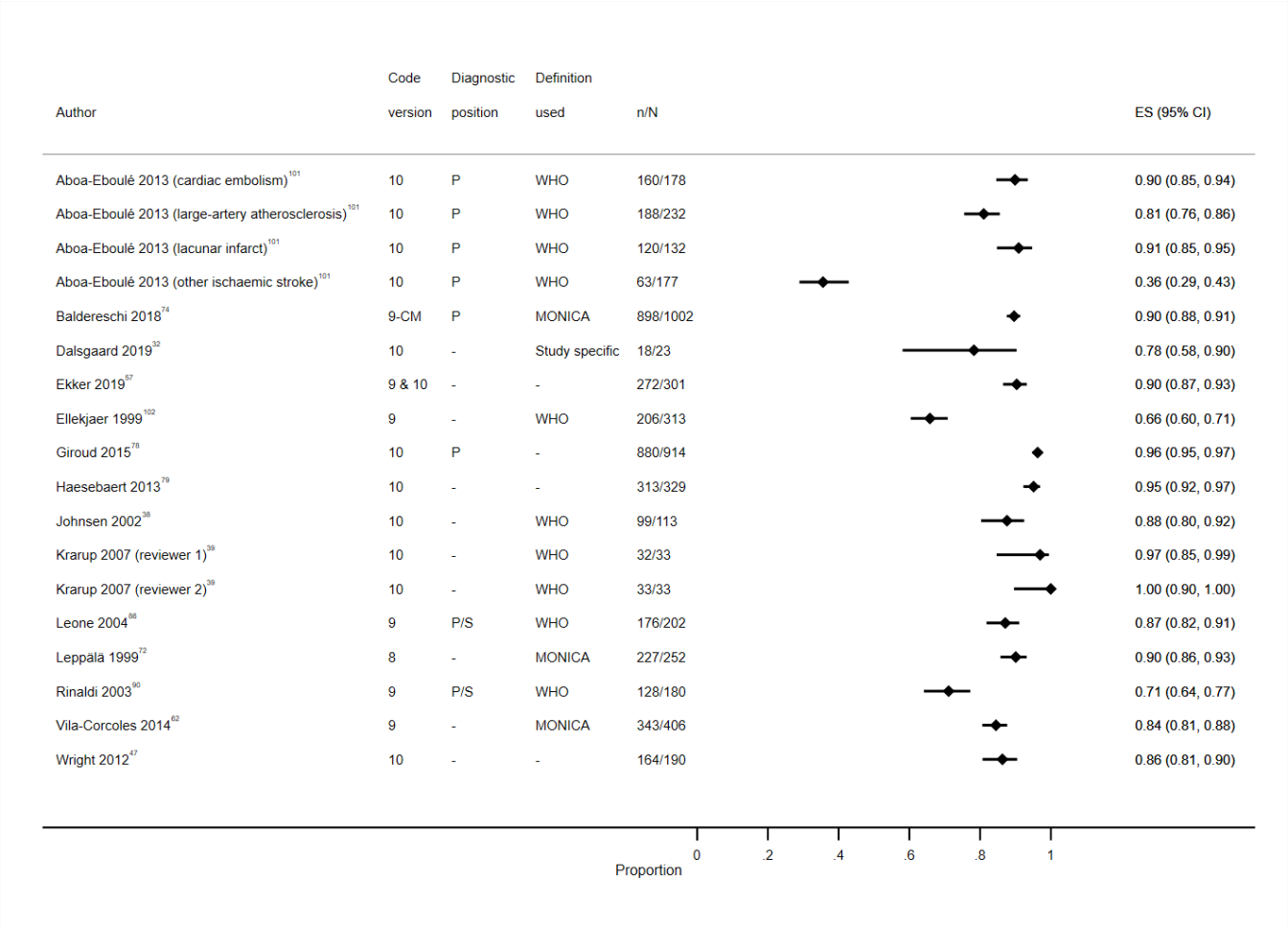
S4 Table. Assessment of quality of evidence for outcomesa

Number of studies[b]	Design	Risk of bias	Inconsistency	Imprecision	Any reason to upgrade	Quality
Sensitivity of HF diagnoses in secondary care EHRs						
6	6 validation studies	not serious	very serious[c]	likely serious[d]	none	⊕◯◯◯ very low
PPV of HF diagnoses in secondary care EHRs						
14	13 validation studies, 1 non-validation study	not serious	Serious[e]	serious	none	⊕◯◯◯ very low
PPV of HF diagnoses in primary care EHRs						
5	5 validation studies	serious	very serious	not serious	none	⊕◯◯◯ very low
Sensitivity of MI diagnoses in secondary care EHRs						
10	9 validation studies, 1 non-validation study	not serious	very serious[f]	not serious	none	⊕⊕◯◯ low
PPV of MI diagnoses in secondary care EHRs						
23	22 validation studies, 1 non-validation study	not serious	very serious[g]	not serious	none	⊕⊕◯◯ low
PPV of MI diagnoses in primary care EHRs						
5	5 validation studies	not serious	Serious[h]	not serious	none	⊕⊕⊕◯ moderate
Sensitivity of stroke diagnoses in secondary care EHRs						
15	15 validation studies	not serious	very serious[i]	serious	none	⊕◯◯◯ very low

PPV of stroke diagnoses in secondary care EHRs						
25	23 validation studies, 2 non-validation studies	not serious	Serious[j]	not serious	none	⊕⊕⊕◯ moderate
PPV of stroke diagnoses in primary care EHRs						
3	1 validation study, 2 non-validation studies	serious	serious, likely very serious[k]	not serious	none	⊕◯◯◯ very low

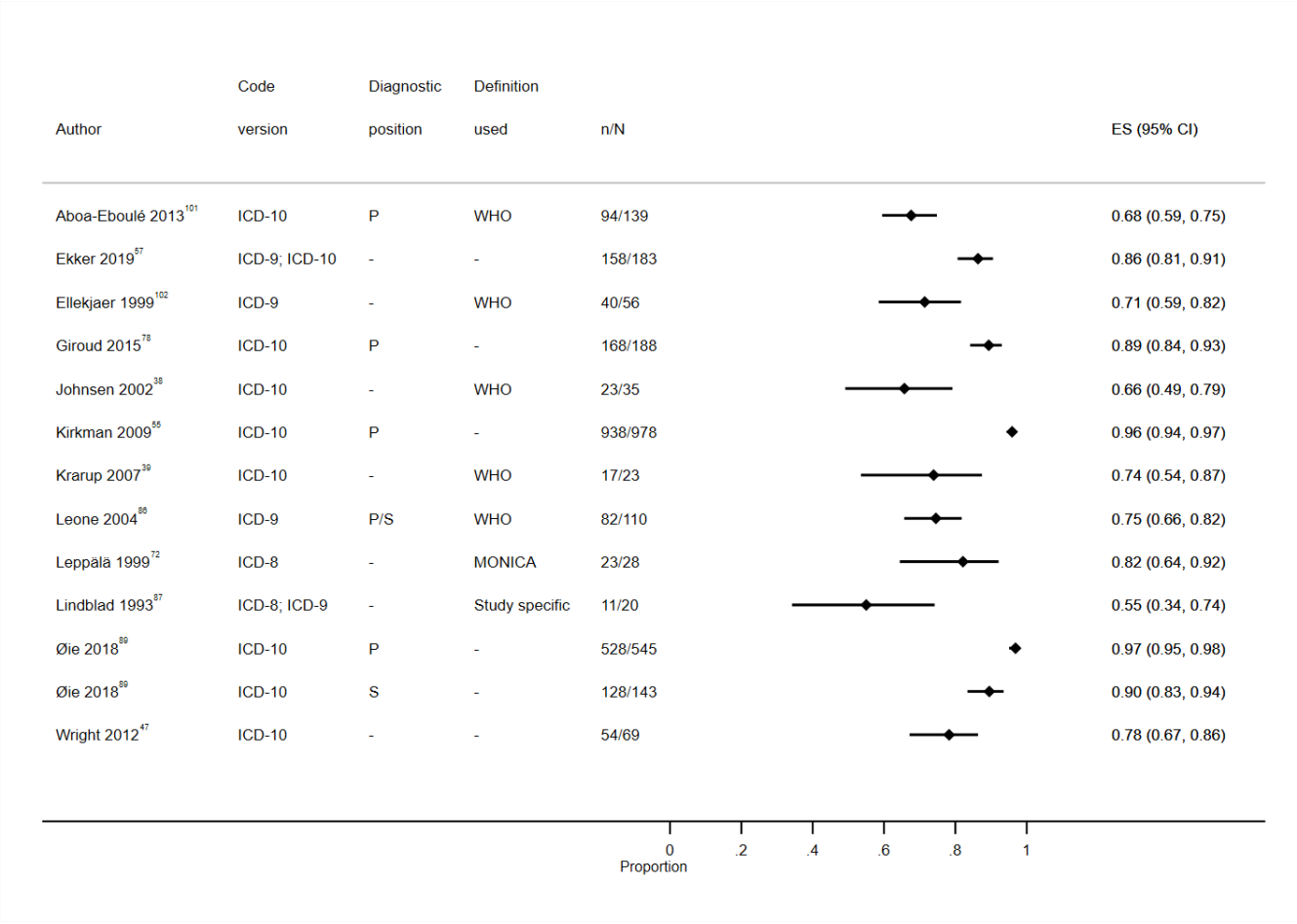
[a] publication bias could not be assessed due to the low number of studies which included the measures of sensitivity and specificity required for assessment and indirectness was not be assessed as this is not relevant to cross-study quality for validation studies; [b] only studies which reported a main result were included; [c] heterogeneity assessed in 5 studies with raw data; [d] confidence intervals only available for 1 study, which were wide; [e] heterogeneity assessed in 11 with raw data; [f] heterogeneity assessed in 4 studies with raw data; [g] heterogeneity assessed in 18 studies with raw data, confidence intervals available for 22 studies; [h] confidence intervals available for 4 studies; [h] heterogeneity assessed in 10 studies with raw data, confidence intervals available for 14 studies; [i] heterogeneity assessed in 18 studies with raw data; heterogeneity undetermined as raw data only available for 1 study

S1 Fig. Positive predictive value for ischaemic stroke diagnoses recorded in secondary care EHRs from studies which reported the number of records confirmed positive and the total number of records



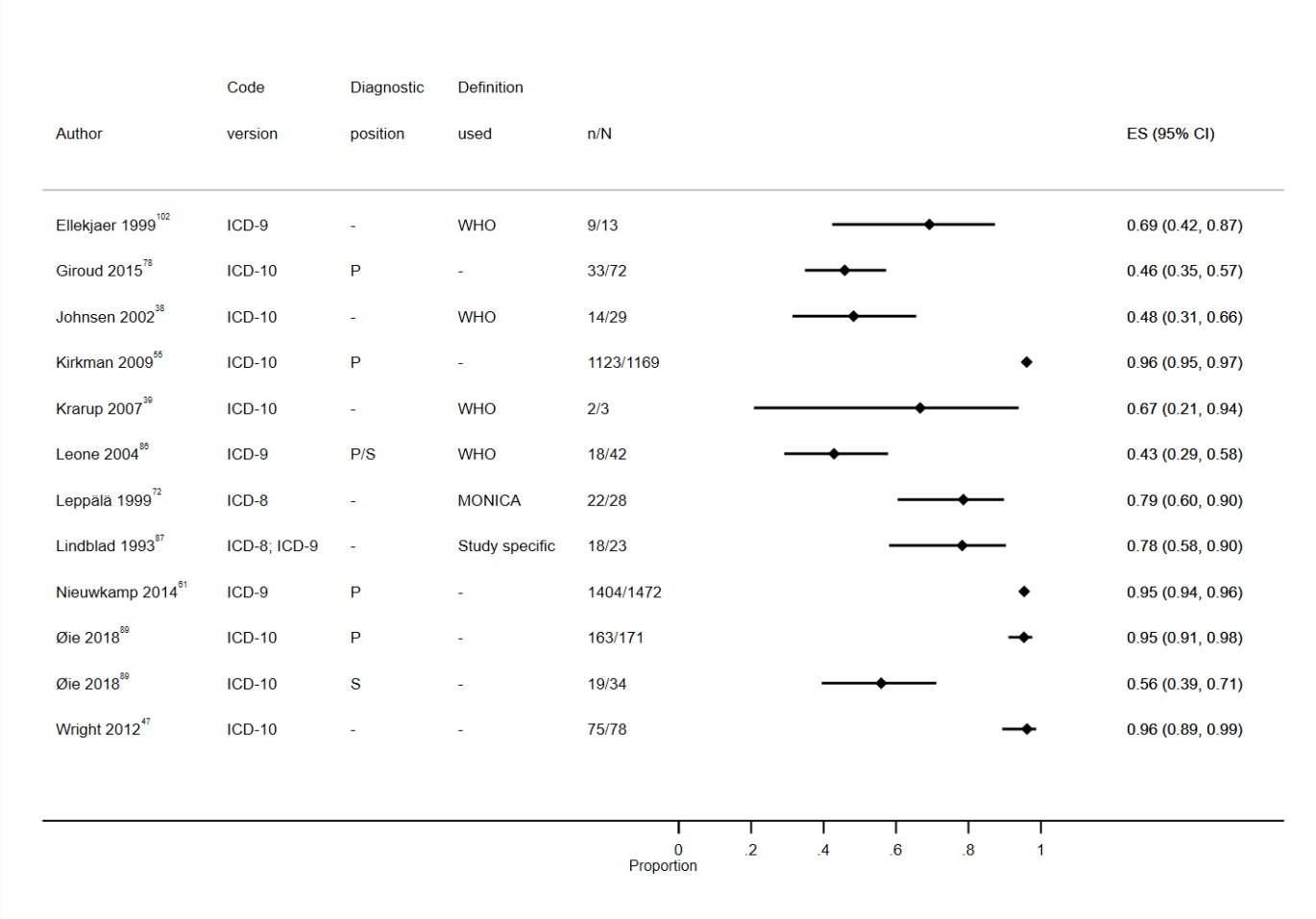
ES, effect size; P, primary; P/S, primary or secondary.

S2 Fig. Positive predictive value for intracerebral haemorrhage diagnoses recorded in secondary care EHRs from studies which reported the number of records confirmed positive and the total number of records



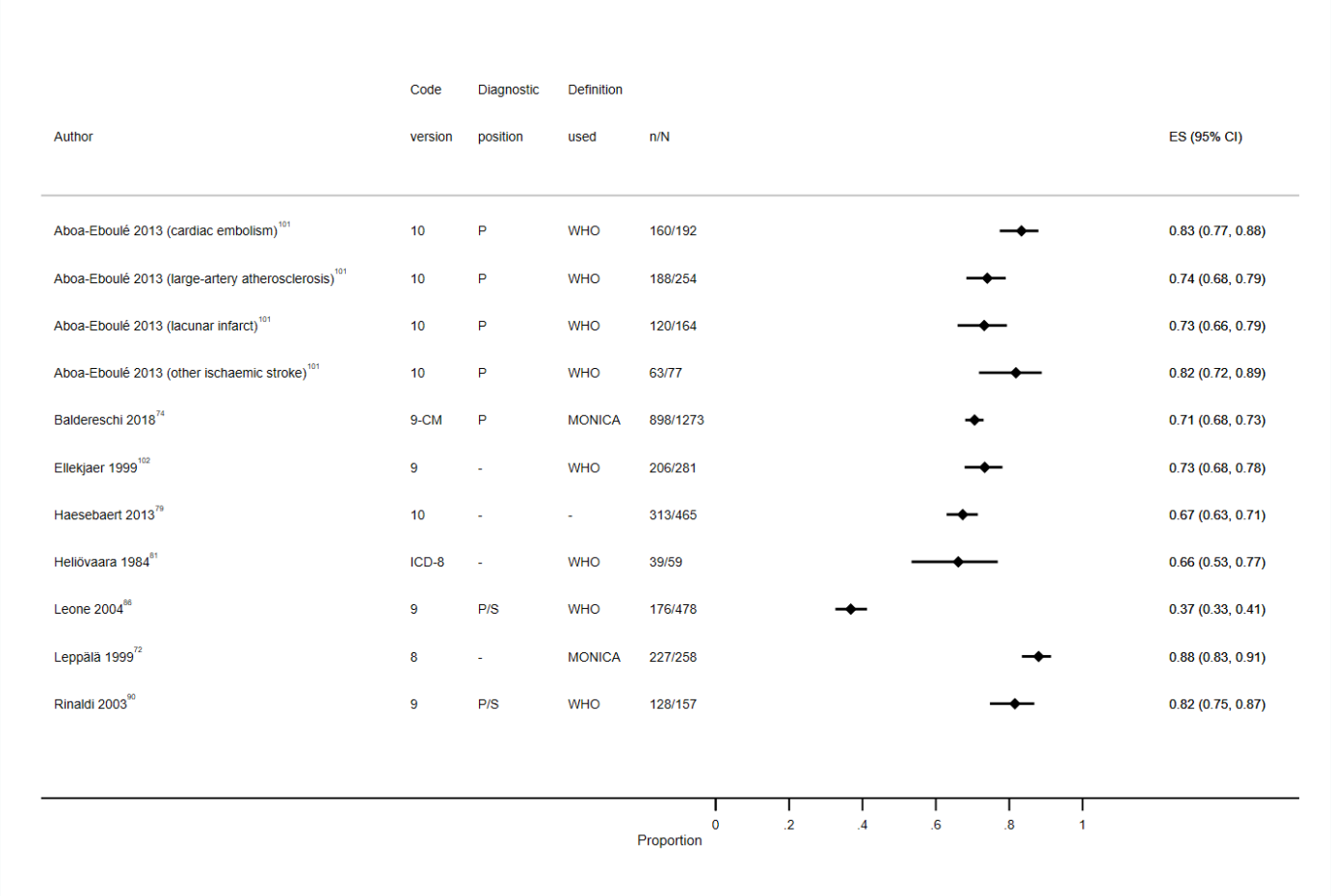
ES, effect size; P, primary; P/S, primary or secondary.

S3 Fig. Positive predictive value for subarachnoid haemorrhage diagnoses recorded in secondary care EHRs from studies which reported the number of records confirmed positive and the total number of records



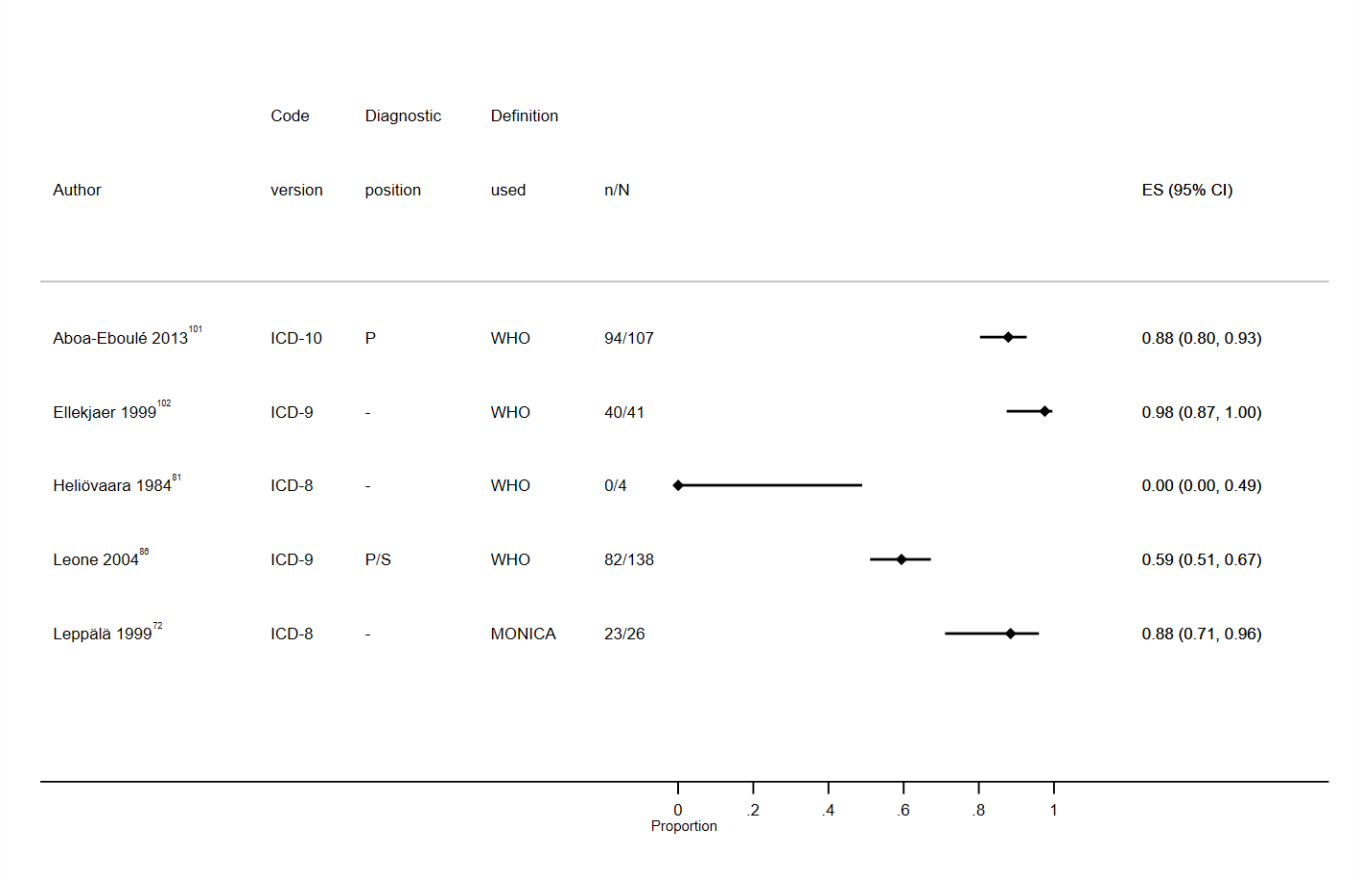
ES, effect size; P, primary; P/S, primary or secondary.

S4 Fig. Sensitivity for ischaemic stroke diagnoses recorded in secondary care EHRs from studies which reported the number of records confirmed positive and the total number of records



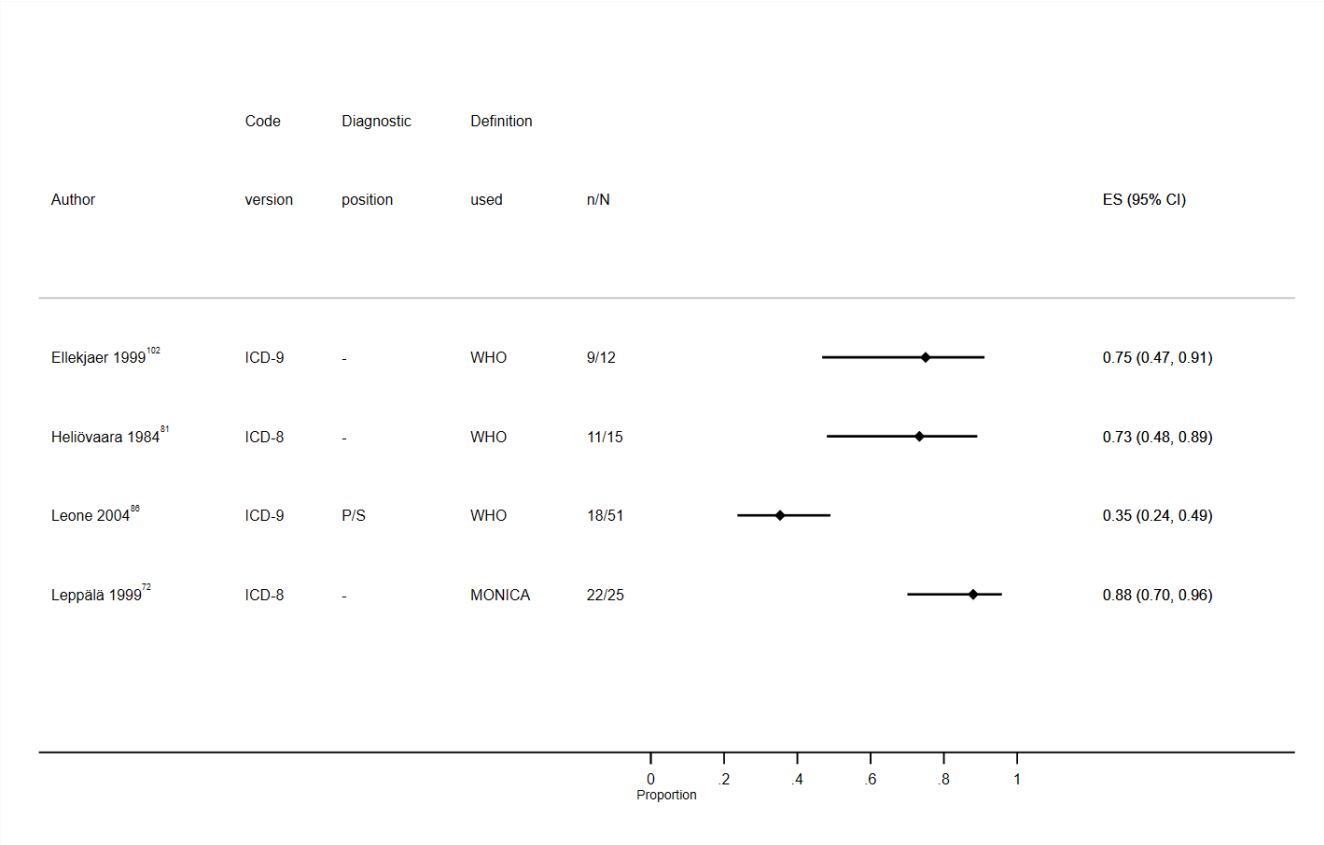
ES, effect size; P, primary; P/S, primary or secondary.

S5 Fig. Sensitivity for intracerebral haemorrhage diagnoses recorded in secondary care EHRs from studies which reported the number of records confirmed positive and the total number of records



ES, effect size; P, primary; P/S, primary or secondary.

S6 Fig. Sensitivity for subarachnoid haemorrhage diagnoses recorded in secondary care EHRs from studies which reported the number of records confirmed positive and the total number of records



ES, effect size; P, primary; P/S, primary or secondary.

THE LANCET
Digital Health

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: Davidson JA, Banerjee A, Smeeth L, et al. Risk of acute respiratory infection and acute cardiovascular events following acute respiratory infection among adults with increased cardiovascular risk in England between 2008 and 2018: a retrospective, population-based cohort study. *Lancet Digit Health* 2021; **3:** e773–83.

Supplementary material

Contents

Supplementary methods

Definitions used for excluded health conditions **2**

Study period and follow-up outline **3**

Acute respiratory infection episode structure **3**

Calculating QRISK2 scores **4**

Covariate and effect modifier selection **4**

Comparing results from recorded and calculated QRISK2 scores **4**

Supplementary tables

Supplementary table 1. Baseline demographic and lifestyle characteristics of included study population separated by database **5**

Supplementary table 2. Database comparison of acute respiratory infection incidence rates and incidence rate ratios **6**

Supplementary table 3. Crude and adjusted incidence rate ratios for the association between cardiovascular risk and ARI among sensitivity analysis study population **7**

Supplementary table 4. Crude and adjusted incidence rate ratios for the association between QRISK2 score and ARI, by QRISK2 score identification method **8**

Supplementary table 5. Acute cardiovascular events after influenza/ILI incidence rates and hazard ratios by cardiovascular risk group **9**

Supplementary table 6. Acute cardiovascular events after pneumonia incidence rates and incidence rate ratios by cardiovascular risk group **10**

Supplementary table 7. Acute cardiovascular events after ARI by anti-hypertensives, statins and antiplatelets prescription status **11**

Supplementary table 8. Database comparison of Acute cardiovascular events after ARI incidence rates and hazard ratios by cardiovascular risk group **12**

Supplementary table 9. Crude and adjusted hazard ratios for the association between cardiovascular risk and Acute cardiovascular events after ARI among sensitivity analysis study population **14**

Supplementary table 10. Crude and adjusted hazard ratios for the association between QRISK2 score and Acute cardiovascular events after acute respiratory infection, by QRISK2 score identification method **15**

Supplementary table 11. MACE after infection sensitivity analysis results **16**

Supplementary table 12. Crude and adjusted incidence rate ratios for the association between cardiovascular risk and M Acute cardiovascular events ACE after ARI among only patients who did not receive influenza or pneumococcal vaccine during follow-up **17**

Supplementary figures

Supplementary figure 1. Age-specific infection rates by cardiovascular risk group **18**

Supplementary figure 2. Timing between acute respiratory infection and major adverse cardiovascular event **19**

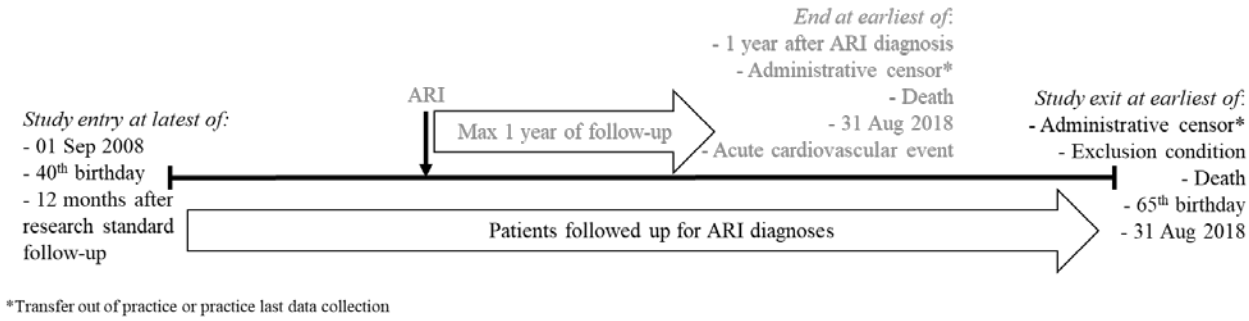
Supplementary material references **20**

Study protocol **21**

Definitions used for excluded health conditions

Health condition	Study definition
Cardiovascular disease (CVD)	Any previous clinical diagnosis, major intervention for, or clinical review specific to CVD including heart disease (congenital or otherwise), heart failure, stroke or transient ischaemic attack.
Chronic liver disease	Any previous clinical diagnosis of, or clinical review specific to, chronic liver disease including cirrhosis, oesophageal varices, biliary atresia and chronic hepatitis.
Chronic kidney disease (CKD)	Any previous clinical diagnosis of, or clinical review specific to, CKD stages 3-5, history of dialysis or renal transplant. Or with estimated glomerular filtration rate to classify CKD stages 3-5.[1] Only stages 4-5 excluded from sensitivity analysis using pneumococcal vaccine recommendations.
Chronic respiratory disease (not asthma)	Any previous clinical diagnosis of, or clinical review specific to, chronic respiratory disease, including chronic obstructive pulmonary disease, emphysema, bronchitis, cystic fibrosis, or fibrosing interstitial lung diseases.
Asthma	Any previous clinical diagnosis of, or clinical review specific to, asthma with at least two prescriptions of inhaled steroids in the year before baseline. Or any previous hospitalisation for asthma. Not excluded from sensitivity analysis using pneumococcal vaccine recommendations.
Chronic neurological disease	Any previous clinical diagnosis of, or clinical review specific to, a neurological disease such as Parkinson's disease, motor neurone disease, multiple sclerosis (MS), cerebral palsy, dementia or a learning/intellectual disability. Not excluded from sensitivity analysis using pneumococcal vaccine recommendations.
Diabetes mellitus	Any previous diagnosis of, or clinical review specific to, diabetes mellitus, or with a prescription for medication used to treat diabetes. Only treated diabetes excluded from sensitivity analysis using pneumococcal vaccine recommendations.
Asplenia/sickle cell disease	Any previous clinical diagnosis of, or clinical review specific to, asplenia or dysfunction of the spleen (including sickle cell disease but not sickle cell trait).
Severe obesity	Latest body mass index before baseline was $\geq$40 kg/m^2. Not excluded from sensitivity analysis using pneumococcal vaccine recommendations.
Immunosuppression	Any previous clinical diagnosis of, or clinical review specific to, HIV, solid organ transplant or other permanent immunosuppression (such as genetic conditions compromising immune function).
	Previous clinical diagnosis of, or clinical review specific to, aplastic anaemia or haematological malignancy, or receiving a bone marrow or stem cell transplant in the 2 years before baseline.
	Previous clinical diagnosis of, or clinical review specific to, other/unspecified immune deficiency or receiving chemotherapy or radiotherapy in the year before baseline.
	Prescription of biological therapy or at least 2 prescriptions for oral steroids or other immunosuppressants including DMARDS, Methotrexate, Azathioprine, or corticosteroid injections in the year before baseline.

Study period and follow-up outline



*Transfer out of practice or practice last data collection

Acute respiratory infection episode structure

Records containing ARI codes were identified in both CPRD and HES. All ARI codes were used to define the primary outcome (ARI) and subsets were used to define the secondary outcomes of influenza/influenza-like-illness (ILI) and pneumonia.

To account for multiple consultations related to one illness, CPRD or HES records dated within 28 days of each other were regarded as part of the same illness-episode. The earliest record in the episode determined the date of illness for the primary outcome of ARI and for the secondary outcome of influenza/ILI. By comparison, for the secondary outcome of pneumonia, the earliest record which was coded as pneumonia determined the date of illness. Figure A below illustrates the episode structure. Influenza/ILI and pneumonia episodes were structured differently to account for the aetiological differences in the conditions. In an episode where there was an influenza/ILI record, but this was not the first record, it is likely that the original presentation was due to influenza/ILI but not identified as such at the time. Conversely, pneumonia can develop and worsen over time from the original infection.

In analysis of ARI outcomes, to account for multiple episodes of illness per patient, an "observation" period ended at the date of ARI and a new "observation" period begin. In analysis of acute cardiovascular events after ARI, an ARI episode triggered a one-year follow-up. Further ARI episodes which occurred in that further up were counted and adjusted for in analysis.

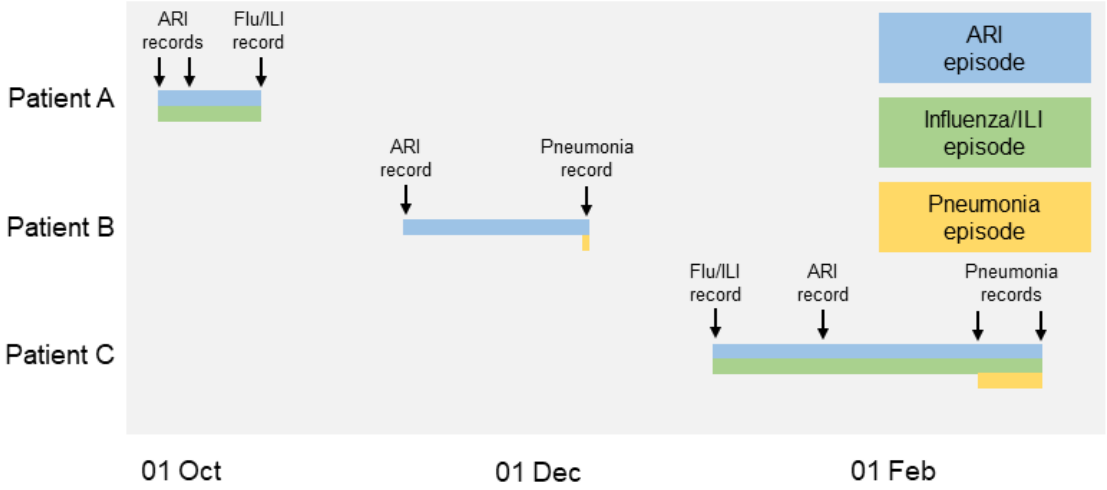


Figure A. Example patient episode classifications

Calculating QRISK2 scores

An individual's QRISK2 score is calculated based on age, sex, ethnicity, deprivation score from linked Townsend data, diabetes, family history of coronary heart disease in a first degree relative <60 years, atrial fibrillation, chronic kidney disease stage 4 or 5, rheumatoid arthritis, ratio of total serum cholesterol to high density lipoprotein cholesterol, systolic blood pressure, treated hypertension, body-mass index, and smoking status.[2] The complete QRISK2 algorithm used to calculate score has never been published but was updated annually during its use from 2008-2017. Using the information which the authors have published online (https://qrisk.org/), the London School of Hygiene and Tropical Medicine Electronic Health Record Research Group wrote Stata program files to calculate scores based on the information published for the 2015 version of QRISK2 and Quality and Outcomes Framework Read and SNOMED codes[3] (with the exception of chronic kidney disease stage 4 or 5 which used all available Read and SNOMED codes). A population-average imputation approach was used to account for missing data, to reflect the QRISK2 algorithm used in clinical practice. Our QRISK2 score process is published at https://zenodo.org/record/3981238.

Of note diabetes, chronic kidney disease and severe obesity are determinants of a higher QRISK2 score, but individuals with these conditions were excluded from our study (existing eligibility for influenza vaccination).

Covariate and effect modifier selection

We adjusted for age and sex as standard. Additionally, older age is associated with increase in cardiovascular risk[4] as well as ARI.[5] Men have higher cardiovascular risk than women,[4] while there is suggestion that women are more likely to experience ARI.[6] Non-White ethnicity is associated with both cardiovascular risk[4] and ARI.[7] Smoking, alcohol intake and BMI are important lifestyle factors, along with socio-economic deprivation, which effect the likelihood of many health conditions. Consultation frequency was adjusted for as hypertension (along with the factors included in the QRISK2 algorithm) and ARI diagnoses are more frequent in patients who regularly attend primary care services. No comorbid health conditions were included due to patients with many possible confounding conditions excluded from our study population.

Antihypertensives, statins and antiplatelets were considered effect modifiers in our analysis of cardiovascular risk and acute cardiovascular events after ARI, with stratified analysis conducted. Patients with diagnosed hypertension are likely to be prescribed antihypertensive treatments. Antihypertensive prescriptions were not used to classify patients with hypertension. Patients prescribed antihypertensives may have better-controlled hypertension and be less likely to experience an acute cardiovascular event following an ARI. Similarly, patients with hypertension may also be prescribed cholesterol-lowering medication. Individuals with a high QRISK2 score, many of whom have hypertension, are offered statin treatment.[8]

Comparing results from recorded and calculated QRISK2 scores

In main analyses, our classification of cardiovascular risk level based on QRISK2 was done using our own algorithm. To validate the results obtained from our algorithm we repeated the main analyses for ARI outcome and acute cardiovascular events after ARI restricted to patients with a QRISK2 score recorded in CPRD directly by GPs in 2015-2017. We chose this time period as our algorithm was based on the 2015 version of QRISK2, and minimal changes were made to QRISK2 from 2015-2017.

In CPRD GOLD we identified patients with QRISK2 scores using Read codes "22W..00" , "38DF.00" or "38DP.00" from the Clinical file. If any of these three codes were recorded for consultations between 2015-2017, then the corresponding results was obtained from the Additional file. In CPRD Aurum, we used SNOMED codes "718087004", "763244005", "1085871000000105", "1656451000006101", "1656461000006104", "810931000000108" from the Observation file.

In our subpopulation of patients with recorded QRISK2 scores we started follow-up from the latest of first recorded QRISK2 on or after 1 January 2015 or start of follow-up date from main analysis. Follow-up ended at the earliest of 31 December 2017 or end of follow-up date from main analysis.

Supplementary table 1. Baseline demographic and lifestyle characteristics of included study population separated by database

	Gold **n = 773,362**	**Aurum** **n = 3,439,568**
Age (years)		
40-44	316,376 (40·9%)	1,603,993 (46·6%)
45-49	152,207 (19·7%)	630,690 (18·3%)
50-54	121,385 (15·7%)	490,817 (14·3%)
55-59	99,557 (12·9%)	391,062 (11·4%)
60-64	83,837 (10·8%)	323,006 (9·4%)
Sex	n = 773,351	n = 3,439,547
Male	405,608 (52·4%)	1,820,953 (52·9%)
Female	367,743 (47·6%)	1,618,594 (47·1%)
Ethnicity	n = 642,421	n = 3,060,297
White	580,960 (90·4%)	2,661,147 (87·0%)
South Asian	24,419 (3·8%)	170,512 (5·6%)
Black	18,343 (2·9%)	135,927 (4·4%)
Mixed/Other	18,699 (2·9%)	92,711 (3·0%)
Townsend quintile	n = 772,985	n = 3,434,620
1 (least deprived)	190,835 (24·7%)	813,835 (23·7%)
2	180,189 (23·3%)	725,502 (21·1%)
3	158,379 (20·5%)	667,300 (19·4%)
4	138,537 (17·9%)	600,650 (17·5%)
5 (most deprived)	105,045 (13·6%)	627,333 (18·3%)
BMI category*	n = 772,985	n = 2,803,225
Underweight (<18·5 kg/m^2)	9,269 (1·4%)	41,733 (1·5%)
Normal weight (18·5-24·9 kg/m^2)	271,221 (42·1%)	1,178,462 (42·0%)
Overweight (25·0-29·9 kg/m^2)	238,478 (37·0%)	1,036,223 (37·0%)
Obese (30·0-39·9 kg/m^2)	125,411 (19·5%)	546,807 (19·5%)
Smoking status*	n = 747,925	n = 3,334,866
Non-smoker	354,909 (47·5%)	1,332,010 (39·9%)
Current smoker	193,384 (25·9%)	883,323 (26·5%)
Ex-smoker	199,632 (26·7%)	1,119,533 (33·6%)
Alcohol consumption*	n = 678,452	n = 2,996,042
Not a heavy drinker	654,536 (96·5%)	2,822,800 (94·2%)
Heavy drinker	26,109 (3·5%)	173,242 (5·8%)

Data are n (%). *Closest measure before start of follow-up.

