Personalised approaches to antithrombotic therapies: insights from linked electronic health records

Laura Pasea

A thesis submitted for the degree of

Doctor of Philosophy

University College London

The Farr Institute for Health Informatics Research

UCL Institute of Health Informatics

December 2017

Declaration

I, Laura Pasea confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Acknowledgements

I would like to thank my supervisors, Professor Harry Hemingway and Dr Sheng-Chia Chung for their invaluable help and support throughout my PhD.

Abstract

Antithrombotic drugs are increasingly used for the prevention of atherothrombotic events in cardiovascular diseases and represent a paradigm for the study of personalised medicine because of the need to balance potential benefits with the substantial risks of bleeding harms. To be effective, personalised medicine needs validated prognostic risk models, rich phenotypes, and patient monitoring over time.

The opportunity to use linked electronic health records has potential advantages; we have rich longitudinal data spanning patients' entire journey through the healthcare system including primary care visits, clinical biomarkers, hospital admissions, hospital procedures and prescribed medication. Challenges include structuring the data into research-ready format and accurately defining clinical endpoints and handling missing data.

The data used in this thesis was from the CALIBER platform: linked routinely-collected electronic health records from general practices, hospitals admissions, myocardial infarction registry and death registry for 2 million patients in England from 1997 to 2010.

In this thesis I (1) developed comprehensive bleeding phenotypes in linked electronic health records, (2) assessed the incidence and prognosis of bleeding in atrial fibrillation and coronary disease patients in England, (3) developed and validated prognostic models for atherothrombotic and bleeding events in stable myocardial infarction survivors pertaining to the benefits and harms of prolonged dual antiplatelet therapy, (4) assessed the predictors and outcomes associated with time in therapeutic range for patients treated with oral anticoagulants (5) assessed the predictive value of novel measures of international normalised ratio control in patients treated with oral anticoagulants for atherothrombotic and bleeding outcomes.

Taken together these findings offer researchers scalable methodological approaches, that may be applied to other diseases and treatments with crucial benefits and harms considerations, and demonstrates how records used in clinical practice maybe harnessed to improve treatment decisions, monitoring and overall care of a cardiovascular disease population treated with a class of drugs.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Explanation
ACS	Acute coronary syndromes
ADP	Adenosine diphosphate
AF	Atrial fibrillation
ATT	Antithrombotic therapies
BMI	Body mass index
BNF	British National Formulary
CABG	Coronary artery bypass graft
CAD	Coronary artery disease
CALIBER	Clinical research using LInked Bespoke studies and Electronic health Records
CI	Confidence interval
COPD	Chronic obstructive pulmonary disease
CPRD	Clinical Practice Research Datalink
CV	Cardiovascular
CVD	Cardiovascular disease
DAPT	Dual antiplatelet therapy
DBP	Diastolic blood pressure
eGFR	estimated Glomerular Filtration Rate
EHR	Electronic health records
ESC	European Society of Cardiology
GI	Gastrointestinal
GPRD	General Practice Research Database
HDL	High density lipoproteins
HES	Hospital Episode Statistics
HR	Hazard ratio
ICD-10	International Statistical Classification of Diseases and Related Health Problems (version 10)
IMD	Index of multiple deprivation
INR	International normalised ratio
IQR	Interquartile range
ISAC	Independent Scientific Advisory Committee
LMWH	Low molecular weight heparin
MI	Myocardial infarction
MICE	Multiple Imputation using Chained Equations
MINAP	Myocardial Ischaemia National Audit Project
NICE	National Institute for Health and Care Excellence
NOAC/ DOAC	Novel oral anticoagulants/ Direct oral anticoagulants
NOS	Not otherwise specified

Abbreviation	Explanation
NSAID	Nonsteroidal anti-inflammatory drugs
NSTEMI	Non-ST elevated myocardial infarction
OAC	Oral anticoagulants
ONS	Office of National Statistics
OPCS-4	Classification of Interventions and Procedures version 4
PAD	Peripheral arterial disease
PCI	Percutaneous coronary intervention
PPV	Positive predictive value
SBP	Systolic blood pressure
SCAD	Stable coronary artery disease
SD	Standard deviation
STEMI	ST elevated myocardial infarction
TIA	Transient ischaemic attack
TTR	Time in therapeutic range
VKA	Vitamin K antagonists
VTE	Venous thromboembolism

1 Introduction

1.1 Background

1.1.1 Antithrombotic therapies and their uses

Antithrombotic therapies are a class of drugs which act to reduce blood clot (thrombus) formation. Blood clots limit the flow of blood through arteries and veins which can lead to conditions such as stroke. Blood clots are formed through the combination of two components: platelets (a type of blood cell) and proteins called coagulation factors. There are two types of antithrombotic therapy, each with a different mechanism of action:

- **Antiplatelets** which act to inhibit the production of a chemical called thromboxane which is used to call for platelets to gather at the site of an injury. Common antiplatelets include aspirin and clopidogrel.
- **Anticoagulants** which act to prevent the production of coagulation factors, for example through inhibiting Vitamin K, which is required to produce coagulation factors. Common anticoagulants include warfarin and heparin.

There are a wide range of indications for antithrombotic therapies as many conditions involve risks of blood clotting. The most common indication is for the prevention of primary and secondary cardiovascular disease. In 2014 in the UK the prevalence of atrial fibrillation (AF) was around 2%, and coronary artery disease (CAD) is prevalent in around 5.5% of men and 2.5% of women.[1] Cardiovascular disease is among the top causes of death in the United Kingdom (UK), accounting for around 30% of deaths.[1] Antiplatelets are among the most commonly prescribed treatments for preventing cardiovascular disease, behind lipid regulating drugs (statins) and antihypertensive drugs, whilst prescribing of anticoagulants is also increasing.[1]

Within the classes of antithrombotic therapies there are a large and increasing number of different agents approved for medical use. Furthermore, there are a variety of antithrombotic therapy treatment strategies. Some conditions are indicated for short term treatment (e.g. venous thromboembolism (VTE)) while others are indicated for lifelong treatment (e.g. AF, acute coronary syndromes (ACS)). Combinations of antithrombotic agents, such as dual antiplatelet therapy and triple therapy, are also indicated in some instances, for patients at greatest atherothrombotic risk.

1.1.2 Antithrombotic therapies: side effects and monitoring considerations

Antithrombotic therapy is associated with an increased risk of bleeding.[2] Antithrombotic therapy related bleeding is a heterogeneous side effect and can range in severity, from

manageable nosebleeds to life-changing or fatal intracranial bleeding. There are a number of bleeding classification schemes used in clinical trials which describe minor and major bleeding, and in some cases a third middle category, such as moderate bleeding. However there are variations between classifications in how bleeding severity is defined and efforts have been made to standardise bleeding definitions.[3] It is important that in studies of antithrombotic therapies, the risks of different bleeding severities are accounted for and a consensus in bleeding definition would allow for easier interpretation and comparability of results between studies.

Vitamin K antagonists (VKAs) are a group of oral anticoagulants commonly prescribed to patients with atrial fibrillation for stroke prevention. VKAs are sensitive to interactions with concomitant treatments (e.g. aspirin) and dietary habits (e.g. vitamin K rich food) and therefore require frequent monitoring and dosage adjustments to ensure the treatment is maintained within the limits of a narrow therapeutic range. If patients fall out of the therapeutic range the treatment may be ineffective to prevent thrombotic risks or may increase the risk of bleeding. The international normalised ratio (INR) is the test used to assess the therapeutic level of patients treated with VKAs, by measuring the time taken for blood to clot. The time in therapeutic range (TTR), calculated using longitudinal INR records, is a commonly used measure for assessing the overall level of patients' VKA treatment control.

1.1.3 Antithrombotic therapies: balancing benefits and harms

Antithrombotic therapies are an important class of drugs; they can prevent potential fatal or life-changing atherothrombotic events in high risk populations but can equally pose serious bleeding harms. There is potential for an iatrogenic epidemic of bleeding if treatment decisions considering potential benefits and harms are not carefully considered. Therefore it is important that clinicians and patients are able to weigh up potential benefits versus potential harms when deciding on treatment strategy.

Current efforts to consider benefits and harms as part of treatment decisions exist for the atrial fibrillation population. Clinical guidelines[4,5] recommend using the CHA_2DS_2-VASc score[6] to assess stroke risk and the HAS-BLED score[7] to assess bleeding risk in the atrial fibrillation patients to determine whether anticoagulation therapy is suitable and whether patients may require extra monitoring for bleeding.

1.1.4 CALIBER linked electronic health records

In this thesis the data used to address the aims and objectives is from the CALIBER[8] (Clinical research using Linked Bespoke studies and Electronic health Records), linked electronic health records platform. CALIBER contains linkages between primary care general practices (Clinical

Practice Research Datalink (CPRD)), hospital admissions (Hospital Episode Statistics (HES)), myocardial infarction disease registry (Myocardial Ischaemic National Audit Project (MINAP)) and cause of death (Office for National Statistics (ONS)) data sources. Primary care data from CPRD contains patient demographic data, diagnoses and procedures coded using Read terms, clinical biomarkers and prescribed drugs. Hospital admissions data from HES includes diagnoses coded using ICD-10 terms and procedures coded using OPCS-4 terms. MINAP is national registry for hospitalised acute coronary syndrome (ACS) events covering 230 hospitals in England and Wales. ONS mortality data includes underlying cause of death and up to 15 secondary causes of death coded using ICD-9 and ICD-10 term.