Supplementary table 2. Database comparison of acute respiratory infection incidence rates and incidence rate ratios

Outcome	Cardiovascular risk	Database	Rate per 1,000 person-years (95% CI)		Crude IRR (95% CI)	Age and sex-adjusted IRR (95% CI)	Fully-adjusted* IRR (95% CI)
			High risk	Low risk			
ARI	Hypertension	GOLD	39·5 (38·7-40·4)	29·9 (29·6-30·1)	1·32 (1·29-1·35)	1·29 (1·26-1·32)	1·08 (1·05-1·11)
		Aurum	40·5 (40·1-40·9)	29·0 (28·9-29·2)	1·39 (1·37-1·40)	1·34 (1·32-1·35)	1·05 (1·03-1·06)
		GOLD and Aurum	40·3 (40·0-40·7)	29·1 (29·0-29·2)	1·38 (1·36-1·39)	1·33 (1·32-1·34)	1·04 (1·03-1·05)
		Meta-analysis of GOLD and Aurum	-	-	1·36 (1·29-1·43)	1·32 (1·27-1·37)	1·06 (1·03-1·09)
	QRISK2 ≥10%	GOLD	43·9 (42·9-44·8)	29·6 (29·3-29·8)	1·49 (1·45-1·52)	-	1·37 (1·34-1·41)
		Aurum	43·8 (43·3-44·2)	28·6 (28·5-28·8)	1·52 (1·51-1·54)	-	1·39 (1·37-1·40)
		GOLD and Aurum	43·8 (43·4-44·2)	28·8 (28·7-28·9)	1·52 (1·50-1·53)	-	1·39 (1·37-1·40)
		Meta-analysis of GOLD and Aurum	-	-	1·51 (1·48-1·54)	-	1·39 (1·37-1·40)
Influenza / ILI	Hypertension	GOLD	6·2 (5·9-6·5)	5·6 (5·5-5·7)	1·10 (1·05-1·16)	1·22 (1·16-1·29)	1·01 (0·96-1·07)
		Aurum	6·3 (6·1-6·4)	5·5 (5·4-5·5)	1·14 (1·12-1·17)	1·25 (1·22-1·28)	0·98 (0·95-1·00)
		GOLD and Aurum	6·3 (6·1-6·4)	5·5 (5·4-5·5)	1·14 (1·11-1·16)	1·25 (1·22-1·27)	0·98 (0·96-1·00)
		Meta-analysis of GOLD and Aurum	-	-	1·13 (1·09-1·16)	1·25 (1·22-1·27)	0·99 (0·96-1·01)
	QRISK2 ≥10%	GOLD	5·1 (4·8-5·3)	5·7 (5·6-5·8)	0·89 (0·84-0·94)	-	0·83 (0·78-0·88)
		Aurum	5·4 (5·3-5·5)	5·5 (5·5-5·6)	0·97 (0·95-0·99)	-	0·89 (0·87-0·91)
		GOLD and Aurum	5·4 (5·3-5·5)	5·6 (5·5-5·6)	0·96 (0·94-0·98)	-	0·88 (0·86-0·90)
		Meta-analysis of GOLD and Aurum	-	-	0·93 (0·86-1·01)	-	0·87 (0·81-0·92)
Pneumonia	Hypertension	GOLD	1·9 (1·7-2·1)	1·3 (1·3-1·4)	1·46 (1·32-1·61)	1·25 (1·13-1·39)	1·07 (0·97-1·19)
		Aurum	2·4 (2·3-2·5)	1·6 (1·5-1·6)	1·61 (1·55-1·68)	1·33 (1·28-1·39)	1·12 (1·07-1·17)
		GOLD and Aurum	2·3 (2·2-2·4)	1·5 (1·5-1·5)	1·59 (1·53-1·65)	1·32 (1·27-1·38)	1·12 (1·07-1·16)
		Meta-analysis of GOLD and Aurum	-	-	1·55 (1·40-1·69)	1·31 (1·25-1·38)	1·11 (1·07-1·16)
	QRISK2 ≥10%	GOLD	3·1 (2·9-3·3)	1·2 (1·2-1·3)	2·62 (2·41-2·86)	-	2·37 (2·17-2·60)
		Aurum	3·5 (3·4-3·6)	1·4 (1·4-1·5)	2·59 (2·50-2·69)	-	2·31 (2·23-2·39)
		GOLD and Aurum	3·5 (3·4-3·6)	1·4 (1·4-1·4)	2·60 (2·52-2·69)	-	2·32 (2·25-2·40)
		Meta-analysis of GOLD and Aurum	-	-	2·60 (2·51-2·68)	-	2·32 (2·24-2·39)

In meta-analysis the between database heterogeneity was assessed using the I^2 statistic. I^2 results for fully-adjusted estimates: ARI and hypertension=67%, ARI and QRISK2=6%, influenza/ILI and hypertension=0%, influenza/ILI and QRISK2=79%, pneumonia and hypertension=0%, pneumonia and QRISK2=0%. *Hypertension models adjusted for: age, sex, ethnicity, socioeconomic status, BMI, alcohol intake, smoking status and consultation frequency. QRISK2 models adjusted for: alcohol intake and consultation frequency.

Supplementary table 3. Crude and adjusted incidence rate ratios for the association between cardiovascular risk and ARI among sensitivity analysis study population

Outcome	Cardiovascular risk	No. of events	Rate per 1,000 person-years (95% CI)	Crude IRR (95% CI)	Age and sex-adjusted IRR (95% CI)	Fully-adjusted* IRR (95% CI)
Acute respiratory infection	Hypertension	96,203	42·7 (42·4-43·1)	1·40 (1·38-1·41)	1·34 (1·33-1·36)	1·04 (1·03-1·05)
	No hypertension	558,302	30·4 (30·3-30·5)	1	1	1
	QRISK2 ≥10%	96,193	45·8 (45·4-46·2)	1·51 (1·50-1·52)	-	1·37 (1·36-1·39)
	QRISK2 <10%	558,312	30·1 (30·0-30·2)	1	-	1
Influenza/ILI	Hypertension	14,436	6·4 (6·3-6·5)	1·14 (1·12-1·17)	1·25 (1·23-1·27)	0·98 (0·96-1·00)
	No hypertension	102,207	5·6 (5·5-5·6)	1	1	1
	QRISK2 ≥10%	11,573	5·5 (5·4-5·6)	0·97 (0·95-0·99)	-	0·88 (0·86-0·90)
	QRISK2 <10%	105,070	5·7 (5·6-5·7)	1	-	1
Pneumonia	Hypertension	5,700	2·5 (2·5-2·6)	1·61 (1·56-1·67)	1·35 (1·30-1·39)	1·11 (1·07-1·16)
	No hypertension	29,987	1·6 (1·6-1·7)	1	1	1
	QRISK2 ≥10%	7,798	3·7 (3·6-3·8)	2·59 (2·51-2·67)	-	2·30 (2·23-2·37)
	QRISK2 <10%	27,889	1·5 (1·5-1·5)	1	-	1

Total person-years per 1,000: hypertension = 2,251·4, no hypertension = 18,373·1, QRISK2 ≥10% = 2,099·5 and QRISK2 <10% = 18,524·0. LRT p-values all <0·0001. *Hypertension models adjusted for: age, sex, ethnicity, socioeconomic status, BMI, alcohol intake, smoking status and consultation frequency. QRISK2 models adjusted for: alcohol intake and consultation frequency.

Supplementary table 4. Crude and adjusted incidence rate ratios for the association between QRISK2 score and ARI, by QRISK2 score identification method

Cardiovascular risk method	No. of events	Rate per 1,000 person-years	Crude IRR (95% CI)	Alcohol intake adjusted IRR (95% CI)
Recorded QRISK2 ≥10%	6,018	27·6 (26·8-28·4)	1·41 (1·37-1·46)	1·42 (1·38-1·47)
Recorded QRISK2 <10%	22,561	19·7 (19·5-20·0)	1	1
Calculated QRISK2 ≥10%	4,296	29·6 (28·6-30·6)	1·48 (1·43-1·54)	1·40 (1·35-1·46)
Calculated QRISK2 <10%	24,283	20·0 (19·7-20·3)	1	1

Total person-years per 1,000: calculated QRISK2 ≥10% = 145·30, calculated QRISK2 <10% = 1,215·49, recorded QRISK2 ≥10% = 218·09 and recorded QRISK2 <10% = 1,142·70.

Supplementary table 5. Acute cardiovascular events after influenza/ILI incidence rates and hazard ratios by cardiovascular risk group

Outcome	Cardiovascular risk	No. of events	Rate per 1,000 person-years (95% CI)	Crude HR (95% CI)	Age and sex-adjusted HR (95% CI)	Fully-adjusted* HR (95% CI)
Any event	Hypertension	99	8·0 (6·6-9·8)	2·24 (1·79-2·80)	2·13 (1·70-2·66)	2·07 (1·60-2·67)
	No hypertension	352	3·7 (3·3-4·1)	1	1	1
	QRISK2 ≥10%	116	11·5 (9·7-13·9)	3·36 (2·72-4·15)	-	3·35 (2·70-4·17)
	QRISK2 <10%	335	3·4 (3·1-3·8)	1	-	1
ACS	Hypertension	34	2·8 (2·0-3·9)	1·99 (1·36-2·90)	1·87 (1·28-2·73)	1·97 (1·30-2·99)
	No hypertension	136	1·4 (1·2-1·7)	1	1	1
	QRISK2 ≥10%	51	5·1 (3·9-6·8)	4·15 (2·99-5·77)	-	4·04 (2·89-5·63)
	QRISK2 <10%	119	1·2 (1·0-1·5)	1	-	1
Heart failure	Hypertension	21	1·7 (1·1-2·7)	2·85 (1·73-4·69)	2·69 (1·63-4·43)	2·82 (1·53-5·21)
	No hypertension	59	0·6 (0·5-0·8)	1	1	1
	QRISK2 ≥10%	22	2·2 (1·5-3·4)	3·68 (2·25-6·01)	-	3·72 (2·23-6·19)
	QRISK2 <10%	58	0·6 (0·5-0·8)	1	-	1
Stroke or TIA	Hypertension	45	3·6 (2·7-5·0)	2·33 (1·67-3·25)	2·24 (1·60-3·13)	2·04 (1·39-2·98)
	No hypertension	153	1·6 (1·4-1·9)	1	1	1
	QRISK2 ≥10%	41	4·1 (3·0-5·6)	2·53 (1·80-3·57)	-	2·49 (1·74-3·57)
	QRISK2 <10%	157	1·6 (1·4-1·9)	1	-	1
CVD death	Hypertension	12	1·0 (0·6-1·8)	2·81 (1·46-5·41)	2·66 (1·38-5·12)	3·31 (1·51-7·25)
	No hypertension	34	0·4 (0·3-0·5)	1	1	1
	QRISK2 ≥10%	13	1·3 (0·8-2·4)	3·82 (2·01-7·26)	-	4·19 (2·17-8·07)
	QRISK2 <10%	33	0·3 (0·2-0·5)	1	-	1

Acute limb ischaemia not included as secondary outcome due to event numbers <10. Total person-years per 1,000: hypertension = 12·3, no hypertension = 95·7, QRISK2 ≥10% = 10·0 and QRISK2 <10% = 98·0. LRT p-values all <0·0001. *Hypertension models adjusted for: age, sex, ethnicity, socioeconomic status, BMI, alcohol intake and smoking status. QRISK2 models adjusted for: alcohol intake. †Results subdivided into MI and unstable angina separately: MI fully-adjusted HR in hypertension model 2·34 (95% CI 1·48-3·71) and in QRISK2 model 4·84 (95% CI 3·36-6·97), angina fully-adjusted HR in hypertension model 0·89 (95% CI 0·30-2·61) and in QRISK2 model 2·17 (95% CI 0·96-4·91). ‡Results subdivided into stroke and TIA separately: stroke fully-adjusted HR in hypertension model 2·40 (95% CI 1·54-3·74) and in QRISK2 model 2·69 (95% CI 1·76-4·12), TIA fully-adjusted HR in hypertension model 1·49 (95% CI 0·76-2·91) and in QRISK2 model 2·47 (95% CI 1·33-4·56).

Supplementary table 6. Acute cardiovascular events after pneumonia incidence rates and incidence rate ratios by cardiovascular risk group

Outcome	Cardiovascular risk	No. of events	Rate per 1,000 person-years (95% CI)	Crude HR (95% CI)	Age and sex-adjusted HR (95% CI)	Fully-adjusted* HR (95% CI)
Any event	Hypertension	341	105·6 (94·5-118·3)	1·62 (1·44-1·82)	1·58 (1·41-1·78)	1·65 (1·44-1·89)
	No hypertension	1,265	63·6 (60·1-67·4)	1	1	1
	QRISK2 ≥10%	574	130·9 (120·1-142·9)	2·17 (1·96-2·40)	-	2·13 (1·92-2·37)
	QRISK2 <10%	1,032	55·1 (51·8-58·7)	1	-	1
ACS	Hypertension	117	36·1 (30·1-43·6)	1·73 (1·41-2·12)	1·68 (1·37-2·06)	1·91 (1·50-2·44)
	No hypertension	405	20·3 (18·4-22·5)	1	1	1
	QRISK2 ≥10%	217	49·5 (43·3-56·9)	2·75 (2·31-3·26)	-	2·82 (2·35-3·38)
	QRISK2 <10%	305	16·2 (14·5-18·2)	1	-	1
Heart failure	Hypertension	145	44·8 (38·1-53·1)	1·59 (1·32-1·90)	1·56 (1·30-1·87)	1·70 (1·37-2·11)
	No hypertension	545	27·4 (25·2-29·8)	1	1	1
	QRISK2 ≥10%	253	57·7 (50·9-65·9)	2·26 (1·94-2·64)	-	2·20 (1·87-2·58)
	QRISK2 <10%	437	23·3 (21·2-25·7)	1	-	1
Acute limb ischaemia	Hypertension	10	3·1 (1·7-6·3)	2·32 (1·12-4·80)	2·30 (1·11-4·76)	3·19 (1·34-7·55)
	No hypertension	25	1·3 (0·9-1·9)	1	1	1
	QRISK2 ≥10%	15	3·4 (2·1-6·0)	3·21 (1·65-6·26)	-	3·26 (1·65-6·42)
	QRISK2 <10%	20	1·1 (0·7-1·7)	1	-	1
Stroke or TIA	Hypertension	124	38·4 (32·2-46·1)	1·85 (1·52-2·26)	1·83 (1·50-2·23)	1·79 (1·43-2·24)
	No hypertension	400	20·1 (18·2-22·2)	1	1	1
	QRISK2 ≥10%	160	36·5 (31·2-42·9)	1·71 (1·42-2·06)	-	1·67 (1·38-2·04)
	QRISK2 <10%	364	19·4 (17·5-21·6)	1	-	1
CVD death	Hypertension	58	17·9 (13·9-23·5)	1·60 (1·20-2·14)	1·56 (1·17-2·08)	1·57 (1·11-2·23)
	No hypertension	219	11·0 (9·6-12·6)	1	1	1
	QRISK2 ≥10%	120	27·4 (22·9-33·0)	3·02 (2·39-3·83)	-	3·11 (2·39-4·04)
	QRISK2 <10%	157	8·4 (7·2-9·8)	1	-	1

Total person-years per 1,000: hypertension = 3·2, no hypertension = 19·9, QRISK2 ≥10% = 4·4 and QRISK2 <10% = 18·7. LRT p-values all <0·0001. *Hypertension models adjusted for: age, sex, ethnicity, socioeconomic status, BMI, alcohol intake and smoking status. QRISK2 models adjusted for: alcohol intake. †Results subdivided into MI and unstable angina separately: MI fully-adjusted HR in hypertension model 1·94 (95% CI 1·48-2·54) and in QRISK2 model 3·04 (95% CI 2·49-3·70), angina fully-adjusted HR in hypertension model 2·42 (95% CI 1·38-4·22) and in QRISK2 model 2·03 (95% CI 1·24-3·33). ‡Results subdivided into stroke and TIA separately: stroke fully-adjusted HR in hypertension model 1·78 (95% CI 1·42-2·23) and in QRISK2 model 1·77 (95% CI 1·45-2·15), TIA fully-adjusted HR in hypertension model 1·74 (95% CI 0·74-4·08) and in QRISK2 model 1·14 (95% CI 0·56-2·34).

Supplementary table 7. Acute cardiovascular events after ARI by anti-hypertensives, statins and antiplatelets prescription status

Prescribed treatment of interest	Cardiovascular risk	No. of events	Total person-years per 1,000	Rate per 1,000 person-years (95% CI)	Crude HR (95% CI)	Age and sex-adjusted HR (95% CI)	Fully-adjusted* HR (95% CI)
Anti-hypertensives (n=57,800)	Hypertension (n=42,888)	532	55·6	9·6 (8·8-10·4)	1·00 (0·84-1·18)	0·86 (0·72-1·02)	1·03 (0·83-1·27)
	No hypertension (n=14,912)	183	19·0	9·6 (8·4-11·2)	1	1	1
	QRISK2 ≥10% (n=22,805)	436	28·9	15·1 (13·7-16·6)	2·45 (2·11-2·85)	-	2·47 (2·12-2·89)
	QRISK2 <10% (n=34,995)	279	45·6	6·1 (5·4-6·9)	1	-	1
No anti-hypertensives (n=469,000)	Hypertension (n=25,843)	453	32·2	14·1 (12·8-15·4)	2·62 (2·37-2·89)	2·58 (2·34-2·85)	2·56 (2·30-2·86)
	No hypertension (n=443,157)	3,001	534·5	5·6 (5·4-5·8)	1	1	1
	QRISK2 ≥10% (n=49,332)	1,090	58·1	18·8 (17·7-19·9)	4·04 (3·76-4·34)	-	3·93 (3·65-4·24)
	QRISK2 <10% (n=419,668)	2,364	508·7	4·6 (4·5-4·8)	1	-	1
Statins (n=33,837)	Hypertension (n=16,865)	246	21·9	11·3 (9·9-12·8)	1·09 (0·91-1·30)	1·06 (0·88-1·26)	1·11 (0·91-1·35)
	No hypertension (n=16,972)	228	22·1	10·3 (9·1-11·8)	1	1	1
	QRISK2 ≥10% (n=16,746)	322	21·5	15·0 (13·4-16·7)	2·22 (1·83-2·69)	-	2·16 (1·77-2·63)
	QRISK2 <10% (n=17,091)	152	22·5	6·8 (5·8-8·0)	1	-	1
No statins (n=492,963)	Hypertension (n=51,866)	739	66·0	11·2 (10·4-12·1)	2·14 (1·98-2·32)	2·09 (1·93-2·27)	2·13 (1·94-2·34)
	No hypertension (n=441,097)	2,956	531·4	5·6 (5·4-5·8)	1	1	1
	QRISK2 ≥10% (n=55,391)	1,204	65·5	18·4 (17·4-19·5)	3·93 (3·67-4·21)	-	3·83 (3·57-4·11)
	QRISK2 <10% (n=437,572)	2,491	531·8	4·7 (4·5-4·9)	1	-	1
Antiplatelets (n=7,670)	Hypertension (n=3,658)	88	4·4	20·0 (16·3-24·9)	0·86 (0·65-1·13)	0·78 (0·59-1·04)	0·94 (0·68-1·28)
	No hypertension (n=4,012)	112	4·9	22·9 (19·1-27·8)	1	1	1
	QRISK2 ≥10% (n=3,970)	131	4·7	28·2 (23·7-33·7)	1·82 (1·36-2·44)	-	1·72 (1·27-2·33)
	QRISK2 <10% (n=3,700)	69	4·6	14·9 (11·8-19·1)	1	-	1
No antiplatelets (n=519,130)	Hypertension (n=65,073)	897	83·4	10·8 (10·1-11·5)	2·05 (1·91-2·21)	1·96 (1·82-2·11)	1·97 (1·81-2·15)
	No hypertension (n=454,057)	3,072	548·6	5·6 (54-5·8)	1	1	1
	QRISK2 ≥10% (n=68,167)	1,395	82·4	16·9 (16·1-17·9)	3·68 (3·45-3·93)	-	3·60 (3·36-3·85)
	QRISK2 <10% (n=450,963)	2,574	549·7	4·7 (4·5-4·9)	1	-	1

*Hypertension models adjusted for: age, sex, ethnicity, socioeconomic status, BMI, alcohol intake and smoking status. QRISK2 models adjusted for: alcohol intake.

Supplementary table 8. Database comparison of acute cardiovascular events after ARI incidence rates and hazard ratios by cardiovascular risk group

Outcome	Cardiovascular risk	Database	Rate per 1,000 person-years (95% CI)		Crude HR (95% CI)	Age and sex-adjusted HR (95% CI)	Fully-adjusted* HR (95% CI)
			High risk	Low risk			
Any event	Hypertension	GOLD	11·5 (9·7-13·6)	5·9 (5·4-6·4)	2·03 (1·68-2·46)	1·92 (1·59-2·33)	2·06 (1·67-2·54)
		Aurum	11·2 (10·4-12·0)	5·7 (5·5-6·0)	2·08 (1·93-2·25)	1·98 (1·84-2·14)	1·97 (1·80-2·16)
		GOLD and Aurum	11·2 (10·5-12·0)	5·8 (5·5-6·0)	2·08 (1·93-2·23)	1·97 (1·84-2·12)	1·98 (1·83-2·15)
		Meta-analysis of GOLD and Aurum	-	-	2·07 (1·93-2·22)	1·97 (1·83-2·11)	1·98 (1·82-2·15)
	QRISK2	GOLD	17·2 (15·0-19·9)	5·1 (4·6-5·6)	3·37 (2·85-4·00)	-	3·35 (2·81-3·99)
		Aurum	17·6 (16·7-18·6)	4·7 (4·5-4·9)	3·80 (3·55-4·07)	-	3·70 (3·44-3·97)
		GOLD and Aurum	17·5 (16·7-18·5)	4·8 (4·6-5·0)	3·65 (3·42-3·89)	-	3·65 (3·42-3·89)
		Meta-analysis of GOLD and Aurum	-	-	3·67 (3·27-4·06)	-	3·63 (3·35-3·91)
ACS	Hypertension	GOLD	4·8 (3·7-6·2)	2·2 (1·9-2·6)	2·24 (1·66-3·02)	2·10 (1·55-2·83)	2·18 (1·54-3·08)
		Aurum	4·1 (3·7-4·6)	2·0 (1·9-2·2)	2·18 (1·92-2·48)	2·06 (1·81-2·34)	2·12 (1·83-2·46)
		GOLD and Aurum	4·2 (3·8-4·7)	2·1 (1·9-2·2)	2·19 (1·95-2·46)	2·06 (1·83-2·32)	2·13 (1·86-2·44)
		Meta-analysis of GOLD and Aurum	-	-	2·19 (1·93-2·45)	2·07 (1·82-2·31)	2·13 (1·84-2·42)
	QRISK2	GOLD	7·0 (5·7-8·8)	1·9 (1·6-2·2)	3·66 (2·80-4·79)	-	3·79 (2·87-5·00)
		Aurum	7·0 (6·5-7·7)	1·6 (1·5-1·7)	4·56 (4·08-5·10)	-	4·48 (3·99-5·03)
		GOLD and Aurum	7·0 (6·5-7·6)	1·6 (1·5-1·7)	4·42 (3·98-4·89)	-	4·37 (3·93-4·86)
		Meta-analysis of GOLD and Aurum	-	-	4·22 (3·36-5·01)	-	4·30 (3·70-4·89)
Heart failure	Hypertension	GOLD	4·2 (3·2-5·6)	1·7 (1·5-2·0)	2·53 (1·83-3·49)	2·39 (1·74-3·30)	2·83 (1·96-4·08)
		Aurum	3·2 (2·8-3·6)	1·7 (1·6-1·9)	1·96 (1·69-2·26)	1·85 (1·60-2·13)	1·96 (1·65-2·32)
		GOLD and Aurum	3·3 (2·9-3·7)	1·7 (1·6-1·9)	2·04 (1·79-2·32)	1·92 (1·69-2·19)	2·08 (1·79-2·42)
		Meta-analysis of GOLD and Aurum	-	-	2·11 (1·62-2·59)	1·99 (1·53-2·45)	2·24 (1·44-3·04)
	QRISK2	GOLD	5·5 (4·3-7·1)	1·6 (1·3-1·9)	3·50 (2·58-4·73)	-	3·43 (2·51-4·70)
		Aurum	5·5 (5·0-6·1)	1·4 (1·3-1·5)	4·10 (3·62-4·63)	-	3·93 (3·45-4·48)
		GOLD and Aurum	5·5 (5·0-6·0)	1·4 (1·3-1·5)	4·00 (3·57-4·49)	-	3·85 (3·42-4·34)
		Meta-analysis of GOLD and Aurum	-	-	3·99 (3·53-4·45)	-	3·84 (3·37-4·31)
Acute limb ischaemia	Hypertension	GOLD	0·0 (0·0-0·0)	0·1 (0·1-0·2)	-	-	-
		Aurum	0·3 (0·2-0·5)	0·1 (0·1-0·1)	3·49 (2·15-5·68)	3·32 (2·03-5·43)	5·72 (3·25-10·08)
		GOLD and Aurum	0·3 (0·2-0·4)	0·1 (0·1-0·1)	2·98 (1·85-4·78)	2·82 (1·74-4·55)	4·63 (2·68-7·99)
		Meta-analysis of GOLD and Aurum	-	-	-	-	-
	QRISK2	GOLD	0·3 (0·1-1·3)	0·1 (0·0-0·2)	4·18 (0·99-17·68)	-	4·21 (0·99-17·85)

Outcome	Cardiovascular risk	Database	Rate per 1,000 person-years (95% CI)		Crude HR (95% CI)	Age and sex-adjusted HR (95% CI)	Fully-adjusted* HR (95% CI)
			High risk	Low risk			
		Aurum	0·5 (0·4-0·7)	0·1 (0·0-0·1)	7·56 (4·76-12·00)	-	7·34 (4·56-11·80)
		GOLD and Aurum	0·5 (0·4-0·7)	0·1 (0·1-0·1)	7·15 (4·62-11·07)	-	6·93 (4·43-10·83)
		Meta-analysis of GOLD and Aurum	-	-	7·03 (3·70-10·35)	-	6·85 (3·53-10·18)
Stroke or TIA	Hypertension	GOLD	4·0 (3·1-5·4)	2·0 (1·7-2·4)	2·06 (1·50-2·85)	1·96 (1·42-2·71)	2·12 (1·50-2·99)
		Aurum	4·1 (3·7-4·6)	2·0 (1·9-2·2)	2·17 (1·91-2·46)	2·10 (1·84-2·38)	1·99 (1·72-2·30)
		GOLD and Aurum	4·1 (3·7-4·6)	2·0 (1·9-2·1)	2·15 (1·91-2·42)	2·08 (1·84-2·34)	2·01 (1·75-2·29)
		Meta-analysis of GOLD and Aurum	-	-	2·15 (1·90-2·41)	2·08 (1·83-2·33)	2·01 (1·74-2·28)
	QRISK2	GOLD	5·6 (4·4-7·2)	1·8 (1·5-2·1)	3·08 (2·29-4·14)	-	2·91 (2·14-3·95)
		Aurum	5·3 (4·9-5·9)	1·8 (1·7-2·0)	2·98 (2·65-3·35)	-	2·93 (2·59-3·31)
		GOLD and Aurum	5·4 (4·9-5·9)	1·8 (1·7-1·9)	2·99 (2·68-3·34)	-	2·93 (2·62-3·28)
		Meta-analysis of GOLD and Aurum	-	-	2·99 (2·67-3·32)	-	2·93 (2·59-3·26)
CVD death	Hypertension	GOLD	1·4 (0·9-2·3)	0·7 (0·5-0·9)	2·06 (1·19-3·56)	1·96 (1·13-3·40)	2·31 (1·13-4·71)
		Aurum	1·5 (1·2-1·8)	0·8 (0·7-0·8)	2·12 (1·72-2·63)	2·00 (1·62-2·47)	2·13 (1·65-2·74)
		GOLD and Aurum	1·5 (1·2-1·8)	0·7 (0·7-0·8)	2·11 (1·73-2·58)	1·99 (1·63-2·43)	2·15 (1·69-2·73)
		Meta-analysis of GOLD and Aurum	-	-	2·11 (1·69-2·54)	2·00 (1·60-2·39)	2·15 (1·62-2·67)
	QRISK2	GOLD	2·5 (1·8-3·7)	0·5 (0·4-0·7)	4·58 (2·87-7·31)	-	4·86 (2·91-8·11)
		Aurum	2·7 (2·3-3·1)	0·6 (0·5-0·6)	4·80 (4·00-5·76)	-	4·80 (3·92-5·87)
		GOLD and Aurum	2·6 (2·3-3·0)	0·6 (0·5-0·6)	4·77 (4·03-5·66)	-	4·81 (3·99-5·81)
		Meta-analysis of GOLD and Aurum	-	-	4·77 (3·95-5·58)	-	4·81 (3·89-5·72)

In meta-analysis the between database heterogeneity was assessed using the I^2 statistic. The I^2 for all pooled estimates was 0%, with the exception of; QRISK2 and acute cardiovascular events (crude I^2=44%, fully-adjusted I^2=11%), QRISK2 and ACS (crude I^2=60%, fully-adjusted I^2=23%), and hypertension and heart failure (crude and age- and sex-adjusted I^2=38%, fully-adjusted I^2=58%). *Hypertension models adjusted for: age, sex, ethnicity, socioeconomic status, BMI, alcohol intake and smoking status. QRISK2 models adjusted for: alcohol intake.

Supplementary table 9. Crude and adjusted hazard ratios for the association between cardiovascular risk and acute cardiovascular events after ARI among sensitivity analysis study population

Outcome	Cardiovascular risk	No. of events	Rate per 1,000 person-years (95% CI)	Crude HR (95% CI)	Age and sex-adjusted HR (95% CI)	Fully-adjusted* HR (95% CI)
Any event	Hypertension	1,263	11·6 (11·0-12·3)	2·12 (1·98-2·26)	2·04 (1·91-2·18)	2·04 (1·89-2·19)
	No hypertension	3,560	5·9 (5·7-6·1)	1	1	1
	QRISK2 ≥10%	1,820	17·8 (17·0-18·6)	3·70 (3·49-3·92)	-	3·61 (3·40-3·84)
	QRISK2 <10%	3,003	4·9 (4·7-5·1)	1	-	1
ACS	Hypertension	437	4·0 (3·7-4·4)	2·10 (1·88-2·34)	2·01 (1·80-2·24)	2·08 (1·83-2·36)
	No hypertension	1,240	2·0 (1·9-2·2)	1	1	1
	QRISK2 ≥10%	695	6·8 (6·3-7·3)	4·31 (3·92-4·75)	-	4·29 (3·88-4·74)
	QRISK2 <10%	982	1·6 (1·5-1·7)	1	-	1
Heart failure	Hypertension	392	3·6 (3·3-4·0)	2·14 (1·91-2·40)	2·06 (1·83-2·31)	2·16 (1·89-2·47)
	No hypertension	1,098	1·8 (1·7-1·9)	1	1	1
	QRISK2 ≥10%	592	5·8 (5·3-6·3)	4·02 (3·62-4·46)	-	3·90 (3·49-4·34)
	QRISK2 <10%	898	1·5 (1·4-1·6)	1	-	1
Acute limb ischaemia	Hypertension	30	0·3 (0·2-0·4)	3·13 (2·01-4·89)	3·00 (1·92-4·70)	5·01 (2·99-8·38)
	No hypertension	56	0·1 (0·1-0·1)	1	1	1
	QRISK2 ≥10%	47	0·5 (0·3-0·6)	7·34 (4·80-11·20)	-	7·17 (4·65-11·07)
	QRISK2 <10%	39	0·1 (0·0-0·1)	1	-	1
Stroke or TIA	Hypertension	450	4·1 (3·8-4·5)	2·14 (1·92-2·38)	2·09 (1·87-2·33)	2·04 (1·80-2·30)
	No hypertension	1,250	2·1 (1·9-2·2)	1	1	1
	QRISK2 ≥10%	555	5·4 (5·0-5·9)	2·95 (2·67-3·27)	-	2·90 (2·61-3·22)
	QRISK2 <10%	1,145	1·9 (1·8-2·0)	1	-	1
CVD death	Hypertension	173	1·6 (1·4-1·9)	2·20 (1·85-2·62)	2·12 (1·78-2·52)	2·23 (1·81-2·75)
	No hypertension	472	0·8 (0·7-0·9)	1	1	1
	QRISK2 ≥10%	270	2·6 (2·3-3·0)	4·39 (3·76-5·14)	-	4·37 (3·68-5·18)
	QRISK2 <10%	375	0·6 (0·6-0·7)	1	-	1

Total person-years per 1,000: hypertension = 108·6, no hypertension = 606·8, QRISK2 ≥10% = 102·4 and QRISK2 <10% = 613·0. LRT p-values all <0·0001.
*Hypertension models adjusted for: age, sex, ethnicity, socioeconomic status, BMI, alcohol intake and smoking status. QRISK2 models adjusted for: alcohol intake.

Supplementary table 10. Crude and adjusted hazard ratios for the association between QRISK2 score and acute cardiovascular events after acute respiratory infection, by QRISK2 score identification method

Cardiovascular risk method	No. of events	Rate per 1,000 person-years	Crude HR (95% CI)	Alcohol intake adjusted HR (95% CI)
Recorded QRISK2 ≥10%	110	22·6 (18·8-27·3)	3·31 (2·57-4·26)	3·34 (2·58-4·32)
Recorded QRISK2 <10%	132	6·8 (5·7-8·1)	1	1
Calculated QRISK2 ≥10%	89	25·3 (20·6-31·2)	3·43 (2·65-4·46)	3·39 (2·60-4·42)
Calculated QRISK2 <10%	153	7·4 (6·3-8·6)	1	1

Total person-years per 1,000: calculated QRISK2 ≥10% = 3·51, calculated QRISK2 <10% = 20·75, recorded QRISK2 ≥10% = 4·87 and recorded QRISK2 <10% = 19·40.

Supplementary table 11. MACE after infection sensitivity analysis results

Infection type	Cardiovascular risk	No. of events	Rate per 1,000 person-years (95% CI)	Crude HR (95% CI)	Age and sex-adjusted HR (95% CI)	Fully-adjusted* HR (95% CI)
ARI	Hypertension	850	9·7 (9·0-10·4)	2·11 (1·95-2·28)	2·00 (1·85-2·16)	2·02 (1·85-2·21)
	No hypertension	2,711	4·9 (4·7-5·1)	1	1	1
	QRISK2 ≥10%	1,321	15·2 (14·4-16·0)	3·82 (3·57-4·09)	-	3·71 (3·46-3·99)
	QRISK2 <10%	2,240	4·0 (3·9-4·2)	1	-	1
Influenza/ILI	Hypertension	84	6·8 (5·5-8·5)	2·50 (1·96-3·19)	2·36 (1·85-3·03)	2·37 (1·79-3·15)
	No hypertension	268	2·8 (2·5-3·2)	1	1	1
	QRISK2 ≥10%	96	9·6 (7·9-11·8)	3·64 (2·88-4·60)	-	3·59 (2·82-4·59)
	QRISK2 <10%	256	2·6 (2·3-3·0)	1	-	1
Pneumonia	Hypertension	326	100·9 (90·2-113·3)	1·65 (1·47-1·87)	1·62 (1·43-1·83)	1·65 (1·44-1·90)
	No hypertension	1,182	59·5 (56·1-63·1)	1	1	1
	QRISK2 ≥10%	544	124·1 (113·6-135·7)	2·20 (1·98-2·44)	-	2·15 (1·93-2·40)
	QRISK2 <10%	964	51·5 (48·3-55·0)	1	-	1

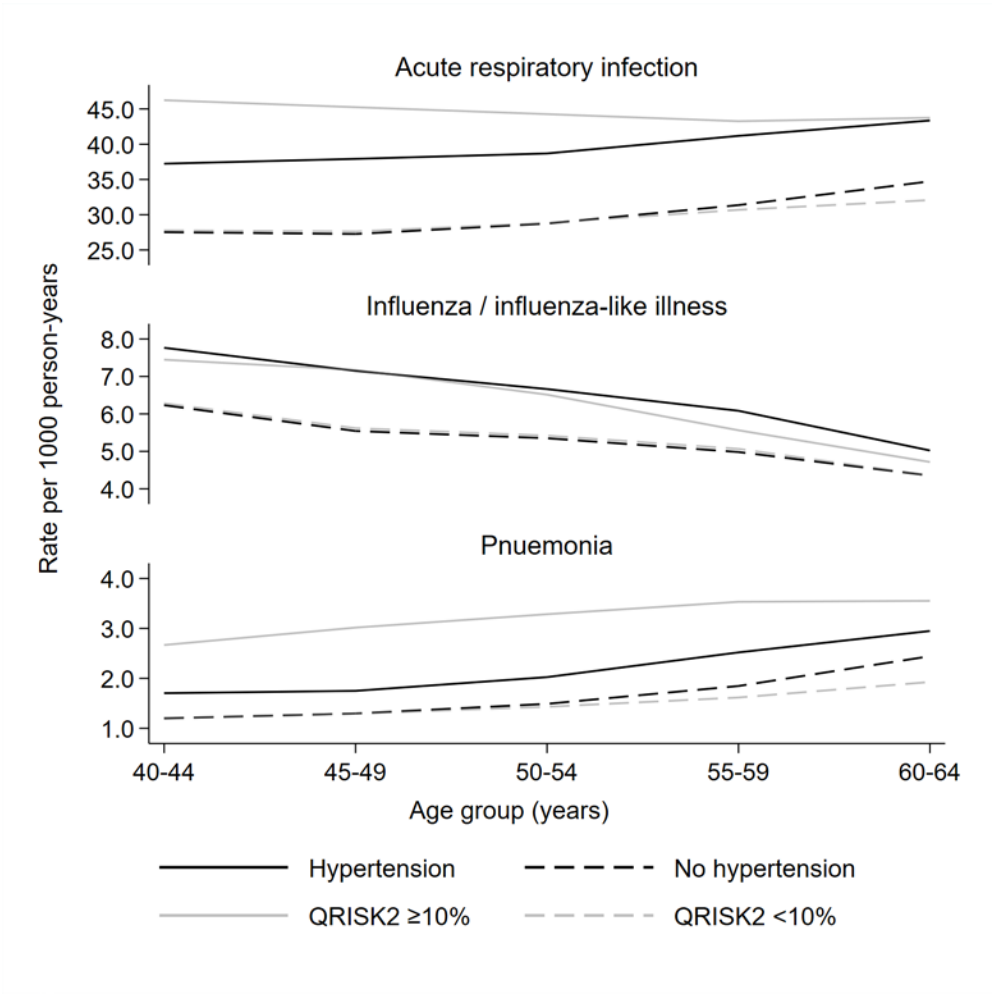
*Hypertension models adjusted for: age, sex, ethnicity, socioeconomic status, BMI, alcohol intake and smoking status. QRISK2 models adjusted for: alcohol intake.