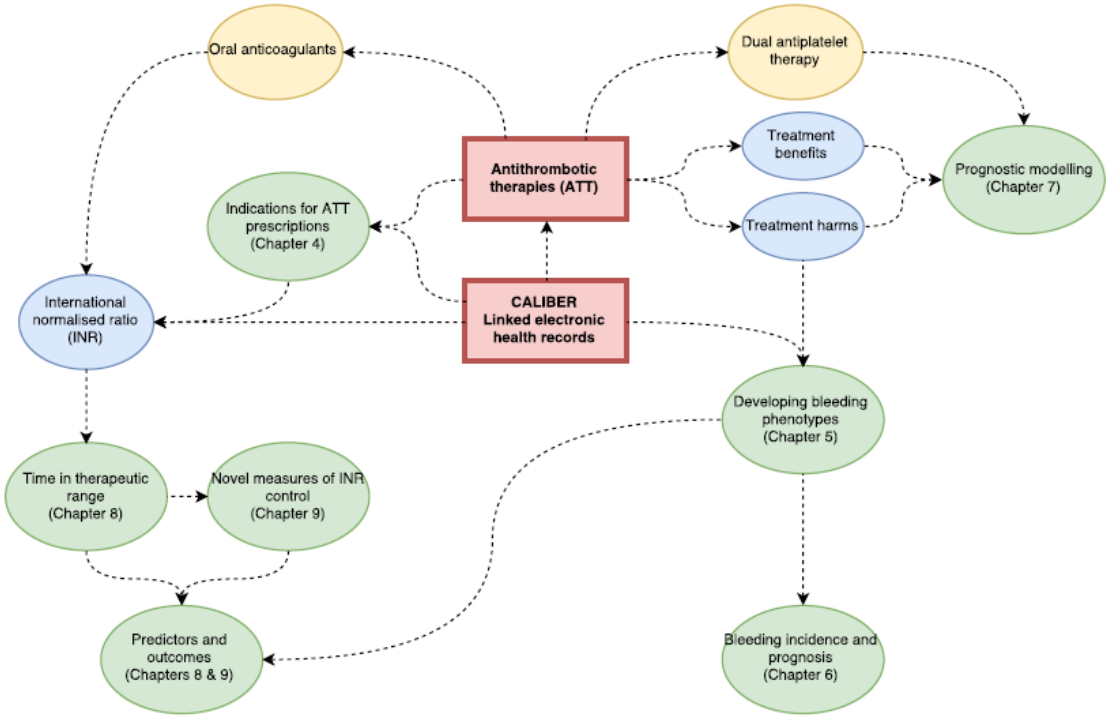
Using CALIBER the platform, reproducible high-resolution phenotypes which utilise records across the linked data sources have been developed and validated for numerous diseases and conditions, including MI and AF. However bleeding phenotypes relevant to antithrombotic use have yet to be developed.

The CALIBER data used in this thesis contains a sample of 2 million patients in England, representative of the general population, registered at CPRD general practices which consented to linkage between 1997 and 2010.

1.2 Thesis objectives

In this section I described the main analytical objectives of the thesis. A mind map of the themes and topics explored in this thesis are shown in **Figure 1.1**.

Figure 1.1: Themes and topics explored in this thesis



Chapter 4: Indications for ATT prescriptions in electronic health records

Within primary care electronic health records there is rich prescribing data, however indications for prescriptions are not captured.

The aims of this chapter were (1) to investigate the prevalence of antithrombotic therapy prescribing in CALIBER; (2) to determine how many of patients first antithrombotic prescription had an indication recorded within a reasonable timeframe.

Chapter 5: Developing bleeding phenotypes using linked electronic health records

Bleeding is the main side effect associated with antithrombotic therapies. Bleeding varies in its manifestation for instance, with respect to anatomical site and severity. As identified in clinical trials and prospective observational studies it is important to have standardised definitions and classifications for bleeding events.

Studies of bleeding relating to antithrombotic use in using EHRs have produced a number of differing bleeding phenotypes. The phenotypes are not comprehensive, for example they use a single healthcare setting (e.g. hospitalisations only) or only use bleeding terms relating to particular anatomical sites (e.g. gastrointestinal) or do not distinguish between bleeding events of different severities. As for studies in clinical settings, it is important to have comprehensive, reproducible and standardised definitions for bleeding for use in EHR to allow for scientific replication and ease comparability between different EHR studies.

Therefore the objectives in this chapter were (1) to determine the extent that population-based EHRs can replicate the definitions of bleeding used in trials; (2) to determine if bleeding severity can be defined within EHRs using information on bleeding diagnosis, anatomical site, fatality, length of stay, haemoglobin, transfusion, endoscopy, surgical interventions across primary and secondary care.

Chapter 6: Population-based estimates of bleeding incidence and prognosis

There are increasing numbers of patients with common heart diseases treated with one or a combination of antithrombotic therapies. The population of burden of bleeding is unknown and there are no studies which compare bleeding across common cardiovascular disorders. Furthermore, bleeding definitions differ across studies.

The objectives of this chapter were (1) to estimate the incidence of major bleeding and any bleeding across patients with atrial fibrillation, acute myocardial infarction, unstable and stable

angina who are on different antiplatelet and anticoagulation regimens; (2) to assess time trends in bleeding incidence along with the changes in antithrombotic management (3) to evaluate the association between major and minor bleeding and long term prognosis in terms of all-cause mortality, a composite of cardiovascular mortality, MI and stroke and recurrent bleeding.

Chapter 7: Developing and validating prognostic models for benefits and harms of prolonged dual antiplatelet therapy in myocardial infarction survivors

Among patients who survived a year since their last acute myocardial infarction (MI), the risks of subsequent major cardiovascular events, all-cause mortality and major bleeding are high. Recent evidence from clinical trials has shown the duration of dual antiplatelet therapy following an MI may be prolonged beyond a year, as currently suggested by guidelines, for effective reduction of atherothrombotic risks. However the decision to prolong dual antiplatelet therapy should be based on balancing patient's risks of further atherothrombotic events and bleeding. How to assess these risks is unknown as there are currently no prognostic models to assess patient risks one year following MI. Therefore I sought to (1) develop and validate prognostic models for atherothrombotic and bleeding risks using a cohort of patients who were atherothrombotic event-free one year following a MI and (2) demonstrate how these models may be used to aid treatment decisions.

Chapter 8: Time in therapeutic range: predictors and outcomes

INR control measured using TTR in atrial fibrillation patients has been well documented, and is a key part of treatment management in atrial fibrillation guidelines. However it is not known how well INR control compares between atrial fibrillation and other disease populations indicated for warfarin use.

The objectives of this chapter were (1) to describe INR and TTR data within CALIBER within 4 distinct populations indicated for VKA therapy (AF, VTE, heart valve replacement or other indication); (2) to assess patient baseline characteristics associated with time in therapeutic range in the study population and whether they agree with the conclusions made in previous literature; (3) to assess the strength of the SAME-TT_2R_2 score for predicting TTR in the study population; (4) to assess all-cause mortality, atherothrombotic and bleeding outcomes following INR spells with TTR below recommended levels (<60, 65, 70)

Chapter 9: Predictive value of measures of INR control

Extending on the previous chapter I evaluate a range of novel, alternative measures of INR control beyond TTR, including INR variability, time above therapeutic range, time below therapeutic range and INR trajectory. It is not known whether TTR or alternative measures of INR may be useful predictors in models for estimating atherothrombotic or bleeding risks. The objectives in this chapter were (1) To illustrate overall distribution of INR control within the study cohort; (2) To estimate predictors of INR variability and (3) To estimate the predictive value of a range of measures of INR control for atherothrombotic and bleeding events; Is TTR the best measure and do measures of INR control work well for predicting both bleeding and atherothrombotic endpoints and add value to standard risk factors?

Thesis structure

In this thesis firstly I will explore the literature relevant to the objectives, secondly I will describe the CALIBER data resource and the approaches and principles I used in preparing the data for analysis, thirdly I present the studies to address the thesis objectives and finally in the conclusion I will summarise key findings and implications arising from the studies, and make recommendations for further research.

2 Literature

In this chapter I described the literature that helped to shape the areas of research in my thesis. I used a non-systematic approach to identify the papers discussed, which provided a broad summary of the topics explored. As this thesis is not focussed around a single research question this scoping approach was more appropriate than carrying out multiple exhaustive systematic reviews, in which finding further older or smaller studies would be likely of little scientific benefit. The purpose of my literature reviews was to clarify and specify my research questions and understand what each individual study that I carried out added to existing knowledge.

I carried out literature reviews in the following areas:

- Indications for antithrombotic therapies and guidelines
- Electronic records and studies of antithrombotic therapies
    - Data quality
    - Phenotyping
    - CALIBER
- Antithrombotic therapies and bleeding
    - Bleeding definitions in trials
    - Bleeding definitions in electronic health records
    - Population-based risks of bleeding
- Prognostic modelling
    - Methods
    - Applications to cardiovascular diseases and bleeding
- International normalised ratio
    - Time in therapeutic range, risk factors and prognosis
    - Further measures of INR control

Methods

For each area I begun with broad search terms (e.g. for the topic of bleeding in electronic health records [“bleeding” OR “haemorrhage”] AND “electronic health records”) to identify initial studies of interest. Studies were limited to those published in English. Following the ascertainment of initial studies I carried out a process of forward and backward citation tracking i.e. looking at studies cited within and studies that have cited the works of interest.

2.1 Antithrombotic therapies

Antithrombotic therapies (ATT) are a class of drugs used to prevent the formation of, or dissolve thrombi, or clots, in the circulatory system. Blockages in arteries and veins can lead to diseases such as myocardial infarction (MI), stroke and pulmonary embolism, causing significant disability and increasing mortality to the sufferers.[9]

ATT consist of anticoagulant and antiplatelet drugs which both act to prevent clotting but with distinct mechanisms.

Anticoagulants can be divided into two groups according to administration method, parenteral and oral. The parenteral anticoagulants consist of unfractionated heparin and low molecular weight heparin (LMWH) and the oral anticoagulants consist of vitamin K antagonists (VKA's) and novel oral anticoagulants.

Novel oral anticoagulants - the newest class of antithrombotic therapy - include rivaroxaban, dabigatran and apixaban. They are advantageous over traditional anticoagulants since they do not require regular monitoring as warfarin does nor do they interact with food or drugs, which is a common drawback of warfarin treatment. However not all drugs in this class have an antidote (in 2016, idarucizumab was approved for the use of reversing the effects of dabigatran), therefore it is difficult to prevent bleeds in over-coagulated patients.[10]

Antiplatelet drugs can be grouped as cyclooxygenase inhibitors, such as aspirin, adenosine diphosphate (ADP) receptor inhibitors (or $P2Y_{12}$ inhibitors), such as clopidogrel, prasugrel and ticagrelor, glycoprotein IIB/IIIA inhibitors, such as abciximab, adenosine reuptake inhibitors, such as dipyridamole, thromboxane inhibitors and phosphodiesterase inhibitors all of which act to reduce platelet aggregation.

The primary adverse event related to antithrombotic use is bleeding which can range from minor bleeds to potentially fatal haemorrhages in the gut or brain.[2] Patients prescribed warfarin, a VKA, must undergo regular blood tests to ensure their international normalised ratio (INR) is kept within a safe range. Thus the patients are neither over-anticoagulated (i.e. at increased risk of haemorrhage), nor under-anticoagulated (i.e. at increased risk of stroke).

2.1.1 Indications for antithrombotic therapy

While antithrombotic therapies are prescribed and recommended as both preventative and therapeutic treatment for a wide range of indications **(Table 2.1)**, they are primarily used in cardiovascular diseases. Numerous large scale randomised control trials have provided evidence towards their wide use. Recent trials involving antithrombotic therapy have been held to investigate the use of novel oral anticoagulants in patients with atrial fibrillation and acute coronary syndromes.[11-16]

Additionally there are large trials to investigate the use of antiplatelets in previously unexplored cardiovascular disease populations. An example of this is the Pegasus-TIMI trial [17] which tested dual antiplatelet therapy in the survivors of myocardial infarction, whilst international treatment guidelines currently suggest patients should only be treated with aspirin monotherapy in the long term.[18] Pegasus-TIMI trial results suggested dual antiplatelet therapy reduces atherothrombotic events whilst increasing bleeding risk, clinicians and

patients will have to balance bleeding risks and survival benefits should this treatment strategy be approved by regulatory bodies. The use of prognostic risk scores is apt to optimise the balance of risk and benefit in new evidence-based treatment strategies that may improve current cardiovascular care practice.

2.1.2 Cardiovascular indications: evidence and guidelines

2.1.2.1 Atrial Fibrillation

What is atrial fibrillation?

Atrial fibrillation (AF) is the most common form of cardiac arrhythmia. AF affects 1-2% of the population and is increasing.[19] AF can be 'silent' – patients with AF may never present to hospitals therefore actual prevalence of AF may be higher. AF has increasing prevalence with age and in men. AF is associated with increased rate of death, stroke and other thromboembolic (blood clot) events, heart failure, hospitalisations, lowered quality of life, reduced exercise capacity and left ventricular dysfunction.

Oral anticoagulants are used to prevent ischemic events in individuals with AF. Prognostic scores such as $CHADS_2$[20] and CHA_2DS_2-VASc [6] are used to determine individuals risk of stroke and the HAS-BLED[7] score is used to estimate risk of bleeding. European Society of Cardiology (ESC) guidelines[21] suggest, based on evidence from clinical trials and using CHA_2DS_2-VASc scoring, oral anticoagulants (OAC) are recommended for a score of 2 or higher, either OAC (preferred) or 75-325mg aspirin for a score of 1 and either 75-325mg aspirin or no ATT (preferred) for a score of 0. Patients should be monitored for bleeding if they have a HAS-BLED score of 3 or higher.

Patients with AF and stable coronary artery disease are usually treated with a VKA (e.g. warfarin) plus an antiplatelet (e.g. aspirin or clopidogrel). Patients with AF and post-acute coronary syndrome (ACS) or undergoing percutaneous coronary intervention (PCI) patients should be treated with triple therapy (e.g. aspirin, clopidogrel and warfarin) in the short term followed by a VKA and single antiplatelet combination for the long term. The length of time a patient is recommended to have each regimen depends on their risk of bleeding (the HAS-BLED risk score is recommend to evaluate bleeding risk), whether they are elective or emergency and the type of stent used in the PCI procedure. Dual antiplatelet therapy is recommended for patients with AF and non-ST elevated myocardial infarction (NSTEMI). Patients with an increased risk of stroke should also be given an oral anticoagulant (i.e. triple therapy). Acute ST-elevated myocardial infarction (STEMI) patients should initially be given triple therapy followed by VKA plus clopidogrel for long term management.

Anticoagulation is not recommended for patients with acute stroke as it may result in intracranial haemorrhage. In AF patients presenting with acute stroke or transient ischemic attack (TIA) haemorrhage should be excluded and uncontrolled hypertension should be managed prior to anticoagulation.

The guidelines were updated in 2012[22] in light of the results of the ROCKET-AF[13] (rivaroxaban vs VKA), ARISTOTLE[14] (apixaban) and AVERROES[16] (apixaban vs. aspirin) trials. Novel oral anticoagulants (NOAC's) are recommended over warfarin, whilst taking into consideration both clinician and patient preference, for anticoagulation. NOAC's do not have food and drug interactions or require regular blood monitoring however not all of the drugs in this class currently have an antidote in the case of a bleeding adverse event. Guidelines have shifted focus to identifying truly 'low-risk' patients who do not need, benefit from or may even be harmed by any ATT. This requires stronger risk scores than the previously advocated simple $CHADS_2$. The HAS-BLED score is preferred over the complicated $HEMORR_2HAGES$ and less practical ATRIA scores for assessing bleeding risk. The 2014 NICE guideline recommends the use of CHA_2DS_2-VASc and HAS-BLED scores to assess stroke and bleeding risk respectively in AF patients.[4] Anticoagulation is recommended for patients with a CHA_2DS_2-VASc score of 2 or higher with bleeding risk also to be simultaneously considered. Anticoagulation may be with apixaban, dabigatran, rivaroxaban or a VKA depending on an individual's clinical features and preference. Antiplatelet monotherapy is not recommended as an intervention to prevent stroke in AF patients.

2.1.2.2 Acute coronary syndromes

What are acute coronary syndromes?

Acute coronary syndromes are a range of acute diseases resulting from complete blockage or reduced flow of blood through the coronary arteries and thus decreased heart function. ACS usually presents as chest pain and can be classified as one of three conditions depending on electrocardiogram (ECG) results and cardiac biomarkers (e.g. troponin):

- **ST-elevation myocardial infarction (STEMI):** If the ECG results show ST-segment elevation and severe rise in cardiac biomarkers this indicates STEMI, which is the complete blockage of a major coronary artery.
- **No ST-elevation myocardial infarction (NSTEMI):** If the ECG results show ST-segment depression and slight rise in cardiac biomarkers this indicates NSTEMI, which is the complete blockage of a minor coronary artery and is less severe compared with STEMI.
- **Unstable angina:** Unstable angina initially presents similarly to NSTEMI – the ECG shows ST-segment depression and is distinguished from NSTEMI by through the lack of

troponin elevation. An unstable angina diagnosis indicates lack of injury to the heart which would otherwise occur with an MI.

Patients with ACS are at heightened risk of stroke and further ACS events and are usually prescribed differing antiplatelet therapy depending on their type of ACS and their risk level.

Acute myocardial infarction with ST-elevation

The *ESC 2012*[23] *and NICE 2013*[24] guidelines advise prasugrel, ticagrelor or clopidogrel in addition to aspirin is recommended for patients undergoing primary PCI. In addition patients are recommended anti-thrombin treatment with heparin or LMWH if they have been treated with prasugrel or ticagrelor, or bivalirudin in combination with clopidogrel. Recent trials of novel oral anticoagulants on post ACS patients (ATLAS ACS 2-TIMI 51 [rivaroxaban] and APPRAISE-2 [apixaban] and phase ii trials with dabigatran) are acknowledged but no recommendations for using this class of drugs as a whole in this population are made due to mixed results. Low dose rivaroxaban may be considered for patients with a low bleeding risk and currently use dual antiplatelet therapy.

Dual antiplatelet therapy (clopidogrel/ticagrelor and aspirin) is recommended for up to 12 months following a STEMI with a minimum of 1 month DAPT if the patient received and bare metal stent during PCI or 6 months if the patient received a drug eluting stent during PCI. Indefinite treatment with low dose aspirin (or clopidogrel for aspirin intolerant) is recommended following a STEMI. Triple therapy is recommended for patients with comorbidity such as AF (i.e. also indicated for oral anticoagulation) with minimum duration of dual antiplatelet use in order to minimise bleeding risk.

Acute coronary syndromes without ST-elevation

ESC 2011[25] and NICE 2010[26] guidelines suggest assessing ischaemic and bleeding risk in ACS patients without ST-elevation using the CRUSADE [27]and GRACE[28] scores, respectively. For the prevention of further ischaemic events, dual antiplatelet therapy (aspirin and a $P2Y_{12}$ inhibitor) is recommended for 12 months except for those at high risk of bleeding. The choice of $P2Y_{12}$ inhibitor depends on the characteristics of the patient. Ticagrelor is recommended for all patients, prasugrel is only recommended for patients who have had a PCI and not previously used clopidogrel and clopidogrel should be chosen if ticagrelor or prasugrel are contraindicated. Aspirin monotherapy should be continued indefinitely (or clopidogrel if aspirin intolerant).

2.1.2.3 Stable coronary artery disease

What is stable coronary artery disease?

Stable coronary artery disease (SCAD) is a chronic condition and includes patients who have stable angina or those who have had a past ACS and have remained stable symptom-wise. Stable angina presents as chest tightness or pain and is caused by the narrowing of coronary arteries and thus insufficient blood supply to the heart muscle. SCAD patients are at high risk of further coronary events and often have long-term antiplatelet treatment to reduce their risk.

ESC 2013[18] and NICE 2011[29] guidelines suggest aspirin (or clopidogrel in aspirin intolerant patients) is recommended for the prevention of ischaemic events in those with SCAD, but it is advised to consider bleeding risk and comorbidities. No cardiovascular or bleeding prognostic risk tools are recommended for this population. Dual antiplatelet therapy is not recommended for stable patients but post-hoc analyses in the CHARISMA study [30] demonstrated a potential benefit in SCAD patients with a history of MI. It is suggested that dual antiplatelet therapy may be beneficial in stable patients who remain at high risk of ischaemic events and low risk of bleeding but no recommendations are made due to lack of evidence at the time of publication.

2.1.2.4 Peripheral arterial disease

What is peripheral arterial disease?

Peripheral arterial disease is a build-up of plaque in arteries carrying blood to limbs and organs which restricts blood-flow. PAD usually refers to restricted blood flow to the legs. Patients with PAD are at increased risk of thrombosis, embolism and stroke.

ESC 2011[31] guidelines state antiplatelet therapy with clopidogrel is recommended for the prevention of ischaemic events in patients with symptoms of PAD. Dual antiplatelet therapy is not recommended due to it not showing large enough increase in benefit relative to the increased bleeding risk when compared with antiplatelet monotherapy. Patients undergoing revascularisation with carotid artery stenting should be given dual antiplatelet therapy with clopidogrel and aspirin. Patients with SCAD and PAD should consider clopidogrel over aspirin for long-term antiplatelet therapy.