Supplementary table 12. Crude and adjusted incidence rate ratios for the association between cardiovascular risk and acute cardiovascular events after ARI among only patients who did not receive influenza or pneumococcal vaccine during follow-up

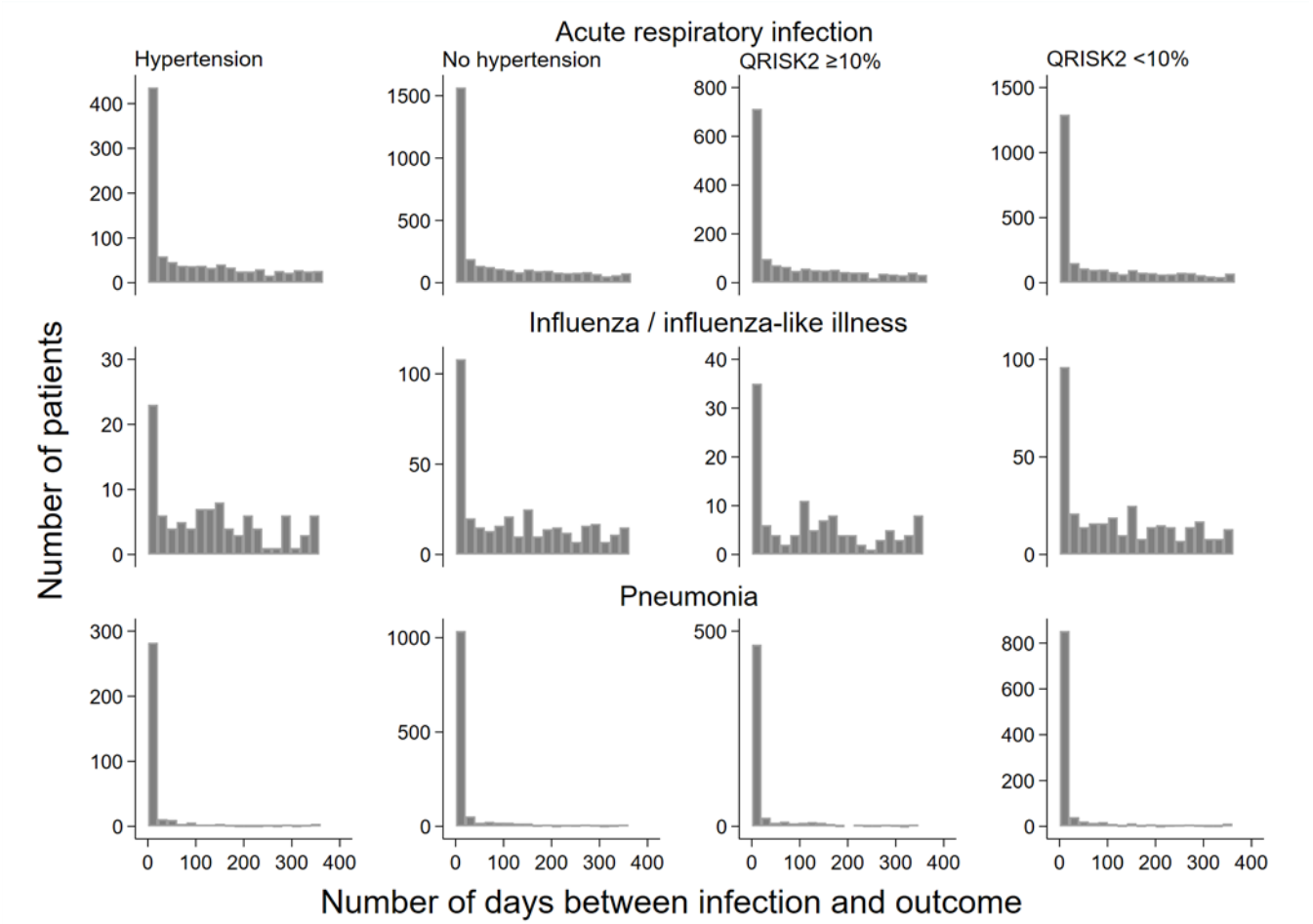
Cardiovascular risk	No. of events	Rate per 1,000 person-years (95% CI)	Crude HR (95% CI)	Age and sex-adjusted HR (95% CI)	Fully-adjusted* HR (95% CI)
Hypertension	985	12·5 (11·8-13·3)	2·17 (2·02-2·33)	2·06 (1·92-2·22)	2·07 (1·90-2·24)
No hypertension	3,184	6·2 (5·9-6·4)	1	1	1
QRISK2 ≥10%	1,526	20·2 (19·2-21·3)	4·06 (3·81-4·32)	-	3·96 (3·70-4·22)
QRISK2 <10%	2,643	5·1 (4·9-5·3)	1	-	1

Patients from 35,505 ARI episodes received influenza or pneumococcal vaccine. Among patients with raised cardiovascular risk, a higher proportion were vaccinated (hypertension=10%, 6,898/68,731 and QRISK2 ≥10%=13%, 9,330/72,137) compared with low cardiovascular risk (no hypertension=6%, 28,607/458,069 and QRISK2 <10%=6%, 26,175/454,663). None of the vaccinated patients had an acute cardiovascular event during follow-up. *Hypertension models adjusted for: age, sex, ethnicity, socioeconomic status, BMI, alcohol intake and smoking status. QRISK2 models adjusted for: alcohol intake.

Supplementary figure 1. Age-specific infection rates by cardiovascular risk group



Supplementary figure 2. Timing between acute respiratory infection and major adverse cardiovascular event



Supplementary material references

1. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med*. 2009;150(9):604-612. doi:10.7326/0003-4819-150-9-200905050-00006

2. 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(7659):1475-1482. doi:10.1136/bmj.39609.449676.25

3. National Health Service. Quality and Outcomes Framework (QOF). https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-collections/quality-and-outcomes-framework-qof. Published 2020. Accessed March 3, 2020.

4. Damen JAAG, Hooft L, Schuit E, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. *BMJ*. 2016;353:i2416. doi:10.1136/bmj.i2416

5. Talbot HK. Influenza in Older Adults. *Infect Dis Clin North Am*. 2017;31(4):757-766. doi:10.1016/j.idc.2017.07.005

6. Klein SL, Hodgson A, Robinson DP. Mechanisms of sex disparities in influenza pathogenesis. *J Leukoc Biol*. 2012;92(1):67-73. doi:10.1189/jlb.0811427

7. Davidson J, Banerjee A, Mathur R, et al. Ethnic differences in the incidence of clinically diagnosed influenza: an England population-based cohort study 2008-2018. *Wellcome Open Res*. 2021;6:49. doi:10.12688/wellcomeopenres.16620.3

8. National Institute for Health and Care Excellence. *CVD Risk Assessment and Management*.; 2019. https://cks.nice.org.uk/cvd-risk-assessment-and-management. Accessed August 14, 2019.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
United Kingdom
Switchboard: +44 (0)20 7636 8636

www.lshtm.ac.uk

Observational / Interventions Research Ethics Committee

Ms Jennifer Davidson
LSHTM

8 November 2019

Dear Ms Jennifer

Study Title: Effect of raised cardiovascular risk on the rates of acute respiratory infections and subsequent cardiovascular complications: a cohort study using electronic health records

LSHTM Ethics Ref: 17894

Thank you for your application for the above research project which has now been considered by the Observational Committee via Chair's Action.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation, subject to the conditions specified below.

Conditions of the favourable opinion

Approval is dependent on local ethical approval having been received, where relevant.

Approved documents

The final list of documents reviewed and approved is as follows:

Document Type	File Name	Date	Version
Protocol / Proposal	85_Respiratory_Infections_CVD_19_209_ISAC_protocol	14/10/2019	1
Protocol / Proposal	85_Respiratory_Infections_CVD_19_209_ISAC_feedback	14/10/2019	1
Investigator CV	CV OCT 2019 Jennifer Davidson	14/10/2019	1
Investigator CV	Short cv_CWG_16-07-2019	30/10/2019	1
Investigator CV	Liam Smeeth 2 page CV 2019	30/10/2019	1
Investigator CV	2 page CV_Oct_2019	05/11/2019	1

After ethical review

The Chief Investigator (CI) or delegate is responsible for informing the ethics committee of any subsequent changes to the application. These must be submitted to the committee for review using an Amendment form. Amendments must not be initiated before receipt of written favourable opinion from the committee.

The CI or delegate is also required to notify the ethics committee of any protocol violations and/or Suspected Unexpected Serious Adverse Reactions (SUSARs) which occur during the project by submitting a Serious Adverse Event form.

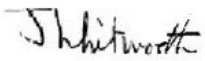
An annual report should be submitted to the committee using an Annual Report form on the anniversary of the approval of the study during the lifetime of the study.

At the end of the study, the CI or delegate must notify the committee using the End of Study form.

All aforementioned forms are available on the ethics online applications website and can only be submitted to the committee via the website at: http://leo.lshtm.ac.uk.

Further information is available at: www.lshtm.ac.uk/ethics.

Yours sincerely,



Professor Jimmy Whitworth
Chair

ethics@lshtm.ac.uk
http://www.lshtm.ac.uk/ethics/

Study protocol

Applicants must complete all sections listed below Sections which do not apply should be completed as '*Not Applicable'* and justification provided
A. Study Title (Max. 255 characters) Effect of raised cardiovascular risk on the rates of acute respiratory infections and subsequent cardiovascular complications: a cohort study using electronic health records
B. Lay Summary (Max. 250 words) Heart disease, stroke and lower respiratory tract infections are among the global leading causes of ill-health and death. In 2017, cardiovascular disease (CVD) accounted for more than 160,000 deaths in the UK. Previous research has shown that people have a short-term risk of heart attack and stroke in the days after a serious respiratory infection, such as flu or pneumonia. This risk has mostly been found in older adults and those with pre-existing CVD. This study aims to establish whether this risk also occurs in people without pre-existing CVD but who have raised cardiovascular risk, for example high blood pressure, which is suggestive of future CVD. Establishing such risk is important for UK vaccine policy; influenza and pneumococcal vaccination recommendations currently include people aged ≥65 years and those with CVD, but not people aged <65 years with raised cardiovascular risk. We will use routinely collected healthcare data in England to compare the occurrence of serious respiratory infections as well as subsequent cardiovascular events, such as heart attack or stroke, in people with a raised cardiovascular risk compared to people without raised risk. The results from this study will inform future research to evaluate the impact of influenza and pneumococcal vaccination in populations with raised cardiovascular risk.
C. Technical Summary (Max. 300 words) In the UK there are an estimated 7 million people living with CVD, for which the annual costs are estimated to be £19 billion. As the population ages and multimorbidity prevalence increases, stratified interventions are ever more important. The risk of cardiovascular complications after an acute systemic respiratory infection in people with raised cardiovascular risk but without established CVD is unknown. Quantifying any such increased risk will inform whether these groups should be considered for influenza and pneumococcal vaccination. Our cohort study will use CPRD data linked to HES and ONS mortality data to increase ascertainment of respiratory and cardiovascular events. We will define cardiovascular risk by hypertension diagnosis and QRISK2 score. QRISK2 is a prediction algorithm for future CVD which utilises a range of risk factors, beyond hypertension, to determine risk. We will first calculate age-specific incidence rates for diagnosis of acute systemic respiratory infections by cardiovascular risk among adults aged 40 to 64 years. We will compare 1) people with hypertension to those without hypertension and 2) people with a QRISK2 score ≥10% in ten years compared to those with a QRISK2 score <10%. We will then use Poisson regression models with Lexis expansion by age group and cardiovascular risk level to estimate incidence rates and rate ratios for 1) acute systemic respiratory infections including influenza-like illness and pneumonia, and 2) major acute cardiovascular events (MACE). Using Cox proportional hazards regression multivariable models, which adjust for potential confounders, we will then estimate the effect of cardiovascular risk on MACE after an acute systemic respiratory infection. In our definition of MACE we will include; myocardial infarction, unstable angina, left ventricular heart failure, stroke, transient ischaemic attack, acute limb ischaemia and cardiovascular death.
D. Outcomes to be Measured ***Primary*** Aim 1 & 2: all-cause acute systemic respiratory tract infections. This includes clinical or confirmed diagnoses such as pneumonia, acute bronchitis, influenza / influenza-like illness (ILI), and other acute infections suggestive of lower respiratory tract improvement.

Aim 3 & 4: all-cause major adverse cardiovascular events (MACE). This includes; cardiovascular death, acute coronary syndrome (ACS) which captures both myocardial infarction (MI) and unstable angina, stroke, transient ischaemic attack (TIA), left ventricular heart failure and acute ischaemic limb.

Secondary

Aim 1 & 2:

- Influenza / ILI
- pneumonia

Aim 3 & 4 cause-specific acute cardiovascular events:

- ACS,
- stroke / TIA,
- left ventricular heart failure,
- acute ischaemic limb, and
- cardiovascular death.

E. Objectives, Specific Aims and Rationale

Research objective

We will use CPRD GOLD and Aurum with linked HES and ONS mortality data to investigate whether the occurrence of cardiovascular complications after acute systemic respiratory infections varies by cardiovascular risk level. Cardiovascular risk will be defined in two separate ways; presence or absence of hypertension diagnosis (exposure group A) and 10-year predicted cardiovascular risk based on calculated QRISK2 score of ≥10% compared to <10% (exposure group B).

Specific aims

- To describe annual age-specific rates of medical attendance for acute respiratory infection (ARI) by cardiovascular risk level in exposure group A and B.
- Compare the effect of cardiovascular risk level (in exposure group A and B) on ARI, particularly ILI and pneumonia, rates using multivariable Poisson regression with Lexis expansion models.
- Compare the effect of cardiovascular risk level (in exposure group A and B) on MACE rates using multivariable Poisson regression with Lexis expansion models.
- By cardiovascular risk level (in exposure group A and B) compare the effect of ARI on cardiovascular complications using multivariable Cox regression models.

Rationale

ARIs are known to trigger cardiovascular events among the elderly and people with existing CVD. Influenza and pneumococcal vaccination are already recommended for these groups. Establishing the risk of cardiovascular complications after an ARI in people with hypertension and with a QRISK2 ≥10% in 10 years (who are not currently recommended for vaccination) could lead to the prevention of premature cardiovascular events and inform whether UK influenza and pneumococcal vaccine policy should be extended to include these population groups.

Hypertension is one of the primary risk factors for future CVD, however it is only one element of risk. QRISK2 provides an individualised CVD risk score based on multiple morbidities and risk factors including blood pressure. Some patients will be defined as high cardiovascular risk based on both blood pressure and QRISK2 score, while others will only be captured by one classification. Ideally, all patients with high cardiovascular risk could be identified based on QRISK2 scores recorded in primary care records. Unfortunately, this score is currently only routinely recorded for a small subset of patients. The morbidities and risk factors on which QRISK2 scores are based are generally well recorded in primary care records allowing scores to be calculated in research datasets.

F. Study Background

Globally, ischaemic heart disease (IHD) and stroke, have been the leading causes of death for more than 15 years, accounting for 15.2 million deaths in 2016 (1). More than 13 million people globally were estimated to have suffered a stroke in 2016 (2), resulting in healthcare expenditure associated with stroke between an estimated 3% and 5% (3–5). While CVD mortality rates in the UK have fallen steadily in the last four decades, in 2017 CVD still accounted for 168,472 deaths, at a rate of 246 per 100,000 people (6). Additionally, the prevalence of CVD has remained stable for the past decade; IHD prevalence was 3.5% in

2007/8 and 3.1% in 2017/8, while stroke and TIA prevalence was 1.6% in 2007/8 and 1.8% in 2017/8, and heart failure at 0.7-0.8% (6). Overall, 7 million people in the UK are estimated to be living with CVD with an estimated annual cost of £19 billion to the UK economy (7).

Lower respiratory tract infections are also among the global leading causes of death (1). Ecological studies have been used throughout the 20th, and into the 21st, century to demonstrate an excess in CVD incidence and mortality during influenza seasons (8–13). Understanding and addressing interactions between diseases, such as ARIs and CVD, is becoming increasingly important to deal with growing multimorbidity, which can be best tackled through stratified and targeted interventions.

Pooled estimates from two systematic reviews have showed that the odds of myocardial infarction (MI) are two times higher in those with ILI (14,15). Compared to ILI, the association between pneumonia and acute cardiovascular complications has been less investigated. However, a systematic review conducted in 2010 found that after community acquired pneumonia, inpatients had a pooled incidence of 14.1% (95% CI 9.3–20.6) for acute heart failure and 5.3% (95% CI 3.2–8.6) for ACS (16). Recent studies have also shown an increased rates of MI (17,18) as well as stroke (18) following confirmed influenza or *Streptococcus pneumoniae* infection. Overall less is known about the specific risk of heart failure following ARI (15), although findings from an ecological analysis using the USA Atherosclerosis Risk in Communities Study cohort showed a temporally association between influenza activity and heart failure hospitalisations (19).

Population studies in the UK, including some which utilised CPRD data, and Canada using a self-controlled case series (SCCS) design have demonstrated that the association between ARI and cardiovascular complications is transient; highest in the first few days after infection (17,18,20,21) but could last for up to one month depending on the infective agent and cardiovascular event (18).

These existing studies were largely conducted in older populations or those with pre-existing CVD. There is a lack of evidence to determine where cardiovascular complications following ARI exist in those aged less than 65 years and at increased cardiovascular risk but without pre-existing diagnosed CVD. One SCCS analysis of laboratory-confirmed respiratory infections conducted analysis stratified by age group, identifying higher rates of first MI (incidence ratio (IR) 16.1, 95% CI 5.12-50.9) and stroke (IR 23.4, 95% CI 5.71-96.3) in the first three days after a respiratory virus infection in those under 65 years of age (18). The authors hypothesise this is due to lower vaccination rates in the younger population. The number of people included in this younger age group was small producing imprecise estimates. Other studies have not found any increased risk of cardiovascular complications following infection in younger age groups (17,21). These studies were all underpowered to specifically estimate relative incidence and risk in younger age groups.

In the UK, influenza and pneumococcal vaccination is recommended for all adults aged over 65 years, and people younger than 65 years if they are in a clinical risk group more likely to experience medical complications following an infection, such as those with chronic heart disease (22,23). Influenza and pneumococcal vaccinations are not currently recommended for primary prevention of acute cardiovascular events, such as MI or stroke, in people with raised cardiovascular risk but without established CVD. People with hypertension are considered to be at raised risk of CVD, while the QRISK2 prediction algorithm estimates an individual's 10-year risk of developing CVD based on a wide range of risk factors (24) and is being increasingly used to guide CVD risk assessment and prevention strategies (25).

This study aims to test the hypothesis that people with raised cardiovascular risk (hypertensive or with calculated QRISK2 ≥10% in 10 years) are at increased risk of cardiovascular events after ARIs compared to those without raised risk. We will produce estimates for this absolute risk to inform future studies which will investigate the impact of influenza and pneumococcal vaccination in these cardiovascular risk groups.

G. Study Type

Descriptive and hypothesis testing

H. Study Design

Cohort study

I. Feasibility counts

In September 2013, the midpoint of our study period, in CPRD GOLD with linked HES data there were 687,479 patients aged 40-64 years with at least 12 months of research standard follow-up who did not have existing CVD or a chronic condition (respiratory disease, diabetes, liver disease, kidney disease and asplenia or splenic dysfunction) already eligible for influenza or pneumococcal vaccination. Based on previous work conducted by the HPRU in Immunisation (ISAC number 18_218) we estimate a further 1% of patients will be immunosuppressed (26) (and therefore already eligible for vaccination), and assuming a maximum of 5% of patients will additionally receive vaccination (i.e. occupational group or paid for vaccination). This would result in 646,574 patients for inclusion from CPRD GOLD linked data. Using Aurum as well as GOLD we should have twice as large a patient population. We, therefore, estimate that in 2013 using both data source we will have 1,293,148 patients.

Among this population, based on CPRD GOLD with linked HES data:

- Hypertension: we would have 81,382 patients with hypertension at baseline, with a ratio of 1:7 patients with hypertension to those without hypertension.
- QRISK2: we would have 161,811 patients with a cardiovascular risk score recorded at baseline. QRISK2 is only recorded for approximately 10% of eligible patients (27). We also expect a ratio of 2:1 patients for scores <10%:≥10% (28).
- ARIs: we will have 7.7% (n=49,838) of patients with a diagnosis of ARI between September 2013 and August 2014.
- MACE: we will have 0.6% (n=4,035) of patients with a MACE event between September 2013 and August 2014.

J. Sample size considerations

Using the results of our feasibility counts to carry our sample size calculations, we estimate the minimum effect estimates we will be able to detect are:
Exposure group A (hypertension): we can detect hazard ratios of 0.97 for ARIs and 0.91 for MACE (with 80% power and alpha 5%).
Exposure group B (QRISK2): we can detect hazard ratios of 0.97 for ARI and 0.88 for MACE (with 80% power and alpha 5%).

K. Planned use of linked data (if applicable):

ARIs and MACE may be recorded in either primary or secondary care records, with major acute events resulting in hospitalisation. Therefore, to increase ascertainment of these diagnoses we will use CPRD linked to HES in-patient data for our entire study population. To include the outcome of cardiovascular death, ONS mortality data is required to identify these events.

Townsend scores will be used in the calculation of QRISK2 scores. We will additionally use this to consider socioeconomic status as a potential confounder.

The above-mentioned linked data will also be used to identify potential confounders as well, further detail is described in Section N on covariates.

L. Definition of the Study population

The study population is adults aged 40 to 64 in CPRD GOLD or Aurum datasets with linked HES data between 01 September 2008 and 31 August 2018. This time period covers the duration of QRISK2 use. We will start follow up from September as this corresponds to when patients would be assessed for seasonal influenza vaccine eligibility. Our analysis will be divided into influenza/non-influenza season (see Section N for further detail). Only patients with at least 12 months of research standard follow up will be included.

Baseline and follow up are defined for each of the study aims in Section O.

<u>Exclusions</u>

Influenza and pneumococcal vaccines are already offered to a range of risk groups. We aim to identify ARI risk and MACE after ARI risk in patients who are not already eligible for influenza or pneumococcal vaccination. We will therefore exclude at baselines patients who have:

- any previous record for diagnosis of CVD included in either influenza or pneumococcal vaccination policy (chronic heart disease or stroke/TIA).

- any previous record for diagnosis of a chronic condition included in influenza vaccination policy. These include a previous record for diagnosis of; chronic respiratory disease, chronic kidney disease, chronic liver disease, diabetes, asplenia / splenic dysfunction, chronic neurological conditions, or morbid obesity (in years where these were included in influenza vaccination clinical risk groups).
- a record of receiving immunosuppressive treatment (chemotherapy, radiotherapy or immunosuppressive drug prescribed in primary care i.e. steroids) in the year prior to baseline.
- a record of ever receiving pneumococcal vaccine. Previous vaccination could reflect membership of another risk group (beyond those previously outlined) such as an occupational group or uptake of private vaccination, therefore these individuals would be unlikely to benefit further from an extension of the current vaccination recommendations.
- a record of receiving influenza vaccination in the year prior to baseline.

We will define our exclusion groups using codelists created by the HPRU in Immunisation (ISAC protocol 18_218RA).

M. Selection of comparison group(s) or controls

Exposure group A: patients with hypertension will be compared to patients without hypertension.
Exposure group B: patients with a recorded/calculate QRISK2 ≥10% in ten years will be compared to patients with QRISK2 <10%.

N. Exposures, Outcomes and Covariates

Exposures

All analysis will be conducted by cardiovascular risk group defined as:

- Exposure group A (hypertension): patients with a Read or SNOMED code in the GOLD clinical or referral files / Aurum observation (problem or referral) files for a hypertension diagnosis (as listed in Appendix 1 for CPRD GOLD, to be mapped to SNOMED Aurum codes) but without any CVD diagnosis in the same GOLD/Aurum files or ICD-10 code in HES (as defined by HPRU codelists). We will only base our definition of hypertension on recorded diagnosis and not include blood pressure readings or treatments to classify patients into a risk group. This represents the practical method by which patients would be identified in primary care systems to offer vaccination. As a comparison group we will include patients with no hypertension or existing CVD defined as absence of any relevant Read or SNOMED code.
- Exposure group B (raised vascular risk): QRISK2 scores will be calculated based on the previous approach taken by Bhaskaran, Gadd et al (ISAC 17_008). QRISK2 scores will be calculated using bulk processing software provided by ClinRisk, using the most recent measures available. Variables in QRISK2 include age, sex, ethnicity, blood pressure, deprivation score from linked data, diabetes, family history of coronary heart disease, atrial fibrillation, chronic kidney disease stage 4 or 5, cholesterol/HDL ratio, rheumatoid arthritis, use of blood pressure lowering drugs, BMI (using weight and height records), smoking status (using smoking-related Read codes and structured data on smoking). Of note, those with diabetes and chronic kidney disease will be excluded from our study (see Section L Exclusions). Time variant variables such as BMI and smoking status at the most recent date to index date will be utilised and if missing the median value will be assigned. We will also use recorded QRISK2 scores based on Read / SNOMED codes which relate to QRISK in GOLD using the clinical file linked to corresponding results in the additional file and in Aurum using appropriate observation file data. Based on the recorded/calculated score we will group patients into those with a score of ≥10% in ten years who are the group of interest and compared these patients to those with a score of <10% in ten years.

For aim 4, the baseline date will be the date of ARI (see section O for further detail). Patients with a record of a Read or SNOMED code or ICD-10 code in HES for a diagnosis of an ARI as listed in Appendix 2 will be included in follow up.

Outcomes

Aim 1 & 2: the primary outcome will be all-cause acute systemic respiratory infections, with separate secondary outcomes of ILI and pneumonia. Infections will be defined by Read and SNOMED codes or ICD-

10 code in HES as listed in Appendix 2. The codes which will be used to define ILI and pneumonia are also indicated in Appendix 2.

Aim 3 & 4: cardiovascular complications can manifest in many forms. We will use the composite outcome of MACE to account for the wide range of outcomes possible and to achieve statistical power. There is no standard definition of MACE, although most definition include cardiovascular death, MI and stroke. A wider definition often used includes left ventricular heart failure and unstable angina. We will take a broad definition and define MACE as; ACS which will include MI and unstable angina, left ventricular heart failure, stroke, TIA, acute ischaemic limb and cardiovascular death. Outcomes will be defined by Read and SNOMED codes or ICD-10 codes in HES (and ONS for deaths), respectively which are listed in Appendix 3. In sensitivity analysis we will use a narrower definition of MACE with the most severe outcomes; MI, left ventricular heart failure, stroke and cardiovascular death. Our secondary outcomes are each of the cardiovascular conditions separately.

Covariates

We will control for a range of potential confounding variables and investigate potential effect modifiers. We expect, based on existing literature and a priori hypotheses, to include the variables listed below.

- Age: categorised into 5-year bands of 40-44, 45-49, 50-54, 55-59, 60-64 from CPRD Gold/Aurum patient file
- Sex: defined as male or female from CPRD Gold/Aurum patient file
- Ethnicity: from CPRD GOLD/Aurum codes, supplemented with HES data where missing from CPRD and available in HES
- Socioeconomic status: from linked Townsend scores grouped into quintiles
- Consultation frequency: based on in-person and telephone consultations in the year prior to baseline
- Lifestyle factors: alcohol consumption and smoking status from CPRD GOLD additional file as well as codes in the clinical or referral files CPRD / Aurum observation files
- Body Mass Index (BMI): recorded (or if not recorded directly then to be calculated from height and weight recorded) in CPRD GOLD additional file / CPRD Aurum observation files.
- Comorbidities/existing health conditions: those not covered by exclusions such as dementia using Read and SNOMED codes recorded in the CPRD GOLD clinical file / CPRD Aurum observation files.
- Primary prevention of CVD: statins and antihypertensives from CPRD GOLD therapy file / CPRD Aurum observation files.
- Season: when outcome occurs as well as when exposure occurs for aim 3, defined as flu season (1 September – 31 March) or non-flu season (1 April – 31 August).

Effect modifier

- Primary prevention of ARI: influenza and pneumococcal vaccination given in the follow-up year from CPRD GOLD immunisation file / CPRD Aurum observation files.
- Antiviral treatment: if a suitable number of patients have been prescribed antivirals, in stratified analysis we will explore the effect of antiviral use given in the follow-up year both where it is suggested to be for prophylactic use (no ARI diagnosis) and treatment (ARI diagnosis) using data from CPRD GOLD therapy file / CPRD Aurum observation files.
- Antiplatelets: again, if a suitable number of patients have been prescribed an antiplatelet, we will explore their impact in stratified analysis using data from CPRD GOLD therapy file / CPRD Aurum observation files.

O. Data/ Statistical Analysis

To achieve the necessary statistical power we will use both CPRD GOLD and Aurum (query number CPRD00040222). This will require deduplication of data where practices have migrated from Vision to EMIS software system, we will use the file of migrating practices provided by CPRD for this deduplication. Where possible we will arrange GOLD and Aurum datasets into a single combined dataset to analyse individual level data. Where this cannot be achieved, we will analyse data from GOLD and Aurum separately. When analysed separately, to calculate single summary estimates, we will either combine the number of events and person-time from each database (aim 1) or use fixed effects meta-analysis (aim 2-4).

Before calculating single summary estimates between-database heterogeneity (based on the I^2 statistic) will be assessed.

For thorough ascertainment of ARIs and MACE all analyses will be conducted using linked HES data. We will clean the data in order to only count events once. Multiple records in CPRD or HES primary diagnostic position for ARI within 28 days of each other will be counted as the same episode. The index date will be the date of the first ARI consultation and the end of the episode will be 28 days after the last ARI consultation. Confounders where data are missing in CPRD will be identified, where possible, using HES.

Aim 1 (age-specific rates of ARI) and aim 2 (association between cardiovascular risk and ARI rates)
An open cohort of patients without any existing CVD, a chronic clinical condition eligible for vaccination or previous pneumococcal vaccination, and without influenza vaccination or immunosuppression diagnosis/treatment in the year prior to baseline (see exclusions in Section L) will be created. Follow-up will begin at the latest of: 01 September 2008, 40th birthday or 12 months after research standard follow up. Eligibility will be updated during follow-up when new measures become available, with follow up ending at the earliest date of: pneumococcal or influenza vaccination (excluded from further follow up), development of CVD (excluded from further follow up), development of a clinical condition or immunosuppressive state eligible for vaccination (excluded from further follow up), death, transfer out, the practice's last data collection, 65th birthday, or 31 August 2018.

Cardiovascular risk level (hypertension status and then again repeated for QRISK2) will be classified at cohort entry. Hypertension status will be updated when a new diagnosis is made. QRISK2 scores will be updated each time a new measure becomes available or every year (to reflect change in age), whichever is sooner.

For patients meeting inclusion criteria we will describe the baseline characteristics stratified by cardiovascular risk; hypertension compared to no hypertension (exposure group A) and QRISK2 ≥10% compared to QRISK2 <10% (exposure group B). In this we will include the extent and pattern of missing data. Additionally, we will conduct a descriptive analysis of patients who progress from low cardiovascular risk at baseline to high risk in follow up.

Using Poisson regression models, we will estimate the crude annual incidence rates and rate ratios for any ARI (primary outcome) and for ILI and pneumonia (secondary outcomes) which occurred during follow up by each baseline cardiovascular risk groups. Lexis expansion by age group and cardiovascular risk level will be used to estimate stratified incidence rates and rate ratios. Multiple ARI episodes in the same patient will be accounted for by fitting random effects models. Models will be separately produced for exposure group A (hypertension) and exposure group B (QRISK2). We will generate multivariable models which adjust first for age and sex, and then for additional confounders (see section N).

Aim 3 (association between cardiovascular risk and MACE)
Our open cohort of patients meeting inclusion criteria described for aim 1 & 2 will again be followed up from the same baseline. Follow up will end at the earliest date of: MACE (outcome of interest), pneumococcal or influenza vaccination (excluded from further follow up), development of a clinical condition or immunosuppressive state eligible for vaccination (excluded from further follow up), death, transfer out, the practice's last data collection or 31 August 2018.

We will again use Poisson regression with Lexis expansion by age group and cardiovascular risk level, this time to estimate annual stratified incidence rates and rate ratios for the primary and secondary MACE outcomes. We will stratify results by season to identify peaks in incidence and MACE risk. Again, separate models will be produced for exposure group A (hypertension) and exposure group B (QRISK2), with multivariable models first adjusting first for age and sex, and then for additional confounders. If a suitable number of records exist we will also build models which stratify by antiviral and antiplatelet treatments, otherwise sensitivity analysis will be conducted (see below).

Pre-specified sensitivity analyses for aims 1-3

1. Repeat main analysis for years 2015-2017 restricting to patients with recorded QRISK2 score (as opposed to our calculated QRISK2 scores), this will allow consideration of the clinical practice that would have occurred for identifying patients eligible for vaccination had a policy based on QRISK2 score been in place and to compare results obtained with recorded score to those from our calculated scores. We have limited the time period as our algorithm is based on the 2015 version of QRISK (minimal changes to the QRISK2 algorithm occurred between 2015 and 2017).
2. Repeat main analysis only excluding patients who are included in both the current influenza and pneumococcal vaccination eligibility recommendations.
3. Repeat main analysis but replacing the binary season variables for one with four categories; autumn (September-November), winter (December-February), spring (March-May) and summer (June-August).

Pre-specified sensitivity analyses for aim 3

1. Repeat main analysis restricting outcome definition of MACE to only include subset of cardiovascular events (as described in Section N, Outcomes).
2. Repeat main analysis excluding any patients who received antiviral treatment during follow-up. This will only be done if too small a number of patients are vaccinated to allow for meaningful stratified analysis.
3. Repeat sensitivity analysis 2 for antiplatelet treatment.

Aim 4 (association between cardiovascular risk and MACE after ARI)

Start of follow-up will be defined as date of ARI diagnosis. End of follow-up will be the earliest of: MACE diagnosis, death, transfer out, the practice's last data collection, one year after ARI diagnosis date or 31st August 2018. Patients with existing CVD, a chronic clinical condition eligible for vaccination or previous pneumococcal vaccination at the date of ARI will be excluded. Patients with seasonal influenza vaccination or immunosuppression diagnosed within one year of ARI diagnosis date will also be exclude. See exclusions in Section L for further detail.

Cardiovascular risk level (hypertension status and then again repeated for QRISK2) will be classified at cohort entry.

We will use Cox proportional hazards regression to generate hazard ratios for primary and secondary MACE outcomes comparing cardiovascular risk groups (separately for exposure group A and B). Calendar time will be used as the underlying time scale. To account for the same patient having multiple ARIs more than one year apart we will conduct analysis using a random effects model. Models will first be adjusted for age and sex, followed by further confounders. The number of ARI episodes which occurred within the year follow up will included in the models. If a suitable number of records exist we will stratify analysis by vaccination status, antiviral treatment and antiplatelet use, otherwise sensitivity analysis will be conducted (see below).

Pre-specified sensitivity analyses for aim 4

1. Repeat main analysis for years 2015-2017 restricting to patients with recorded QRISK2 score.
2. Repeat main analysis only excluding patients who are included in both the current influenza and pneumococcal vaccination eligibility recommendations.
3. Repeat main analysis but replacing the binary season variables for one with four categories; autumn (September-November), winter (December-February), spring (March-May) and summer (June-August).
4. Repeat main analysis restricting outcome definition of MACE to only include subset of cardiovascular events (as described in Section N, Outcomes).
5. Repeat main analysis excluding any patients who received influenza and/or pneumococcal vaccination during follow-up. This will only be done if too small a number of patients are vaccinated to allow for meaningful stratified analysis.
6. Repeat sensitivity analysis 5 for antiviral treatment.
7. Repeat sensitivity analysis 5 for antiplatelet use.

P. Plan for addressing confounding

We will use multivariable regression to adjust for the potential covariates we have hypothesised (listed in section N).

Q. Plans for addressing missing data

We will describe all missing data and consider findings as a limitation in any outputs resulting from our study.

Missing data on acute systemic respiratory infection: People who experience mild and short-lived respiratory illness are unlikely to seek healthcare, this would lead to us underestimating not only the incidence of ARIs but also their association with cardiovascular outcomes. We are interested in severe infections, for which it is more biologically plausible to have systemic complications, which are more likely to result in healthcare attendance. However, we will first conduct our analysis (aim 3) without the inclusion of ARIs to analyse the effect of cardiovascular risk on MACE stratifying by season to determine whether the pattern of MACE follows the seasonality of some ARIs i.e. ILI. Patients who regularly attend their GP may be more likely to present with an ARI, we will therefore include consultation frequency in our models where this is identified as a confounder.

Missing data on cardiovascular outcomes: We will use HES data to supplement primary care recording of all cardiovascular events with our analysis limited to linked data.

Missing data on the classification of cardiovascular risk: While blood pressure and therefore hypertension are generally well recorded in primary care data this requires patients to have presented to the GP. For patients who do not often attend the GP the diagnosis will not be made and therefore those with hypertension recorded may be healthier with better managed hypertension. The algorithm to calculate QRISK2 score is based on many variables. Missing data for the chronic conditions included in the algorithm indicates that the patient does not have the risk factor in question although, similar to hypertension, there may be undiagnosed or under-recording. Missing data for variables such as ethnicity, BMI, alcohol intake and smoking status are due to non-recording. HES will be used to increase the ascertainment of ethnicity. Our QRISK2 algorithm will replace missing values for alcohol intake, smoking status and BMI with the median value, in line with the QRISK2 calculator. We will compare our calculated QRISK2 scores to those recorded by GPs to validate our classifications and conduct a sensitivity analysis which only includes patients with a recorded QRISK2 score. Our algorithm to calculate QRISK2 scores is based on the 2015 version of QRISK2, we will therefore only include patients with QRISK2 score which were recorded in 2015-2017 (during which time changes to the QRISK2 algorithm were minimal).

Missing data on covariates (such as smoking status, alcohol intake, BMI and ethnicity): as data are unlikely to be missing at random i.e. the assumption required for multiple imputation to be valid, we will conduct complete case analysis for variables not set by the presence or absence of a code.

R. Patient or user group involvement (if applicable)

Patients and user groups have not been involved in the development of this research.

S. Plans for disseminating and communicating study results, including the presence or absence of any restrictions on the extent and timing of publication

The study findings will be submitted for publication in a peer-reviewed scientific journal. Additionally, results will be presented at conferences and other meetings as appropriate. We will work with the LSHTM press office to publicise our study and its findings. The electronic health records research group at LSHTM also have a webpage with research news and social media accounts which will be used. We will work with the British Heart Foundation, who are the funders of this research, to publicise our findings to a wider audience including the general public and patient groups.

Conflict of interest statement: The work is supported by a British Heart Foundation PhD Studentship (FS/18/71/33938). There are no conflicts of interest to declare

T. Limitations of the study design, data sources, and analytic methods

As discussed in Section Q on missing data, there is a risk of misclassifying whether patients have had an ARI when using routinely healthcare records data as a large number of people will not attend healthcare for this, particularly those with mild or short-lived illness. Those who do attend are likely to be patients who

have higher rates of consultation overall and those with underlying illness. We have outlined our plan to address this in Section Q.

There is also a risk of misclassification of MACE, however previous validation of CPRD and HES for MI has identified these sources to be of high quality, with positive predictive values of >90% in each (29).

Missing data, particularly for some covariates to be included in models, presents a limitation to our analysis and therefore interpretation of our results. As described in Section Q, we will describe missing data and conduct complete case analysis.

As our analysis aims to estimate the effect of ARI and subsequent MACE which could be avoided by vaccination, we are excluding many groups who are already eligible for vaccination under existing policy. This relies on the ability to identify eligible populations, we will exclude major risk groups with chronic lung disease, chronic kidney disease, chronic liver disease, diabetes and neurological conditions. There may be smaller risk groups which we have missed. We will stratify for vaccination which occurs during follow-up. However, if vaccination was conducted outside of primary care, for example if given through occupational health or paid for at a pharmacy, then we may not be able to reliably identify these unless GPs record the vaccine as having been provided by other healthcare provider.

U. References

1. World Health Organization. The top 10 causes of death [Internet]. Available from: https://www.who.int/en/news-room/fact-sheets/detail/the-top-10-causes-of-death
2. GBD 2016 Stroke Collaborators CO, Nguyen M, Roth GA, Nichols E, Alam T, Abate D, et al. Global, regional, and national burden of stroke, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol. 2019 Mar 11;0(0).
3. Evers SMAA, Struijs JN, Ament AJHA, van Genugten MLL, Jager J (Hans) C, van den Bos GAM. International Comparison of Stroke Cost Studies. Stroke. 2004 May 1;35(5):1209–15.
4. Saka O, McGuire A, Wolfe C. Cost of stroke in the United Kingdom. Age Ageing. 2008 Nov 13;38(1):27–32.
5. Chevreul K, Durand-Zaleski I, Gouépo A, Fery-Lemonnier E, Hommel M, Woimant F. Cost of stroke in France. Eur J Neurol. 2013 Jul 1;20(7):1094–100.
6. British Heart Foundation. Cardiovascular Disease Statistics 2019 [Internet]. 2019. Available from: https://www.bhf.org.uk/what-we-do/our-research/heart-statistics/heart-statistics-publications/cardiovascular-disease-statistics-2018
7. British Heart Foundation. Heart statistics publications BHF Statistics Factsheet UK [Internet]. 2018. Available from: https://www.bhf.org.uk/what-we-do/our-research/heart-statistics/heart-statistics-publications
8. Collins SD. Excess Mortality from Causes Other than Influenza and Pneumonia during Influenza Epidemics. Public Heal Reports. 1932;47(46):2159.
9. Tillet HE, Smith JWG, Gooch CD. Excess Deaths Attributable to Influenza in England and Wales: Age at Death and Certified Cause. Int J Epidemiol. 1983 Sep 1;12(3):344–52.
10. Reichert TA, Simonsen L, Sharma A, Pardo SA, Fedson DS, Miller MA. Influenza and the Winter Increase in Mortality in the United States, 1959-1999. Am J Epidemiol. 2004 Sep 1;160(5):492–502.
11. Warren-Gash C, Bhaskaran K, Hayward A, Leung GM, Lo S-V, Wong C-M, et al. Circulating Influenza Virus, Climatic Factors, and Acute Myocardial Infarction: A Time Series Study in England and Wales and Hong Kong. J Infect Dis. 2011 Jun 15;203(12):1710–8.
12. Nguyen JL, Yang W, Ito K, Matte TD, Shaman J, Kinney PL. Seasonal Influenza Infections and Cardiovascular Disease Mortality. JAMA Cardiol. 2016 Jun 1;1(3):274.
13. Blackburn R, Zhao H, Pebody R, Hayward A, Warren-Gash C. Laboratory-Confirmed Respiratory Infections as Predictors of Hospital Admission for Myocardial Infarction and Stroke: Time-Series Analysis of English Data for 2004–2015. Clin Infect Dis. 2018 Jun 18;67(1):8–17.
14. Barnes M, Heywood AE, Mahimbo A, Rahman B, Newall AT, Macintyre CR. Acute myocardial infarction and influenza: a meta-analysis of case–control studies. Heart. 2015 Nov 1;101(21):1738–47.

15. Kwok CS, Aslam S, Kontopantelis E, Myint PK, Zaman MJS, Buchan I, et al. Influenza, influenza-like symptoms and their association with cardiovascular risks: a systematic review and meta-analysis of observational studies. Int J Clin Pract. 2015 Sep;69(9):928–37.
16. Corrales-Medina VF, Suh KN, Rose G, Chirinos JA, Doucette S, Cameron DW, et al. Cardiac Complications in Patients with Community-Acquired Pneumonia: A Systematic Review and Meta-Analysis of Observational Studies.
17. Kwong JC, Schwartz KL, Campitelli MA, Chung H, Crowcroft NS, Karnauchow T, et al. Acute Myocardial Infarction after Laboratory-Confirmed Influenza Infection. N Engl J Med. 2018 Jan 25;378(4):345–53.
18. Warren-Gash C, Blackburn R, Whitaker H, McMenamin J, Hayward AC. Laboratory-confirmed respiratory infections as triggers for acute myocardial infarction and stroke: a self-controlled case series analysis of national linked datasets from Scotland. Eur Respir J. 2018 Mar;51(3):1701794.
19. Kytömaa S, Hegde S, Claggett B, Udell JA, Rosamond W, Temte J, et al. Association of Influenza-like Illness Activity With Hospitalizations for Heart Failure. JAMA Cardiol. 2019 Apr 1;4(4):363.
20. Smeeth L, Thomas SL, Hall AJ, Hubbard R, Farrington P, Vallance P. Risk of myocardial infarction and stroke after acute infection or vaccination. N Engl J Med. 2004 Dec 16;351(25):2611–8.
21. Warren-Gash C, Hayward AC, Hemingway H, Denaxas S, Thomas SL, Timmis AD, et al. Influenza Infection and Risk of Acute Myocardial Infarction in England and Wales: A CALIBER Self-Controlled Case Series Study. J Infect Dis. 2012 Dec 1;206(11):1652–9.
22. Public Health England. Influenza: the green book, chapter 19 [Internet]. 2019. Available from: https://www.gov.uk/government/publications/influenza-the-green-book-chapter-19
23. Public Health England. Pneumococcal: the green book, chapter 25 [Internet]. 2018. Available from: https://www.gov.uk/government/publications/pneumococcal-the-green-book-chapter-25
24. Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. BMJ. 2017 May 23;357:j2099.
25. National Institute for Health and Care Excellence. CVD risk assessment and management [Internet]. 2019.
26. Grint D, Thomas SL, Evans D, Walker JL, Andrews NJ, Tessier E, et al. What can research primary care datasets contribute to routine monitoring of vaccine coverage and uptake? In Public Health Research & Science Conference, Manchester UK.; 2019.
27. Finnikin S, Ryan R, Marshall T. Statin initiations and QRISK2 scoring in UK general practice: a THIN database study. Br J Gen Pract. 2017 Dec;67(665):e881–7.
28. Robson J, Dostal I, Sheikh A, Eldridge S, Madurasinghe V, Griffiths C, et al. The NHS Health Check in England: an evaluation of the first 4 years. BMJ Open. 2016 Jan 13;6(1):e008840.
29. Herrett E, Shah AD, Boggon R, Denaxas S, Smeeth L, van Staa T, 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 May 21;346(may20 3):f2350–f2350.

Appendix 4 Chapter 7 protocol and ethics approval

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
United Kingdom
Switchboard: +44 (0)20 7636 8636

www.lshtm.ac.uk

Observational / Interventions Research Ethics Committee

Ms Jennifer Davidson
LSHTM

10 June 2021

Dear Jennifer

Submission Title: Investigating the effect of influenza vaccine on acute cardiovascular events by cardiovascular risk status

LSHTM Ethics Ref: 26191

Thank you for responding to the Observational Committee Chair's request for further information on the above research and submitting revised documentation.

The further information has been considered on behalf of the Committee by the Chair.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised, subject to the conditions specified below.

Conditions of the favourable opinion

Approval is dependent on local ethical approval having been received, where relevant.

Approved documents

The final list of documents reviewed and approved is as follows:

Document Type	File Name	Date	Version
Protocol / Proposal	Vacc_CV_ISAC_protocol_sccs_14042021_v1.0	14/04/2021	1
Investigator CV	CV DEC 2020 JD	19/04/2021	1
Investigator CV	Short cv_CWG_16-07-2019	26/04/2021	1
Investigator CV	2 page CV_Oct_2019	26/04/2021	1
Other	Research_Ethics_online_training_certificate	26/05/2021	1
Local Approval	ISACapproval_printscreen	07/06/2021	1
Covering Letter	Cover Letter_07-06-2021	07/06/2021	1

After ethical review

The Chief Investigator (CI) or delegate is responsible for informing the ethics committee of any subsequent changes to the application. These must be submitted to the committee for review using an Amendment form. Amendments must not be initiated before receipt of written favourable opinion from the committee.

The CI or delegate is also required to notify the ethics committee of any protocol violations and/or Suspected Unexpected Serious Adverse Reactions (SUSARs) which occur during the project by submitting a Serious Adverse Event form.

An annual report should be submitted to the committee using an Annual Report form on the anniversary of the approval of the study during the lifetime of the study

At the end of the study, the CI or delegate must notify the committee using the End of Study form.

All aforementioned forms are available on the ethics online applications website and can only be submitted to the committee via the website at: http://leo.lshtm.ac.uk.

Further information is available at: www.lshtm.ac.uk/ethics.

Yours sincerely,

Professor Jimmy Whitworth

Medicines & Healthcare products Regulatory Agency

CPRD

INDEPENDENT SCIENTIFIC ADVISORY COMMITTEE (ISAC) PROTOCOL APPLICATION FORM

PART 1: APPLICATION FORM

IMPORTANT

Both parts of this application must be completed in accordance with the guidance note 'Completion of the ISAC Protocol Application Form', which can be found on the CPRD website cprd.com/research-applications

FOR ISAC USE ONLY	
Protocol No. -	**Submission date -**

GENERAL INFORMATION ABOUT THE PROPOSED RESEARCH STUDY

1. Study Title (Max. 255 characters)

Investigating the effect of influenza vaccine on acute cardiovascular events by cardiovascular risk status

2. Research Area (place '**X**' in all boxes that apply)

Drug Safety		Economics	
Drug Utilisation		Pharmacoeconomics	
Drug Effectiveness	X	Pharmacoepidemiology	
Disease Epidemiology	X	Methodological	
Health Services Delivery			

3. Chief Investigator

Title:	Dr
Full name:	Charlotte Warren-Gash
Job title:	Associate Professor of Epidemiology
Affiliation/organisation:	London School of Hygiene and Tropical Medicine
Email address:	Charlotte.Warren-Gash1@lshtm.ac.uk
CV Number (if applicable):	815_16

4. Corresponding Applicant

Title:	Ms
Full name:	Jennifer Davidson
Job title:	PhD Student
Affiliation/organisation:	London School of Hygiene and Tropical Medicine
Email address:	Jennifer.Davidson@lshtm.ac.uk
CV Number (if applicable):	503_19

5. List of all investigators/collaborators

Title:	Dr
Full name:	Amitava Banerjee
Job title:	Associate Professor in Clinical Data Science and Honorary Consultant in Cardiology
Affiliation/organisation:	University College London
Email address:	ami.banerjee@ucl.ac.uk
CV Number (if applicable):	090_16
Will this person be analysing the data? (Y/N)	N

Title:	Professor
Full name:	Liam Smeeth
Job title:	Professor of Pharmacoepidemiology
Affiliation/organisation:	London School of Hygiene and Tropical Medicine
Email address:	Liam.Smeeth@lshtm.ac.uk
CV Number (if applicable):	
Will this person be analysing the data? (Y/N)	N

Title:	Professor
Full name:	Ian Douglas
Job title:	Professor of Clinical Epidemiology
Affiliation/organisation:	London School of Hygiene and Tropical Medicine
Email address:	ian.douglas@lshtm.ac.uk
CV Number (if applicable):	045_15CEPSL
Will this person be analysing the data? (Y/N)	N

Title:	Dr
Full name:	Emily Herrett
Job title:	Assistant Professor
Affiliation/organisation:	London School of Hygiene and Tropical Medicine
Email address:	Emily.Herrett@lshtm.ac.uk
CV Number (if applicable):	085_15
Will this person be analysing the data? (Y/N)	N

Title:	Dr
Full name:	Helen McDonald
Job title:	Clinical Assistant Professor
Affiliation/organisation:	London School of Hygiene and Tropical Medicine
Email address:	Helen.McDonald@lshtm.ac.uk
CV Number (if applicable):	320_15CES
Will this person be analysing the data? (Y/N)	N

Title:	Dr
Full name:	Harriet Forbes
Job title:	Assistant Professor
Affiliation/organisation:	London School of Hygiene and Tropical Medicine
Email address:	Harriet.Forbes@lshtm.ac.uk
CV Number (if applicable):	465_15
Will this person be analysing the data? (Y/N)	N

Title:	Dr
Full name:	Clémence Leyrat
Job title:	Assistant Professor

Affiliation/organisation:	London School of Hygiene and Tropical Medicine
Email address:	Clemence.Leyrat@lshtm.ac.uk
CV Number (if applicable):	332_18
Will this person be analysing the data? (Y/N)	N

Title:	Dr
Full name:	Richard Pebody
Job title:	Consultant Epidemiologist & Honorary Professor
Affiliation/organisation:	University College London
Email address:	pebodyr@who.int / r.pebody@ucl.ac.uk
CV Number (if applicable):	504_19
Will this person be analysing the data? (Y/N)	N

6. Experience/expertise available

List below the member(s) of the research team who have experience with CPRD data.

Name:	**Protocol Number/s:**
Charlotte Warren-Gash	20_000135, 19_209, 19_096, 18_278, 18_207, 18_134R, 18_009, 17_176R
Liam Smeeth	Selected examples: 19_209, 19_096, 18_278, 18_207, 16_174, 16_113A, 15_257, 12_065, 12_027RA
Amitava Banerjee	20_000135, 19_209
Emily Herrett	20_000135, 19_209, 17_008, 17_156, 17_196
Harriet Forbes	20_000135, 18_278, 16_100, 12_090
Helen McDonald	19_209, 19_084, 18_278, 11_033A
Clémence Leyrat	20_012, 18_060, 17-197R
Ian Douglas	
Jennifer Davidson	20_000135, 19_209

List below the member(s) of the research team who have statistical expertise.

Name(s):
Clémence Leyrat

List below the member(s) of the research team who have experience of handling large datasets (greater than 1 million records).

Name(s):
Emily Herrett
Helen McDonald
Jennifer Davidson
Harriet Forbes
Ian Douglas

List below the member(s) of the research team, or supporting the research team, who have experience of practicing in UK primary care.

Name(s):
Liam Smeeth
Helen McDonald

ACCESS TO THE DATA

7. Sponsor of the study

Institution/Organisation:	London School of Hygiene and Tropical Medicine
Address:	Keppel St, Bloomsbury, London WC1E 7HT

8. Funding source for the study

Same as Sponsor?	Yes		No	X
Institution/Organisation:	British Heart Foundation			
Address:	Greater London House, 180 Hampstead Road, London NW1 7AW			

9. Institution conducting the research

Same as Sponsor?	Yes	X	No	
Institution/Organisation:	London School of Hygiene and Tropical Medicine			
Address:	Keppel St, Bloomsbury, London WC1E 7HT			

10. Data Access Arrangements

Indicate with an '**X**' the method that will be used to access the data for this study:

Study-specific Dataset Agreement	

Institutional Multi-study Licence	X
Institution Name	London School of Hygiene and Tropical Medicine
Institution Address	Keppel St, Bloomsbury, London WC1E 7HT

Will the dataset be extracted by CPRD?

Yes		No	X

If yes, provide the reference number:

11. Data Processor(s):

Processing	LSHTM
Accessing	LSHTM
Storing	LSHTM
Processing area (UK/EEA/Worldwide)	UK
Organisation name	London School of Hygiene and Tropical Medicine
Organisation address	Keppel St, Bloomsbury, London WC1E 7HT

INFORMATION ON DATA

12. Primary care data (place '**X**' in all boxes that apply)

CPRD GOLD		CPRD Aurum	X

13. Please select any linked data or data products being requested

Patient Level Data (place '**X**' in all boxes that apply)

ONS Death Registration Data	X	CPRD Mother Baby Link	
HES Admitted Patient Care	X	Pregnancy Register	

HES Outpatient		NCRAS (National Cancer Registration and Analysis Service) Cancer Registration Data	
HES Accident and Emergency		NCRAS Cancer Patient Experience Survey (CPES) data	
HES Diagnostic Imaging Dataset		NCRAS Systemic Anti-Cancer Treatment (SACT) data	
HES PROMS (Patient Reported Outcomes Measure)		NCRAS National Radiotherapy Dataset (RTDS) data	
		Mental Health Services Data Set (MHDS)	

Area Level Data (place '**X**' in all boxes that apply)

Practice level (UK)		**Patient level (England only)**	
Practice Level Index of Multiple Deprivation (Standard)		Patient Level Index of Multiple Deprivation	
Practice Level Index of Multiple Deprivation (Non-standard)		Patient Level Townsend Score	X
Practice Level Index of Multiple Deprivation Domains (Non-standard)			
Practice Level Carstairs Index for 2011 Census (Excluding Northern Ireland) (Standard)			
2011 Rural-Urban Classification at LSOA level (Non-standard)			

14. Are you requesting linkage to a dataset not listed above?

Yes		No	X

If yes, provide the reference number:

15. Does any person named in this application already have access to any of these data in a patient identifiable form, or associated with an identifiable patient index?

Yes		No	X

If yes, provide further details:

VALIDATION/VERIFICATION

16. Does this protocol describe an observational study using purely CPRD data?

Yes	X	No	

17. Does this protocol involve requesting any additional information from GPs, or contact with patients?

Yes		No	X

If yes, provide the reference number:

Medicines & Healthcare products Regulatory Agency

CPRD

PART 2: PROTOCOL INFORMATION

Applicants must complete all sections listed below Sections which do not apply should be completed as '*Not Applicable*' and justification provided
A. Study Title (Max. 255 characters) Investigating the effect of influenza vaccine on acute cardiovascular events by cardiovascular risk status
B. Lay Summary (Max. 250 words) In the days after severe influenza infection, people are at higher risk of an acute cardiovascular event, such as a heart attack or stroke. Influenza vaccine can prevent influenza infection or reduce the severity of illness. Vaccination can, therefore, also prevent any acute cardiovascular event which may occur after infection. Clinical trials have previously shown that in people who have pre-existing cardiovascular disease (e.g., heart disease or a previous stroke), the influenza vaccine reduces the chance of further cardiovascular events. It is unknown whether the vaccine offers the same protection to people who have risk factors for cardiovascular disease (without having pre-existing cardiovascular disease), such as in those with high blood pressure. Anonymised GP and hospital records are often used to measure the real-world effectiveness of vaccines. We will use records from adults aged 40-84 years to estimate the occurrence of first acute cardiovascular events following influenza vaccination compared to an event without recent influenza vaccination. We will do this separately for people with and without risk factors for cardiovascular disease. In England, routine seasonal influenza vaccine is offered to all people aged ≥65 years and those aged <65 years in a clinical risk group more likely to experience infection-related complications. Our study will include all adults aged 40-84 who receive influenza vaccine as individuals <65 years with risk factors for cardiovascular disease (our population of interest) are not typically targeted for influenza vaccine.
C. Technical Summary (Max. 300 words) Observational studies have identified that acute respiratory infections can trigger acute cardiovascular events, particularly myocardial infarction and stroke. In self-controlled case series (SCCS) studies, which implicitly control for the effect of fixed confounding factors using within-individual comparisons, myocardial infarction and stroke risk were elevated two- to six-fold in the days following clinically diagnosed respiratory infections and influenza-like illness, with this elevated risk remaining for up to one month. Consistent cardiovascular triggering effects have been found for laboratory-confirmed infections, including the influenza virus. Randomized controlled trials and observational studies have also demonstrated the cardiovascular benefit of the seasonal influenza vaccine. However, randomized controlled trial data is largely limited to individuals with established cardiovascular disease (CVD). While observational studies have considered the vaccine's benefit among those without a previous cardiovascular event, the benefit specifically in people with raised cardiovascular risk has not been considered. Quantifying any protective effects of influenza vaccine in people with raised cardiovascular risk, predictive of future CVD, will further understanding of the vaccine's cardiovascular benefits and whether it is effective for primary prevention of CVD. We aim to investigate the association between influenza vaccination and major adverse cardiovascular events in a SCCS study, focusing on the effect of raised cardiovascular risk level, defined by QRISK2 score ≥10% and diagnosed hypertension, on any association. We will use CPRD Aurum and HES admissions data from 2008-2019 and include all adults aged 40-84 years who had seasonal influenza vaccine and experienced their first major adverse cardiovascular event in the same year as vaccination was given. SCCS design removes between-person confounding, a major issue in vaccine effectiveness studies using observational data. In conditional Poisson regression models adjusted for season, we will calculate incidence ratios. We will also stratify by patient cardiovascular risk status.

D. Outcomes to be Measured

We will use the same primary and secondary cardiovascular outcomes as detailed in our ISAC protocol no. 19_209, with the exception of cardiovascular death (note death excluded as an outcome due to SCCS design in which an event of interest must not censor the observation period, see "Data/statistical analysis" and "Limitations" sections for further details).

Primary outcome
First major adverse cardiovascular events (MACE). This includes; acute coronary syndrome (ACS) which captures both myocardial infarction (MI) and unstable angina, stroke, transient ischaemic attack (TIA), left ventricular heart failure and acute limb ischaemia.

Secondary outcomes
Individual cause-specific acute cardiovascular events:

- acute coronary syndrome (subdivided into MI and unstable angina),
- stroke / TIA,
- left ventricular heart failure.

E. Objectives, Specific Aims and Rationale

Research objective: investigate the association between influenza vaccination and MACE.

Specific aims:
1. using a SCCS study design estimate the relative incidence of MACE after influenza vaccination within individuals aged 40-84 years by comparing risk during baseline (unexposed) time periods to following influenza vaccination (exposed). Exposed and unexposed time periods are defined in the "Data/statistical analysis" section.
2. use the methodology from aim 1 and stratify by patient cardiovascular risk status. Cardiovascular risk will first be defined by QRISK2 score (raised risk ≥10% and low risk <10%) and in a separate analysis by the presence or absence of diagnosed hypertension.

Rationale:
We have previously used CPRD data (ISAC no. 19_209) to show that people aged 40-64 years who are not currently eligible for influenza vaccination but who have raised cardiovascular risk have a higher risk of MACE after an acute respiratory infection.

Randomized controlled trials have shown that influenza vaccination can reduce the risk of MACE among patients with established cardiovascular disease (CVD). SCCS studies have previously shown a reduction in the risk of MI and stroke after influenza vaccination among middle-aged and older adults. Identifying a reduction in risk of MACE after influenza vaccination specifically in people with raised cardiovascular risk, but without established CVD, could prevent premature CVD and inform whether UK influenza vaccine policy should be extended.

We are unable to directly investigate the impact of influenza vaccination on MACE incidence in the same study population aged 40-64 years used in our previous study as influenza vaccine is only recommended for ≥65s and in <65s in a clinical risk group (1). Therefore, we will include all adults aged 40-84 to provide a large enough study population, among who we can stratify results by cardiovascular risk status as well as by age, and if necessary, by health condition, to investigate the vaccine's effect in a study population biologically similar to our target population.

Additionally, as patients already eligible for influenza vaccination are included in our study population our results may help improve uptake if we demonstrate that vaccination in those without CVD could prevent first MACE.

F. Study Background

Acute respiratory infections and cardiovascular complications

Acute respiratory infections (ARIs) can trigger a transient increased risk of acute cardiovascular events. Population-level studies illustrate that trends in cardiovascular mortality correspond to the circulation of respiratory viruses, particularly influenza, even after controlling for temporal and environmental factors (2–4). In several self-controlled case series (SCCS) studies, two to six-fold increases in the risk of cardiovascular complications – particularly myocardial infarction (MI) and stroke – in the days following ARI have been found, with elevated risk persisting for around one month (5–8). In SCCS studies, cases act as their own controls, accounting for fixed confounders, during periods of non-exposure (9). The method was developed to investigate associations between acute outcomes and transient exposures.

Established cardiovascular disease (CVD) has long been an indication for influenza vaccination (1), however, raised cardiovascular risk is not. In a recent study (ISAC protocol 19_209, data yet to be published) we have shown that adults aged 40-64 years with raised cardiovascular risk, but without established CVD, had a higher incidence of both ARI and a major adverse cardiovascular event (MACE) following ARI compared to people at low cardiovascular risk. The risk of MACE after ARI was two-fold higher in those with hypertension and nearly four-fold higher in those with a QRISK2 10-year risk score ≥10%.

Influenza vaccination to prevent cardiovascular events

Randomised controlled trials (RCTs) and observational studies have also quantified the cardiovascular benefits of influenza vaccination. A meta-analysis of four secondary prevention RCTs comparing influenza vaccine to placebo or no vaccination among individuals with mean ages of 55-65 years with recent acute coronary syndrome (ACS) or a coronary intervention showed a significant reduction in cardiovascular mortality among those vaccinated (risk ratio (RR) 0.45, 95% CI 0.26-0.76) (10). Another meta-analysis of five small RCTs, in which three RCTs participants had recent ACS or stable coronary artery disease, found influenza vaccine reduced the risk of composite cardiovascular outcomes within one year of follow up (RR 0.64, 0.48-0.86) (11). Self-controlled case series studies, using CPRD data, have demonstrated similar effects. In an analysis of influenza vaccination in heart failure patients, the vaccine was associated with a lower risk of hospitalization for cardiovascular outcomes (incidence ratio (IR) 0.73, 0.71-0.76) (12). Smeeth *et al* found reduced relative incidence of MI and stroke up to a month after influenza vaccination (IR 0.87, 0.79-0.96 and 0.88, 0.80-0.97, respectively, for 15-28 days post-vaccination). More recent studies found the relative incidence of MI and stroke were significantly reduced in the 60 days following vaccination (IR 0.82, 0.75-0.90 and 0.83, 0.77-0.89, respectively, for 29-59 days post-vaccination) (13,14). A recent study using Norwegian electronic health record data, compared the relative incidence of MI and stroke during the 2009 H1N1 influenza pandemic among vaccinated adults who were deemed at high and low cardiovascular risk (15). While the study found the pandemic influenza vaccine reduced the relative incidence of MI and stroke in those with high cardiovascular risk, an increase in risk was identified among people at low cardiovascular risk. The discordant results among high and low cardiovascular risk may be due to bias. The study defined cardiovascular risk using prescription of anti-diabetic, anti-obesity, anti-thrombotic, pulmonary or cardiovascular medications at the time of vaccination, which was after follow-up had started. Such study design would have resulted in any individual who had a cardiovascular event before vaccination likely prescribed cardiovascular medications and, therefore, classified as high risk by vaccination.

Mechanisms

Influenza infection induce a range of haemodynamic, inflammatory and pro-coagulant effects which exacerbate underlying CVD that may predispose to thrombotic events (16,17). Infection can trigger systemic inflammatory processes including release of pro-inflammatory cytokines, which are key mediators in atherosclerosis, and directly impact plaque rupture through local inflammation. Raised cholesterol, diabetes and hypertension cause endothelial dysfunction, a key early stage of atherosclerosis, suggesting increased risk in individuals with raised cardiovascular risk but without established CVD (17). Additionally, during severe infection, influenza virus can invade the myocardium inducing cardiotoxicity (18). Influenza vaccine prevents infection and, therefore, the occurrence of the previously outlined processes. It has also been theorized that the vaccine modulate an adaptive immunoinflammatory response to provide a cardioprotective effect (19). One animal study found influenza vaccine increased the secretion of anti-inflammatory cytokines (20). A molecular study suggested influenza vaccine antibodies could be agonists on atheroprotective pathways (21), which would lead to the oxidization of low-density

lipoproteins and reduce their uptake by macrophages, thereby decelerating atherosclerotic plaque formation and progression to rupture (19).

Public Health importance

Seasonal influenza poses significant healthcare pressure in England each winter (22) and is associated with an average of 10,000 deaths annually, the majority of which are in adults aged ≥65 years (23). Deaths due to influenza vary in number each year depending on the severity of the flu season.

Widespread uptake of influenza vaccine is, therefore, important to reduce the pressure on health services during the winter months and protect individuals at risk of severe influenza. In England, 2019/20 influenza vaccine uptake (which is similar to previous years) was high (72%) among adults aged ≥65 years but low (45%) among those aged <65 years who are recommended to receive the vaccine due to underlying health conditions (24). The vaccine is not currently specifically recommended to people with raised cardiovascular risk.

Evidence to show the cardiovascular benefits of influenza vaccine in people who are not currently eligible for routine vaccination, will inform any future expansion of vaccine recommendations. This is particularly important in the context of the COVID-19 pandemic. Influenza vaccination can help to minimise co-circulation of influenza virus and SARS-CoV-2, as well as the associated increased pressures on health services, including hospital admission.

G. Study Type

Hypothesis testing: test the null hypothesis that influenza vaccination does not affect the relative incidence of MACE within 120 days of vaccination, and that there is no difference in the relative incidence of MACE after influenza vaccination by cardiovascular risk status.

H. Study Design

Self-controlled case series study

I. Feasibility counts

There were 15,503 patients aged 65-84 in CPRD Aurum with at least 12 months post registration follow-up who were eligible for HES linkage and had a first MACE in 2013-14, the midpoint of our study period.

Based on other published work, we assume 70% of these patients will have a seasonal influenza vaccination (24), providing an estimated 10,852 CPRD Aurum patients with a MACE and influenza vaccine in 2013/14.

1 in 4 adults in the UK are reported to have hypertension, we assume at least as many to have a QRISK2 ≥10% (25,26). Given a proportion may have established CVD and some under-reporting in CPRD, we estimate that one-fifth of our study population will have hypertension or QRISK2 ≥10% (i.e., 2,170 for 2013/14).

Note we limited our feasibility count to adults aged 65-84 as this population is covered by universal influenza vaccination. In those <65 years selective vaccination applies and uptake is suboptimal. While we will additionally include adults aged 40-64 years we have based our sample size calculations only on those aged 65-84 to ensure we have adequate power among those who are not a selective study population.

J. Sample size considerations

The user-written Stata command 'sampsi_sccs' was used to calculate sample size requirements.

The sample size for each cardiovascular risk level required to identify a range of incidence ratio in the post-vaccination interval of 15-28 days, when the observation period is 365 days and all subjects are exposed (assuming 90% power at the 5% significance level) is shown below.

Incidence ratio	Sample size
0.75	4,222
0.80	6,812
0.85	1,2488
0.90	28,938

Assuming a constant trend in MACE incidence and vaccine uptake, based on our feasibility counts for the midpoint of our study period, we should be able to detect an incidence ratio >0.90 with all years combined in our raised cardiovascular risk population.

K. Planned use of linked data (if applicable):

We require linkage to HES admitted patient care data, ONS mortality data, and patient level Townsend twentile data.

HES admitted patient care data: identify MACE outcomes, given the majority of MACE result in hospitalisation. We will also use HES APC data to exclude patients with established CVD. Additionally, ethnicity, which will be included in our baseline characteristics description of the study population, will be supplemented with HES data when missing in CPRD.

ONS mortality data: ascertain date of death to censor follow-up as well as to exclude patients who died from a sensitivity analysis (see "Data/statistical analysis" and "Limitations" sections for further details).

Townsend twentiles: socioeconomic status is included in the calculation of QRISK2 scores.

L. Definition of the Study population

We will include CPRD Aurum patients with linked HES APC data aged 40-84 years at first MACE who have at least 12 months of follow-up post-registration from 01/09/2008-31/08/2019. To ensure we only include an individual's first MACE we will exclude anyone with established CVD. We will define established CVD as a history of coronary heart disease (myocardial infarction, angina, revascularisation procedure or coronary heart disease not otherwise specified), stroke or TIA, or heart failure. Among individuals with a MACE, we will only include those who received an influenza vaccine within the same year (01 September-31 August). For each individual follow-up will start on 01 September and end on the earliest of: death (censor), transfer out of CPRD practice (censor) or 31 August of the following year. Given we are only including first MACE this means each individual will only contribute to one year of follow-up.

Although we are interested in the cardiovascular benefit of influenza vaccine in adults not already eligible for influenza vaccine i.e., aged <65 years with raised cardiovascular risk, we are limited to investigating the benefit among those who currently receive the vaccine. ≥65s are all eligible for influenza vaccination providing a large and non-selected population among whom uptake is high. 84 years is the oldest age that CVD risk prediction (QRISK2) should be applied as those aged ≥85 years are all considered to be at high risk of CVD (27). Additionally, immunosenescencing will impact inclusion of older ages. We will include those aged 40-64 years to have a biologically similar study population to our target group. However, those aged 40-64 years who are vaccinated will be a selective population often with underlying health conditions and low vaccine uptake.

We will start follow-up on 01 September as this corresponds to when patients eligible for seasonal influenza vaccine are identified by GPs and is similar to how we defined follow-up in our previous study which investigated the impact of acute respiratory infections on MACE.

Our study will end in 2019 to avoid introducing any bias in results due to COVID-19 circulation in 2020.

M. Selection of comparison group(s) or controls

As our study is a SCCS, comparisons will be within individuals i.e., patients will act as their own controls during different time periods. The effect of influenza vaccination on the relative incidence of MACE occurring up to 120 days after vaccine receipt compared to baseline periods for each individual will be presented for each stratum of cardiovascular risk level. This will be defined first as QRISK2 score <10% (low), QRISK2 ≥10% (raised), then by the presence or absence of hypertension.

N. Exposures, Outcomes and Covariates

We will utilise the codelists developed for our previous study (ISAC protocol 19_209) to define our exposure and outcomes as well as stratifying variable of cardiovascular risk.

Influenza vaccination (exposure)
Influenza vaccination is given annually from 01 September, with the majority of vaccines given by the end of December in preparation for the influenza season which usually occurs between December and March (28). We will identify those who have received influenza vaccine in CPRD data (Appendix 2 – codelist).

MACE (outcome)
We will use the composite outcome of MACE (defined in "Outcomes to be Measured" section, Appendix 3 – codelist) to account for the wide range of outcomes possible and to achieve statistical power. Our secondary outcomes are each of the cardiovascular conditions separately.

Stratification by cardiovascular risk group
Patients will be stratified based on cardiovascular risk at study entry. This will be classified by QRISK2 score ≥10% (raised risk, with further breakdown into 10-19% and ≥20% in sensitivity analysis, see "Data/statistical analysis" section for further details) or <10% (low risk) and separately by the presence or absence of diagnosed hypertension (Appendix 1 – codelist).

Covariates
SCCS analysis has the advantage of implicitly controls for fixed between-person confounding effects. Additionally, we will only include one year of follow-up for each individual so will not adjust for age. Season (warmer months = April-September and cooler months = October-March and in a sensitivity analysis four seasons of the year September-November, December-February, March-May, June-August) will be adjusted for.
We will also consider stratifying by potential effect modifiers:

- Age
- Sex
- Timing of vaccination (early = 01 September-15 November, or later = 16 November-31 March)
- Year and/or years grouped into matched and non-matched seasons (matched = vaccine strains correspond to circulating strains and non-matched = vaccine strains did not correspond to circulating strains)
- Acute respiratory infection / influenza-like illness (within 28 days prior to MACE)

O. Data/ Statistical Analysis

We will describe the socio-demographic and co-morbid characteristics of our study population, including stratification by cardiovascular risk status (low or raised).

We will divide follow-up into risk and baseline intervals. After vaccination, it can take up to 14 days for an individual to be effectively protected by the vaccine (10). Therefore, we will consider individuals as effectively vaccinated from day 15 post-vaccination and ending 120 days after vaccination. We will subdivide the risk period into intervals of 15-28, 29-59, 60-90, and 91-120 days. We will consider days 1-14 after vaccination as a separate interval, excluding this time from baseline, given protection against influenza will develop over these 14 days. The 14 days before vaccination will also be considered as a separate interval since MACE during this period is likely to affect the subsequent likelihood of receiving an influenza vaccine (a violation of SCCS assumptions). We will exclude individuals who have a MACE and influenza vaccination on the same day given the sequence of the events cannot be determined. All other time will be defined as baseline, with the number of events in the pre-vaccination and post-vaccination baseline period reported. Our baseline and risk periods are outlined in Figure 1 (Appendix 5).

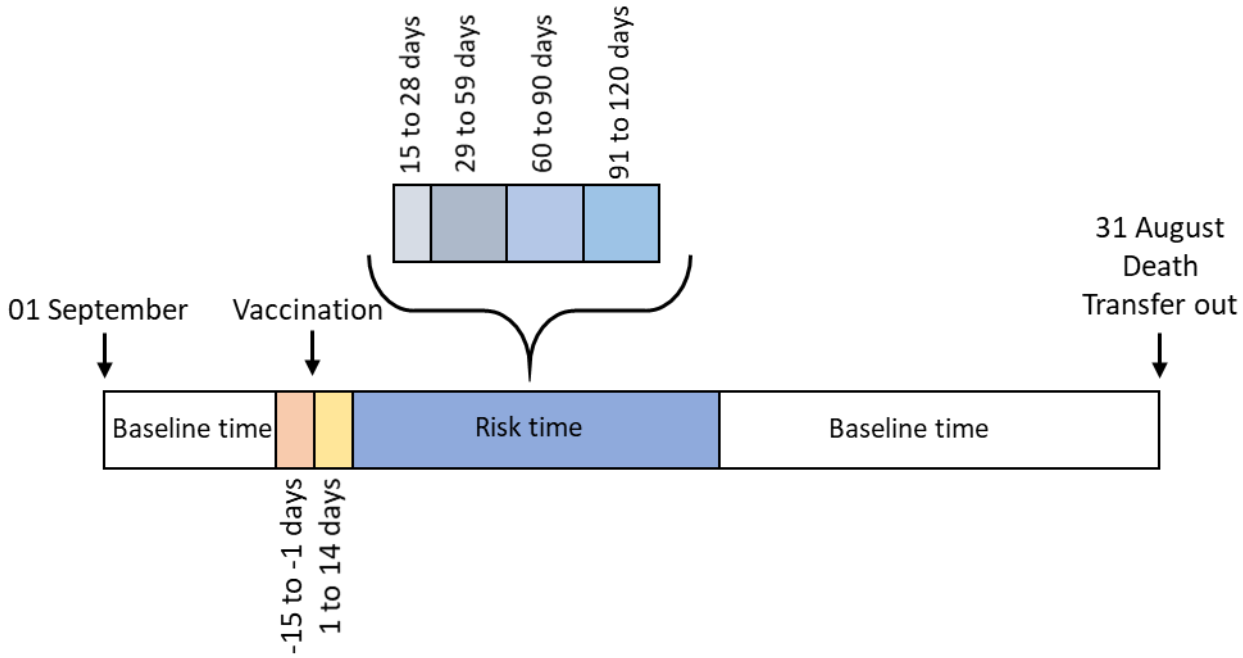


Figure 1. Study design *[note will be added as appendix on online application as figures cannot be added directly]*

Using conditional Poisson regression, we will calculate the relative incidence of MACE occurring in time periods after influenza vaccination compared to baseline time periods for each individual, adjusting for season. We will present case numbers and incidence ratios overall and stratified by cardiovascular risk level defined by i) QRISK2 score ii) hypertension status. Additionally, we will present yearly incidence ratios and compare trends to the corresponding year's vaccine effectiveness estimates published by Public Health England (29) to aid interpretation of our results. We will present stratified analyses for potential effect modifiers: age (split into <65 (selective vaccination), 65-74 and 74-84 years), sex, year/matching season, timing of vaccination and acute respiratory infection/ influenza-like illness.

Pre-specified sensitivity analysis

1. Death can occur as a direct result of our outcome of interest, MACE. This violates an assumption of the SCCS: observation periods should end independent of event timing. To address this, we will repeat our analysis excluding non-survivors i.e., those who die, and compare estimate obtained to our main analysis. If results of our sensitivity analysis are substantially different to our main results, we will conduct a further analysis using an extension of the SCCS method to correct for non-random censoring of follow-up modelling post-event survival times (30). The extension works by conditioning explicitly on the age at

censoring, with cases weighted by the density of the time from event to censoring. Under this new model, the interpretation of the exposure effect is the same as for the standard model, but the age effect takes on a different interpretation, representing the combined effect of age-specific relative incidence and censoring.

2. As MACE can affect the subsequent likelihood of vaccination, we will repeat analysis starting follow-up and defining cardiovascular risk status on the date of vaccination, with fixed follow-up, regardless of survival given no further exposure can occur and outcome can only be after exposure, to 31st August. All baseline time will be the time from 121 days after vaccination to 31st August. Our follow-up is outlined in Figure 2 (Appendix 5).

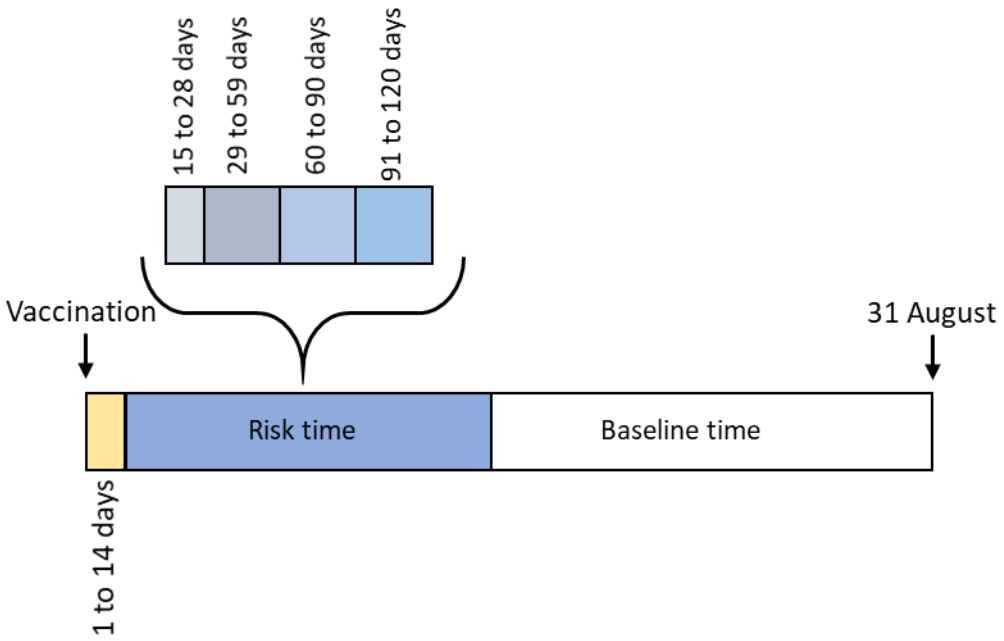


Figure 2. Sensitivity analysis study design

3. Influenza circulation peaks during winter month just after/during influenza vaccine administration. MACE have also been shown to peak during winter months. Estimates will be adjusted by season in all our analyses, but exposed and unexposed time in each seasonal category is needed to distinguish the vaccine effects from seasonal effects. Therefore, in our main analysis we will use a broad definition of season (warmer months = April-September and cooler months = October-March) but we will repeat analysis attempting to use as narrow a definition of season as possible, with the four seasons of the year (September-November, December-February, March-May, June-August) considered.
4. We will explore the effect of using finer stratifications of QRISK2 score with ≥10% further broken down into 10-19% and ≥20%.

P. Plan for addressing confounding

A major advantage of the SCCS study design is that it implicitly controls for fixed between-person confounding effects (31). Additionally, in our analysis we will adjust for season. See "Data/statistical analysis" section for details.

Q. Plans for addressing missing data

For some variables included in QRISK2 such as BMI, cholesterol, blood pressure, smoking status and ethnicity, there are likely to be missing values. The QRISK2 algorithm replaces missing data with average values. GP systems operate in the same way as our algorithm and will calculate a score using the same process.

As our study uses a SCCS design i.e., comparisons are within-person, we do not need to include many potential confounders, only those which are time-varying such as age, because they will be implicitly controlled for in the design.

R. Patient or user group involvement (if applicable)

Patient and patient involvement and engagement (PPIE) groups have not been involved in the development of this research. We will consider the use of either the Health Protection Research Unit (HPRU) Vaccines and Immunisation (a research partnership between Public Health England and the London School of Hygiene and Tropical Medicine) or any British Heart Foundation PPIE groups in the use and disseminating of our results, for example to explore whether people are aware that influenza vaccine can prevent MI/stroke, and whether that might affect their attitudes towards value of influenza vaccine.