2.1.2.5 Cardiovascular disease prevention

There is an increasing prevalence in cardiovascular disease (CVD) and it is a leading cause of early death. Known risk factors for developing CVD include medical risk factors such as diabetes, renal disease and hypertension and lifestyle factors such as weight, smoking and alcohol consumption.

Antithrombotic therapies can be prescribed as preventative treatment in patients with presenting with known CVD risk factors. ESC 2012[32] guidelines suggest aspirin may be considered for hypertensive patients with no history of CVD and poor renal function but also considering their bleeding risk. Due to inconsistent study results there are no longer firm recommendations for aspirin in diabetic patients with no prior CVD. Antiplatelet therapy with clopidogrel in patients without overt CVD has not shown any significant benefit but increases bleeding risk and therefore is not recommended.

2.1.2.6 Venous thromboembolism

What is venous thromboembolism?

A venous thromboembolism (VTE) is a blood clot that has broken free and travels in the bloodstream. An example of this is where a deep vein thrombosis (DVT), a blood clot in the deep veins of the leg moves to the lungs, a pulmonary embolism (PE). Both PE and DVT can be fatal or lead to chronic conditions. Anticoagulation is recommended to treat VTE and prevent further VTE events.

NICE 2012[33] and ESC 2014[34] guidelines recommend patients with acute VTE should initially be administered with a parental anticoagulant. Low molecular weight heparins or fondaparinux are preferred to unfractionated heparins due lower bleeding risk. Oral anticoagulation should immediately follow parental anticoagulation. If VKA's are used for oral anticoagulation, patients should continue using parental anticoagulants until they reach a stable INR 2-3 for two consecutive days. Additionally, the novel oral anticoagulants, rivaroxaban, dabigatran and apixaban are approved for treating VTE. Oral anticoagulation should continue for at least 3 months following a VTE related event. Anticoagulation can be indefinite for individuals at high risk of further VTE events such as cancer patients as long as they are not at high risk of bleeding.

2.2 Electronic health records and antithrombotic research

Electronic health records are computerised collections of data. Such data may be collected for clinical care, administrative or auditing and quality assurance purposes. Information captured in electronic health records can include patient demographics, disease diagnoses and medical history, issued prescriptions, procedures and interventions and centre level information (e.g. general practices or hospitals).

Linkage between different electronic health records, for example between primary care and hospital data, can be performed using unique patient identifiers or through probabilistic methods. Data linkage enhances the use of electronic health records for research as generally

different data sets complement each other in terms of data collected or can improve case ascertainment for variables in common.

2.2.1 Assessment of data quality and validity in electronic health records

Before performing research using electronic health records it is important to understand the quality and validity of the data, particularly in routinely collected or administrative patient data not specifically intended for research purposes. Research performed in poor quality data could potentially misinform if the limitations of the data are not addressed and interpreted correctly.

The gold standard method for validating electronic health records is to compare them with the original patient medical charts. Herrett et al performed a systematic review of validation studies for diagnoses recorded in the General Practice Research Database (GPRD).[35] The authors identified 212 studies in which 183 disease diagnoses were validated using a variety of methods. Internal validation methods included 1) diagnostic algorithms where diagnoses are made based on the presence of certain codes including supporting data such as symptoms, prescriptions and test results, 2) manual review of computerised free text, 3) sensitivity analysis in which disease definitions with broad criteria were compared with definitions with more restricted criteria. External validation methods included 1) sending questionnaires to general practitioners, 2) requesting paper medical records from general practitioners, 3) comparison of rates disease prevalence, incidence with known estimated from other sources. The most common validation methods were requesting additional information from general practitioners (160/357 validations) and comparison of rates (143/357 validations). The most common quantitative measure for assessing validity of disease diagnoses was the positive predictive value (PPV); the proportion of cases identified in GPRD that were found to be true cases. Disease diagnoses in GPRD were generally good, a median of positive predictive value of 89% (range: 24-100%); however certain diseases had the tendency to be over or under reported in GPRD.

Herrett et al[36] sought to validate myocardial infarction diagnosis records captured in CALIBER (linked data from CPRD (primary care), HES (hospital admissions) and MINAP (national myocardial infarction disease registry)). The authors demonstrated that no single data source provided complete coverage of myocardial infarction events - only a third of events were captured in primary care, hospital admissions and disease registry data. Therefore the incidence of myocardial infarction events is underestimated when only a single data source was used.

Rodriguez et al[37] reviewed the use of GPRD for pharmacoepidemiology studies. The authors cited the fact that GPRD is a large population based and representative sample of the UK to be

a major advantage. Aspects of GPRD identified for improvement at the time of the study included improved capture of socioeconomic data and over-the-counter medication use, linkage to further healthcare data resources and wider use of GPRD data in research groups. The authors identified five key areas for which the GPRD may be applied to pharmacoepidemiology: 1) Studies of drug utilisation within disease groups, 2) drug safety surveillance, 3) drug efficacy estimation, 4) improving the balance of benefits and harms of drugs and 5) to perform economic evaluation of treatments.

2.2.2 Phenotype development using electronic health records

Studies have demonstrated methodology for developing disease phenotypes in electronic health records. Morley et al[38] developed atrial fibrillation phenotypes, in linked primary care and hospital admissions data. Further atrial fibrillation cases were inferred in patients without atrial fibrillation diagnosis records in primary or secondary care, through records of prescribed anticoagulation or antiarrhythmic treatments whilst excluding for other indications for treatment. No single data source had complete coverage of atrial fibrillation diagnosis and 40% of atrial fibrillation cases were captured in both primary and secondary care data. The characteristics of patients in terms of age, gender and comorbidities varied across atrial fibrillation patients depending on the source of their diagnosis record, primary care only, secondary care only or both primary and secondary care.

2.2.3 CALIBER

The CALIBER[8] (Clinical research using Linked Bespoke studies and Electronic health Records) platform is a series of linkages between UK electronic healthcare databases: Clinical Practice Research Datalink (CPRD) a longitudinal primary care database, Hospital Episode Statistics (HES) a database of hospital admissions and procedures, Myocardial Ischaemia National Audit Project (MINAP) a national acute coronary disease registry and the Office of National Statistics (ONS) for cause-specific mortality and social deprivation data.

2.2.4 Exemplar CALIBER studies

The scope of research performed using CALIBER is wide ranging. A core set of studies have assessed the associations between risk factors, including socio-economic deprivation,[39] blood pressure,[40] gender,[41] smoking status,[41] type 2 diabetes,[42] rheumatoid arthritis,[43] alcohol consumption,[44] and neutrophil count,[45] and the incidence of twelve cardiovascular diseases (stable angina, unstable angina, myocardial infarction, unheralded coronary death, cardiac arrest, heart failure, transient ischaemic attack, ischaemic stroke, subarachnoid haemorrhage, intracerebral haemorrhage, peripheral arterial disease and abdominal aortic aneurysm). These studies have emphasised the validity of the phenotyping of risk factors and cardiovascular

diseases within in the CALIBER platform, whilst also providing useful results that can aid and inform policies, approaches to clinical screening and risk factors for prognostic models.

2.3 Bleeding and antithrombotic therapies

2.3.1 Background

Bleeding is the most common adverse effect of antithrombotic therapy.[2,46] Antithrombotic therapies increase risk of bleeding through reduced platelet activity in the case of antiplatelet drugs and reduced clotting ability with oral anticoagulants. There is a wide range of severity of bleeding events associated with antithrombotic therapy use - for example, from minor nosebleeds to fatal intracranial haemorrhage. Given the heterogeneity of bleeding events it is important to have clear definitions for bleeding in research. It is also important to understand the incidence of bleeding in populations of patients prescribed antithrombotic therapies, and whether these events may be prevented through personalised treatment decisions that balance patients' bleeding risks with their atherothrombotic risks.

2.3.2 Bleeding classification in randomised trials and prospective cohort studies

Due to the heterogeneity in the manifestation of bleeding events, there have been a number of bleeding definitions used in trials. The components used in bleeding classifications used to define severity are shown in **Table 2.2.** In trials and prospective studies bleeding severity is generally adjudicated by a number of study personnel at the time of the event.

The Thrombolysis in Myocardial Infarction (TIMI) bleeding classification was developed in conjunction with trials conducted by the TIMI study group, which primarily investigated interventions for patients with ACS. The original TIMI bleeding definition[47] classified bleeding events as major if they were intracranial or resulted in a haemoglobin drop >5g/dL. Bleeding events were classified as minor if it was spontaneous and observed as gross haematuria or hematemesis or had a haemoglobin drop >3g/dL. The current iteration of TIMI bleeding[48] has three levels of severity, major, minor and minimal. Major TIMI bleeding events are fatal, intracranial, or bleeding with a haemoglobin drop ≥5g/dL. Minor TIMI bleeding events are bleeding with a haemoglobin drop 3 to 5g/dL, requires intervention to stop bleeding, leads to hospitalisation or prompts medical evaluation. Minimal TIMI bleeding events are any that do not meet the criteria of major or minor. In the TRITON-TIMI 38 trial[49] which compared the use of prasugrel versus clopidogrel in ACS patients the risk of TIMI major bleeding was 2.4% for patients randomised to prasugrel and 1.8% for patients randomised to clopidogrel. In the PEGASUS-TIMI 54 trial[17] which assessed the use of prolonged dual antiplatelet therapy (aspirin and ticagrelor) in myocardial infarction survivors the risk of TIMI major bleeding was 2.3% for patients randomised to ticagrelor 60mg and 1.06% for patients randomised to placebo. TIMI

minor bleeding was reported in 1.18% patients randomised to ticagrelor 60mg and 0.36% for patients randomised to placebo.

The Global Utilization of Streptokinase and Tissue Plasminogen Activator for Occluded Coronary Arteries (GUSTO) bleeding definition[50] classifies bleeding as severe or life-threatening, moderate or mild. Severe or life-threatening events include intracerebral haemorrhage or bleeding resulting in haemodynamic compromise and moderate bleeding events are those that require transfusion with no haemodynamic compromise. All other bleeding events are classified as mild. GUSTO is the only bleeding classification assessed here that does not account for haemoglobin drop **(Table 2.2).** In the GUSTO trial randomised patients with acute MI to streptokinase and subcutaneous heparin, streptokinase and intravenous heparin, t-PA and intravenous heparin or streptokinase, t-PA and intravenous heparin. 30 day GUSTO severe or life-threatening bleeding occurred in 0.3%, 0.5%, 0.4% and 0.6% of patients in the streptokinase and subcutaneous heparin, streptokinase and intravenous heparin, t-PA and intravenous heparin and streptokinase, t-PA and intravenous heparin groups respectively. 30 day GUSTO moderate bleeding occurred in 5.6%, 5.8%, 5.1% and 5.6% of patients in the streptokinase and subcutaneous heparin, streptokinase and intravenous heparin, t-PA and intravenous heparin and streptokinase, t-PA and intravenous heparin groups respectively.