S. Plans for disseminating and communicating study results, including the presence or absence of any restrictions on the extent and timing of publication

The study findings will be submitted for publication as a preprint on medRxiv and in a peer-reviewed scientific journal. Additionally, results will be presented at conferences and other meetings as appropriate. We will work with the LSHTM press office to publicise our study and its findings. The electronic health records research group at LSHTM also have a webpage with research news and social media accounts which will be used. We will work with the British Heart Foundation, who are the funders of this research, to publicise our findings to a wider audience including the general public and patient groups.

Conflict of interest statement: The work is supported by a British Heart Foundation PhD Studentship (FS/18/71/33938). There are no conflicts of interest to declare.

T. Limitations of the study design, data sources, and analytic methods

SCCS assumptions:

1. Recurrent events must be independent i.e., the chance of a second event is not influenced by the first event. A MACE increases the likelihood of future MACE. To avoid violating this assumption we have excluded patients with established CVD before the study period and only considered the first MACE.

2. Occurrence of an event (MACE) should not alter the probability of subsequent exposure. Evidence exists to show influenza vaccine may be a useful secondary prevention for CVD. Therefore, while influenza vaccine is universally offered to all people aged ≥65 years, vaccination may be targeted at those who are perceived to be at increased risk of MACE. As patients' health declines vaccination effectiveness may be lower due to frailty. We plan to start follow-up at the start of each influenza vaccination season so will only include a short pre-exposure baseline. We will also conduct a sensitivity analysis starting follow-up at vaccination, so only events which occur after vaccination are included. MACE may also affect the timing of vaccination i.e., may be more or less likely to get vaccinated if a current inpatient at the start of flu vaccine rollout. Again, we have addressed this by including a pre-exposure window.

3. The event of interest must not censor the observation period. While MACE increases the risk of death, we will conduct a sensitivity analysis excluding patients who die and consider the use of an extension to the SCCS method to correct for non-random censoring.

Paradoxical result: sometimes in SCCS studies, a paradoxical result is obtained when stratifying by a risk factor, whereby the relative incidence of the event is higher in those considered least at risk. For example, when stratifying by age, the relative incidence of MACE is higher after ARI in younger versus older individuals (8,32). This is likely because younger individuals have a much lower baseline risk of MACE but by definition must have had a MACE to be included in the SCCS study. Therefore, the relative incidence after a triggering event such as ARI appears higher (even though the absolute incidence would still be low). Whereas older individuals are more likely to have MACE all year round, so the relative incidence after ARI is lower (but the absolute incidence is high). It is possible that we may find a similar paradoxical result: influenza vaccine protects overall against MACE, but the relative incidence of MACE after influenza vaccine may appears more protective among those at lowest CVD risk because of the differing baseline risk levels. To address this, we will first conduct our analysis without stratification by cardiovascular risk status and compare these results to those with stratification to interpret our findings.

Underascertainment of vaccination status: All people aged ≥65 years are eligible for influenza vaccination, so the majority of vaccinations will be given at GP surgeries rather than occupational settings (with the majority of ≥65s retired) or paid for in pharmacies. Therefore, a large amount of missing vaccination data is not anticipated.

Misclassification of cardiovascular outcomes: We will use HES data to supplement primary care recording of all cardiovascular events with our analysis limited to linked data. Our systematic review of ACS, HF and stroke diagnosis – the three major components of MACE – validation in European EHRs found overall diagnoses are accurate for use in research (33). Sensitivity was ≥80% for 91% of MI studies and ≥70% for 73% of stroke studies, although ≤66% in all but one heart failure study. Positive predictive value was ≥80% in 74% of heart failure, 88% of MI, and 70% of stroke studies.

Misclassification of CVD risk: While blood pressure and therefore hypertension are generally well recorded in primary care data this requires patients to have presented to the GP. For patients who do not often attend the GP the diagnosis will not be made. Those with hypertension recorded may, therefore, be healthier with better managed hypertension. The algorithm to calculate QRISK2 score is based on many variables. Missing data for the chronic conditions included in the algorithm indicates that the patient does not have the risk factor in question although, similar to hypertension, there may be underdiagnosis or under-recording. Missing data for variables such as ethnicity, BMI, alcohol consumption and smoking status are due to non-recording. Our QRISK2 algorithm will replace missing values for smoking status and BMI with the median value, in line with the QRISK2 calculator. We have previously compared our calculated QRISK2 scores to those recorded by GPs to validate our classifications, with consistent results obtained.

Healthy vaccinee effect: We would have ideally used a cohort or case-control study design to estimate influenza vaccine effectiveness against cardiovascular complications among people with raised cardiovascular risk compare to low cardiovascular risk. However, vaccinated patients tend to have health, lifestyle and behavioural differences to those who are not vaccinated. The SCCS method is more robust at accounting for the healthy vaccinee effect than case-control or cohort study methods as it controls for unmeasured confounders which may differ between individuals. However, as our study will only include people who received the vaccine (at least once) our results will not be representative to all adults aged 40-84. However, it will give an unbiased estimate of the effect of the vaccine among people likely to get the vaccine.

Power: Although there will be sufficient power to look at the primary outcome, there may be insufficient power for individual secondary outcomes, particularly in stratified analysis.

U. References

1. Public Health England. Influenza: the green book, chapter 19 [Internet]. 2020. Available from: https://www.gov.uk/government/publications/influenza-the-green-book-chapter-19
2. Blackburn R, Zhao H, Pebody R, Hayward A, Warren-Gash C. Laboratory-Confirmed Respiratory Infections as Predictors of Hospital Admission for Myocardial Infarction and Stroke: Time-Series Analysis of English Data for 2004–2015. Clin Infect Dis. 2018 Jun 18;67(1):8–17.
3. Imai C, Barnett A, Hashizume M, Honda Y, Imai C, Barnett AG, et al. The Role of Influenza in the Delay between Low Temperature and Ischemic Heart Disease: Evidence from Simulation and Mortality Data from Japan. Int J Environ Res Public Health. 2016 Apr 28;13(5):454.
4. Warren-Gash C, Bhaskaran K, Hayward A, Leung GM, Lo S-V, Wong C-M, et al. Circulating Influenza Virus, Climatic Factors, and Acute Myocardial Infarction: A Time Series Study in England and Wales and Hong Kong. J Infect Dis. 2011 Jun 15;203(12):1710–8.
5. Smeeth L, Thomas SL, Hall AJ, Hubbard R, Farrington P, Vallance P. Risk of myocardial infarction and stroke after acute infection or vaccination. N Engl J Med. 2004 Dec 16;351(25):2611–8.
6. Warren-Gash C, Hayward AC, Hemingway H, Denaxas S, Thomas SL, Timmis AD, et al. Influenza Infection and Risk of Acute Myocardial Infarction in England and Wales: A CALIBER Self-Controlled Case Series Study. J Infect Dis. 2012 Dec 1;206(11):1652–9.
7. Kwong JC, Schwartz KL, Campitelli MA, Chung H, Crowcroft NS, Karnauchow T, et al. Acute Myocardial Infarction after Laboratory-Confirmed Influenza Infection. N Engl J Med. 2018 Jan 25;378(4):345–53.
8. Warren-Gash C, Blackburn R, Whitaker H, McMenamin J, Hayward AC. Laboratory-confirmed respiratory infections as triggers for acute myocardial infarction and stroke: a self-controlled case series analysis of national linked datasets from Scotland. Eur Respir J. 2018 Mar;51(3):1701794.
9. Whitaker HJ, Farrington CP, Spiessens B, Musonda P. Tutorial in biostatistics: The self-controlled case series method. Stat Med. 2006 May 30;25(10):1768–97.
10. Clar C, Oseni Z, Flowers N, Keshtkar-Jahromi M, Rees K. Influenza vaccines for preventing cardiovascular disease. Cochrane Database Syst Rev. 2015 May 5;(5):CD005050.
11. Udell JA, Zawi R, Bhatt DL, Keshtkar-Jahromi M, Gaughran F, Phrommintikul A, et al. Association Between Influenza Vaccination and Cardiovascular Outcomes in High-Risk Patients. JAMA. 2013 Oct 23;310(16):1711.
12. Mohseni H, Kiran A, Khorshidi R, Rahimi K. Influenza vaccination and risk of hospitalization in patients with heart failure: a self-controlled case series study. Eur Heart J. 2017 Sep 22;38(5):326–33.
13. Gwini SM, Coupland CAC, Siriwardena AN. The effect of influenza vaccination on risk of acute myocardial infarction: Self-controlled case-series study. Vaccine. 2011 Feb 1;29(6):1145–9.
14. Asghar Z, Coupland C, Siriwardena N. Influenza vaccination and risk of stroke: Self-controlled case-series study. Vaccine. 2015 Oct 5;33(41):5458–63.
15. Sen A, Bakken IJ, Govatsmark RES, Varmdal T, Bønaa KH, Mukamal KJ, et al. Influenza vaccination and risk for cardiovascular events: a nationwide self-controlled case series study. BMC Cardiovasc Disord. 2021 Dec 1;21(1):31.
16. Bazaz R, Marriott HM, Francis SE, Dockrell DH. Mechanistic links between acute respiratory tract infections and acute coronary syndromes. J Infect. 2013 Jan;66(1):1–17.
17. Corrales-Medina VF, Madjid M, Musher DM. Role of acute infection in triggering acute coronary syndromes. Lancet Infect Dis. 2010 Feb 1;10(2):83–92.
18. Estabragh ZR, Mamas MA. The cardiovascular manifestations of influenza: a systematic review. Int J Cardiol. 2013 Sep 10;167(6).
19. Aidoud A, Marlet J, Angoulvant D, Debacq C, Gavazzi G, Fougère B. Influenza vaccination as a novel means of preventing coronary heart disease: Effectiveness in older adults. Vol. 38, Vaccine. Elsevier Ltd; 2020. p. 4944–55.
20. Bermúdez-Fajardo A, Oviedo-Orta E. Influenza vaccination promotes stable atherosclerotic plaques in apoE knockout mice. Atherosclerosis. 2011 Jul 1;217(1):97–105.
21. Veljkovic V, Glisic S, Veljkovic N, Bojic T, Dietrich U, Perovic VR, et al. Influenza vaccine as prevention for cardiovascular diseases: Possible molecular mechanism. Vaccine. 2014 Nov 12;32(48):6569–75.
22. Iacobucci G. Winter pressure: high demand, pension crisis, and flu could be “perfect storm” for NHS, warns BMA. BMJ. 2019 Nov 6;367:l6399.
23. Public Health England. Surveillance of influenza and other respiratory viruses in the UK: Winter 2019 to 2020. 2020.
24. Public Health England. Seasonal influenza vaccine uptake in GP patients: winter season 2019 to 2020.

2020.
25. Robson J, Dostal I, Sheikh A, Eldridge S, Madurasinghe V, Griffiths C, et al. The NHS Health Check in England: an evaluation of the first 4 years. BMJ Open. 2016 Jan 13;6(1):e008840.
26. Public Health England. Tackling high blood pressure: an update [Internet]. 2018.
27. National Institute for Health and Care Excellence. CVD risk assessment and management [Internet]. 2019.
28. Public Health England. Weekly national flu reports [Internet]. Available from: https://www.gov.uk/government/collections/weekly-national-flu-reports
29. Public Health England. Influenza vaccine effectiveness: seasonal estimates [Internet]. Available from: https://www.gov.uk/government/publications/influenza-vaccine-effectiveness-seasonal-estimates
30. Paddy Farrington C, Anaya-Izquierdo K, Whitaker HJ, Hocine MN, Douglas I, Smeeth L. Self-controlled case series analysis with event-dependent observation periods. J Am Stat Assoc. 2011 Jun;106(494):417–26.
31. Whitaker HJ, Hocine MN, Farrington CP. The methodology of self-controlled case series studies. Stat Methods Med Res. 2009;18(1):7–26.
32. Ohland J, Warren-Gash C, Blackburn R, Mølbak K, Valentiner-Branth P, Nielsen J, et al. Acute myocardial infarctions and stroke triggered by laboratory-confirmed respiratory infections in Denmark, 2010 to 2016. Eurosurveillance. 2020 Apr 30;25(17):1900199.
33. Davidson J, Banerjee A, Muzambi R, Smeeth L, Warren-Gash C. Validity of Acute Cardiovascular Outcome Diagnoses Recorded in European Electronic Health Records: A Systematic Review. Clin Epidemiol. 2020 Oct 14;Volume 12:1095–111.

List of Appendices

Appendix 1 Hypertension Aurum codes
Appendix 2 Influenza vaccine Aurum codes
Appendix 3 Cardiovascular disease Aurum codes
Appendix 4 Major adverse cardiovascular event (MACE) Aurum codes

Appendix 5 Chapter 7 supplementary material

Supplementary Table 1. Incidence ratios for first acute cardiovascular event in risk periods following influenza vaccination by cardiovascular risk, age group and sex

Risk period	All		QRISK2				Hypertension			
			Raised risk		Low risk		Raised risk		Low risk	
	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)
All ages		p <0.0001		p <0.0001		p <0.0001		p = 0.0003		p <0.0001
Women										
15-28 days	3,377	0.76 (0.73-0.79)	2,909	0.79 (0.76-0.82)	468	0.62 (0.56-0.68)	1,986	0.78 (0.75-0.82)	1,391	0.72 (0.68-0.77)
29-59 days	7,699	0.80 (0.78-0.82)	6,709	0.85 (0.82-0.87)	990	0.60 (0.56-0.65)	4,500	0.82 (0.79-0.85)	3,199	0.77 (0.74-0.81)
60-90 days	7,773	0.84 (0.82-0.86)	6,762	0.89 (0.86-0.92)	1,011	0.64 (0.60-0.69)	4,494	0.86 (0.83-0.89)	3,279	0.82 (0.79-0.86)
91-120 days	7,519	0.87 (0.85-0.90)	6,554	0.93 (0.90-0.95)	965	0.66 (0.61-0.70)	4,345	0.89 (0.86-0.92)	3,174	0.85 (0.82-0.89)
Baseline	59,150	ref	50,051	Ref	9,099	ref	34,098	ref	25,052	ref
Men										
15-28 days	3,707	0.69 (0.67-0.72)	3,518	0.75 (0.72-0.78)	189	0.31 (0.27-0.36)	1,962	0.76 (0.73-0.80)	1,745	0.63 (0.60-0.66)
29-59 days	8,334	0.72 (0.70-0.74)	7,858	0.78 (0.76-0.80)	476	0.36 (0.33-0.40)	4,406	0.79 (0.77-0.82)	3,928	0.66 (0.63-0.68)
60-90 days	8,436	0.76 (0.74-0.78)	8,016	0.82 (0.80-0.85)	420	0.34 (0.30-0.37)	4,595	0.86 (0.83-0.89)	3,841	0.67 (0.64-0.69)
91-120 days	8,379	0.81 (0.79-0.83)	7,911	0.87 (0.85-0.89)	468	0.42 (0.38-0.46)	4,494	0.90 (0.87-0.93)	3,885	0.73 (0.70-0.76)
Baseline	67,223	ref	61,240	Ref	5,983	ref	34,360	ref	32,863	ref
40-64 years		p <0.0001		p <0.0001		p <0.0001		p = 0.0021		p <0.0001
Women										
15-28 days	597	0.64 (0.59-0.70)	220	0.74 (0.64-0.86)	377	0.59 (0.53-0.66)	258	0.72 (0.63-0.82)	339	0.59 (0.52-0.66)
29-59 days	1,337	0.66 (0.62-0.70)	532	0.83 (0.75-0.91)	805	0.58 (0.54-0.63)	589	0.75 (0.68-0.83)	748	0.60 (0.55-0.65)
60-90 days	1,350	0.69 (0.65-0.73)	522	0.83 (0.75-0.92)	828	0.62 (0.58-0.67)	559	0.73 (0.67-0.81)	791	0.66 (0.61-0.71)
91-120 days	1,262	0.69 (0.65-0.73)	483	0.82 (0.74-0.91)	779	0.63 (0.58-0.68)	519	0.73 (0.66-0.80)	743	0.67 (0.62-0.72)
Baseline	11,639	ref	4,207	Ref	7,432	ref	4,725	ref	6,914	ref
Men										
15-28 days	805	0.49 (0.45-0.52)	617	0.59 (0.54-0.64)	188	0.31 (0.27-0.36)	334	0.58 (0.52-0.65)	471	0.44 (0.40-0.48)
29-59 days	1,784	0.50 (0.47-0.52)	1,313	0.58 (0.55-0.62)	471	0.36 (0.32-0.40)	786	0.63 (0.58-0.68)	998	0.43 (0.40-0.46)
60-90 days	1,707	0.50 (0.47-0.52)	1,288	0.59 (0.55-0.63)	419	0.34 (0.30-0.37)	780	0.64 (0.59-0.69)	927	0.42 (0.39-0.45)
91-120 days	1,772	0.56 (0.53-0.59)	1,307	0.65 (0.61-0.69)	465	0.42 (0.38-0.46)	821	0.72 (0.67-0.78)	951	0.47 (0.44-0.51)

Risk period	All		QRISK2				Hypertension			
			Raised risk		Low risk		Raised risk		Low risk	
	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)
Baseline	18,383	ref	12,421	Ref	5,962	ref	7,245	ref	11,138	ref
65-74 years		p = 0.0176		p = 0.0308		p = 0.2686		p = 0.0022		p = 0.5397
Women										
15-28 days	1,143	0.82 (0.77-0.87)	1,052	0.82 (0.77-0.88)	91	0.80 (0.64-1.00)	642	0.84 (0.77-0.92)	501	0.79 (0.72-0.87)
29-59 days	2,491	0.82 (0.79-0.86)	2,306	0.83 (0.79-0.87)	185	0.75 (0.63-0.89)	1,310	0.79 (0.74-0.84)	1,181	0.86 (0.81-0.93)
60-90 days	2,534	0.86 (0.82-0.91)	2,351	0.87 (0.83-0.92)	183	0.76 (0.64-0.90)	1,341	0.83 (0.78-0.89)	1,193	0.90 (0.84-0.96)
91-120 days	2,479	0.90 (0.86-0.94)	2,293	0.91 (0.87-0.95)	186	0.81 (0.69-0.96)	1,335	0.89 (0.83-0.95)	1,144	0.92 (0.85-0.98)
Baseline	20,010	ref	18,343	Ref	1,667	ref	10,959	ref	9,051	ref
Men										
15-28 days	1,374	0.79 (0.74-0.84)	1,373	0.79 (0.74-0.84)	1	0.70 (0.08-5.97)	704	0.81 (0.74-0.88)	670	0.77 (0.71-0.84)
29-59 days	3,208	0.85 (0.81-0.89)	3,203	0.85 (0.81-0.89)	5	1.63 (0.48-5.56)	1,651	0.87 (0.82-0.93)	1,557	0.83 (0.78-0.88)
60-90 days	3,331	0.91 (0.87-0.95)	3,330	0.91 (0.87-0.95)	1	0.36 (0.04-3.07)	1,766	0.97 (0.91-1.02)	1,565	0.86 (0.81-0.91)
91-120 days	3,290	0.96 (0.92-1.00)	3,287	0.96 (0.92-1.00)	3	1.13 (0.28-4.63)	1,679	0.97 (0.92-1.03)	1,611	0.94 (0.89-1.00)
Baseline	24,698	ref	24,677	Ref	21	ref	12,708	ref	11,990	ref
75-84 years		p = 0.2316						p = 0.1814		p = 0.8404
Women										
15-28 days	1,637	0.78 (0.74-0.83)					1,086	0.77 (0.73-0.83)	551	0.80 (0.73-0.88)
29-59 days	3,871	0.86 (0.83-0.90)					2,601	0.87 (0.83-0.91)	1,270	0.86 (0.80-0.92)
60-90 days	3,889	0.91 (0.88-0.95)					2,594	0.91 (0.87-0.95)	1,295	0.92 (0.86-0.98)
91-120 days	3,778	0.96 (0.92-1.00)					2,491	0.95 (0.90-0.99)	1,287	0.98 (0.92-1.05)
Baseline	27,501	ref					18,414	ref	9,087	ref
Men										
15-28 days	1,528	0.83 (0.78-0.87)					924	0.84 (0.78-0.91)	604	0.80 (0.73-0.87)
29-59 days	3,342	0.85 (0.81-0.88)					1,969	0.84 (0.80-0.89)	1,373	0.85 (0.80-0.91)
60-90 days	3,398	0.90 (0.87-0.94)					2,049	0.92 (0.87-0.97)	1,349	0.88 (0.83-0.94)
91-120 days	3,317	0.95 (0.92-0.99)					1,994	0.97 (0.92-1.02)	1,323	0.94 (0.88-1.00)
Baseline	24,142	ref					14,407	ref	9,735	ref

QRISK2 score results are not included for those aged 75-84 years as all individuals were high risk. P-values in table are for sex interaction

Supplementary Table 2. Incidence ratios for first acute cardiovascular event in risk periods following influenza vaccination by cardiovascular risk, age group and vaccination timing

Risk period	All		QRISK2				Hypertension			
			Raised risk		Low risk		Raised risk		Low risk	
	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)
All ages		p <0.0001		p <0.0001		p <0.0001		p <0.0001		p = 0.0004
Vaccinated on or before 15 November										
15-28 days	5,291	0.73 (0.71-0.75)	4,833	0.76 (0.73-0.78)	458	0.57 (0.51-0.62)	2,990	0.76 (0.73-0.79)	2,301	0.70 (0.67-0.73)
29-59 days	12,046	0.77 (0.76-0.79)	11,038	0.80 (0.79-0.82)	1,008	0.58 (0.54-0.62)	6,830	0.81 (0.78-0.83)	5,216	0.73 (0.71-0.76)
60-90 days	12,303	0.82 (0.80-0.84)	11,288	0.86 (0.84-0.87)	1,015	0.61 (0.57-0.65)	7,023	0.86 (0.84-0.89)	5,280	0.77 (0.75-0.80)
91-120 days	12,098	0.87 (0.85-0.88)	11,113	0.90 (0.88-0.92)	985	0.63 (0.59-0.68)	6,834	0.90 (0.88-0.93)	5,264	0.83 (0.80-0.85)
Baseline	94,592	ref	84,764	Ref	9,828	ref	52,666	ref	41,926	ref
Vaccinated after 15 November										
15-28 days	1,793	0.69 (0.66-0.73)	1,594	0.80 (0.76-0.84)	199	0.33 (0.29-0.38)	958	0.81 (0.76-0.87)	835	0.59 (0.55-0.63)
29-59 days	3,987	0.71 (0.68-0.73)	3,529	0.82 (0.79-0.85)	458	0.35 (0.31-0.38)	2,076	0.82 (0.78-0.86)	1,911	0.62 (0.59-0.65)
60-90 days	3,906	0.72 (0.69-0.74)	3,490	0.84 (0.81-0.87)	416	0.32 (0.29-0.36)	2,066	0.84 (0.80-0.89)	1,840	0.61 (0.58-0.65)
91-120 days	3,800	0.76 (0.73-0.78)	3,352	0.87 (0.83-0.90)	448	0.39 (0.35-0.43)	2,005	0.88 (0.84-0.93)	1,795	0.65 (0.62-0.69)
Baseline	31,781	ref	26,527	Ref	5,254	ref	15,792	ref	15,989	ref
40-64 years		p <0.0001		p = 0.0053		p <0.0001		p = 0.0186		p <0.0001
Vaccinated on or before 15 November										
15-28 days	961	0.61 (0.57-0.66)	572	0.68 (0.62-0.74)	389	0.54 (0.49-0.60)	420	0.70 (0.63-0.78)	541	0.56 (0.51-0.62)
29-59 days	2,138	0.63 (0.60-0.66)	1,275	0.70 (0.66-0.74)	863	0.55 (0.51-0.60)	960	0.74 (0.68-0.79)	1,178	0.57 (0.53-0.60)
60-90 days	2,143	0.66 (0.63-0.69)	1,268	0.72 (0.67-0.76)	875	0.59 (0.55-0.64)	962	0.76 (0.71-0.82)	1,181	0.60 (0.56-0.64)
91-120 days	2,115	0.70 (0.66-0.73)	1,279	0.77 (0.73-0.82)	836	0.61 (0.57-0.66)	962	0.81 (0.75-0.87)	1,153	0.63 (0.59-0.67)
Baseline	19,828	ref	11,234	Ref	8,594	ref	8,201	ref	11,627	ref
Vaccinated after 15 November										
15-28 days	441	0.40 (0.36-0.44)	265	0.50 (0.44-0.57)	176	0.31 (0.27-0.36)	172	0.48 (0.41-0.57)	269	0.36 (0.32-0.41)
29-59 days	983	0.41 (0.38-0.44)	570	0.49 (0.45-0.54)	413	0.33 (0.30-0.37)	415	0.53 (0.48-0.60)	568	0.35 (0.32-0.38)
60-90 days	914	0.39 (0.36-0.42)	542	0.48 (0.43-0.53)	372	0.30 (0.27-0.34)	377	0.49 (0.44-0.55)	537	0.34 (0.31-0.37)
91-120 days	919	0.43 (0.40-0.47)	511	0.50 (0.45-0.55)	408	0.37 (0.34-0.42)	378	0.54 (0.49-0.61)	541	0.38 (0.35-0.42)
Baseline	10,194	Ref	5,394	Ref	4,800	ref	3,769	ref	6,425	ref

Risk period	All		QRISK2				Hypertension			
			Raised risk		Low risk		Raised risk		Low risk	
	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)
65-74 years		p = 0.1323		p = 0.0561		p = 0.5096		p = 0.0043		p = 0.2365
Vaccinated on or before 15 November										
15-28 days	1,884	0.79 (0.75-0.83)	1,815	0.79 (0.75-0.83)	69	0.84 (0.65-1.08)	1,003	0.79 (0.74-0.85)	881	0.79 (0.73-0.85)
29-59 days	4,297	0.83 (0.80-0.86)	4,152	0.83 (0.80-0.86)	145	0.81 (0.67-0.98)	2,309	0.84 (0.80-0.88)	1,988	0.82 (0.78-0.87)
60-90 days	4,419	0.88 (0.85-0.92)	4,279	0.89 (0.86-0.92)	140	0.80 (0.66-0.97)	2,360	0.89 (0.85-0.93)	2,059	0.88 (0.84-0.93)
91-120 days	4,398	0.94 (0.90-0.97)	4,249	0.94 (0.91-0.97)	149	0.89 (0.74-1.08)	2,313	0.93 (0.88-0.97)	2,085	0.95 (0.90-1.00)
Baseline	33,871	ref	32,637	Ref	1,234	ref	18,273	ref	15,598	ref
Vaccinated after 15 November										
15-28 days	633	0.84 (0.77-0.91)	610	0.85 (0.77-0.92)	23	0.69 (0.44-1.08)	343	0.94 (0.83-1.06)	290	0.75 (0.66-0.85)
29-59 days	1,402	0.86 (0.80-0.91)	1,357	0.87 (0.81-0.93)	45	0.62 (0.44-0.88)	652	0.82 (0.75-0.90)	750	0.89 (0.81-0.97)
60-90 days	1,446	0.91 (0.85-0.97)	1,402	0.92 (0.86-0.98)	44	0.62 (0.44-0.87)	747	0.97 (0.89-1.06)	699	0.85 (0.77-0.93)
91-120 days	1,371	0.92 (0.86-0.98)	1,331	0.93 (0.87-0.99)	40	0.60 (0.42-0.86)	701	0.97 (0.88-1.06)	670	0.87 (0.79-0.95)
Baseline	10,837	ref	10,383	Ref	454	ref	5,394	ref	5,443	ref
75-84 years		p <0.0001						p <0.0001		p = 0.0426
Vaccinated on or before 15 November										
15-28 days	2,446	0.76 (0.73-0.80)					1,567	0.77 (0.73-0.81)	879	0.76 (0.70-0.81)
29-59 days	5,611	0.82 (0.79-0.84)					3,561	0.81 (0.78-0.85)	2,050	0.83 (0.78-0.87)
60-90 days	5,741	0.88 (0.85-0.91)					3,701	0.89 (0.85-0.92)	2,040	0.86 (0.82-0.91)
91-120 days	5,585	0.93 (0.90-0.96)					3,559	0.93 (0.89-0.96)	2,026	0.93 (0.88-0.98)
Baseline	40,893	ref					26,192	ref	14,701	ref
Vaccinated after 15 November										
15-28 days	719	0.98 (0.90-1.06)					443	0.98 (0.88-1.09)	276	0.98 (0.86-1.12)
29-59 days	1,602	1.02 (0.96-1.09)					1,009	1.05 (0.97-1.13)	593	0.98 (0.89-1.09)
60-90 days	1,546	1.03 (0.97-1.10)					942	1.03 (0.95-1.11)	604	1.05 (0.95-1.16)
91-120 days	1,510	1.09 (1.03-1.16)					926	1.09 (1.01-1.18)	584	1.09 (0.99-1.21)
Baseline	10,750	ref					6,629	ref	4,121	ref

QRISK2 score results are not included for those aged 75-84 years as all individuals were high risk. P-values in table are for timing interaction

Supplementary Table 3. Incidence ratios for first acute cardiovascular event in risk periods following influenza vaccination by year

Risk period	2008/09		2009/10		2010/11		2011/12	
	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)
15-28 days	641	0.79 (0.72-0.86)	605	0.72 (0.66-0.78)	631	0.76 (0.69-0.82)	586	0.71 (0.65-0.78)
29-59 days	1,313	0.76 (0.71-0.81)	1,383	0.76 (0.71-0.81)	1,320	0.74 (0.69-0.79)	1,332	0.75 (0.70-0.80)
60-90 days	1,489	0.90 (0.84-0.95)	1,367	0.79 (0.74-0.84)	1,377	0.80 (0.75-0.86)	1,360	0.80 (0.75-0.85)
91-120 days	1,375	0.89 (0.84-0.95)	1,281	0.79 (0.75-0.85)	1,284	0.81 (0.76-0.86)	1,360	0.85 (0.80-0.91)
Baseline	10,095	ref	10,611	Ref	10,461	ref	10,675	ref

Risk period	2012/13		2013/14		2014/15		2015/16	
	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)
15-28 days	604	0.69 (0.64-0.76)	622	0.71 (0.65-0.78)	635	0.71 (0.66-0.78)	667	0.74 (0.68-0.81)
29-59 days	1,484	0.79 (0.75-0.84)	1,389	0.74 (0.69-0.78)	1,512	0.79 (0.74-0.84)	1,486	0.76 (0.72-0.81)
60-90 days	1,490	0.83 (0.78-0.88)	1,401	0.77 (0.73-0.82)	1,469	0.80 (0.75-0.85)	1,495	0.80 (0.75-0.85)
91-120 days	1,357	0.81 (0.76-0.86)	1,478	0.87 (0.82-0.93)	1,491	0.87 (0.82-0.92)	1,499	0.86 (0.81-0.91)
Baseline	11,052	ref	11,198	Ref	11,479	ref	11,762	ref

Risk period	2016/17		2017/18		2018/19	
	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)
15-28 days	628	0.66 (0.60-0.71)	752	0.71 (0.66-0.77)	713	0.73 (0.67-0.79)
29-59 days	1,563	0.76 (0.71-0.80)	1,652	0.72 (0.68-0.76)	1,599	0.75 (0.71-0.80)
60-90 days	1,524	0.76 (0.72-0.81)	1,667	0.76 (0.72-0.80)	1,570	0.76 (0.72-0.81)
91-120 days	1,481	0.79 (0.75-0.84)	1,663	0.81 (0.77-0.86)	1,629	0.85 (0.80-0.90)
Baseline	12,409	ref	13,545	Ref	13,086	ref

Supplementary Table 4. Incidence ratios for first acute cardiovascular event in the 14 days after influenza vaccination

Time period	All		QRISK2				Hypertension			
			Raised risk		Low risk		Raised risk		Low risk	
	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)
1-7 days	3,370	0.68 (0.66-0.71)	3,048	0.72 (0.69-0.75)	322	0.48 (0.43-0.53)	1,830	0.71 (0.68-0.74)	1,540	0.66 (0.62-0.69)
8-14 days	3,495	0.71 (0.68-0.73)	3,160	0.75 (0.72-0.78)	335	0.49 (0.44-0.55)	1,946	0.75 (0.72-0.79)	1,549	0.66 (0.62-0.69)

Supplementary Table 5. Incidence ratios for non-fatal first acute cardiovascular event in risk periods following influenza vaccination among individuals by cardiovascular risk and age group

Risk period	All		QRISK2				Hypertension			
			Raised risk		Low risk		Raised risk		Low risk	
	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)
All ages										
15-28 days	6,541	0.76 (0.74-0.78)	5,916	0.81 (0.79-0.84)	625	0.49 (0.45-0.53)	3,638	0.82 (0.79-0.85)	2,903	0.70 (0.67-0.72)
29-59 days	14,634	0.77 (0.76-0.79)	13,261	0.83 (0.81-0.85)	1,373	0.49 (0.46-0.52)	8,130	0.84 (0.81-0.86)	6,504	0.71 (0.69-0.73)
60-90 days	14,735	0.79 (0.77-0.80)	13,389	0.85 (0.83-0.87)	1,346	0.49 (0.46-0.52)	8,236	0.86 (0.84-0.88)	6,499	0.72 (0.70-0.74)
91-120 days	14,491	0.82 (0.80-0.83)	13,162	0.88 (0.86-0.89)	1,329	0.52 (0.49-0.55)	8,036	0.88 (0.86-0.90)	6,455	0.75 (0.73-0.78)
Baseline	118,508	ref	103,929	Ref	14,579	ref	63,880	ref	54,628	ref
40-64 years										
15-28 days	1,336	0.55 (0.52-0.58)	795	0.63 (0.58-0.68)	541	0.46 (0.42-0.50)	567	0.64 (0.59-0.70)	769	0.49 (0.46-0.53)
29-59 days	2,937	0.55 (0.53-0.57)	1,742	0.63 (0.59-0.66)	1,195	0.46 (0.44-0.49)	1,290	0.67 (0.63-0.71)	1,647	0.48 (0.46-0.51)
60-90 days	2,878	0.55 (0.53-0.57)	1,706	0.63 (0.59-0.66)	1,172	0.47 (0.44-0.50)	1,269	0.67 (0.63-0.71)	1,609	0.49 (0.46-0.51)
91-120 days	2,859	0.59 (0.56-0.61)	1,705	0.67 (0.63-0.70)	1,154	0.50 (0.47-0.53)	1,270	0.71 (0.66-0.75)	1,589	0.52 (0.49-0.55)
Baseline	29,067	ref	16,091	Ref	12,976	ref	11,571	ref	17,496	ref
65-74 years										
15-28 days	2,347	0.84 (0.80-0.88)	2,263	0.84 (0.80-0.88)	84	0.81 (0.64-1.02)	1,254	0.86 (0.81-0.91)	1,093	0.82 (0.77-0.87)
29-59 days	5,287	0.86 (0.83-0.88)	5,109	0.86 (0.83-0.89)	178	0.77 (0.65-0.92)	2,759	0.86 (0.82-0.89)	2,528	0.86 (0.82-0.90)
60-90 days	5,388	0.88 (0.85-0.91)	5,214	0.89 (0.86-0.91)	174	0.76 (0.64-0.90)	2,850	0.89 (0.85-0.93)	2,538	0.87 (0.83-0.91)
91-120 days	5,358	0.92 (0.89-0.95)	5,183	0.92 (0.89-0.95)	175	0.79 (0.67-0.94)	2,809	0.92 (0.88-0.96)	2,549	0.91 (0.87-0.95)
Baseline	42,262	ref	40,659	Ref	1,603	ref	22,386	ref	19,876	ref
75-84 years										
15-28 days	2,858	0.88 (0.84-0.92)					1,817	0.89 (0.84-0.93)	1,041	0.87 (0.81-0.93)
29-59 days	6,410	0.90 (0.87-0.92)					4,081	0.91 (0.87-0.94)	2,329	0.88 (0.84-0.93)
60-90 days	6,469	0.92 (0.89-0.95)					4,117	0.93 (0.89-0.96)	2,352	0.90 (0.86-0.95)
91-120 days	6,274	0.94 (0.91-0.97)					3,957	0.94 (0.90-0.98)	2,317	0.94 (0.89-0.98)
Baseline	47,179	ref					29,923	ref	17,256	ref

QRISK2 score results are not included for those aged 75-84 years as all individuals were high risk.

Supplementary Table 6. Incidence ratios for first acute cardiovascular event in risk periods following influenza vaccination by further QRISK2 stratification

	QRISK2					
	≥20%		10-19%		<10%	
Risk period	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)
All ages						
15-28 days	4,756	0.79 (0.77-0.82)	1,671	0.71 (0.67-0.75)	657	0.83 (0.67-1.02)
29-59 days	10,912	0.85 (0.83-0.87)	3,655	0.72 (0.69-0.74)	1,466	0.78 (0.67-0.92)
60-90 days	11,107	0.90 (0.88-0.92)	3,671	0.74 (0.71-0.77)	1,431	0.86 (0.73-1.00)
91-120 days	10,875	0.95 (0.93-0.97)	3,590	0.77 (0.75-0.80)	1,433	0.87 (0.75-1.03)
Baseline	80,450	ref	30,841	Ref	15,082	ref
40-64						
15-28 days	310	0.76 (0.67-0.86)	527	0.57 (0.52-0.62)	565	0.45 (0.41-0.49)
29-59 days	690	0.78 (0.71-0.85)	1,155	0.57 (0.53-0.61)	1,276	0.47 (0.44-0.50)
60-90 days	679	0.79 (0.72-0.86)	1,131	0.58 (0.54-0.62)	1,247	0.48 (0.45-0.51)
91-120 days	676	0.83 (0.76-0.91)	1,114	0.62 (0.58-0.66)	1,244	0.52 (0.49-0.56)
Baseline	5,615	ref	11,013	Ref	13,394	ref
65-74						
15-28 days	1,387	0.79 (0.74-0.83)	1,038	0.82 (0.77-0.88)	92	0.80 (0.64-1.00)
29-59 days	3,225	0.85 (0.81-0.88)	2,284	0.83 (0.79-0.88)	190	0.76 (0.64-0.90)
60-90 days	3,368	0.91 (0.88-0.95)	2,313	0.87 (0.83-0.91)	184	0.75 (0.64-0.89)
91-120 days	3,320	0.96 (0.92-1.00)	2,260	0.90 (0.86-0.95)	189	0.82 (0.69-0.96)
Baseline	24,915	ref	18,105	Ref	1,688	ref
75-84						
15-28 days	3,059	0.80 (0.77-0.83)	106	0.83 (0.67-1.02)		
29-59 days	6,997	0.86 (0.83-0.88)	216	0.78 (0.67-0.92)		
60-90 days	7,060	0.91 (0.88-0.94)	227	0.86 (0.73-1.00)		
91-120 days	6,879	0.96 (0.93-0.99)	216	0.87 (0.75-1.03)		
Baseline	49,920	ref	1,723	Ref		

QRISK2 score <10% results are not included for those aged 75-84 years as all individuals had a score ≥10%

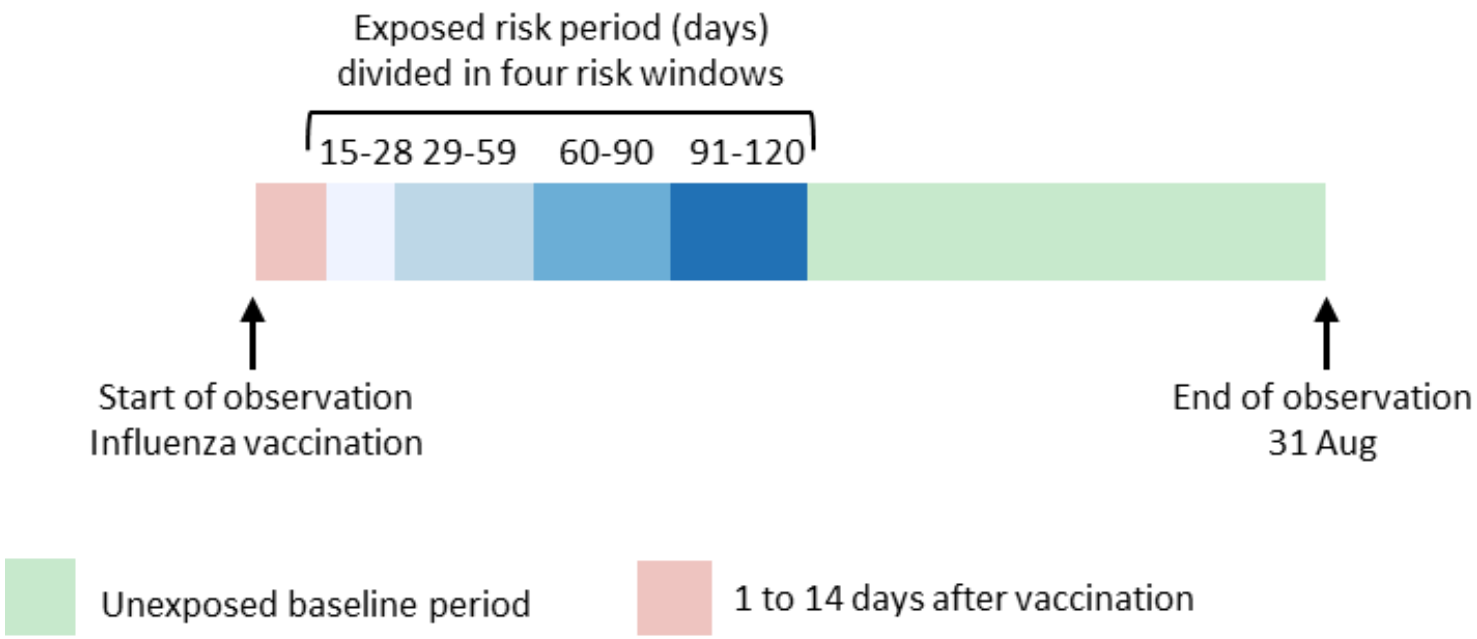
Supplementary Table 7. Baseline characteristics of sensitivity analysis study design population

	All	QRISK2		Hypertension	
		Raised risk	Low risk	Raised risk	Low risk
	n=160,906	n=147,023	n=13,883	n=90,992	n=69,914
Sex					
Female (%)	77,467 (48.1%)	67,783 (46.1%)	9,684 (69.8%)	45,747 (50.3%)	31,720 (45.4%)
Age group (years)					
40-64	29,927 (18.6%)	18,127 (12.3%)	11,800 (85.0%)	13,523 (14.9%)	16,404 (23.5%)
65-74	59,168 (36.8%)	57,085 (38.8%)	2,083 (15.0%)	31,631 (34.8%)	27,537 (39.4%)
75-84	71,811 (44.6%)	71,811 (48.8%)	0 (0.0%)	45,838 (50.4%)	25,973 (37.1%)
Associated hospital stay					
Yes	112,994 (70.2%)	103,455 (70.4%)	9,539 (68.7%)	64,192 (70.5%)	48,802 (69.8%)
Median (IQR) stay	4.0 (2.0-11.0)	4.0 (2.0-11.0)	3.0 (1.0-8.0)	4.0 (2.0-11.0)	4.0 (2.0-10.0)
Died ≤30 days after event					
Yes	13,098 (8.1%)	12,260 (8.3%)	838 (6.0%)	7,584 (8.3%)	5,514 (7.9%)
Died in study period					
Yes	18,644 (11.6%)	17,525 (11.9%)	1,119 (8.1%)	10,853 (11.9%)	7,791 (11.1%)

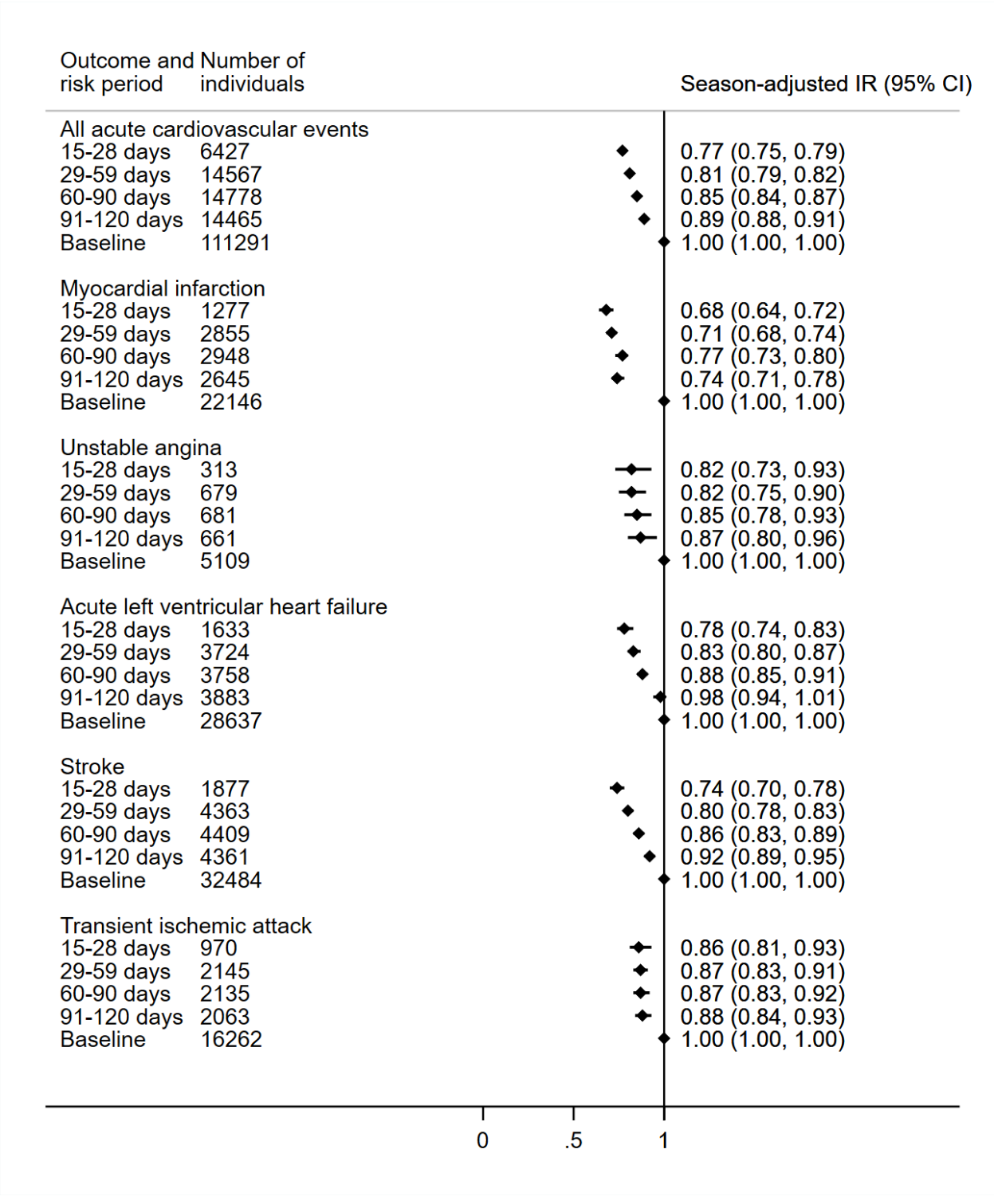
Supplementary Table 8. Sensitivity analysis study design incidence ratios for first acute cardiovascular event in risk periods following influenza vaccination by cardiovascular risk and age group

	All		QRISK2				Hypertension			
			Raised risk		Low risk		Raised risk		Low risk	
Risk period	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)	N events	IR (95% CI)
All ages										
15-28 days	6,923	0.94 (0.91-0.96)	6,313	0.94 (0.91-0.96)	610	0.95 (0.86-1.04)	3,885	0.93 (0.90-0.97)	3,038	0.95 (0.91-0.99)
29-59 days	15,752	0.96 (0.95-0.98)	14,356	0.96 (0.94-0.98)	1,396	0.98 (0.91-1.05)	8,797	0.95 (0.93-0.98)	6,955	0.98 (0.95-1.01)
60-90 days	15,999	0.98 (0.96-1.00)	14,637	0.98 (0.96-1.00)	1,362	0.96 (0.89-1.02)	9,032	0.98 (0.95-1.00)	6,967	0.98 (0.95-1.01)
91-120 days	15,740	1.00 (0.98-1.02)	14,363	1.00 (0.98-1.02)	1,377	1.00 (0.93-1.07)	8,819	0.99 (0.96-1.01)	6,921	1.01 (0.98-1.04)
Baseline	99,885	ref	91,346	ref	8,539	ref	56,780	ref	43,105	ref
40-64 years										
15-28 days	1,353	1.01 (0.95-1.08)	827	1.04 (0.96-1.13)	526	0.97 (0.88-1.07)	585	0.98 (0.90-1.08)	768	1.03 (0.95-1.12)
29-59 days	3,036	1.03 (0.98-1.08)	1,821	1.03 (0.97-1.10)	1,215	1.01 (0.94-1.09)	1,357	1.03 (0.96-1.11)	1,679	1.02 (0.96-1.09)
60-90 days	2,989	1.01 (0.96-1.06)	1,804	1.02 (0.96-1.09)	1,185	0.99 (0.92-1.06)	1,327	1.01 (0.94-1.08)	1,662	1.01 (0.95-1.08)
91-120 days	2,986	1.04 (0.99-1.09)	1,791	1.05 (0.99-1.11)	1,195	1.03 (0.96-1.11)	1,336	1.04 (0.97-1.12)	1,650	1.04 (0.97-1.10)
Baseline	18,211	ref	11,055	ref	7,156	ref	8,300	ref	9,911	ref
65-74 years										
15-28 days	2,470	0.91 (0.87-0.95)	2,386	0.92 (0.87-0.96)	84	0.83 (0.65-1.06)	1,328	0.92 (0.86-0.98)	1,142	0.91 (0.85-0.97)
29-59 days	5,605	0.93 (0.90-0.97)	5,424	0.94 (0.91-0.97)	181	0.81 (0.67-0.97)	2,934	0.92 (0.87-0.96)	2,671	0.96 (0.91-1.01)
60-90 days	5,789	0.97 (0.93-1.00)	5,612	0.97 (0.94-1.01)	177	0.79 (0.66-0.95)	3,088	0.96 (0.92-1.01)	2,701	0.97 (0.92-1.02)
91-120 days	5,715	0.99 (0.95-1.02)	5,533	0.99 (0.96-1.02)	182	0.84 (0.70-1.01)	3,014	0.97 (0.93-1.02)	2,701	1.00 (0.95-1.05)
Baseline	37,202	ref	35,819	ref	1,383	ref	20,007	ref	17,195	ref
75-84 years										
15-28 days	3,100	0.93 (0.90-0.97)					1,972	0.93 (0.88-0.98)	1,128	0.94 (0.88-1.01)
29-59 days	7,111	0.97 (0.94-1.00)					4,506	0.96 (0.92-0.99)	2,605	0.98 (0.93-1.03)
60-90 days	7,221	0.98 (0.95-1.01)					4,617	0.98 (0.95-1.02)	2,604	0.98 (0.93-1.03)
91-120 days	7,039	0.99 (0.96-1.02)					4,469	0.98 (0.95-1.02)	2,570	1.00 (0.95-1.05)
Baseline	44,472	ref					28,473	ref	15,999	ref

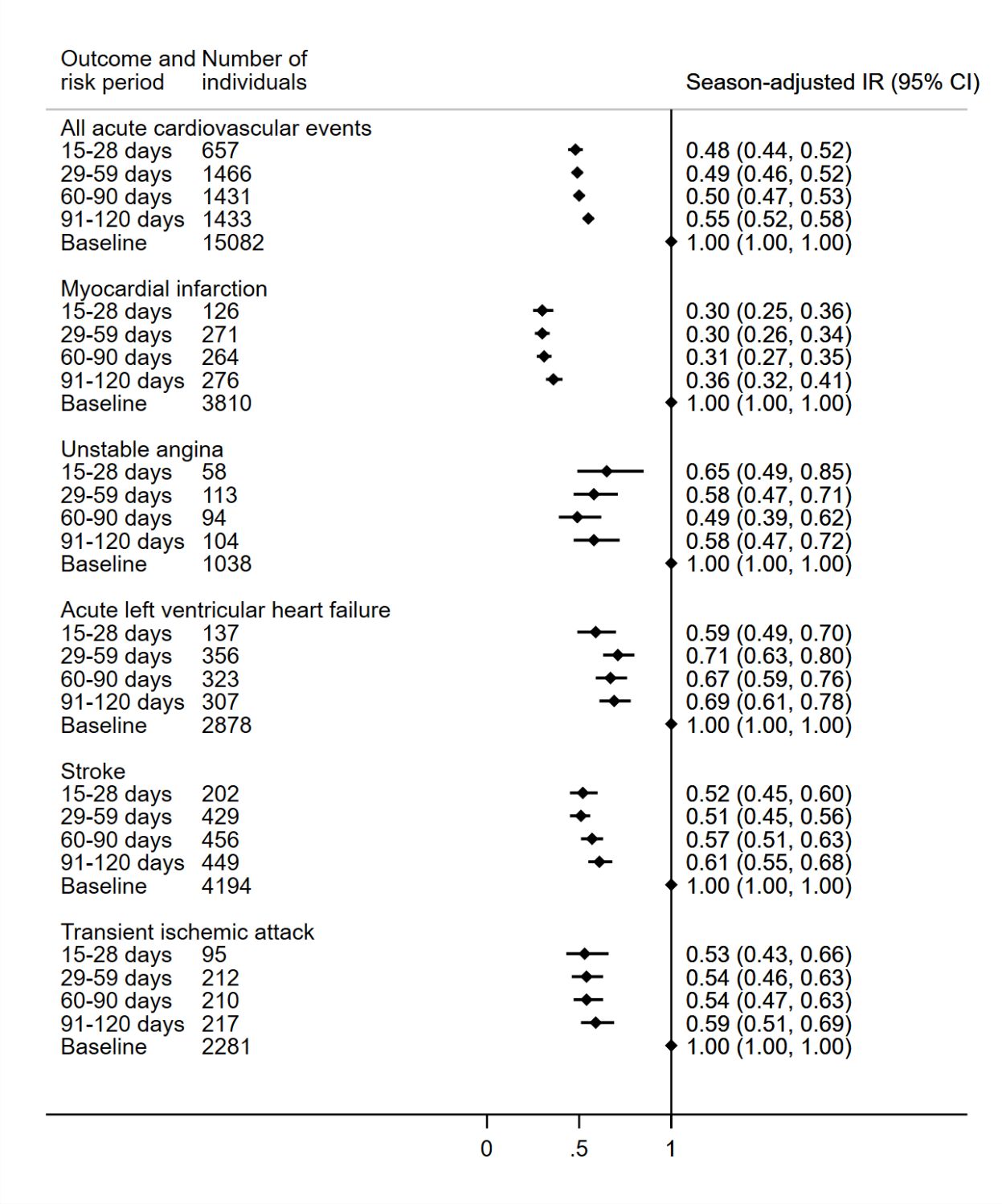
QRISK2 score results are not included for those aged 75-84 years as all individuals were high risk.



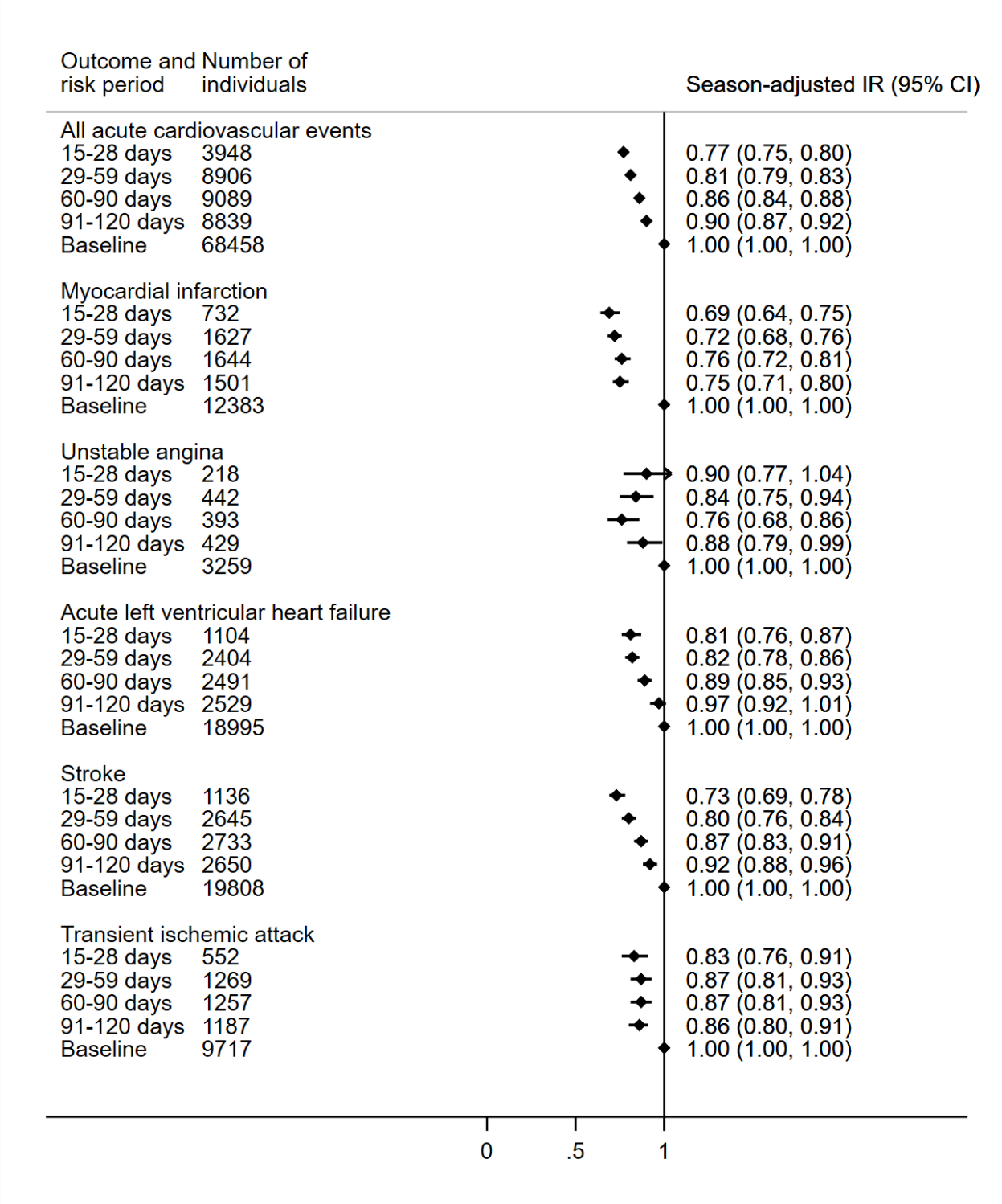
Supplementary Figure 1. Overview of sensitivity analysis study design



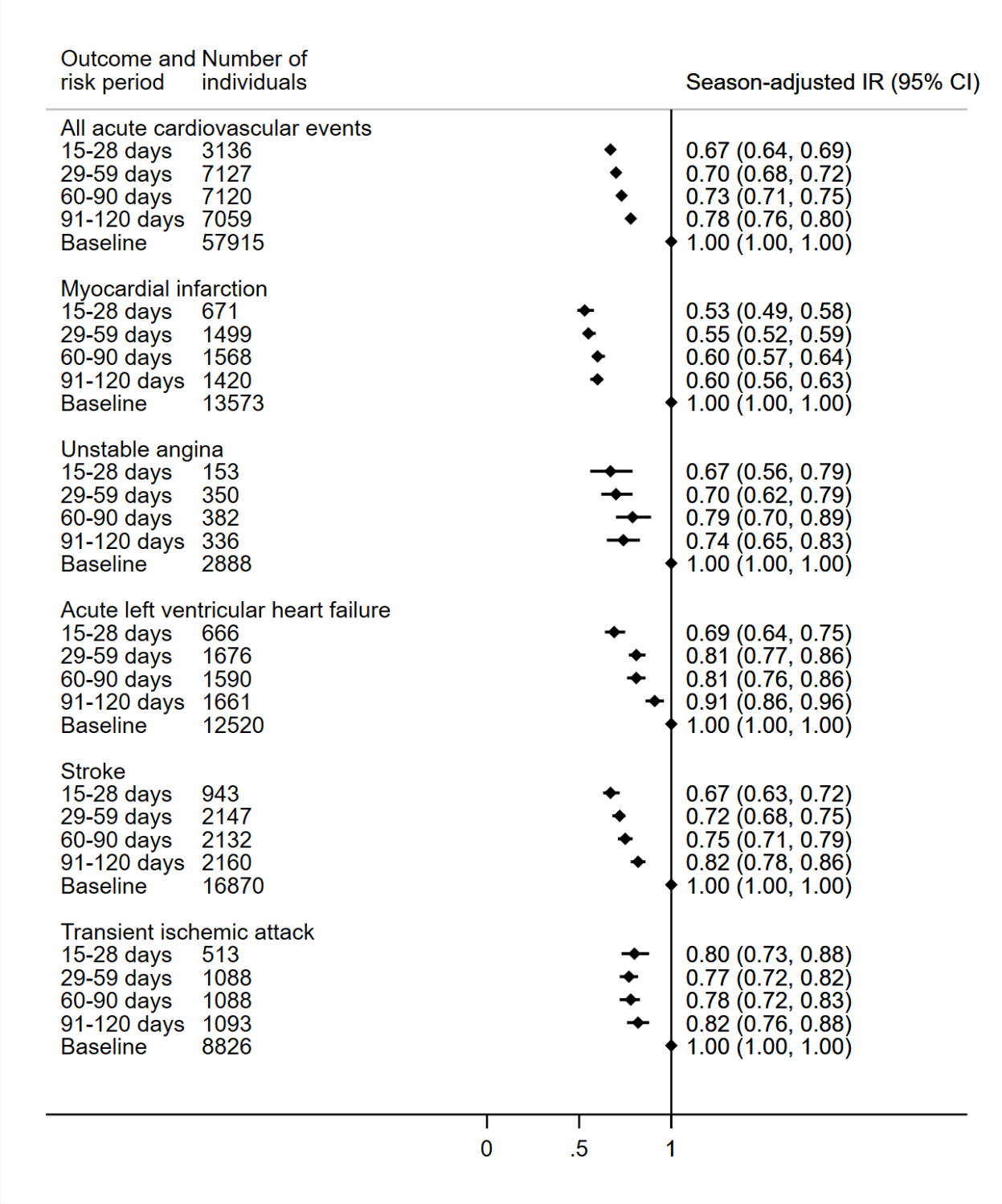
Supplementary Figure 2. Incidence ratios for first acute cardiovascular events in risk periods following influenza vaccination among individuals with a QRISK2 score ≥10% by cardiovascular event type



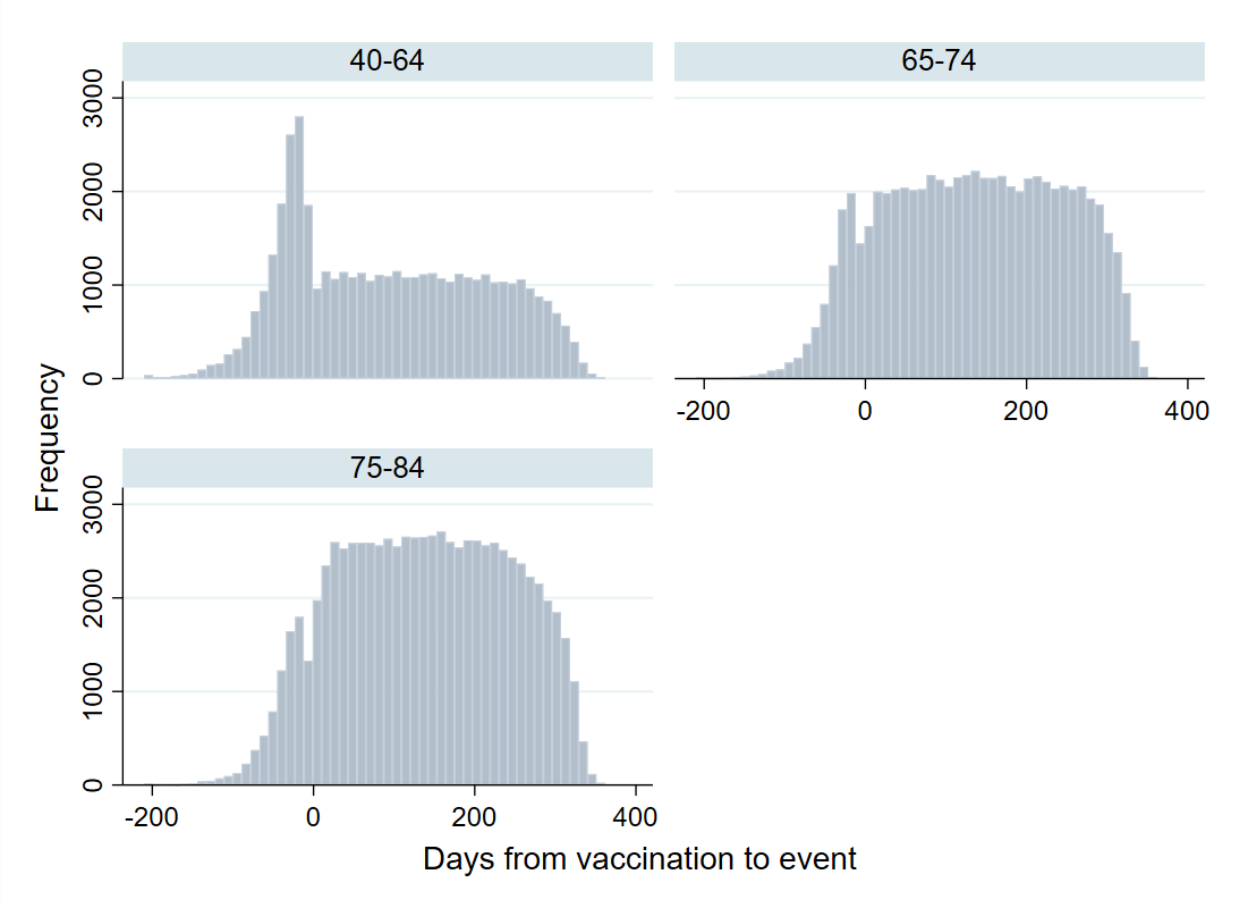
Supplementary Figure 3. Incidence ratios for first acute cardiovascular events in risk periods following influenza vaccination among individuals with a QRISK2 score <10% by cardiovascular event type



Supplementary Figure 4. Incidence ratios for first acute cardiovascular events in risk periods following influenza vaccination among individuals with diagnosed hypertension by cardiovascular event type



Supplementary Figure 5. Incidence ratios for first acute cardiovascular events in risk periods following influenza vaccination among individuals without diagnosed hypertension by cardiovascular event type



Supplementary Figure 6. Difference in time (days) between vaccination and acute cardiovascular event by age group

Appendix 6 Chapter 8 and 9 protocol and ethics approval

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
United Kingdom
Switchboard: +44 (0)20 7636 8636

www.lshtm.ac.uk

Observational / Interventions Research Ethics Committee

Dr. Charlotte Warren-Gash
LSHTM

7 January 2021

Dear Dr. Charlotte Warren-Gash

Study Title: Investigating the effect of cardiovascular risk level on outcomes of COVID-19

LSHTM Ethics Ref: 22717

Thank you for your application for the above research project which has now been considered by the Observational Committee via Chair's Action.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation, subject to the conditions specified below.

Conditions of the favourable opinion

Approval is dependent on local ethical approval having been received, where relevant.

Approved documents

The final list of documents reviewed and approved is as follows:

Document Type	File Name	Date	Version
Other	Research ethics committee online training_26-01-2018	26/01/2018	1
Investigator CV	Herrett_CV_2019	01/12/2019	1
Investigator CV	Liam Smeeth 2 page CV 2019	01/12/2019	1
Investigator CV	LSHTM CV Template v7 (Feb 2020) - Miss Helen Strongman - generated 2020-03-03 15-12	03/03/2020	1
Investigator CV	Full_cv_Charlotte Warren-Gash_Oct2020	01/10/2020	1
Protocol / Proposal	ISAC Protocol Application Form_CVDriskCOVID_v2.0_FINAL	06/11/2020	2.0
Investigator CV	CV DEC 2020 JD	01/12/2020	1
Local Approval	Print screen of ISAC aproval_18-12-2020	18/12/2020	1
Covering Letter	Cover Letter_05-01-2020	05/01/2021	1

After ethical review

The Chief Investigator (CI) or delegate is responsible for informing the ethics committee of any subsequent changes to the application. These must be submitted to the committee for review using an Amendment form. Amendments must not be initiated before receipt of written favourable opinion from the committee.

The CI or delegate is also required to notify the ethics committee of any protocol violations and/or Suspected Unexpected Serious Adverse Reactions (SUSARs) which occur during the project by submitting a Serious Adverse Event form.

An annual report should be submitted to the committee using an Annual Report form on the anniversary of the approval of the study during the lifetime of the study.

At the end of the study, the CI or delegate must notify the committee using the End of Study form.

All aforementioned forms are available on the ethics online applications website and can only be submitted to the committee via the website at: http://leo.lshtm.ac.uk.

Further information is available at: www.lshtm.ac.uk/ethics.

Yours sincerely,

Medicines & Healthcare products Regulatory Agency

CPRD

INDEPENDENT SCIENTIFIC ADVISORY COMMITTEE (ISAC) PROTOCOL APPLICATION FORM

PART 1: APPLICATION FORM

IMPORTANT

Both parts of this application must be completed in accordance with the guidance note 'Completion of the ISAC Protocol Application Form', which can be found on the CPRD website (https://cprd.com/research-applications).

FOR ISAC USE ONLY	
Protocol No. -	**Submission date -**

GENERAL INFORMATION ABOUT THE PROPOSED RESEARCH STUDY

1. Study Title (Max. 255 characters including spaces)

Investigating the effect of cardiovascular risk level on severe outcomes of COVID-19

2. Research Area (place '**X**' in all boxes that apply)

Drug Safety		Economics	
Drug Utilisation		Pharmacoeconomics	
Drug Effectiveness		Pharmacoepidemiology	
Disease Epidemiology	X	Methodological	
Health Services Delivery			

3. Chief Investigator

Title:	Dr
Full name:	Charlotte Warren-Gash
Job title:	Associate Professor of Epidemiology
Affiliation/organisation:	London School of Hygiene & Tropical Medicine
Email address:	charlotte.warren-gash1@lshtm.ac.uk
CV Number (if applicable):	815_16
Will this person be analysing the data? (Y/N)	N

4. Corresponding Applicant

Title:	Dr	
Full name:	Charlotte Warren-Gash	
Job title:	Associate Professor of Epidemiology	
Affiliation/organisation:	London School of Hygiene & Tropical Medicine	
Email address:	charlotte.warren-gash1@lshtm.ac.uk	
CV Number (if applicable):	815_16	
Will this person be analysing the data? (Y/N)	N	

5. List of all investigators/collaborators

Title:	Dr
Full name:	Emily Herrett
Job title:	Assistant Professor
Affiliation/organisation:	London School of Hygiene & Tropical Medicine
Email address:	emily.herrett@lshtm.ac.uk
CV Number (if applicable):	085_15
Will this person be analysing the data? (Y/N)	Y

Title:	Dr
Full name:	Helen Strongman
Job title:	Assistant Professor
Affiliation/organisation:	London School of Hygiene & Tropical Medicine
Email address:	helen.strongman@lshtm.ac.uk
CV Number (if applicable):	
Will this person be analysing the data? (Y/N)	

Title:	Dr
Full name:	Harriet Forbes
Job title:	Assistant Professor
Affiliation/organisation:	London School of Hygiene & Tropical Medicine
Email address:	harriet.forbes@lshtm.ac.uk
CV Number (if applicable):	465_15
Will this person be analysing the data? (Y/N)	

Title:	Ms.
Full name:	Jennifer Davidson
Job title:	PhD student
Affiliation/organisation:	London School of Hygiene & Tropical Medicine
Email address:	jennifer.davidson@lshtm.ac.uk
CV Number (if applicable):	
Will this person be analysing the data? (Y/N)	

Title:	Professor
Full name:	Liam Smeeth
Job title:	Professor of Clinical Epidemiology
Affiliation/organisation:	London School of Hygiene & Tropical Medicine
Email address:	liam.smeeth@lshtm.ac.uk
CV Number (if applicable):	045_15CEPSL
Will this person be analysing the data? (Y/N)	N

Title:	Dr
Full name:	Amitava Banerjee
Job title:	Associate Professor in Clinical Data Science and Honorary Consultant in Cardiology
Affiliation/organisation:	Institute of Health Informatics, Faculty of Population Health Sciences, University College London
Email address:	ami.banerjee@ucl.ac.uk
CV Number (if applicable):	090_16
Will this person be analysing the data? (Y/N)	N

Title:	Professor
Full name:	Judy Breuer

Job title:	Professor of Virology
Affiliation/organisation:	Institute of Infection & Immunity, University College London
Email address:	j.breuer@ucl.ac.uk
CV Number (if applicable):	277_18
Will this person be analysing the data? (Y/N)	N

[Add more investigators/collaborators as necessary by copy and pasting a new table for each investigator/collaborator]

6. Experience/expertise available

List below the member(s) of the research team who have experience with CPRD data.

Name(s):
Charlotte Warren-Gash
Emily Herrett
Helen Strongman
Harriet Forbes
Amitava Banerjee
Liam Smeeth

List below the member(s) of the research team who have statistical expertise.

Name(s):
Emily?
Helen?
Harriet?

List below the member(s) of the research team who have experience of handling large datasets (greater than 1 million records).

Name(s):
Emily Herrett
Helen Strongman
Harriet Forbes

List below the member(s) of the research team, or supporting the research team, who have experience of practicing in UK primary care.

Name(s):
Liam Smeeth

ACCESS TO THE DATA

7. Sponsor of the study

Institution/Organisation:	London School of Hygiene and Tropical Medicine
Address:	Keppel Street, London, WC1E 7HT

8. Funding source for the study

Same as Sponsor?	Yes		No	X
Institution/Organisation:	Rosetrees Trust			
Address:	Russell House, Middlesex, 140 High St, London, Edgware HA8 7LW			

Medicines & Healthcare products Regulatory Agency

CPRD

9. Institution conducting the research

Same as Sponsor?	Yes	X	No	
Institution/Organisation:				
Address:				

10. Data Access Arrangements

Indicate with an '**X**' the method that will be used to access the data for this study:

Study-specific Dataset Agreement	

Institutional Multi-study Licence	X
Institution Name	
Institution Address	

Will the dataset be extracted by CPRD?

Yes		No	X

If yes, provide the reference number:

11. Data Processor(s):

Processing	x
Accessing	x
Storing	x
Processing area (UK/EEA/Worldwide)	UK
Organisation name	London School of Hygiene and Tropical Medicine
Organisation address	Keppel Street, London, WC1E 7HT

Processing	x
Accessing	x
Storing	x
Processing area (UK/EEA/Worldwide)	UK
Organisation name	London School of Hygiene and Tropical Medicine
Organisation address	Keppel Street, London, WC1E 7HT

[Add more processors as necessary by copy and pasting a new table for each processor]

INFORMATION ON DATA

12. Primary care data (place '**X**' in all boxes that apply)

CPRD GOLD		CPRD Aurum	x

Reference number (if applicable):

13. Please select any linked data or data products being requested

Patient Level Data (place '**X**' in all boxes that apply)

ONS Death Registration Data	x
HES Admitted Patient Care	x

HES Outpatient			
HES Accident and Emergency		NCRAS Cancer Registration Data	
HES Diagnostic Imaging Dataset		NCRAS Cancer Patient Experience Survey (CPES) data	
HES PROMS (Patient Reported Outcomes Measure)		NCRAS Systemic Anti-Cancer Treatment (SACT) data	
CPRD Mother Baby Link		NCRAS National Radiotherapy Dataset (RTDS) data	
Pregnancy Register		NCRAS Quality of Life Cancer Survivors Pilot (QOLP)	
Mental Health Data Set (MHDS)		NCRAS Quality of Life Colorectal Cancer Survivors (QOLC)	

Area Level Data (place '**X**' in <u>one</u> Practice / Patient level box that may apply)

Practice level (UK)		**Patient level (England only)**	
Practice Level Index of Multiple Deprivation		Patient Level Index of Multiple Deprivation	
Practice Level Index of Multiple Deprivation (index other than the most recent)		Patient Level Index of Multiple Deprivation Domains	
Practice Level Index of Multiple Deprivation Domains		Patient Level Carstairs Index for 2011 Census	
Practice Level Carstairs Index for 2011 Census (Excluding Northern Ireland)		Patient Level Townsend Score	X
2011 Rural-Urban Classification at LSOA level		2011 Rural-Urban Classification at LSOA level	

Reference / Protocol number (where applicable):

14. Are you requesting linkage to a dataset not listed above?

Yes	**x**	No	

If yes, provide the Non-Standard Linkage reference number: 00062241

COVID-19 linkages to SGSS and CHESS.

15. Does any person named in this application already have access to any of these data in a patient identifiable form, or associated with an identifiable patient index?

Yes		No	**x**

If yes, provide further details:

VALIDATION/VERIFICATION

16. Does this protocol describe an observational study using purely CPRD data?

Yes	**X**	No	

17. Does this protocol involve requesting any additional information from GPs, or contact with patients?

Yes		No	**X**

If yes, provide the reference number:

Medicines & Healthcare products Regulatory Agency

CPRD

PART 2: PROTOCOL INFORMATION

Applicants must complete all sections listed below Applications with sections marked 'Not applicable' without justification will be returned as invalid
A. Study Title (Max. 255 characters, including spaces) Investigating the effect of cardiovascular risk level on severe outcomes of COVID-19
B. Lay Summary (Max. 250 words) The COVID-19 pandemic has led to more than one million deaths worldwide. Many deaths from COVID-19 infection have occurred in people with underlying health conditions such as heart disease. Damage to the heart has also been seen among people who have died from COVID-19. We do not yet know exactly which groups of people are most at risk of developing heart problems or dying after COVID-19. However, it is likely that people whose cardiovascular risk level is already raised, e.g. because they have high blood pressure, may be at higher risk. Currently though, these people are not included in any lists of vulnerable groups and so will not be prioritised to receive a future COVID-19 vaccine. In this study, we aim to investigate whether adults aged 40 years or over who are at raised cardiovascular risk because of factors such as high blood pressure have a greater chance of dying or developing heart problems after COVID-19 infection than those with low cardiovascular risk levels. We will use anonymous data from GP and hospital records and laboratories to compare rates of heart problems and deaths after COVID-19 infection, controlling for factors such as other underlying health conditions. We will carry out two research studies with different designs to improve confidence in our results. This will help policymakers to understand which groups should be prioritised for COVID-19 vaccine.
C. Technical Summary (Max. 300 words) The COVID-19 pandemic has led to more than one million deaths worldwide. Hospitalisations and deaths are common among patients with cardiovascular disease or risk factors such as hypertension, while myocardial injury is frequently observed among severe COVID-19 cases and is strongly associated with mortality. A wide range of other respiratory viruses and bacteria can trigger acute cardiovascular events among vulnerable groups. Nevertheless, people at raised cardiovascular risk but without existing cardiovascular disease have not been identified as a risk group for severe COVID-19 by Public Health England. Information is urgently needed on the risk of adverse outcomes after infection for those at raised cardiovascular risk to ensure that policymakers have the best evidence to target a future COVID-19 vaccine appropriately. We aim to investigate the effect of having a raised cardiovascular risk level, defined initially by QRISK3 score and then by hypertension status, or by the presence of existing cardiovascular disease, on severe outcomes of COVID-19. We will use linked data from general practices, hospital admissions and national laboratory and hospital surveillance for COVID-19 to compare mortality rates and the rates of major adverse cardiovascular events after laboratory-confirmed infection among those aged 40 years or over with differing levels of cardiovascular risk. After describing incidence rates of the outcomes by cardiovascular risk group, we will carry out a cohort study using multivariable Cox proportional hazards regression to calculate hazard ratios for each outcome after COVID-19 infection in the different risk groups. We will then conduct a complementary self-controlled case series study, which implicitly controls for the effect of fixed confounding factors, by level of cardiovascular risk to triangulate results. We will finally carry out a series of sensitivity analyses including exploring the effect of clinical and suspected COVID-19 infection among those without laboratory test results to check the robustness of our models.

D. Outcomes to be Measured

Cohort study:
Primary outcome: Death attributable to COVID-19 (using ONS definitions i.e. death within 28 days of a positive test)

Secondary outcomes:

- Hospitalisation due to COVID-19 (defined by COVID-19 in the primary diagnosis field of any episode recorded in HES)
- ICU admission due to COVID-19 (defined by laboratory-confirmed SARS-CoV-2 during or prior to ICU admission recorded in CHESS)
- Need for respiratory support due to COVID-19 (oxygen by mask or prongs, NIV, mechanical ventilation, ECMO defined in CHESS)
- Major adverse cardiovascular event (composite of acute coronary syndrome i.e. myocardial infarction and unstable angina, ischaemic stroke, acute left ventricular failure, major ventricular arrhythmia)
- Acute coronary syndrome (subdivided into myocardial infarction and unstable angina)
- Ischaemic stroke
- Acute left ventricular failure
- Major ventricular arrhythmia

Self-controlled case series study:
Primary outcome: major adverse cardiovascular event (composite of acute coronary syndrome i.e. myocardial infarction and unstable angina, ischaemic stroke, acute left ventricular failure, major ventricular arrhythmia)

Secondary outcomes:

- Acute coronary syndrome (subdivided into myocardial infarction and unstable angina)
- Ischaemic stroke
- Acute left ventricular failure
- Major ventricular arrhythmia

D. Objectives, Specific Aims and Rationale

The overall aim is to investigate the effect of having a raised cardiovascular risk level on mortality and severe outcomes after laboratory-confirmed SARS-CoV-2 infection in a UK population. This will help to inform the definition of target groups for a future vaccine.