The ACUITY-HORIZONS bleeding classification[51,52] defines major bleeding as intracranial or intraocular haemorrhage, bleeding that requires intervention, ≥5cm haematoma, retroperitoneal, ≥4g/dL haemoglobin drop without an overt source of bleeding, ≥3g/dL haemoglobin drop with an overt source of bleeding, any transfusion or reoperation for bleeding. The Acute Catheterization and Urgent Intervention Triage Strategy (ACUITY) trial[53] randomised ACS patients to unfractionated heparin or enoxaparin plus a glycoprotein IIb/IIIa inhibitor, bivalirudin plus a glycoprotein IIb/IIIa inhibitor, or bivalirudin alone. Major and minor bleeding according to the ACUITY-HORIZONS and the TIMI bleeding classification were reported. In the unfractionated heparin or enoxaparin plus a glycoprotein IIb/IIIa inhibitor group the risk of 30 day ACUITY-HORIZONS major and minor bleeding was 5.7% and 21.6% respectively and TIMI major and minor bleeding was 1.9% and 6.4% respectively. In the bivalirudin plus a glycoprotein IIb/IIIa inhibitor group the risk of 30 day ACUITY-HORIZONS major and minor bleeding was 5.3% and 21.7% respectively and TIMI major and minor bleeding was 1.7% and 6.1% respectively. In the bivalirudin alone group the risk of 30 day ACUITY-HORIZONS major and minor bleeding was 3.0% and 12.8% respectively and TIMI major and minor bleeding was 0.9% and 3.7% respectively. The Harmonising Outcomes with Revascularization and Stents in Acute Myocardial Infarction (HORIZONS-AMI)[52] trial

randomised STEMI patients undergoing PCI to heparin plus a glycoprotein IIb/IIIa inhibitor or to bivalirudin alone. ACUITY-HORIZONS major, TIMI major, TIMI minor, and GUSTO life-threatening or severe and GUSTO moderate bleeding events after 30 days follow-up were reported. In the heparin plus a glycoprotein IIb/IIIa inhibitor group 30 day ACUITY-HORIZONS major bleeding risk was 8.3%, TIMI major bleeding 5.0%, TIMI minor bleeding 4.6%, GUSTO life threatening or severe bleeding 0.6% and GUSTO moderate bleeding 5.0%. In bivalirudin group 30 day ACUITY-HORIZONS major bleeding risk was 4.9%, TIMI major bleeding 3.1%, TIMI minor bleeding 2.8%, GUSTO life threatening or severe bleeding 0.4% and GUSTO moderate bleeding 3.1%.

The Clopidogrel in Unstable Angina to prevent Recurrent Events (CURE) bleeding classification[54] has two levels of severity, major and minor. CURE major bleeding includes fatal bleeding, intracranial haemorrhage, bleeding that requires surgical intervention, bleeding resulting in hypotension, haemoglobin drop ≥5g/dL, bleeding requiring transfusion of 2-3units and intraocular bleeding. Minor bleeding events are those that led to discontinuation of treatment. The CURE trial[54] randomised NSTE-ACS patients to clopidogrel or placebo in addition to aspirin for 3 to 12 months. In the clopidogrel group risk CURE major and minor bleeding was 3.7% and 5.1% respectively. In the placebo group risk of CURE major and minor bleeding was 2.7% and 2.4% respectively. In a study of ACS patients from CURE, OASIS, OASIS-2 studies combined the risk of major bleeding at 6 months was 2%.[55]

CURRENT-OASIS 7 bleeding classification[56] defines severe, major and minor bleeding. Severe bleeding is defined as fatal, requiring transfusion ≥4units, haemoglobin drop ≥5g/dL, leading to hypotension, requiring surgery or intracranial. Major bleeding is defined as requiring transfusion 2-3 units, disabling, intraocular. Minor bleeding refers to events requiring modification of drug regimens. The trial compared double dose versus standard dose clopidogrel and high dose versus low dose aspirin in ACS patients undergoing PCI and reported CURRENT-OASIS 7 major, severe and minor bleeding as well as TIMI major bleeding events.[57] The risk of 30 day CURRENT-OASIS 7 major bleeding, CURRENT-OASIS 7 severe bleeding, TIMI major bleeding and minor bleeding in patients randomised to double dose clopidogrel was 1.6%, 1.1%, 1.0% and 5.1% respectively. The risk of 30 day CURRENT-OASIS 7 major bleeding, CURRENT-OASIS 7 severe bleeding, TIMI major bleeding and minor bleeding in patients randomised to standard dose clopidogrel was 1.1%, 0.8%, 0.7% and 4.3% respectively. The risk of 30 day CURRENT-OASIS 7 major bleeding, CURRENT-OASIS 7 severe bleeding, TIMI major bleeding and minor bleeding in patients randomised to high dose aspirin was 1.5%, 1.1%, 0.9% and 5.0% respectively. The risk of 30 day CURRENT-OASIS 7 major bleeding, CURRENT-OASIS 7

severe bleeding, TIMI major bleeding and minor bleeding in patients randomised to low dose aspirin was 1.3%, 0.9%, 0.7% and 4.3% respectively.

The Safety and Efficacy of Enoxaparin in PCI Patients, an International Randomized Evaluation (STEEPLE) bleeding classification[58] defines major as fatal bleeding, retroperitoneal, intracranial or intraocular bleeding, bleeding that causes haemodynamic compromise, bleeding that requires surgical intervention, bleeding requiring transfusion ≥1 unit or bleeding with haemoglobin drop ≥3g/dL. STEEPLE minor bleeding is gross haematuria, prolonged epistaxis, gastrointestinal haemorrhage, haemoptysis, subconjunctival haemorrhage, haematoma ≥5cm, haemoglobin drop 2-3g/dL or uncontrolled bleeding requiring protamine sulphate. The STEEPLE trial[58] randomised patients undergoing elective PCI to enoxaparin or unfractionated heparin. As well as STEEPLE bleeding, TIMI major, TIMI minor and GUSTO moderate or severe bleeding events were reported. In the heparin group STEEPLE major bleeding risk was 2.8%, STEEPLE minor bleeding 5.9%, TIMI major bleeding 0.3%, TIMI minor bleeding 1.9% and GUSTO moderate or severe bleeding 1.5%. In the enoxaparin 0.5mg per kilogram group STEEPLE major bleeding risk was 1.2%, STEEPLE minor bleeding 4.8%, TIMI major bleeding 0.3%, TIMI minor bleeding 1.8% and GUSTO moderate or severe bleeding 0.6%.

The PLATelet inhibition and patient Outcomes (PLATO) bleeding classification[59] has three defined levels of severity, major, minor and minimal bleeding. PLATO major bleeding includes fatal, intracranial, intrapericardial with cardiac tamponade, bleeding resulting in severe hypotension, haemoglobin drop≥5g/dL, transfusion of ≥4 units or significantly disabling events. PLATO minor bleeding events are those that require medical intervention to treat or stop bleeding, 3-5g/dL haemoglobin drop or transfusion of 2-3 units. Minimal bleeding events are those that do not meet the major or minor criteria. The PLATO trial[60] randomised ACS patients to treatment with ticagrelor or clopidogrel. In the ticagrelor group the risk of PLATO and TIMI major bleeding was 4.5% and 2.8% respectively. In the clopidogrel group the risk of PLATO and TIMI major bleeding was 3.8% and 2.2% respectively.

The Global Registry of Acute Coronary Events (GRACE) bleeding classification[61] defines major bleeding as events that are fatal or result in stroke, intracerebral haemorrhage, require a transfusion of ≥2 units, or a decrease in haematocrit of≥10%. GRACE is multinational registry of ACS patients.[62] Of the 24045 patients in GRACE hospitalised for ACS, 3.9% developed major bleeding.[61] A later study of patients in the GRACE registry identified 2.8% of acute MI patients with major bleeding while hospitalised.[63]

The International Society of Thrombosis and Haemostasis (ISTH) bleeding classification[64] defines major bleeding in non-surgical patients as fatal bleeding, symptomatic bleeding in a

critical area or organ or bleeding causing a fall in haemoglobin ≥2g/dL or transfusion of 2 or more units. The ISTH bleeding endpoint has been used in a number of clinical trials involving novel oral anticoagulants in atrial fibrillation and venous thromboembolism populations. In the ARISTOTLE trial population atrial fibrillation patients were randomised to apixaban or warfarin and the primary safety endpoint was ISTH major bleeding whilst GUSTO severe and TIMI major bleeding events were also reported.[14] The risk of ISTH major bleeding was 2.13% in the apixaban group and 2.09% in the warfarin group. The risk of GUSTO severe bleeding was 0.52% in the apixaban group and 1.13% in the warfarin group. The risk of TIMI major bleeding was 0.96% in the apixaban group and 1.69% in the warfarin group. Therefore in this setting ISTH major bleeding was shown to be a less conservative measure of major bleeding than the GUSTO and TIMI criteria.

The Bleeding Academic Research Consortium (BARC) bleeding classification[3] was developed in an effort to standardise the definition of bleeding in research. The BARC bleeding classification has 5 levels of severity: Type 1 – bleeding that does not result in hospitalisation or consultation of healthcare professionals; Type 2 – overt haemorrhage with either nonsurgical intervention, hospitalisation or prompts evaluation; Type 3a – Overt bleeding with haemoglobin drop 3 to 5g/dL or transfusion; Type 3b – overt bleeding with haemoglobin drop ≥5g/dL, cardiac tamponade, or requires surgical or pharmacological intervention; Type 3c – Intracranial haemorrhage or intraocular haemorrhage; Type 4 – CABG related bleeding; Type 5 – fatal bleeding. In a study of coronary artery disease patients undergoing PCI pooled from 6 clinical trials, the overall risk of BARC type ≥ 2, BARC type≥ 3 and TIMI major bleeding 5.4%, 4.0% and 0.9% respectively.[65] In a prospective cohort study of 4149 stable coronary artery disease patients, after two years follow up the risk of BARC type≥3 was 1.2%.[66]

2.3.2.1 Comparisons of bleeding classifications

There have been studies in prospective cohorts which sought to compare the incidence of bleeding events according to the various classification schemes and analysed the risks of subsequent mortality in a form of validation.