Specific objectives:
Among individuals aged ≥40 years, with differing levels of underlying cardiovascular risk defined by (i) QRISK3 score, (ii) hypertension status, or (iii) presence of existing CVD (objective 1 only)

1. To describe the incidence of laboratory-confirmed SARS-CoV-2, and the incidence of mortality and severe outcomes including major adverse cardiovascular events after laboratory-confirmed SARS-CoV-2.
2. To investigate the effect of raised cardiovascular risk on mortality and severe outcomes including major adverse cardiovascular events after laboratory-confirmed SARS-CoV-2 using a cohort study design.
3. To quantify the association between laboratory-confirmed SARS-CoV-2 and risk of major adverse cardiovascular events using a self-controlled case series design.

Rationale: The COVID-19 pandemic has caused severe cardiorespiratory illness among individuals with a range of underlying comorbidities. While early hospital-based studies suggest that both established cardiovascular disease and the presence of cardiovascular risk factors are associated with severe outcomes after infection with SARS-CoV-2 infection, this has not been comprehensively investigated in large population-based cohort studies. The original list of vulnerable groups identified by Public Health England, which was based upon those recommended to receive seasonal influenza vaccination, did not include individuals with some cardiovascular risk factors such as hypertension. It is likely that any emerging vaccine against SARS-CoV-2 will be targeted towards groups included on the vulnerable list, which therefore needs to be as accurate and comprehensive as possible.

E. Study Background

Acute respiratory infections and cardiovascular disease

Acute respiratory infections (ARIs) cause a transient increase in the risk of acute cardiovascular events including myocardial infarction (MI) and stroke. At a population level, using time series analysis, we showed that circulation of various respiratory viruses is associated with hospitalisation for and deaths from cardiovascular outcomes, after controlling for temporal and environmental factors such as temperature and humidity(1,2). In individual level self-controlled case series studies, cardiovascular complications were seen after GP-diagnosed systemic respiratory infections(3) and influenza-like illnesses(4); in the 3 days following presentation with a clinically defined ARI, MI risk was elevated four-fold above baseline (incidence ratio 4.19, 95% CI 3.18-5.53), with elevated CVD risk persisting for around one month. We have since shown similar, even more marked, triggering effects for a range of laboratory-confirmed respiratory virus infections and *Streptococcus pneumoniae* on acute vascular outcomes such as MI and stroke in different populations and settings using large nationwide surveillance datasets linked to EHRs(5,6). While existing CVD has long been an indication for influenza vaccination, the effect of having a raised cardiovascular risk level on severe outcomes after ARI is not yet known. People at raised cardiovascular risk are therefore not specifically targeted for influenza or pneumococcal vaccines, and this is a subject of ongoing research by our group (ISAC protocol 19_209).

The COVID-19 pandemic and cardiovascular disease

By 2nd November 2020, more than 46 million cases of confirmed SARS-CoV-2 infection had been recorded globally with over 1.2 million deaths(7). Early reports and case series from the original epicentre in Wuhan showed high levels of comorbidities, including diabetes, hypertension and cardiovascular disease among patients hospitalised with COVID-19(8). Confirmed cases with cardiovascular disease or risk factors had a higher case fatality rate (10.5% for those with CVD and 6.0% for hypertension compared to 2.3% overall(9)). Biomarkers of cardiac injury such as troponin T levels were also higher among COVID-19 patients with vascular comorbidities and risk factors(10), and were associated with a range of severe outcomes including acute respiratory distress syndrome (ARDS), ventricular arrhythmias, coagulopathy, the need for mechanical ventilation and death(10,11). Over half of those who died in hospital had an elevated troponin T(10), suggesting that myocardial ischaemia and injury is an important mechanism leading to adverse outcomes of COVID-19. There are unanswered questions about the absolute incidence of short and longer-term cardiovascular complications following COVID-19(12), and the role of background levels of cardiovascular risk on severe outcomes(13).

Mechanisms

Several mechanisms are proposed to explain the observed association between ARIs and acute cardiovascular events. First, infections can lead to systemic inflammation which induces a range of haemodynamic effects (such as increased metabolic demand, coronary vasoconstriction and hypoxaemia), and haemostatic effects (increased platelet aggregation and increased plasma viscosity)(14). Second, some organisms such as *S.pneumoniae* may invade the myocardium during episodes of severe infection causing cardiac injury and scarring(15). In COVID-19, an additional mechanism of cardiac damage involves the Angiotensin-2 (ACE2) receptor. SARS-CoV-2 enters the lung by binding to ACE2 receptors, which leads to their downregulation(16). Such downregulation may increase levels of the pro-inflammatory vasoconstricting angiotensin2, and thereby worsen heart failure and lung injury during the acute phase of COVID-19(17). It is not yet clear whether longer-term CVD complications may result from other mechanisms including cardiomyopathy and fibrosis.

Importance

While clinical comorbidities are important predictors of severe outcomes of SARS-CoV-2 infection, the contribution of different patterns of comorbidities including cardiovascular risk factors is unclear. Those with cardiovascular disease or cardiovascular risk factors did not appear on Public Health England's list of extremely vulnerable groups recommended for shielding measures during the pandemic(18). Existing CVD, but not risk factors alone, was included on a broader list of conditions predisposing to severe COVID-19 disease, which is based upon recommendations for influenza vaccination. It is imperative that the likelihood of severe outcomes among those with differing levels of cardiovascular risk is properly characterised. This will inform clinical management of COVID-19, as well as national and international policy on targeting of a future vaccine within around 12-18 months. Evidence from the research will also benefit future care by informing inclusion criteria for clinical trials e.g. of interventions during COVID-19 to reduce adverse cardiorespiratory outcomes.

F. Study Type

Hypothesis testing: people with a raised cardiovascular risk level have more severe outcomes after SARS-CoV-2 infection than those who are not at raised cardiovascular risk

G. Study Design

Cohort and self-controlled case series studies

H. Feasibility counts

As of October 2020, CPRD Aurum covered 12.9 million currently registered and acceptable patients from 1,350 currently contributing practices. Based upon published feasibility counts for SARS-CoV-2 related research, by 30 Sept 2020 there were 733,611 unique patients in CPRD Aurum with evidence of SARS-CoV-2 codes in their primary care record. Interpolation of age groups presented on the CPRD website suggests that 421,352 of these patients were aged 40 years or more, and it is assumed that all are eligible for data linkage. Linkage to SGSS data on laboratory-confirmed SARS-CoV-2, and inclusion of the time period to 31 December 2020, will substantially increase the numbers eligible to take part in the study.

Based on other published work, we assume:

- A ratio of 1:4 for patients with QRISK3 ≥10% compared to<10% (19). All patients without existing CVD will be eligible for QRISK calculation.
- A ratio of 1:7 patients with hypertension to those without hypertension, based on feasibility figures calculated for our previous study (ISAC protocol 19_209). All patients without existing CVD (which will form a separate comparison group) will be eligible to have hypertension recorded.
- An age-standardised annual incidence of MACE of 0.7% (based on BHF statistics for stroke/TIA, coronary heart disease and heart failure incidence) (20).
- A 2% case fatality rate among those aged 40+ years with COVID-19(21).

I. Sample size considerations

For the cohort study, we estimate that, among those with existing records of SARS-CoV-2 infection, assuming a case fatality of 2% and a ratio of 1:4 for QRISK3 scores ≥10% compared to<10% (the main comparison), we would be able to detect a HR of 0.92 with 90% power at the 5% significance level. We will base actual eligibility for inclusion in the study upon an extra three months of data and will include positive results recorded in SGSS, of which there are likely to be far greater numbers than those recorded in CPRD alone. Assuming that, conservatively, the SGSS linkage doubles the number eligible for inclusion, the study would be powered to detect a HR of 0.94 with 90% power. For the MACE outcomes, the equivalent HRs are 0.84 and 0.89. Given that the daily UK figure for numbers testing positive is rising rapidly (more than 20,000 cases per day as of 31 October 2020), adding an extra 3 months of data collection will substantially increase power to detect differences in all outcomes by cardiovascular risk level.

For the self-controlled case series study, our previous studies suggest that the relative incidence of acute CVD events is roughly 5-fold higher in the week after respiratory-confirmed virus infection. Conservatively, to detect a 2-fold increase in MACE with 90% power at the 5% significance level assuming a risk period of 28 days and 9 months of follow up requires 181 subjects. Shortening the length of the risk period to 7 days requires 622 subjects and increasing it to 91 days requires 91 subjects. By applying the expected annual incidence rate of MACE to SARS-CoV-2 positive subjects over 9 months, 2,212 would be eligible to take part in this study, thus allowing adequate power for these differing scenarios.

J. Planned use of linked data (if applicable):

HES: HES admitted patient care data is being requested in order to identify COVID-19 related hospital admissions (outcome for objective 2), as well as to increase the sensitivity of recording of cardiovascular outcome diagnoses (outcomes for objectives 2 and 3). Using HES APC data will help to accurately define a history of previous cardiovascular disease (exposure for objective 1 and exclusion criterion for objectives 2 and 3). Additional information on clinical confounders may be identified using HES data (confounders for objective 2).

ONS: COVID-related mortality is the primary outcome for objective 2, and death will be used to censor follow up for the other outcomes.

Townsend data: Socioeconomic status is needed for calculation of QRISK3. It is also a potential confounder of the association between hypertension and severe outcomes after COVID-19.

SGSS data: Linked data on laboratory-confirmed SARS-CoV2 is essential to classify accurately the start of follow up for objective 2, and exposure status for objective 3.

CHESS data: Linked surveillance data from CHESS will enable ICU admissions and respiratory support for COVID-19 to be identified accurately (outcomes for objective 2).

Patients in England and Wales will benefit from this research as the results will enable improved understanding of the likely outcomes after COVID-19 among people with differing levels of cardiovascular risk. This will inform prevention e.g. targeting of a future COVID-19 vaccine to those most at risk of adverse outcomes, treatment e.g. consideration of anticoagulants for preventing thrombotic events, and health service planning e.g. modelling of the likely cardiovascular burden of COVID-19 on the NHS including bed and treatment requirements.

K. Definition of the Study population

People aged 40 years or over in CPRD Aurum with at least one year of follow up post-registration who are eligible for linkage to HES and the COVID linked datasets will be eligible for inclusion. We will initially define cardiovascular risk level based on QRISK3 score or the presence of existing CVD using data from the 5 years prior to the start of follow up. We will then redefine cardiovascular risk level based on the presence or absence of hypertension, or existing CVD. Follow up will begin at the latest of age 40 years, one year post-registration or 12 March 2020.

The whole population will be included in initial descriptive analyses of incidence rates of SARS-CoV-2 and outcomes by CVD risk level (objective 1). In the cohort study, those who experienced SARS-CoV-2 will then be followed from the date of first positive test to investigate the effect of cardiovascular risk level (defined first by QRISK3 and second by hypertension status among those without existing CVD) on the outcomes listed in section D (objective 2). For the self-controlled case series analysis, the study population will be a subset of the cohort study population who experience both SARS-CoV-2 and an outcome of interest at any time during follow up from 12 March 2020 to 31 Dec 2020 (objective 3) stratified by cardiovascular risk level.

L. Selection of comparison group(s)

For objective 1: Comparisons will be by cardiovascular risk level i.e. individuals will be categorised as having a QRISK3 score <10% (low), QRISK3 ≥10% (raised) or existing CVD, or by the presence or absence of hypertension or existing CVD.

For objective 2: Comparisons will be by cardiovascular risk level among those without existing CVD. Individuals will be categorised as having a QRISK3 score <10% (low), QRISK3 ≥10% (raised), or by the presence or absence of hypertension.

For objective 3: Comparisons will be within individuals i.e. individuals will act as their own controls during different time periods. The effect of lab-confirmed SARS-CoV-2 on the relative incidence of major adverse cardiovascular events occurring up to 3 months after the positive SARS-CoV-2 test compared to a baseline period for each individual will be presented for each stratum of cardiovascular risk level. This will be defined first as QRISK3 score <10% (low), QRISK3 ≥10% (raised), then by the presence or absence of hypertension.

M. Exposures, Outcomes and Covariates

We will develop codelists for variables listed in this section by building on codelists developed for our previous study (ISAC protocol 19_209) which have been reviewed by clinicians with public health, primary care and cardiology expertise, as well as using publicly available codelists published by the OpenSAFELY project (http://codelists.opensafely.org/), of which LSHTM is one of the main collaborators.

Incidence study (objective 1) and Cohort studies (objective 2)
Exposure

For objectives 1 and 2, we will first identify patients with existing CVD, defined as a history of coronary heart disease (myocardial infarction, angina, revascularisation procedure or CHD not otherwise specified), stroke or transient ischaemic attack, peripheral arterial disease or heart failure. For patients without CVD, cardiovascular risk level will be defined by calculated QRISK3 score – a validated measure of cardiovascular risk in the UK population(22). We will calculate QRISK3 scores at the index date (latest of age 40 years, one year post-registration, or 12 March 2020) using existing algorithms developed by our group, available at https://zenodo.org/badge/latestdoi/286983792. Briefly, this involves extracting relevant codes and measures from each patient's record and applying the freely available algorithm to calculate their ten-year risk of CVD. The variables included in QRISK3 calculations are listed below. In sensitivity analyses, we will explore the effect of using finer stratifications of QRISK3 score (<10%, 10-19%, 20-29%, ≥30%) to define cardiovascular risk level. We will also redefine cardiovascular risk level based on QRISK2 as this is the currently available risk calculator in GP software systems.

Among patients without existing CVD, we will then redefine cardiovascular risk level using hypertension status. Hypertension will be identified by the presence of any of the following in the 5 years before cohort entry (mirroring the time period during which components of the QRISK3 score will be collected) (i) a clinical diagnostic code, (ii) a flag on the GP hypertension register, (iii) most recent recorded systolic BP to cohort entry of ≥140mmHg, (iii) most recent recorded diastolic BP to cohort entry of ≥90mmHg. Data on blood pressure medications will also be extracted to stratify those with hypertension into treated and untreated groups.

Outcomes

Outcomes are listed in Section D. COVID-related mortality will be ascertained from linked ONS records. Hospital or ICU admission due to COVID-19 will be ascertained from linked HES and CHESS data. The need for respiratory support due to COVID-19 will be determined using CHESS. Cardiovascular outcomes will be identified by the earliest record of a diagnostic code in Aurum or an ICD-10 record in linked HES data. We have recently conducted a systematic review of the validity of three major cardiovascular outcome diagnoses in European health records(23) which will be used to inform our choice of included codes.

Covariates

The following variables will be defined at cohort entry based upon data recorded in the past 5 years in CPRD Aurum records, linked HES data and linked Townsend data (socioeconomic deprivation only). Those marked with a '*' are required for calculation of QRISK3. As they are drivers of cardiovascular risk which are already accounted for by inclusion in QRISK3, we will not consider them as potential confounders for the QRISK3 analysis. However, they will be included as covariates for the hypertension analysis.

- Age*
- Sex*
- Ethnicity*
- Townsend score*
- Smoking status*
- Diabetes*
- Family history of angina or heart attack in a first degree relative aged under 60 years*
- Chronic kidney disease*
- Atrial fibrillation*
- Blood pressure medication*
- Migraines*
- Rheumatoid arthritis*
- Systemic lupus erythematosus*

- Severe mental illness*
- Antipsychotic medication*
- Regular steroid therapy*
- Erectile dysfunction*
- Cholesterol/ HDL ratio*
- Blood pressure*
- BMI*
- Geographic region
- Healthcare utilisation frequency
- Alcohol misuse
- Chronic respiratory disease
- Chronic liver disease
- Chronic neurological disease/ learning disability
- Previous or current malignancy
- Immunosuppression (based upon a definition developed by the LSHTM Health Protection Research Unit in Vaccines and Immunisation[∞])
- Statin use
-

We will consider age, sex and infection severity (defined by hospitalisation or ICU admission due to COVID-19) as potential effect modifiers.

Self-controlled case series studies (objective 3)

Exposure
SARS-CoV-2 infection will be identified using linked SGSS on laboratory-confirmed infection with the date taken from the first record. In sensitivity analysis, we will explore the effect of broadening the definition to include clinically diagnosed or suspected COVID-19 among those without available laboratory test results.

Outcomes
Outcomes are listed in Section D. Mortality will not be included as an outcome in this analysis because this would violate the assumption of self-controlled case series that occurrence of an outcome is independent of the probability of subsequent exposure and should not result in censoring of observation time. As the MACE outcomes may lead to death, we will conduct a sensitivity analysis to assess the possible impact of this informative censoring on the validity of our results.

Stratifying variable
Cardiovascular risk level, defined at the index date as described above.

Covariates
Self-controlled case series analysis has the major advantage of implicitly controlling for fixed between-person confounding effects. Nevertheless, we will include the following time-varying confounders in our models:

1. Age* (hypertension analysis only as already included in calculation of QRISK3)
2. Season

We will also consider stratifying by the following variables as potential effect modifiers:

1. Age
2. Sex
3. Infection severity defined by hospitalisation or ICU admission due to COVID-19

[∞] Immunosuppression definition = ever history of HIV, solid organ transplant or permanent cellular immune deficiency, history in the last 24 months of aplastic anaemia, bone marrow or stem cell transplant or haematological malignancy, history in the last 12 months of chemo- or radiotherapy, biologics or other immunosuppressant therapy, other or unspecified cellular immune deficiency. Note corticosteroid use considered separately.

N. Data/ Statistical Analysis

For the descriptive analysis in objective 1, follow-up will start at the index date and end at the earliest of the following dates:

1. First occurrence of the specific outcome under study (event)
2. Death (censor, apart from mortality outcome where it is an event)
3. Transfer out of CPRD (censor)
4. End of the available follow-up period (censor)
5. 31 December 2020

For objective 2 (cohort study), follow up will start at the date of SARS-CoV-2 infection and end at the earliest of the following dates:

1. First occurrence of the specific outcome under study (event)
2. Death (censor, apart from mortality outcome where it is an event)
3. Transfer out of CPRD (censor)
4. End of the available follow-up period (censor)
5. 31 December 2020

For objective 3 (self-controlled case series study), follow up will start at the index date and end at the earliest of the following dates:

1. Death (censor)
2. Transfer out of CPRD (censor)
3. End of the available follow-up period (censor)
4. 31 December 2020

First, we will describe baseline characteristics of the whole cohort by cardiovascular risk level, defined by i) QRISK3 score, ii) hypertension status, iii) presence of existing CVD. We will calculate overall and age- and sex-specific incidence rates of laboratory-confirmed SARS-CoV-2 infection by cardiovascular risk group, and will explore the effect of broadening the COVID-19 definition to include those with clinically diagnosed or suspected COVID-19 without laboratory test results. We will also describe rates of COVID-19 related deaths, hospitalisations, ICU admissions and use of respiratory support by cardiovascular risk group, as well as rates of major adverse cardiovascular events following SARS-CoV-2 infection.

Among those with evidence of laboratory-confirmed SARS-CoV-2 infection, we will then investigate the effect of underlying cardiovascular risk level defined as above on each of the outcomes under study. We will first generate Kaplan-Meier curves to describe crude associations between cardiovascular risk level and outcomes. We will then use Cox proportional hazards regression with time since diagnosis as the underlying timescale to generate hazard ratios for the association between CVD risk level and each outcome, initially adjusting for age and sex, and then in a full model adjusted for confounders described in section M. We will test the proportional hazards assumption and adapt our approach if it does not hold, e.g. by stratifying on calendar time. For variables found to modify the association between cardiovascular risk level and outcomes, we will present stratified hazard ratios and absolute incidence rates. Sensitivity analyses will include (i) stratifying QRISK3 levels more finely e.g. by <10% 10-20% 20-30%, (ii) redefining cardiovascular risk level using QRISK2, (iii) repeating analyses using a broader definition of SARS-CoV-2 infection.

Among those with evidence of laboratory-confirmed SARS-CoV-2 infection *and* a major adverse cardiovascular event (MACE) occurring at any time during the study period, we will then carry out a self-controlled case series study. Using conditional Poisson regression, we will calculate the relative incidence of MACE occurring in time periods after SARS-CoV-2 with the risk period divided into 1-7, 8-14, 15-28, 29-91 days, compared to baseline time periods for each individual, adjusting for age and season. This design has the major advantage of implicitly controlling for fixed between-person confounding effects(24). We will present relative incidence stratified by cardiovascular risk level defined by i) QRISK3 score ii) hypertension status and will test for interactions with potential effect modifiers age, sex, cardiovascular risk level and infection severity. We will carry out the same sensitivity analyses as above.

O. Plan for addressing confounding

For objective 2, we will use multivariable Cox models for all outcomes to address confounding, with final confounders selected based upon analysis of DAGs. For objective 3, fixed between-person confounding effects are controlled for implicitly by the self-controlled case series study design. We will however additionally adjust for the time-varying confounders age and season. See section M for details.

P. Plans for addressing missing data

For covariates including BMI, cholesterol, blood pressure, smoking status and ethnicity, there are likely to be missing values. First, this may affect the accuracy of QRISK3 calculation. When calculating a QRISK3 score, missing data are replaced with average values. However, GP systems operate in the same way as our algorithm and will calculate a score using the same process, even in the absence of complete data. Therefore, the score that we generate should be the same as the one generated in general practice.

Second, for the cohort study we plan to use complete case analysis, which is shown to be valid if missingness is independent of the outcome(25). We will not include variables with large amounts of missing data (defined as >10% missing) in our final models, but will rather explore the effect of additionally adjusting for those variables in a further model. It is unlikely that data on BMI, smoking status and ethnicity are missing at random, the assumption required for multiple imputation.

Third, for the self-controlled case series using within-person comparisons we do not need to include these variables as potential confounders because they will be implicitly controlled for in the design. Comparing results between the cohort and self-controlled case series analyses will provide additional reassurance that the cohort study has not been adversely affected either by incomplete control for confounding or by bias introduced through the use of complete case analysis.

Q. Patient or user group involvement

Patients and user groups have not been involved in the development of this research.

R. Plans for disseminating and communicating study results, including the presence or absence of any restrictions on the extent and timing of publication

We will publish our results in peer-reviewed scientific journals and as a pre-print on the medRxiv platform or similar. We will use LSHTM Data Compass to publish the codelists used for this research. Results will also be presented at relevant scientific conferences. We will communicate our findings to patient groups and healthcare professionals through relevant healthcare charities such as the British Heart Foundation, which has funded us to do similar work on cardiovascular risk level and other acute respiratory infections. We will disseminate key findings on the LSHTM Electronic Health Records website and Twitter channel as well as working with the LSHTM press office and Science Media Centre to publicise findings to a wider audience.

Conflict of interest statement: None

S. Limitations of the study design, data sources, and analytic methods

Limitations:

- In the earlier part of the pandemic, laboratory testing for SARS-CoS-2 was targeted towards those with clinical need and a relevant epidemiological link such as a travel history to a high incidence area. Testing criteria have changed over time, and it is likely that using a laboratory-confirmed definition will lack sensitivity for identifying SARS-CoV-2 cases occurring in the earlier part of the pandemic before mass testing became widely available. We will explore this in a sensitivity analysis in which we expand the definition to include clinically diagnosed and suspected COVID-19 cases without laboratory results. We will also stratify by calendar time period should the proportional hazards assumption not hold, as it is possible that any effect of cardiovascular risk level on outcomes has changed between the first and second waves of the pandemic with advances in clinical management of COVID-19 patients. For self-controlled case series analyses, this will only limit the numbers entering the study but should not introduce bias. We will discuss these issues as potential limitations to our study.

- There may be issues with residual confounding in the cohort study, due to factors that are either not measured or are sub-optimally recorded in EHRs. We will triangulate findings from the cohort and self-controlled case series analyses to explore any likely effect of residual confounding and will discuss this explicitly as a limitation.

- As described above, missing data may affect variables such as ethnicity, cholesterol, blood pressure, BMI and smoking status which are needed to generate an accurate QRISK3 score. They might also affect the cohort analysis in which such variables are potential confounders. We will first describe patterns of missing data including carefully considering likely implications for our study. We will use standard approaches including imputing with average values for our calculation of QRISK3 scores and will conduct complete case analysis for our multivariable models. We will also carry out sensitivity analyses where appropriate to generate additional models including variables with missing data.

T. References

1. Warren-Gash C, Bhaskaran K, Hayward A, Leung GM, Lo S-V, Wong C-M, et al. Circulating Influenza Virus, Climatic Factors, and Acute Myocardial Infarction: A Time Series Study in England and Wales and Hong Kong. J Infect Dis. 2011 May 23;203(12):1710–8.

2. Blackburn RM, Zhao H, Pebody R, Hayward AC, Warren-Gash C. Laboratory-confirmed respiratory infections as predictors of hospital admission for myocardial infarction and stroke: time-series analysis of English data for 2004-2015. Clin Infect Dis Off Publ Infect Dis Soc Am. 2018 Jan 6;

3. Smeeth L, Thomas SL, Hall AJ, Hubbard R, Farrington P, Vallance P. Risk of Myocardial Infarction and Stroke after Acute Infection or Vaccination. N Engl J Med. 2004;351(25):2611–8.

4. Warren-Gash C, Hayward AC, Hemingway H, Denaxas S, Thomas SL, Timmis AD, et al. Influenza Infection and Risk of Acute Myocardial Infarction in England and Wales: A CALIBER Self-Controlled Case Series Study. J Infect Dis. 2012 Dec;206(11):1652–9.

5. Warren-Gash C, Blackburn R, Whitaker H, McMenamin J, Hayward AC. Laboratory-confirmed respiratory infections as triggers for acute myocardial infarction and stroke: a self-controlled case series analysis of national linked datasets from Scotland. Eur Respir J. 2018 Mar 21;

6. Ohland J, Warren-Gash C, Blackburn R, Molbak K, Valentiner-Branth P, Nielsen J, et al. Acute myocardial infarctions and stroke triggered by laboratory-confirmed respiratory infections in Denmark. EUROSURVEILLANCE. 2020;In press.

7. European Centre for Disease Prevention and Control. Situation update worldwide, as of 30 March 2020 08:00 [Internet]. 2020. Available from: https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases

8. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet Lond Engl. 2020 15;395(10223):497–506.

9. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. JAMA. 2020 Feb 24;

10. Guo T, Fan Y, Chen M, Wu X, Zhang L, He T, et al. Cardiovascular Implications of Fatal Outcomes of Patients With Coronavirus Disease 2019 (COVID-19). JAMA Cardiol. 2020 Mar 27;

11. Lippi G, Lavie CJ, Sanchis-Gomar F. Cardiac troponin I in patients with coronavirus disease 2019 (COVID-19): Evidence from a meta-analysis. Prog Cardiovasc Dis. 2020 Mar 10;

12. Mitrani RD, Dabas N, Goldberger JJ. COVID-19 cardiac injury: Implications for long-term surveillance and outcomes in survivors. Heart Rhythm. 2020 Jun;S1547527120306251.

13. Sabatino J, Rosa SD, Salvo GD, Indolfi C. Impact of cardiovascular risk profile on COVID-19 outcome. A meta-analysis. PLOS ONE. 2020 Aug 14;15(8):e0237131.

14. Bazaz R, Marriott HM, Francis SE, Dockrell DH. Mechanistic Links Between Acute Respiratory Tract Infections and Acute Coronary Syndromes. J Infect [Internet]. 2012 Oct [cited 2012 Oct 12]; Available from: http://linkinghub.elsevier.com/retrieve/pii/S0163445312002769

15. Reyes LF, Restrepo MI, Hinojosa CA, Soni NJ, Anzueto A, Babu BL, et al. Severe Pneumococcal Pneumonia Causes Acute Cardiac Toxicity and Subsequent Cardiac Remodeling. Am J Respir Crit Care Med. 2017 Sep 1;196(5):609–20.

16. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell. 2020 Mar 4;

17. Sanchis-Gomar F, Lavie CJ, Perez-Quilis C, Henry BM, Lippi G. Angiotensin-Converting Enzyme 2 and Anti-Hypertensives (Angiotensin Receptor Blockers and Angiotensin Converting Enzyme Inhibitors) in Coronavirus Disease 2019 (COVID-19). Mayo Clin Proc. 2020;19.

18. NHS Digital. Coronavirus (COVID-19): Shielded patients list [Internet]. [cited 2020 Mar 31]. Available from: https://digital.nhs.uk/coronavirus/shielded-patient-list

19. Robson J, Dostal I, Sheikh A, Eldridge S, Madurasinghe V, Griffiths C, et al. The NHS Health Check in England: an evaluation of the first 4 years. BMJ Open. 2016 Jan;6(1):e008840.

20. British Heart Foundation. Heart and Circulatory Disease Statistics 2020: chapter 2a Morbidity incidence [Internet]. [cited 2020 Sep 22]. Available from: https://www.bhf.org.uk/what-we-do/our-research/heart-statistics/heart-statistics-publications/cardiovascular-disease-statistics-2020

21. Oke J, Heneghan C. Global Covid-19 Case Fatality Rates [Internet]. 2020 [cited 2020 Sep 22]. Available from: https://www.cebm.net/covid-19/global-covid-19-case-fatality-rates/

22. Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. BMJ. 2017 May 23;357:j2099.

23. Davidson J, Banerjee A, Muzambi R, Smeeth L, Warren-Gash C. Validity of Acute Cardiovascular Outcome Diagnoses Recorded in European Electronic Health Records: A Systematic Review. Clin Epidemiol. In press.

24. Whitaker HJ, Hocine MN, Farrington CP. The methodology of self-controlled case series studies. Stat Methods Med Res. 2009 Feb;18(1):7–26.

25. White IR, Carlin JB. Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. Stat Med. 2010 Dec 10;29(28):2920–31.

List of Appendices

Appendix 7 Chapter 8 supplementary material

Supplementary table 1. Baseline characteristics of the incidence study population by cardiovascular risk

	All	Established CVD	QRISK3 score		Hypertension	
			Raised risk	Low risk	Raised risk	Low risk
	N=6,059,055	N=741,913	N=1,929,627	N=3,387,515	N=1,881,654	N=3,435,488
Age (years), Mean (SD)*	57.6 (12.1)	68.1 (10.7)	67.4 (8.9)	49.7 (7.1)	60.8 (11.6)	53.6 (10.7)
Age group (years)*						
40-54	2,759,591 (45.5%)	95,579 (12.9%)	167,366 (8.7%)	2,496,646 (73.7%)	623,005 (33.1%)	2,041,007 (59.4%)
55-64	1,488,257 (24.6%)	160,677 (21.7%)	526,344 (27.3%)	801,236 (23.7%)	526,946 (28.0%)	800,634 (23.3%)
65-74	1,125,307 (18.6%)	238,877 (32.2%)	796,797 (41.3%)	89,633 (2.6%)	459,164 (24.4%)	427,266 (12.4%)
75-84	685,900 (11.3%)	246,780 (33.3%)	439,120 (22.8%)	0 (0.0%)	272,539 (14.5%)	166,581 (4.8%)
Sex*						
Women	3,016,430 (49.8%)	303,691 (40.9%)	803,754 (41.7%)	1,908,985 (56.4%)	907,170 (48.2%)	1,805,569 (52.6%)
Men	3,042,578 (50.2%)	438,216 (59.1%)	1,125,832 (58.3%)	1,478,530 (43.6%)	974,469 (51.8%)	1,629,893 (47.4%)
Unknown	47 (0.0%)	6 (0.0%)	41 (0.0%)	0 (0.0%)	15 (0.0%)	26 (0.0%)
Ethnicity*						
White or not stated	4,520,954 (74.6%)	592,469 (79.9%)	1,525,861 (79.1%)	2,402,624 (70.9%)	1,441,447 (76.6%)	2,487,038 (72.4%)
South Asian	284,496 (4.7%)	34,849 (4.7%)	98,843 (5.1%)	150,804 (4.5%)	75,778 (4.0%)	173,869 (5.1%)
Black	122,348 (2.0%)	7,544 (1.0%)	16,844 (0.9%)	97,960 (2.9%)	42,751 (2.3%)	72,053 (2.1%)
Mixed/Other	408,571 (6.7%)	42,050 (5.7%)	95,746 (5.0%)	270,775 (8.0%)	123,701 (6.6%)	242,820 (7.1%)
Unknown	722,686 (11.9%)	65,001 (8.8%)	192,333 (10.0%)	465,352 (13.7%)	197,977 (10.5%)	459,708 (13.4%)
Townsend quintile*						
1 (most affluent)	1,377,227 (22.7%)	158,440 (21.4%)	442,185 (22.9%)	776,602 (22.9%)	440,270 (23.4%)	778,517 (22.7%)
2	1,270,784 (21.0%)	156,147 (21.0%)	414,389 (21.5%)	700,248 (20.7%)	407,867 (21.7%)	706,770 (20.6%)
3	1,153,335 (19.0%)	147,442 (19.9%)	368,505 (19.1%)	637,388 (18.8%)	361,033 (19.2%)	644,860 (18.8%)
4	1,075,138 (17.7%)	139,288 (18.8%)	336,485 (17.4%)	599,365 (17.7%)	327,735 (17.4%)	608,115 (17.7%)
5 (least affluent)	1,177,997 (19.4%)	140,093 (18.9%)	366,651 (19.0%)	671,253 (19.8%)	343,397 (18.2%)	694,507 (20.2%)
Unknown	4,574 (0.1%)	503 (0.1%)	1,412 (0.1%)	2,659 (0.1%)	1,352 (0.1%)	2,719 (0.1%)
Region of residence						
North East	195,161 (3.2%)	29,489 (4.0%)	67,385 (3.5%)	98,287 (2.9%)	65,030 (3.5%)	100,642 (2.9%)
North West	1,127,967 (18.6%)	163,745 (22.1%)	366,543 (19.0%)	597,679 (17.6%)	364,348 (19.4%)	599,874 (17.5%)

Yorkshire and the Humber	199,901 (3.3%)	26,077 (3.5%)	68,746 (3.6%)	105,078 (3.1%)	64,774 (3.4%)	109,050 (3.2%)
East Midlands	112,527 (1.9%)	13,988 (1.9%)	36,263 (1.9%)	62,276 (1.8%)	36,110 (1.9%)	62,429 (1.8%)
West Midlands	991,220 (16.4%)	127,718 (17.2%)	341,713 (17.7%)	521,789 (15.4%)	338,837 (18.0%)	524,665 (15.3%)
East of England	266,088 (4.4%)	27,946 (3.8%)	80,234 (4.2%)	157,908 (4.7%)	79,741 (4.2%)	158,401 (4.6%)
South West	1,155,120 (19.1%)	109,644 (14.8%)	321,836 (16.7%)	723,640 (21.4%)	319,626 (17.0%)	725,850 (21.1%)
South Central	1,275,851 (21.1%)	140,308 (18.9%)	403,213 (20.9%)	732,330 (21.6%)	381,420 (20.3%)	754,123 (22.0%)
London	725,286 (12.0%)	101,616 (13.7%)	240,047 (12.4%)	383,623 (11.3%)	228,485 (12.1%)	395,185 (11.5%)
Unknown	9,934 (0.2%)	1,382 (0.2%)	3,647 (0.2%)	4,905 (0.1%)	3,283 (0.2%)	5,269 (0.2%)
BMI category*†						
Underweight (<18·5 kg/m^2)	54,108 (0.9%)	9,692 (1.3%)	18,173 (0.9%)	26,243 (0.8%)	10,520 (0.6%)	33,896 (1.0%)
Normal (18·5-24·9 kg/m^2)	1,204,515 (19.9%)	145,084 (19.6%)	371,520 (19.3%)	687,911 (20.3%)	283,091 (15.0%)	776,340 (22.6%)
Overweight (25·0-29·9 kg/m^2)	1,554,717 (25.7%)	233,765 (31.5%)	580,257 (30.1%)	740,695 (21.9%)	516,934 (27.5%)	804,018 (23.4%)
Obese (30·0-39·9 kg/m^2)	1,191,415 (19.7%)	214,291 (28.9%)	465,390 (24.1%)	511,734 (15.1%)	499,738 (26.6%)	477,386 (13.9%)
Severely obese (≥40.0 kg/m^2)	171,358 (2.8%)	32,283 (4.4%)	62,506 (3.2%)	76,569 (2.3%)	83,298 (4.4%)	55,777 (1.6%)
Unknown	1,882,942 (31.1%)	106,798 (14.4%)	431,781 (22.4%)	1,344,363 (39.7%)	488,073 (25.9%)	1,288,071 (37.5%)
Cholesterol:HDL, Mean (SD)*†	3.6 (1.2)	3.4 (1.1)	3.8 (1.2)	3.6 (1.1)	3.7 (1.2)	3.7 (1.2)
Systolic blood pressure, Mean (SD)*†‡	132.4 (6176.4)	131.0 (186.4)	143.1 (10446.7)	125.5 (13.8)	140.1 (140.3)	127.8 (8547.6)
Smoking status*†						
Non-smoker	2,604,078 (43.0%)	303,876 (41.0%)	814,910 (42.2%)	1,485,292 (43.8%)	872,183 (46.4%)	1,428,019 (41.6%)
Ex-smoker	1,400,144 (23.1%)	280,913 (37.9%)	538,682 (27.9%)	580,549 (17.1%)	472,013 (25.1%)	647,218 (18.8%)
Current smoker	765,115 (12.6%)	105,704 (14.2%)	301,860 (15.6%)	357,551 (10.6%)	228,312 (12.1%)	431,099 (12.5%)
Unknown	1,289,718 (21.3%)	51,420 (6.9%)	274,175 (14.2%)	964,123 (28.5%)	309,146 (16.4%)	929,152 (27.0%)
Alcohol consumption†						
No heavy drinking	3,420,295 (56.4%)	541,921 (73.0%)	1,228,590 (63.7%)	1,649,784 (48.7%)	1,168,514 (62.1%)	1,709,860 (49.8%)
Heavy drinking	514,459 (8.5%)	62,349 (8.4%)	176,737 (9.2%)	275,373 (8.1%)	172,181 (9.2%)	279,929 (8.1%)
Unknown	2,124,301 (35.1%)	137,643 (18.6%)	524,300 (27.2%)	1,462,358 (43.2%)	540,959 (28.7%)	1,445,699 (42.1%)
Family history of CHD*	447,753 (7.4%)	43,028 (5.8%)	180,543 (9.4%)	224,182 (6.6%)	138,688 (7.4%)	266,037 (7.7%)
Consultation frequency in prior 12 months, Median (IQR)	4 (1-8)	8 (4-14)	5 (2-9)	2 (0-6)	4 (2-9)	3 (0-6)
Medication use$						
Regular corticosteroids*	75,385 (1.2%)	25,110 (3.4%)	38,463 (2.0%)	11,812 (0.3%)	24,988 (1.3%)	25,287 (0.7%)
Antihypertensives*	1,823,538 (30.1%)	528,503 (71.2%)	759,312 (39.4%)	535,723 (15.8%)	829,575 (44.1%)	465,460 (13.5%)