Kikkert et al[67] followed-up a prospective cohort of 2002 STEMI patients and used the TIMI, ISTH, GUSTO and BARC classification schemes to define bleeding events. Overall, 19% patients had ISTH major bleeding, 6.3% had GUSTO moderate bleeding, 2.7% had GUSTO severe bleeding, 7.9% had TIMI minor bleeding, 5.0% had TIMI major bleeding, 4.4% had BARC type 2 bleeding, 8.5% had BARC Type 3a bleeding, 5.5% had BARC type 3b/3c bleeding, 1.3% had BARC Type 4 bleeding and 0.3% had BARC type 5 bleeding. In adjusted analysis the risk of one year mortality did not differ significantly between the levels of ISTH or GUSTO bleeding

severity. Patients with TIMI major bleeding had 2 (95% CI: 1.32, 3.01) times the risk of mortality compared with patients that had TIMI minimal or no bleeding. Patients with BARC type 3b/3c bleeding had 1.84 (95% CI: 1.23, 2.77) times the risk of mortality compared with patients who had BARC type 0 or 1 bleeding. The risk of mortality in patients who had BARC type 4 bleeding was not significantly different to those with BARC type 0 or 1. In an analysis of the individual components of the bleeding classification scheme it was found that a haemoglobin drop of ≥5 g/dL was associated with increased risk of mortality.

Vranckx et al[68] validated the BARC bleeding classification in the TRACER trial population of NSTE-ACS patients. In follow up 15.3% of the patients had BARC type≥2 bleeding, 3.7% had TIMI major or minor bleeding and 4.0% had GUSTO severe or moderate bleeding. The rate of mortality following bleeding was highest in those who has GUSTO severe bleeding (30.0% at 2 years post-bleeding), followed by BARC type 3 bleeding (% at 2 years post-bleeding) and TIMI major bleeding (22.4% at 2 years post-bleeding). For the BARC classification, patients with Type ≥ 2 bleeding had increased mortality risk up to one year post-bleeding compared with patients that did not have a bleeding event. Patients with major TIMI bleeding had increased 1 year mortality risk compared with patients that had no TIMI bleeding. Patients with moderate or severe GUSTO bleeding had increased 1 year mortality compared with patients that had no GUSTO bleeding.

Kaatz et al[69] highlighted the important of clinically relevant non-major bleeding, e.g. costs and management of such events, and sought to standardise the definition of non-major bleeding for patients treated with anticoagulants. Through reviewing trials of novel oral anticoagulants in atrial fibrillation and venous thromboembolism patients they found 47 distinct components used to define non-major bleeding. With this information they defined clinically relevant non-major bleeding as a sign or symptom of bleeding that does not meet the ISTH criteria for major bleeding[64] but requires medical intervention, leads to hospitalisation or prompts face to face evaluation.

2.3.3 Bleeding phenotyping in electronic health records

Bleeding phenotypes have been developed for use in electronic health care record systems, and there a large degree of heterogeneity was found between them, as summarised in **Table 2.3**.

Of the 8 phenotypes identified 5 were restricted to upper gastrointestinal bleeding events.[70-74] The three remaining phenotypes[75-77] each explored at least 4 anatomical sites including intracranial, genitourinary, respiratory and gastrointestinal bleeding sites.

All bleeding phenotype studies used bleeding records captured hospital admissions data, except de Abajo et al[71] who used bleeding diagnosis records from primary care general practices only. Only two studies assessed bleeding records captured in death registries.[73,77] Both primary care and hospital admissions data were used by Crooks et al[73] (linked data) and Valkhoff et al[74] (non-linked data).

Most bleeding phenotypes comprised of ICD-9 or ICD-10 codes, with the exception of the studies in UK primary care data which used Read codes and IPCI primary care data which used in IPCI codes. The numbers of codes used in the bleeding definitions was variable amongst the studies. This was particularly noticeable when comparing the studies of upper GI bleeding in hospital admissions data. To define upper GI bleeding, Raiford et al[70] used 30 ICD-9 codes, Wahl et al[72] initially used 75 codes and refined to 33 codes in a secondary analysis, Crooks et al[73] used 22 ICD-10 codes and Valkhoff et al[74] used 26 ICD-9 codes and 16 ICD-10 codes.

Three studies sought to differentiate between the severities of bleeding events. Arnason et al[75] and Cunningham et al[76] 'manually' assigned severity using information recorded on patient charts. Friberg et al[77] classified severity using information in the electronic health records. Initially, fatal bleeding included bleeding records captured in the national death registry or hospital admissions with a discharge code indicating death; non-fatal major bleeding comprised of hospitalisations with bleeding as the primary diagnosis or secondary diagnosis with a coinciding record for transfusion or intracranial bleeding; hospitalised bleeding included any extracranial bleeding event with no coinciding transfusion record and minor bleeding was defined as any extracranial bleeding without hospitalisation or transfusion. However major bleeding was later redefined to combine the non-fatal major and hospitalised bleeding categories, due to a high number of missing transfusion records leading to missed major bleeding events.

The gold standard for assessing the accuracy of electronic health records is to perform a manual review of patient charts. Most studies of bleeding phenotypes developed in the electronic health records had the facility to do this, for at least a sample of the bleeding events. The most commonly reported measure of accuracy was the positive predictive value which is the percentage of electronic bleeding records that are found to be true according to the manual chart review. The positive predictive value for bleeding events in electronic health records was generally high, >70%. A low positive predictive value (21%) was found for upper GI bleeding events defined using IPCI codes which is a much less granular coding system than ICD-10.

Friberg et al[77] was the only study to develop a phenotype in a wide range of anatomical sites, use supporting electronic health record data to define levels of severity and perform the gold standard chart review. However they did not use primary care bleeding records where clinically important events may also be captured.

A major limitation of using electronic health records to define bleeding events is that a number of the components to classify bleeding severity in trials and prospective studies are unavailable or difficult to apply using retrospective cohorts from electronic health records. For example, change in haemoglobin and units of blood transfused are generally not captured in electronic health records, but has been shown to be a prognostically important component in defining severity of events.[67]

2.3.4 Antithrombotic therapy and population-based risks of bleeding

A number of studies have estimated the incidence and risks of bleeding in a variety of population-based cohorts of patients indicated for antithrombotic use, summarised in **Table 2.4**. Understanding previous population-based estimates of bleeding incidence can help to inform the validity of phenotypes developed in electronic health records.

For patients prescribed oral anticoagulants from population-based electronic health records in a Danish region[78] and the United Kingdom[79] the estimated fatal or hospitalised bleeding incidence was 3.5 and 6 per 100 patient years, respectively. A multicentre study of patients in Japan treated with antithrombotic therapies for stroke and cardiovascular disease (n=4,009) found higher incidence of bleeding events of all severities to be associated with the intensity of prescribed antithrombotic therapy.[80] A large population-based patient registry in Denmark linked to prescribing and cause of death registries found increasing levels of antithrombotic therapies prescribing for atrial fibrillation patients (n=118,606) within 90 days of hospital discharge, from 21.4% in 1997 to 44.8% in 2006.[81] The authors also identified 180 day bleeding incidence almost doubling from 4.7% to 9.0% in the same time period.[81] In the same linked data a study of 11,480 patients with atrial fibrillation admitted to hospital for acute myocardial infarction, those treated with triple therapy (warfarin and dual antiplatelet therapy) had the highest bleeding incidence(14.2 per 100 person years) compared with other treatment groups.[82]

A study of 16,513 atrial fibrillation patients using linked primary care, hospital admission and cause of mortality registry data found the incidence of any bleeding to be highest in patients currently exposed to vitamin K antagonists (3.9 per 100 patient years) compared to those with past vitamin K antagonist exposure or no vitamin K exposure (2.7 and 2.9 per 100 patient years respectively).[83]

The standardised incidence of intracerebral haemorrhage was estimated in patients with previous vascular disease using the Oxford stroke and vascular disease registers, OCSP and OXVASC.[84] In both OCSP and OXVASC, the risk of intracerebral haemorrhage was markedly lower in patients aged under 75 (0.10 and 0.06 per 1000 patient years, respectively) compared with patients aged 75 years or above (1.55 and 1.44 per 1000 patient years, respectively).[84]

In an Icelandic hospital database, the crude incidence of upper gastrointestinal bleeding was estimated to be 0.87 per 1000 patient years, and that the use of warfarin, non-steroidal anti-inflammatory drugs and aspirin were associated with increased bleeding risk.[85] A matched case control study of patients with serious upper gastrointestinal bleeding in Danish linked electronic health records also found increased bleeding odds associated with increased intensity of antithrombotic therapy regimen. Patients treated with aspirin and patients treated with triple therapy had 1.8 and 5.3 times the odds of upper gastrointestinal bleeding, respectively, compared with patients not treated with antithrombotic therapy.[86]

2.3.5 Conclusion and implications for this thesis

Bleeding is as serious side effect associated with the use of antithrombotic therapies and is heterogeneous in its manifestation, particularly in terms of site and severity. Numerous bleeding classification applied to different disease populations inhibits the ability to compare bleeding events across studies. While there have been efforts to standardise bleeding definitions in prospective studies and randomised clinical trials, such definitions are not always applicable in electronic health records. Therefore there is a need for a comprehensive standardised bleeding phenotype to be developed in linked electronic health records.

The use of data beyond code lists to define severity of bleeding in electronic health records should be investigated because bleeding severity is prognostically important. A well-defined bleeding phenotype applied to population-based cohorts will help us to accurately estimate bleeding incidence, across severity levels, in different disease populations that are indicated for antithrombotic therapies. Following understanding bleeding incidence we can estimate the prognosis for all-cause mortality, atherothrombotic events and further bleeding in patients that experience bleeding of different severities.

2.4 Prognostic modelling

2.4.1 Background

Prognostic models are used to assess patient risks of clinical events given their current health status and characteristics.[87] Through assessing patient risks of diseases or clinical events, clinicians are able to make recommendations that may improve patient prognosis, for example

changing modifiable risk factors (smoking, weight etc.) or prescribing appropriate treatments. Therefore prognostic models play a vital role in the field of personalised medicine.

Prognostic models can take various forms ranging from simple point-based calculations to more complex statistical models. Point-based scores are usually derived from statistical models which are then simplified to consist of few risk factors where patients are assigned a weighted score (e.g. 1 or 2 points) for each risk factor and their total score corresponds to a level of risk. Point-based scores have the advantage of ease of use and due to their simplicity they may be more likely to be adopted in clinical practice; whereas their main disadvantage is that they may lack the granularity to effectively discriminate between levels of risk.

Prognostic models in the form of multivariable regression models, whilst more complex, have the advantage of assessing a wide range of risk factors and more flexible modelling of associations between risk factors and the events of interest.[88] Regression models may be considered to be more difficult to adopt in clinical practice; however, with the increasing use of electronic health record systems in primary and secondary care, more complex prognostic models can be integrated within the EHR, or accessed using apps or web calculators with minimal computational effort from users. The output of these models depends on the type of statistical models used (e.g. logistic regression for binary outcomes, Cox regression for time to event outcomes, linear regression for continuous outcomes) and is usually a quantifiable estimate of risk, for example a percentage chance of events within a given timeframe.

Prognostic models integrated within electronic health records may be advantageous as they provide the opportunity to monitor and dynamically update patient risks according to changing risk factors captured in primary and hospital care. For example, if a patient's risk of clinical events crosses a predetermined threshold they could be flagged for further assessment and a change in treatment strategy.

2.4.2 Model development methods

Study population

Study populations used for developing prognostic models can come from randomised controlled trials or observational cohorts. Randomised control trial populations and observational disease registry cohorts will generally have rich data with respect to treatments, interventions and patient characteristics or clinical biomarkers relevant and specific to the disease population. Unselected populations from routinely collected linked electronic health records, for example primary care and hospital admissions records, could provide cohorts that are more representative of patients in real-world settings.

Prognostic risk factors and multivariable modelling

Risk factors commonly used in prognostic models include age, sex, ethnicity, behaviours, socioeconomic factors, medical history, clinical biomarkers and medications previously and currently used. Genetic markers have also been incorporated in prognostic models.

In selecting risk factors to include in prognostic models, researchers should consider those risk factors known to be associated with the outcome of interest from the literature and also novel risk factors with previously unassessed utility that may be clinically relevant. It is important to consider the functional form of risk factors, for example, log transformations and categorisation. Furthermore, clinically plausible interactions between prognostic factors should be investigated.

Linear regression, logistic regression and proportional hazard models are examples of types of models for continuous, binary and time to event (survival) outcomes respectively. For time to event outcomes models, fully parametric models are ideal for prognostic models, particularly if the goal is to be able to produce predictions for individuals as well as estimate the relative risks (hazard ratios) for the risk factors, because the baseline hazard function is estimated. Care must be taken when choosing the functional form for underlying baseline hazards. Common choices of baseline hazard function include exponential, Weibull and gamma. Flexible parametric baseline hazards can incorporate different baseline hazard shapes (e.g. using restricted cubic splines) but are more computationally intensive.

In the case of semi-parametric models, i.e. Cox proportional hazards models, because no assumptions are made about the underlying hazard function and it is not estimated, it is not possible to produce individual predictions, only hazard ratios, for the risk factors in the model. Proportional hazards assumptions can be checked using log(-log) plots and residual plots.

To ensure the most parsimonious model is arrived at, model selection procedures may be used. This includes backwards selection, where a model with all potential prognostic factors are included and iteratively checked for the significance of prognostic risk factors. These checks or selection rules are generally pre-determined p-values or Akaike information criterion (AIC) values.

2.4.3 Model validation methods

In order to be considered useful, prognostic models are required to undergo validation. Model validation can be performed internally, i.e. using the same cohort in which the model was developed using methods such as bootstrapping and k-fold cross-validation. Model validation

can also be performed externally by applying the prognostic model in a new cohort not involved in the model development process.

The main measures used to assess model validation are discrimination and calibration.

Model discrimination is how well models distinguish between patients with different risk levels. Harrell's C-index (also known as the area under receiver operator characteristic curve) is the most commonly used measure of discrimination. The c-index can take values from 0 to 1 and is interpreted as the probability that a patient that has an event is assigned a higher risk score than a patient that does not have an event. Therefore a c-index of 0.5 suggests the model as no discriminative ability i.e. no better than a coin flip.

Model calibration is the accuracy of model predictions when compared with observed events. An example of model calibration is to graphically compare observed events with prognostic model predictions and 95% confidence intervals, overall and across risk strata. The Hosmer-Lemeshow goodness of fit test is an example of a statistical measure of calibration that can be applied to logistic regression models. For the external validation of Cox proportional hazards prognostic models, Royston and Altman[89] summarise step-by-step approaches with research examples.

2.4.4 Prognostic models and electronic health records

Using electronic health records of 20 million patients from more than 1000 primary care general practices linked with hospital admissions data from HES and cause-specific mortality data from ONS, the QRESEARCH group have developed prognostic models for use in the general population such as QRISK[90-92] for predicting cardiovascular risk, QBLEED[93] for predicting intracranial and gastrointestinal bleeding risk and scores for numerous other clinical outcomes such as type 2 diabetes,[94] stroke[95] and chronic kidney disease.[96] The models developed and validated by the QRESEARCH group have been implemented and integrated in clinical computer systems used by general practitioners (EMIS) and can flag up patients at high risk of events.

Certain regression modelling methods are ideal for electronic health records. Dynamic prognostic models take account for the effects of changes in risk factors over time and update predicted risks for patients accordingly. Such models would fully harness the longitudinal data available within electronic health records, as opposed to using baseline risk factors alone.

2.4.5 Prognostic models for cardiovascular diseases

Prognostic models have been used to estimate long term cardiovascular disease risk in otherwise healthy patients. Multivariable risk models for cardiovascular events commonly use

prognostic factors such as patient characteristics (e.g. age, sex, and ethnicity), socioeconomic status, patient behaviours (smoking, alcohol consumption, physical activity and diet), comorbidities (e.g. diabetes), family history of cardiovascular disease, clinical biomarkers (e.g. body mass index, cholesterol, blood pressure) and prior and currently prescribed medications. These models are useful to identify patients at high risk of cardiovascular events, allowing for clinicians to advise lifestyle changes for modifiable risk factors (e.g. smoking and diet) or prescribe treatments that may lower cardiovascular risk. The QRISK[90-92] models are examples of cardiovascular prognostic models used in clinical practice.

Prognostic models for estimating risk of secondary events in patients with existing cardiovascular disease have also been developed and are used in clinical practice. For example atrial fibrillation patients are at increased risk of ischaemic stroke and may be prescribed with aspirin or warfarin in order to reduce their risk. $CHADS_2$[20] which was later expanded to CHA_2DS_2-VASc[6] are examples of risk scores developed to estimate the risk of stroke in atrial fibrillation patients and help inform the appropriate intensity of treatment to reduce stroke risk. They are simple point based scores; each prognostic factor is worth 1 or 2 points and the sum of the points is the measure of the patients risk.

2.4.6 Prognostic models for bleeding events

Prognostic models that have been developed for estimating risk of bleeding events are summarised in **Table 2.5.**

The models were developed using disease registries (HAS-BLED,[7] REACH[97]), clinical trial participants (CRUSADE,[27] ACUITY/HORIZONS-AMI[28]), population-based cohorts (QBLEED[93]) or derived from previous models ($HAEMORR_2HAGES$[98]) to estimate bleeding risk at the point of cardiovascular event diagnosis to aid antithrombotic therapy treatment decisions for acute myocardial infarction and atrial fibrillation patients. Since there is no universal major bleeding definition the bleeding outcomes were defined differently in each of the models, limiting their comparability. Prognostic factors used across the models include age, cardiovascular comorbidities, creatinine and antiplatelet or oral anticoagulant use. All of the bleeding prognostic models were externally validated and had good discrimination (c-index ranging from 0.62 to 0.86) and were calibrated well.

2.4.7 Using prognostic models to balance the benefits and harms of antithrombotic therapies

With the availability of developed and validated prognostic models for outcomes relating to the benefits and harms of antithrombotic therapies, atherothrombotic event prevention and bleeding respectively, it is possible to use such models in tandem to fully assess the suitability

of treatment for patients. Using the methodology described by Pocock et al[99] we can consider predicted treatment efficacy and predicted harms along with patient and clinician values regarding the benefits and harms in question. An example of this methodology is illustrated in **Figure 2.1**. In the figure, 'X' represents an individual who may benefit from treatment A as their predicted treatment benefit is greater than their predicted treatment harms. 'O' represents an individual who may not benefit from treatment A as their predicted harms will outweigh the benefits. This is under the rule of equal trade off, where treatment benefits and harms are valued and weighted equally. Other scenarios may apply, for example if clinicians or patients believe the treatment harms are considered to carry more weight than benefits and vice versa.

2.4.8 Prognostic models recommended in clinical guidelines

The use of prognostic models is recommended in a number of clinical guidelines. The CHA_2DS_2-VASc score for stroke risk and the HAS-BLED score for bleeding risk have frequently been implemented in clinical guidelines for atrial fibrillation, with the advice that they be used to aid determining the appropriate antithrombotic therapy for patients (i.e. warfarin vs. aspirin) based on their stroke and bleeding risks.[4] NICE recommends the use of QRISK2 to assess patient's cardiovascular risk before prescribing lipid intervention.[100]

Internationally, cardiovascular risk is assessed by primary care practitioners in New Zealand for general population who are 40 years and older, depending on the risk score, recommendation for interventions are suggested. [PREDICT programme: https://www.fmhs.auckland.ac.nz/en/soph/about/our-departments/epidemiology-and-biostatistics/research/view-study/research.html]

2.4.9 Initiatives to improve the quality of prognostic research

The PROGRESS group [www.progress-partnership.org] introduced a framework with a view to improving the study of prognosis.[101] The framework has four key themes: (1) Understanding health conditions and their nature and quality of care; (2) Identifying risk factors associated with prognosis; (3) Developing and validating prognostic models; (4) Using prognostic models to aid treatment/intervention decisions

The TRIPOD (Transparent Reporting of multivariable prediction model for Individual Prognosis Or Diagnosis) statement[102] is a set of recommendations developed to help improve the quality of reporting on development and validation of prognostic models and to standardise prognostic research methods. It is developed through convening of healthcare professionals, clinicians, researchers and questionnaires. Details of the 22 items comprising the TRIPOD checklist are listed in **Table 2.6**.