Statins	1,270,675 (21.0%)	451,780 (60.9%)	660,072 (34.2%)	158,823 (4.7%)	492,588 (26.2%)	326,307 (9.5%)
Antiplatelets	698,994 (11.5%)	409,772 (55.2%)	211,453 (11.0%)	77,769 (2.3%)	163,743 (8.7%)	125,479 (3.7%)
Anticoagulants	237,101 (3.9%)	122,558 (16.5%)	86,070 (4.5%)	28,473 (0.8%)	59,399 (3.2%)	55,144 (1.6%)
Comorbid condition						
Atrial fibrillation*	162,317 (2.7%)	97,233 (13.1%)	61,788 (3.2%)	3,296 (0.1%)	36,815 (2.0%)	28,269 (0.8%)
Migraines*	159,222 (2.6%)	17,573 (2.4%)	33,288 (1.7%)	108,361 (3.2%)	42,268 (2.2%)	99,381 (2.9%)
Diabetes*	534,471 (8.8%)	163,196 (22.0%)	320,106 (16.6%)	51,169 (1.5%)	206,788 (11.0%)	164,487 (4.8%)
CKD stage 3-5*	619,694 (10.2%)	233,756 (31.5%)	325,064 (16.9%)	60,874 (1.8%)	241,312 (12.8%)	144,626 (4.2%)
Chronic liver disease	64,828 (1.1%)	18,131 (2.4%)	23,261 (1.2%)	23,436 (0.7%)	19,150 (1.0%)	27,547 (0.8%)
Chronic respiratory disease (not asthma)	302,464 (5.0%)	106,770 (14.4%)	147,122 (7.6%)	48,572 (1.4%)	94,365 (5.0%)	101,329 (2.9%)
Asthma with recent OCS use[$]	294,042 (4.9%)	57,789 (7.8%)	110,528 (5.7%)	125,725 (3.7%)	102,130 (5.4%)	134,123 (3.9%)
Asthma with no recent OCS use	539,665 (8.9%)	70,166 (9.5%)	143,006 (7.4%)	326,493 (9.6%)	160,391 (8.5%)	309,108 (9.0%)
Severe mental illness / antipsychotic use*	71,742 (1.2%)	12,428 (1.7%)	30,410 (1.6%)	28,904 (0.9%)	19,970 (1.1%)	39,344 (1.1%)
Dementia	86,965 (1.4%)	34,969 (4.7%)	42,323 (2.2%)	9,673 (0.3%)	22,768 (1.2%)	29,228 (0.9%)
Chronic neurological disease	97,037 (1.6%)	26,735 (3.6%)	37,752 (2.0%)	32,550 (1.0%)	28,232 (1.5%)	42,070 (1.2%)
Learning / intellectual disability	27,561 (0.5%)	3,620 (0.5%)	7,933 (0.4%)	16,008 (0.5%)	7,176 (0.4%)	16,765 (0.5%)
Non-haematological cancer						
Diagnosed <1 year ago	180,797 (3.0%)	47,481 (6.4%)	87,792 (4.5%)	45,524 (1.3%)	64,248 (3.4%)	69,068 (2.0%)
Diagnosed 1-4.9 years ago	238,061 (3.9%)	54,265 (7.3%)	112,826 (5.8%)	70,970 (2.1%)	83,171 (4.4%)	100,625 (2.9%)
Diagnosed ≥5 years ago	390,468 (6.4%)	68,501 (9.2%)	158,495 (8.2%)	163,472 (4.8%)	130,151 (6.9%)	191,816 (5.6%)
Haematological malignancy						
Diagnosed <1 year ago	20,007 (0.3%)	6,322 (0.9%)	9,781 (0.5%)	3,904 (0.1%)	6,475 (0.3%)	7,210 (0.2%)
Diagnosed 1-4.9 years ago	16,610 (0.3%)	4,622 (0.6%)	8,029 (0.4%)	3,959 (0.1%)	5,427 (0.3%)	6,561 (0.2%)
Diagnosed ≥5 years ago	14,591 (0.2%)	3,187 (0.4%)	6,073 (0.3%)	5,331 (0.2%)	4,577 (0.2%)	6,827 (0.2%)
Rheumatoid arthritis*	61,299 (1.0%)	13,551 (1.8%)	29,068 (1.5%)	18,680 (0.6%)	21,866 (1.2%)	25,882 (0.8%)
Systemic lupus erythematosus*	6,897 (0.1%)	1,581 (0.2%)	2,487 (0.1%)	2,829 (0.1%)	2,006 (0.1%)	3,310 (0.1%)
HIV*	7,910 (0.1%)	703 (0.1%)	1,913 (0.1%)	5,294 (0.2%)	2,460 (0.1%)	4,747 (0.1%)
Immunosuppression#	70,397 (1.2%)	16,191 (2.2%)	29,410 (1.5%)	24,796 (0.7%)	24,119 (1.3%)	30,087 (0.9%)
Erectile dysfunction*	323,765 (10.6%)	84,360 (19.3%)	186,601 (16.6%)	52,804 (3.6%)	120,368 (12.4%)	119,037 (7.3%)

*In QRISK3 algorithm, but non-imputed version included here (for smoking status, cholesterol:HDL ratio, systolic BP and BMI). †most recent measure before baseline. ‡Used on hypertension definition. $ at least 1 prescription in the 12 months before baseline. Other than corticosteroids which was defined as at least 2 prescriptions prior to baseline with the most recent ≤28 days before baseline. #ever history of solid organ transplant or permanent cellular immune deficiency; history in the 24 months before baseline for aplastic anaemia, bone marrow or stem cell transplant; history in the 12 months before baseline for biologics or other immunosuppressant therapy (excluding corticosteroids), other or unspecified cellular immune deficiency

Supplementary table 2. Incidence of laboratory-confirmed SARS-CoV-2 and clinically reported COVID-19 and outcomes of interest during wave one of the COVID-19 pandemic

	All		Established CVD		QRISK3 score				Hypertension			
					Raised risk		Low risk		Raised risk		Low risk	
	N	Rate (95% CI) per 1,000	N	Rate (95% CI) per 1,000	N	Rate per 1,000	N	Rate (95% CI) per 1,000	N	Rate per 1,000	N	Rate (95% CI) per 1,000
All individuals	5,955,940		731,214		1,903,772		3,320,954		1,854,236		3,370,490	
COVID-19 death*	4,653	1.9 (1.9-2.0)	2,440	8.1 (7.7-8.4)	1,908	2.4 (2.3-2.5)	305	0.2 (0.2-0.3)	1,155	1.5 (1.4-1.6)	1,058	0.8 (0.7-0.8)
Hospitalization$	13,757	5.6 (5.5-5.7)	5,579	18.2 (17.7-18.7)	5,217	6.5 (6.3-6.7)	2,961	2.2 (2.1-2.3)	4,013	5.1 (5.0-5.3)	4,165	3.0 (2.9-3.1)
Major adverse cardiovascular event	39,725	16.3 (16.1-16.4)	28,882	95.4 (94.3-96.5)	8,310	10.4 (10.2-10.7)	2,533	1.9 (1.8-2.0)	6,303	8.2 (8.0-8.4)	4,540	3.3 (3.2-3.4)
Laboratory-confirmed SARS-CoV-2	26,708	11.0 (10.8-11.1)	7,059	23.3 (22.8-23.9)	7,574	9.5 (9.3-9.7)	12,075	9.0 (8.9-9.2)	7,750	10.1 (9.8-10.3)	11,899	8.7 (8.6-8.9)
COVID-19 death*	3,525	50.3 (48.7-52.0)	1,881	196.3 (187.6-205.4)	1,424	84.5 (80.2-89.0)	220	5.0 (4.4-5.8)	885	42.5 (39.8-45.4)	759	19.1 (17.8-20.5)
ICU admission†	1,277	18.2 (17.3-19.3)	300	31.3 (28.0-35.1)	591	35.1 (32.3-38.0)	386	8.8 (8.0-9.8)	527	25.3 (23.3-27.6)	450	11.3 (10.3-12.4)
Respiratory support‡	815	11.6 (10.9-12.5)	161	16.8 (14.4-19.6)	395	23.4 (21.2-25.9)	259	5.9 (5.3-6.7)	365	17.5 (15.8-19.4)	289	7.3 (6.5-8.2)
Hospitalization$	7,794	111.2 (108.8-113.7)	3,071	320.5 (309.4-332.0)	3,032	179.8 (173.5-186.3)	1,691	38.8 (37.0-40.6)	2,349	112.9 (108.4-117.6)	2,374	59.8 (57.5-62.3)
Major adverse cardiovascular event	1,026	14.6 (13.8-15.6)	658	68.7 (63.6-74.1)	277	16.4 (14.6-18.5)	91	2.1 (1.7-2.6)	213	10.2 (9.0-11.7)	155	3.9 (3.3-4.6)
Clinically reported COVID-19	41,151	16.8 (16.7-17.0)	8,597	28.4 (27.8-29.0)	12,242	15.4 (15.1-15.6)	20,312	15.2 (15.0-15.4)	12,608	16.4 (16.1-16.6)	19,946	14.6 (14.4-14.8)
COVID-19 death*	584	20.1 (18.5-21.8)	293	52.8 (47.1-59.2)	257	30.7 (27.2-34.7)	34	2.2 (1.6-3.1)	135	15.2 (12.8-18.0)	156	10.6 (9.1-12.5)
Hospitalization$	2,421	83.2 (80.0-86.6)	872	157.1 (147.0-167.9)	925	110.4 (103.5-117.8)	624	41.1 (38.0-44.5)	749	84.2 (78.4-90.5)	800	54.6 (51.0-58.5)
Major adverse cardiovascular event	1,161	39.9 (37.7-42.3)	757	136.4 (127.0-146.5)	316	37.7 (33.8-42.1)	88	5.8 (4.7-7.2)	244	27.4 (24.2-31.1)	160	10.9 (9.4-12.8)

*Ascertained from ONS death certificate data in which the COVID related ICD-10 codes U07.1 or U07.2 were present in the record

$Ascertained from presence in CHESS dataset or HES APC record coded with primary diagnosis of U07.1 or U07.2

†Ascertained from CHESS records coded with ICU/HDU admission, only available for those with laboratory confirmed SARS-CoV-2

‡Ascertained from CHESS record coded with use of respiratory support via invasive mechanical ventilation, only available for those with laboratory confirmed SARS-CoV-2

Supplementary table 3. Incidence of laboratory-confirmed SARS-CoV-2 and clinically reported COVID-19 and outcomes of interest during wave two of the COVID-19 pandemic

	All		Established CVD		QRISK3 score				Hypertension			
					Raised risk		Low risk		Raised risk		Low risk	
	N	Rate (95% CI) per 1,000	N	Rate (95% CI) per 1,000	N	Rate per 1,000	N	Rate (95% CI) per 1,000	N	Rate per 1,000	N	Rate (95% CI) per 1,000
All individuals	5,862,260		707,279		1,865,383		3,289,598		1,824,647		3,330,334	
COVID-19 death*	3,213	0.7 (0.7-0.7)	1,724	3.1 (2.9-3.2)	1,295	0.9 (0.8-0.9)	194	0.1 (0.1-0.1)	859	0.6 (0.6-0.6)	630	0.2 (0.2-0.3)
Hospitalization$	14,256	3.1 (3.0-3.2)	5,301	9.4 (9.2-9.7)	5,577	3.8 (3.7-3.9)	3,378	1.3 (1.3-1.4)	4,468	3.1 (3.0-3.2)	4,487	1.7 (1.7-1.8)
Major adverse cardiovascular event	31,310	6.8 (6.7-6.9)	20,436	36.3 (35.8-36.8)	8,294	5.6 (5.5-5.7)	2,580	1.0 (1.0-1.1)	6,330	4.4 (4.3-4.5)	4,544	1.8 (1.7-1.8)
Laboratory-confirmed SARS-CoV-2	147,421	32.0 (31.9-32.2)	17,720	31.5 (31.0-32.0)	33,842	22.8 (22.5-23.0)	95,859	37.6 (37.3-37.8)	43,104	29.8 (29.6-30.1)	86,597	33.4 (33.2-33.6)
COVID-19 death*	2,950	22.2 (21.4-23.1)	1,612	92.0 (87.7-96.6)	1,173	37.4 (35.3-39.6)	165	2.0 (1.7-2.3)	779	19.9 (18.5-21.3)	559	7.4 (6.8-8.0)
ICU admission†	747	5.6 (5.2-6.0)	199	11.4 (9.9-13.1)	339	10.8 (9.7-12.0)	209	2.5 (2.2-2.9)	284	7.2 (6.4-8.1)	264	3.5 (3.1-3.9)
Respiratory support‡	269	2.0 (1.8-2.3)	69	3.9 (3.1-5.0)	131	4.2 (3.5-5.0)	69	0.8 (0.7-1.0)	110	2.8 (2.3-3.4)	90	1.2 (1.0-1.5)
Hospitalization$	10,099	76.1 (74.6-77.6)	3,484	198.9 (192.4-205.7)	4,015	128.0 (124.1-132.0)	2,600	31.0 (29.9-32.2)	3,279	83.7 (80.8-86.6)	3,336	43.9 (42.5-45.4)
Major adverse cardiovascular event	1,225	9.2 (8.7-9.8)	764	43.6 (40.6-46.8)	339	10.8 (9.7-12.0)	122	1.5 (1.2-1.7)	273	7.0 (6.2-7.8)	188	2.5 (2.1-2.9)
Clinically reported COVID-19	29,549	6.4 (6.4-6.5)	5,071	9.0 (8.8-9.3)	7,998	5.4 (5.3-5.5)	16,480	6.5 (6.4-6.6)	8,766	6.1 (5.9-6.2)	15,712	6.1 (6.0-6.2)
COVID-19 death*	139	2.6 (2.2-3.0)	72	7.1 (5.7-9.0)	61	4.0 (3.1-5.1)	6	0.2 (0.1-0.5)	43	2.6 (1.9-3.5)	24	0.9 (0.6-1.3)
Hospitalization$	1,271	23.5 (22.3-24.8)	464	45.9 (41.9-50.3)	460	29.8 (27.2-32.7)	347	12.2 (10.9-13.5)	375	22.7 (20.5-25.2)	432	15.7 (14.3-17.3)
Major adverse cardiovascular event	841	15.6 (14.5-16.7)	525	52.0 (47.7-56.6)	253	16.4 (14.5-18.6)	63	2.2 (1.7-2.8)	174	10.5 (9.1-12.2)	142	5.2 (4.4-6.1)

*Ascertained from ONS death certificate data in which the COVID related ICD-10 codes U07.1 or U07.2 were present in the record

$Ascertained from presence in CHESS dataset or HES APC record coded with primary diagnosis of U07.1 or U07.2

†Ascertained from CHESS records coded with ICU/HDU admission, only available for those with laboratory confirmed SARS-CoV-2

‡Ascertained from CHESS record coded with use of respiratory support via invasive mechanical ventilation, only available for those with laboratory confirmed SARS-CoV-2

Supplementary table 4. Baseline characteristics of the clinically reported COVID-19 study population by cardiovascular risk

	All	QRISK3 score		Hypertension	
		Raised risk	Low risk	Raised risk	Low risk
	N=56,197	N=19,528	N=36,669	N=21,417	N=34,780
Age (years), Mean (SD)*	56.0 (10.9)	66.4 (9.6)	50.5 (6.7)	59.7 (11.2)	53.8 (10.1)
Age group (years)*					
40-54	28,447 (50.6%)	2,303 (11.8%)	26,144 (71.3%)	7,933 (37.0%)	20,514 (59.0%)
55-64	15,755 (28.0%)	6,114 (31.3%)	9,641 (26.3%)	6,629 (31.0%)	9,126 (26.2%)
65-74	7,627 (13.6%)	6,743 (34.5%)	884 (2.4%)	4,133 (19.3%)	3,494 (10.0%)
75-84	4,368 (7.8%)	4,368 (22.4%)	0 (0.0%)	2,722 (12.7%)	1,646 (4.7%)
Sex*					
Women	33,566 (59.7%)	8,722 (44.7%)	24,844 (67.8%)	11,952 (55.8%)	21,614 (62.1%)
Men	22,628 (40.3%)	10,803 (55.3%)	11,825 (32.2%)	9,463 (44.2%)	13,165 (37.9%)
Unknown	3 (0.0%)	3 (0.0%)	0 (0.0%)	2 (0.0%)	1 (0.0%)
Ethnicity*					
White or not stated	39,786 (70.8%)	14,577 (74.6%)	25,209 (68.7%)	15,252 (71.2%)	24,534 (70.5%)
South Asian	4,781 (8.5%)	1,993 (10.2%)	2,788 (7.6%)	1,623 (7.6%)	3,158 (9.1%)
Black	1,790 (3.2%)	282 (1.4%)	1,508 (4.1%)	765 (3.6%)	1,025 (2.9%)
Mixed/Other	5,500 (9.8%)	1,432 (7.3%)	4,068 (11.1%)	2,210 (10.3%)	3,290 (9.5%)
Unknown	4,340 (7.7%)	1,244 (6.4%)	3,096 (8.4%)	1,567 (7.3%)	2,773 (8.0%)
Townsend quintile*					
1 (most affluent)	10,083 (17.9%)	3,224 (16.5%)	6,859 (18.7%)	3,807 (17.8%)	6,276 (18.0%)
2	9,865 (17.6%)	3,309 (16.9%)	6,556 (17.9%)	3,845 (18.0%)	6,020 (17.3%)
3	10,042 (17.9%)	3,420 (17.5%)	6,622 (18.1%)	3,848 (18.0%)	6,194 (17.8%)
4	10,881 (19.4%)	3,781 (19.4%)	7,100 (19.4%)	4,077 (19.0%)	6,804 (19.6%)
5 (least affluent)	15,290 (27.2%)	5,784 (29.6%)	9,506 (25.9%)	5,823 (27.2%)	9,467 (27.2%)
Unknown	36 (0.1%)	10 (0.1%)	26 (0.1%)	17 (0.1%)	19 (0.1%)
Region of residence					
North East	1,236 (2.2%)	474 (2.4%)	762 (2.1%)	516 (2.4%)	720 (2.1%)
North West	11,964 (21.3%)	4,340 (22.2%)	7,624 (20.8%)	4,845 (22.6%)	7,119 (20.5%)
Yorkshire and the Humber	1,679 (3.0%)	652 (3.3%)	1,027 (2.8%)	650 (3.0%)	1,029 (3.0%)
East Midlands	778 (1.4%)	255 (1.3%)	523 (1.4%)	305 (1.4%)	473 (1.4%)

West Midlands	7,879 (14.0%)	2,853 (14.6%)	5,026 (13.7%)	3,178 (14.8%)	4,701 (13.5%)
East of England	1,893 (3.4%)	584 (3.0%)	1,309 (3.6%)	689 (3.2%)	1,204 (3.5%)
South West	14,852 (26.4%)	4,874 (25.0%)	9,978 (27.2%)	5,245 (24.5%)	9,607 (27.6%)
South Central	9,054 (16.1%)	3,027 (15.5%)	6,027 (16.4%)	3,282 (15.3%)	5,772 (16.6%)
London	6,793 (12.1%)	2,448 (12.5%)	4,345 (11.8%)	2,686 (12.5%)	4,107 (11.8%)
Unknown	69 (0.1%)	21 (0.1%)	48 (0.1%)	21 (0.1%)	48 (0.1%)
BMI category*†					
Underweight (<18·5 kg/m^2)	598 (1.1%)	285 (1.5%)	313 (0.9%)	159 (0.7%)	439 (1.3%)
Normal (18·5-24·9 kg/m^2)	11,078 (19.7%)	3,776 (19.3%)	7,302 (19.9%)	3,014 (14.1%)	8,064 (23.2%)
Overweight (25·0-29·9 kg/m^2)	15,335 (27.3%)	5,919 (30.3%)	9,416 (25.7%)	5,921 (27.6%)	9,414 (27.1%)
Obese (30·0-39·9 kg/m^2)	13,824 (24.6%)	5,722 (29.3%)	8,102 (22.1%)	6,824 (31.9%)	7,000 (20.1%)
Severely obese (≥40.0 kg/m^2)	2,588 (4.6%)	1,018 (5.2%)	1,570 (4.3%)	1,508 (7.0%)	1,080 (3.1%)
Unknown	12,774 (22.7%)	2,808 (14.4%)	9,966 (27.2%)	3,991 (18.6%)	8,783 (25.3%)
Cholesterol:HDL, Mean (SD)*†	3.7 (1.2)	3.8 (1.3)	3.7 (1.1)	3.8 (1.2)	3.7 (1.2)
Systolic blood pressure, Mean (SD)*†‡	128.5 (51.7)	134.4 (83.3)	125.0 (13.8)	138.1 (13.9)	122.2 (64.8)
Smoking status*†					
Non-smoker	26,893 (47.9%)	8,285 (42.4%)	18,608 (50.7%)	10,577 (49.4%)	16,316 (46.9%)
Ex-smoker	13,659 (24.3%)	6,143 (31.5%)	7,516 (20.5%)	5,841 (27.3%)	7,818 (22.5%)
Current smoker	7,829 (13.9%)	3,544 (18.1%)	4,285 (11.7%)	2,798 (13.1%)	5,031 (14.5%)
Unknown	7,816 (13.9%)	1,556 (8.0%)	6,260 (17.1%)	2,201 (10.3%)	5,615 (16.1%)
Alcohol consumption†					
No heavy drinking	33,730 (60.0%)	13,298 (68.1%)	20,432 (55.7%)	14,048 (65.6%)	19,682 (56.6%)
Heavy drinking	5,343 (9.5%)	2,013 (10.3%)	3,330 (9.1%)	2,168 (10.1%)	3,175 (9.1%)
Unknown	17,124 (30.5%)	4,217 (21.6%)	12,907 (35.2%)	5,201 (24.3%)	11,923 (34.3%)
Family history of CHD*	5,435 (9.7%)	2,174 (11.1%)	3,261 (8.9%)	1,966 (9.2%)	3,469 (10.0%)
Consultation frequency in prior 12 months, Median (IQR)	6 (3-11)	8 (4-14)	5 (2-10)	7 (4-13)	5 (2-10)
Medication use$					
Regular corticosteroids*	1,400 (2.5%)	976 (5.0%)	424 (1.2%)	660 (3.1%)	740 (2.1%)
Antihypertensives*	15,636 (27.8%)	8,018 (41.1%)	7,618 (20.8%)	9,590 (44.8%)	6,046 (17.4%)
Statins	8,875 (15.8%)	6,701 (34.3%)	2,174 (5.9%)	5,159 (24.1%)	3,716 (10.7%)
Antiplatelets	3,797 (6.8%)	2,377 (12.2%)	1,420 (3.9%)	2,068 (9.7%)	1,729 (5.0%)
Anticoagulants	1,430 (2.5%)	979 (5.0%)	451 (1.2%)	710 (3.3%)	720 (2.1%)

Comorbid condition					
Atrial fibrillation*	807 (1.4%)	767 (3.9%)	40 (0.1%)	447 (2.1%)	360 (1.0%)
Migraines*	2,830 (5.0%)	614 (3.1%)	2,216 (6.0%)	933 (4.4%)	1,897 (5.5%)
Diabetes*	5,626 (10.0%)	4,672 (23.9%)	954 (2.6%)	3,077 (14.4%)	2,549 (7.3%)
CKD stage 3-5*	4,931 (8.8%)	3,826 (19.6%)	1,105 (3.0%)	3,031 (14.2%)	1,900 (5.5%)
Chronic liver disease	962 (1.7%)	482 (2.5%)	480 (1.3%)	408 (1.9%)	554 (1.6%)
Chronic respiratory disease (not asthma)	4,191 (7.5%)	2,926 (15.0%)	1,265 (3.4%)	1,971 (9.2%)	2,220 (6.4%)
Asthma with recent OCS use[$]	4,585 (8.2%)	1,929 (9.9%)	2,656 (7.2%)	1,942 (9.1%)	2,643 (7.6%)
Asthma with no recent OCS use	7,435 (13.2%)	2,331 (11.9%)	5,104 (13.9%)	2,800 (13.1%)	4,635 (13.3%)
Severe mental illness / antipsychotic use*	1,101 (2.0%)	613 (3.1%)	488 (1.3%)	400 (1.9%)	701 (2.0%)
Dementia	1,121 (2.0%)	921 (4.7%)	200 (0.5%)	479 (2.2%)	642 (1.8%)
Chronic neurological disease	975 (1.7%)	523 (2.7%)	452 (1.2%)	413 (1.9%)	562 (1.6%)
Learning / intellectual disability	444 (0.8%)	190 (1.0%)	254 (0.7%)	139 (0.6%)	305 (0.9%)
Non-haematological cancer					
Diagnosed <1 year ago	1,951 (3.5%)	1,273 (6.5%)	678 (1.8%)	944 (4.4%)	1,007 (2.9%)
Diagnosed 1-4.9 years ago	2,044 (3.6%)	1,102 (5.6%)	942 (2.6%)	916 (4.3%)	1,128 (3.2%)
Diagnosed ≥5 years ago	3,610 (6.4%)	1,537 (7.9%)	2,073 (5.7%)	1,446 (6.8%)	2,164 (6.2%)
Haematological malignancy					
Diagnosed <1 year ago	261 (0.5%)	186 (1.0%)	75 (0.2%)	123 (0.6%)	138 (0.4%)
Diagnosed 1-4.9 years ago	161 (0.3%)	93 (0.5%)	68 (0.2%)	69 (0.3%)	92 (0.3%)
Diagnosed ≥5 years ago	137 (0.2%)	66 (0.3%)	71 (0.2%)	49 (0.2%)	88 (0.3%)
Rheumatoid arthritis*	831 (1.5%)	488 (2.5%)	343 (0.9%)	382 (1.8%)	449 (1.3%)
Systemic lupus erythematosus*	134 (0.2%)	67 (0.3%)	67 (0.2%)	54 (0.3%)	80 (0.2%)
HIV*	119 (0.2%)	35 (0.2%)	84 (0.2%)	50 (0.2%)	69 (0.2%)
Immunosuppression#	883 (1.6%)	471 (2.4%)	412 (1.1%)	389 (1.8%)	494 (1.4%)
Erectile dysfunction*	2,995 (13.2%)	2,270 (21.0%)	725 (6.1%)	1,522 (16.1%)	1,473 (11.2%)

*In QRISK3 algorithm, but non-imputed version included here (for smoking status, cholesterol:HDL ratio, systolic BP and BMI)
†most recent measure before baseline
‡Used on hypertension definition
$ at least 1 prescription in the 12 months before baseline. Other than corticosteroids which was defined as at least 2 prescriptions prior to baseline with the most recent ≤28 days before baseline
#ever history of solid organ transplant or permanent cellular immune deficiency; history in the 24 months before baseline for aplastic anaemia, bone marrow or stem cell transplant; history in the 12 months before baseline for biologics or other immunosuppressant therapy (excluding corticosteroids), other or unspecified cellular immune deficiency

Supplementary table 5. Hazard ratios for the effect of raised cardiovascular risk on severe outcomes after laboratory-confirmed SARS-CoV-2

	N events	Rate (95% CI) per 1,000 person-years	Crude HR (95% CI)	Age- and sex-adjusted HR (95% CI)	Fully-adjusted[#] HR (95% CI)
COVID-19 death*					
QRISK3 ≥10%	2,183	310.4 (297.7-323.7)	16.33 (14.61-18.24)	NA	8.77 (7.62-10.10)
QRISK3 <10%	365	20.4 (18.4-22.6)	1 (ref)	NA	1 (ref)
Hypertension	1,384	156.2 (148.2-164.6)	2.27 (2.10-2.45)	1.08 (1.00-1.17)	1.05 (0.94-1.18)
No hypertension	1,164	72.6 (68.6-76.9)	1 (ref)	1 (ref)	1 (ref)
ICU admission[†]					
QRISK3 ≥10%	876	120.0 (112.1-128.6)	4.27 (3.83-4.76)	NA	3.66 (3.18-4.21)
QRISK3 <10%	573	29.9 (27.5-32.6)	1 (ref)	NA	1 (ref)
Hypertension	768	82.2 (76.3-88.4)	2.16 (1.94-2.41)	1.55 (1.38-1.73)	1.15 (0.98-1.36)
No hypertension	681	40.2 (37.2-43.5)	1 (ref)	1 (ref)	1 (ref)
Respiratory support[‡]					
QRISK3 ≥10%	498	66.9 (61.1-73.3)	4.30 (3.72-4.98)	NA	3.73 (3.10-4.49)
QRISK3 <10%	320	16.6 (14.8-18.7)	1 (ref)	NA	1 (ref)
Hypertension	452	47.9 (43.5-52.7)	2.39 (2.07-2.75)	1.73 (1.49-2.01)	1.20 (0.97-1.48)
No hypertension	366	21.3 (19.2-23.7)	1 (ref)	1 (ref)	1 (ref)
Hospitalisation[$]					
QRISK3 ≥10%	6,547	1212.4 (1183.4-1242.1)	4.41 (4.24-4.59)	NA	3.38 (3.22-3.56)
QRISK3 <10%	4,247	256.2 (248.6-264.0)	1 (ref)	NA	1 (ref)
Hypertension	5,325	713.0 (694.1-732.4)	1.91 (1.84-1.99)	1.26 (1.21-1.31)	1.05 (0.99-1.11)
No hypertension	5,469	377.0 (367.1-387.1)	1 (ref)	1 (ref)	1 (ref)
Major adverse cardiovascular event					
QRISK3 ≥10%	570	82.4 (75.9-89.4)	7.51 (6.40-8.81)	NA	5.43 (4.44-6.64)
QRISK3 <10%	204	11.5 (10.0-13.1)	1 (ref)	NA	1 (ref)
Hypertension	450	51.3 (46.8-56.3)	2.63 (2.28-3.04)	1.62 (1.40-1.87)	1.49 (1.20-1.85)
No hypertension	324	20.3 (18.2-22.6)	1 (ref)	1 (ref)	1 (ref)

*Ascertained from ONS death certificate data in which the COVID related ICD-10 codes U07.1 or U07.2 were present in the record. [$]Ascertained from presence in CHESS dataset or HES APC record coded with primary diagnosis of U07.1 or U07.2. [†]Ascertained from CHESS records coded with ICU/HDU admission, only available for those with laboratory confirmed SARS-CoV-2. [‡]Ascertained from CHESS record coded with use of respiratory support via invasive mechanical ventilation, only available for those with laboratory confirmed SARS-CoV-2. [#]Hypertension models were adjusted for age, sex, ethnicity, socioeconomic status, body-mass index, alcohol consumption, smoking status, total cholesterol: high density lipoprotein cholesterol ratio, family history of coronary heart disease, treatment with corticosteroids, antiplatelets, or anticoagulants, diagnosis of atrial fibrillation, migraine, diabetes, chronic kidney disease stage 3-5, chronic liver disease, chronic lung disease, asthma, severe mental illness, dementia, chronic neurological disease, learning disability, or malignancy, and

treatment or diagnosis of a immunosuppressive condition; and QRISK3 models were adjusted for alcohol consumption, treatment with antiplatelets or anticoagulants, diagnosis of chronic liver disease, chronic lung disease, asthma, dementia, chronic neurological disease, learning disability, or malignancy, and treatment or diagnosis of a immunosuppressive condition (with are not included in the QRISK3 algorithm).

Supplementary table 6. Hazard ratios for the effect of raised cardiovascular risk on severe outcomes after laboratory-confirmed SARS-CoV-2 stratified by wave of the pandemic

	Wave 1				Wave 2			
	N events	Rate (95% CI) per 1,000 person-years	Crude HR (95% CI)	Fully-adjusted[#] HR (95% CI)	N events	Rate (95% CI) per 1,000 person-years	Crude HR (95% CI)	Fully-adjusted[#] HR (95% CI)
COVID-19 death*								
QRISK3 ≥10%	1249	356.0 (336.8-376.3)	11.10 (9.61-12.83)	7.04 (5.84-8.50)	934	265.1 (248.6-282.6)	18.28 (15.39-21.72)	9.32 (7.52-11.55)
QRISK3 <10%	215	28.3 (24.7-32.3)	1 (ref)	1 (ref)	150	14.6 (12.5-17.2)	1 (ref)	1 (ref)
Hypertension	774	185.2 (172.6-198.8)	1.81 (1.64-2.01)	1.05 (0.90-1.22)	610	130.3 (120.3-141.0)	2.53 (2.24-2.85)	1.14 (0.95-1.36)
No hypertension	690	99.6 (92.4-107.3)	1 (ref)	1 (ref)	474	52.1 (47.6-57.0)	1 (ref)	1 (ref)
ICU admission[†]								
QRISK3 ≥10%	561	154.3 (141.5-168.3)	2.74 (2.38-3.14)	2.88 (2.41-3.44)	315	87.5 (78.3-97.9)	4.75 (3.96-5.69)	4.02 (3.20-5.05)
QRISK3 <10%	374	45.5 (40.9-50.6)	1 (ref)	1 (ref)	199	18.7 (16.2-21.5)	1 (ref)	1 (ref)
Hypertension	505	114.5 (104.4-125.5)	1.86 (1.62-2.12)	1.24 (1.01-1.53)	263	54.5 (48.2-61.7)	2.08 (1.75-2.48)	1.14 (0.88-1.46)
No hypertension	430	58.3 (52.8-64.3)	1 (ref)	1 (ref)	251	26.8 (23.6-30.4)	1 (ref)	1 (ref)
Respiratory support[‡]								
QRISK3 ≥10%	374	100.0 (89.9-111.2)	2.63 (2.22-3.11)	2.83 (2.28-3.51)	124	35.0 (29.3-41.8)	5.59 (4.14-7.55)	4.81 (3.32-6.98)
QRISK3 <10%	255	30.8 (27.0-35.0)	1 (ref)	1 (ref)	65	6.3 (5.0-8.1)	1 (ref)	1 (ref)
Hypertension	349	78.0 (69.9-87.1)	1.99 (1.68-2.35)	1.31 (1.02-1.68)	103	21.8 (18.0-26.5)	2.36 (1.77-3.14)	1.08 (0.72-1.60)
No hypertension	280	37.1 (32.8-42.0)	1 (ref)	1 (ref)	86	9.5 (7.7-11.7)	1 (ref)	1 (ref)
Hospitalisation[$]								
QRISK3 ≥10%	2870	1299.2 (1252.5-1347.6)	3.10 (2.92-3.29)	2.83 (2.61-3.06)	3677	1152.3 (1115.6-1190.1)	4.41 (4.19-4.64)	3.36 (3.15-3.58)
QRISK3 <10%	1687	257.1 (245.2-269.7)	1 (ref)	1 (ref)	2560	255.6 (245.9-265.7)	1 (ref)	1 (ref)
Hypertension	2261	737.0 (707.3-768.0)	1.57 (1.48-1.67)	1.08 (0.99-1.18)	3064	696.3 (672.0-721.4)	1.96 (1.86-2.05)	1.07 (1.00-1.16)
No hypertension	2296	402.7 (386.5-419.5)	1 (ref)	1 (ref)	3173	360.4 (348.0-373.1)	1 (ref)	1 (ref)
Major adverse cardiovascular event								
QRISK3 ≥10%	304	89.1 (79.6-99.7)	5.74 (4.58-7.18)	5.04 (3.77-6.74)	266	75.9 (67.3-85.6)	7.69 (6.12-9.66)	5.21 (3.93-6.90)
QRISK3 <10%	102	13.5 (11.1-16.4)	1 (ref)	1 (ref)	102	10.0 (8.2-12.1)	1 (ref)	1 (ref)
Hypertension	239	58.3 (51.4-66.2)	2.31 (1.90-2.82)	1.62 (1.20-2.19)	211	45.2 (39.5-51.7)	2.66 (2.16-3.27)	1.42 (1.03-1.95)
No hypertension	167	24.3 (20.9-28.2)	1 (ref)	1 (ref)	157	17.3 (14.8-20.2)	1 (ref)	1 (ref)

*Ascertained from ONS death certificate data in which the COVID related ICD-10 codes U07.1 or U07.2 were present in the record. [$]Ascertained from presence in CHESS dataset or HES APC record coded with primary diagnosis of U07.1 or U07.2. [†]Ascertained from CHESS records coded with ICU/HDU admission, only available for those with laboratory confirmed SARS-CoV-2. [‡]Ascertained from CHESS record coded with use of respiratory support via invasive mechanical ventilation, only available for those with laboratory confirmed SARS-CoV-2. [#]Hypertension models were adjusted for age, sex, ethnicity, socioeconomic status, body-mass index,

alcohol consumption, smoking status, total cholesterol: high density lipoprotein cholesterol ratio, family history of coronary heart disease, treatment with corticosteroids, antiplatelets, or anticoagulants, diagnosis of atrial fibrillation, migraine, diabetes, chronic kidney disease stage 3-5, chronic liver disease, chronic lung disease, asthma, severe mental illness, dementia, chronic neurological disease, learning disability, or malignancy, and treatment or diagnosis of a immunosuppressive condition; and QRISK3 models were adjusted for alcohol consumption, treatment with antiplatelets or anticoagulants, diagnosis of chronic liver disease, chronic lung disease, asthma, dementia, chronic neurological disease, learning disability, or malignancy, and treatment or diagnosis of a immunosuppressive condition (with are not included in the QRISK3 algorithm).

Supplementary table 7. Hazard ratios for the effect of raised cardiovascular risk on severe outcomes after clinically reported COVID-19

	N events	Rate (95% CI) per 1,000 person-years	Crude HR (95% CI)	Age- and sex-adjusted HR (95% CI)	Fully-adjusted[#] HR (95% CI)
COVID-19 death*					
QRISK3 ≥10%	286	34.6 (30.8-38.8)	13.82 (9.89-19.31)	NA	7.39 (4.83-11.31)
QRISK3 <10%	39	2.5 (1.8-3.5)	1 (ref)	NA	1 (ref)
Hypertension	162	17.6 (15.1-20.6)	1.61 (1.29-2.00)	0.82 (0.66-1.03)	1.07 (0.75-1.53)
No hypertension	163	11.2 (9.6-13.1)	1 (ref)	1 (ref)	1 (ref)
Hospitalisation[$]					
QRISK3 ≥10%	1,312	168.7 (159.8-178.1)	2.57 (2.36-2.79)	NA	2.05 (1.84-2.28)
QRISK3 <10%	965	64.4 (60.5-68.6)	1 (ref)	NA	1 (ref)
Hypertension	1,087	124.3 (117.2-132.0)	1.49 (1.37-1.61)	1.16 (1.07-1.26)	1.00 (0.88-1.13)
No hypertension	1,190	84.9 (80.2-89.8)	1 (ref)	1 (ref)	1 (ref)
Major adverse cardiovascular event					
QRISK3 ≥10%	531	65.4 (60.1-71.2)	6.81 (5.67-8.18)	NA	5.09 (4.06-6.37)
QRISK3 <10%	147	9.6 (8.1-11.2)	1 (ref)	NA	1 (ref)
Hypertension	400	44.1 (40.0-48.7)	2.32 (1.99-2.70)	1.50 (1.28-1.75)	1.26 (1.00-1.58)
No hypertension	278	19.3 (17.1-21.7)	1 (ref)	1 (ref)	1 (ref)

*Ascertained from ONS death certificate data in which the COVID related ICD-10 codes U07.1 or U07.2 were present in the record

[$]Ascertained from presence in CHESS dataset or HES APC record coded with primary diagnosis of U07.1 or U07.2

[#]Hypertension models were adjusted for age, sex, ethnicity, socioeconomic status, body-mass index, alcohol consumption, smoking status, total cholesterol: high density lipoprotein cholesterol ratio, family history of coronary heart disease, treatment with corticosteroids, antiplatelets, or anticoagulants, diagnosis of atrial fibrillation, migraine, diabetes, chronic kidney disease stage 3-5, chronic liver disease, chronic lung disease, asthma, severe mental illness, dementia, chronic neurological disease, learning disability, or malignancy, and treatment or diagnosis of a immunosuppressive condition; and QRISK3 models were adjusted for alcohol consumption, treatment with antiplatelets or anticoagulants, diagnosis of chronic liver disease, chronic lung disease, asthma, dementia, chronic neurological disease, learning disability, or malignancy, and treatment or diagnosis of a immunosuppressive condition (with are not included in the QRISK3 algorithm).

Supplementary table 8. Hazard ratios for the effect of raised cardiovascular risk with refined QRISK3 score categories on severe outcomes after laboratory-confirmed SARS-CoV-2

	N events	Rate (95% CI) per 1,000 person-years	Crude HR (95% CI)	Fully-adjusted[#] HR (95% CI)
COVID-19 death*				
QRISK3 ≥20%	1,589	561.7 (534.7-590.0)	30.81 (27.49-34.52)	15.15 (13.05-17.59)
QRISK3 10-<20%	594	141.3 (130.4-153.2)	7.23 (6.35-8.24)	5.32 (4.54-6.23)
QRISK3 <10%	365	20.4 (18.4-22.6)	1 (ref)	1 (ref)
ICU admission[†]				
QRISK3 ≥20%	411	137.2 (124.0-151.8)	5.07 (4.45-5.79)	4.21 (3.54-5.02)
QRISK3 10-<20%	465	108.5 (98.8-119.1)	3.76 (3.31-4.27)	3.40 (2.91-3.96)
QRISK3 <10%	573	29.9 (27.5-32.6)	1 (ref)	1 (ref)
Respiratory support[‡]				
QRISK3 ≥20%	230	75.9 (66.3-86.8)	5.06 (4.24-6.04)	4.63 (3.68-5.83)
QRISK3 10-<20%	268	60.9 (53.8-68.9)	3.82 (3.23-4.52)	3.31 (2.70-4.07)
QRISK3 <10%	320	16.6 (14.8-18.7)	1 (ref)	1 (ref)
Hospitalisation[$]				
QRISK3 ≥20%	3,769	1903.1 (1843.3-1964.8)	6.73 (6.44-7.04)	4.77 (4.50-5.07)
QRISK3 10-<20%	2,778	812.4 (782.7-843.1)	3.01 (2.87-3.16)	2.70 (2.55-2.86)
QRISK3 <10%	4,247	256.2 (248.6-264.0)	ref	ref
Major adverse cardiovascular event				
QRISK3 ≥20%	343	124.2 (111.7-138.0)	11.54 (9.71-13.73)	8.11 (6.46-10.17)
QRISK3 10-<20%	227	54.6 (48.0-62.2)	4.92 (4.07-5.94)	4.10 (3.27-5.14)
QRISK3 <10%	204	11.5 (10.0-13.1)	ref	ref

*Ascertained from ONS death certificate data in which the COVID related ICD-10 codes U07.1 or U07.2 were present in the record
[$]Ascertained from presence in CHESS dataset or HES APC record coded with primary diagnosis of U07.1 or U07.2
[†]Ascertained from CHESS records coded with ICU/HDU admission, only available for those with laboratory confirmed SARS-CoV-2
[‡]Ascertained from CHESS record coded with use of respiratory support via invasive mechanical ventilation, only available for those with laboratory confirmed SARS-CoV-2
[#]Adjusted for alcohol consumption, treatment with antiplatelets or anticoagulants, diagnosis of chronic liver disease, chronic lung disease, asthma, dementia, chronic neurological disease, learning disability, or malignancy, and treatment or diagnosis of a immunosuppressive condition (with are not included in the QRISK3 algorithm)