2.4.10 Conclusion and implications for this thesis

Prognostic modelling is an established and still growing field of research and has demonstrated clinical utility, as evidenced by the recommendation for their use in treatment guidelines. Prognostic models have been developed for the outcomes relating to benefits and harms of antithrombotic therapy, for use prior to patients first prescription. There is a need to assess patient's benefits and harms for novel uses of antithrombotic therapies, for example, prolonged dual antiplatelet therapy in myocardial infarction survivors.[17] Patients who have been treated with dual antiplatelet therapy for a year may have different risk profiles from patients at the time of their acute myocardial infarction, therefore existing prognostic models for acute patients may not calibrate well in stable coronary disease populations.

2.5 International normalised ratio control

2.5.1 Background

Oral anticoagulants are a widely used class of drugs primarily for stroke prevention in atrial fibrillation, treating venous thromboembolism and prevent atherothrombotic events in patients who underwent heart valve replacement. Vitamin K antagonists (VKA) are currently the class of oral anticoagulants most commonly used for long-term treatment. The therapeutic range for VKA treatment is particularly narrow and there are severe implications for patients who consistently fall outside of it. If over treated with VKAs i.e. dosage too high, patients are at increased risk of major bleeding or if undertreated patients (VKA dosage too low) may be at increased risk of atherothrombotic events. VKAs are particularly sensitive drugs and their effectiveness may be altered by comorbidities, concomitant medications, diet and alcohol consumption[103]. Therefore there is a wide personal variation for optimal VKA dosage which distinguishes it from other drugs commonly used to treat and prevent cardiovascular diseases which have a known optimal dose.

The international normalised ratio (INR) is widely used biomarker for patients being treated with VKAs in order to measure treatment control and requires patients to undergo regular blood tests at specialised anticoagulation clinics. INR is a standardised calculation of prothrombin time which is the time taken for blood to clot. In healthy people not treated with oral anticoagulants a normal INR is between 0.8 and 1.2. For patients being treated with oral anticoagulants the therapeutic range for most diseases is between 2 and 3. Therefore patients with INR values greater than 3 may be at increased of bleeding and patients with INR values below 2 may be receiving no benefit from their VKA therapy and even at increased risk of atherothrombotic events.

Given the concern of bleeding for patients above the therapeutic range and atherothrombotic events for patients below the therapeutic range, patients are monitored regularly for the duration of their treatment with oral anticoagulants. The frequency of INR tests is dependent on the stage of anticoagulation, presence of bleeding risk factors, concomitant prescriptions with drugs known to interact with oral anticoagulants, and poor INR.[104] Depending on the indication for oral anticoagulation patients may be treated for 3 months (e.g. venous thromboembolism) to lifelong (e.g. for atrial fibrillation with high stroke risk).

2.5.2 Time in therapeutic range

There are many ways repeated INR tests may be summarised for analysis, the most common being percent time in therapeutic range (TTR). In calculating TTR a commonly used assumption is linear trajectory between consecutive INR measures, known as Roosendaal's method.[105]

The guidelines for VKA therapy control in atrial fibrillation, venous thromboembolism and heart valve replacement patients are summarised in **Table 2.7**. The NICE 2014 guidelines for atrial fibrillation management[4] state anticoagulation should be reassessed if a patient's INR value exceeds 5 on more than one occasion, if their INR drops below on more than one occasion or if their TTR is below 65%. The ESC 2016 guidelines for atrial fibrillation management[5] recommend TTR is maintained above 70%.

Guidelines for treatment of patients with venous thromboembolism[33,34,106] suggests an initial 3 months of VKA therapy with the potential to extend the duration for long term secondary prevention of atherothrombotic events in high risk patients. Patients with mechanical heart valve replacements are recommended to undergo lifelong VKA therapy.[107,108] The intensity of treatment defined by the therapeutic range is dependent on the presence of additional risk factors for atherothrombotic events. Patients with bio-prosthetic heart valve replacements are recommended to undergo 3months of VKA therapy.[107,108] No target time in therapeutic range is recommended in guidelines for treating venous thromboembolism or heart valve prostheses.

2.5.3 Risk factors associated with time in therapeutic range

A number of studies have assessed risks factors associated with INR control measured using time in therapeutic range. An overview of the studied risk factors is shown in **Figure 2.2,** and described in more detail in the following tables, sorted by study setting:

- The studies of electronic health records are summarised in **Table 2.8**,
- The studies of non-EHR observational cohorts are summarised in **Table 2.9**
- The studies of randomised control trial populations are summarised in **Table 2.10**

2.5.3.1 Patient demographics

Across the studies most found increasing age to be associated with higher TTR.[109-118] For example in UK population based linked electronic health records from primary care and hospital admissions, Macedo et al estimated that atrial fibrillation patients aged above 80 years 50% lower odds of having a TTR<70% compared with patients aged between 18 and 44 years and a similar estimate for venous thromboembolism patients.[109] Three studies, with smaller sample sizes, found the increased age to be associated with lower TTR[119-121] and several studies found no significant association between age and TTR.[122-132] In most studies women were found to have lower TTR compared with men.[109,110,112,114-120,122,123,126,128,133] One study found women to have higher TTR compared with men.[129]

Four studies assessed the effects of ethnicity on TTR and all found patients with non-white ethnicity to have lower TTR compared with white patients.[110,116-118] Two studies investigated the association between deprivation and TTR, and both found higher levels of deprivation to be associated with lower TTR.[110,114]

Singer et al[116] studied AF patients recruited in an international randomised control trial and found patients in Latin America, South Africa, Eastern Europe, East Asia and India had lower TTR compared with patients in North America. Kooistra et al[132] studied VTE patients recruited in an international randomised control trial and found of the patients randomised to warfarin, those in Eastern Europe, Asia or South America had lower TTR compared with patients from Western Europe, Israel or South Africa whilst patients from Australia or New Zealand had higher TTR. In a smaller scale international cohort study Ansell et al found patients in Spain and Italy had higher mean TTR compared with patients in USA and France.[134]

2.5.3.2 Patient behaviours

Smoking status has been shown to have a negative impact on INR control and time in therapeutic range.[109,117,135] However in two smaller observational cohorts positive associations between smoking and TTR were observed.[125,129] Alcohol abuse and substance abuse has also been shown to be associated with lower TTR.[110,116] However in smaller cohorts significant associations were not observed.[125,128,129]

2.5.3.3 INR indication and target therapeutic range

Three studies compared TTR in populations with more than one indication or target therapeutic range. Witt et al found patients with a target INR>3 had less stable TTR compared with patients with an INR target of 2.5.[111] The authors also found no significant difference in TTR stability between patients with atrial fibrillation, venous thromboembolism, heart valve replacement or other indications.[111] In a study of veteran healthcare data, Rose et al found

patients with venous thromboembolism compared with AF patients had higher TTR in the first 6 months of anticoagulation, but worse TTR after 6 months.[110] Witt et al found patients with lower target INR (2 vs. 2.5) had higher odds of very stable INR (100% TTR) than patients with higher target INR (3 vs. 2.5), but there were no significant differences INR stability between patients with different indications for anticoagulation (venous thromboembolism, heart valve replacement and other indications versus atrial fibrillation).[112]

Four studies investigated the frequency of INR monitoring and all found an association between higher frequency of monitoring and worse TTR.[114,119,120,131] Dlott et al found a negative quadratic association between INR monitoring frequency and TTR i.e. patients with the least frequent and most frequent monitoring had lower TTR.[114]

2.5.3.4 Medical history: Cardiovascular diseases

Patients with history of stroke were found to have lower TTR in three studies. Melamed et al found 7.4% of patients with excellent anticoagulation control (TTR>75%) had prior stroke compared with 20.3% of patients with poor anticoagulation control (TTR<60%).[120] Nelson et al found that patients with prior stroke had 1.15 (95% CI: 1.04, 1.27) times the odds of having a TTR<55% compared to patients with no history of stroke.[123] Sanchez et al found patients with prior stroke had 31% lower odds of achieving TTR≥65%.[129] However, in a small single centre study patients with a prior history of stroke were found to have TTR 26.23% higher than patients without.[130] A number of studies included stroke in their analyses but found no significant association with TTR.

Three studies assessed the association between myocardial infarction and TTR. Two studies found a negative association: In a population based analysis of atrial fibrillation patients in Stockholm, Szummer et al included myocardial infarction as part of a composite vascular disease risk factor and found those patients to have lower TTR[124] and Apostolakis et al found patients with a history of myocardial infarction had mean TTR of 61% compared to 65% for patients without.[117] Barrios et al found no significant association.[128]

Of the seven studies assessing coronary artery disease as a risk factor for TTR, two studies found a positive association: Singer et al found patients with coronary artery disease had 2.4% higher TTR than patients without[116] and Wypasek et al found patients with a history of coronary artery disease had 11.8% higher TTR than those without. Two studies found a negative association: Rose et al found patients with a history of coronary artery disease had lower TTR in the long term, after 6 months of oral anticoagulation[110] and Gotsman et al found patients with prior coronary artery disease had 9.2% lower TTR than patients without.[135] Three studies found no association between coronary artery disease and TTR.[127,129,131]

Three studies in electronic health records have found that patients with hypertension had higher TTR[110,115,123] whilst a small single centre cohort study found a negative association between hypertension and TTR,[127] otherwise studies of hypertension as a risk factor found no significant association.

Heart failure has consistently been found to be negatively associated with TTR and INR stability.[109-112,115-118,120,122-124,126,128] For example, in a study of a large healthcare database for veterans Rose et al found patients with heart failure had 1.0% (95% CI: 0.7, 1.3) lower TTR in the long term.[110] Szummer et al estimated that those with heart failure had 17.1% (95% CI: 7.7, 26.4) lower TTR than those without.[124] Macedo et al found heart failure to be associated with lower TTR for patients with atrial fibrillation but not those with venous thromboembolism.[109]

Studies have shown history of peripheral arterial disease to be associated with lower TTR.[110,118,124,128] For example, in a study of a large healthcare database for veterans Rose et al found patients with peripheral arterial disease had 0.5% (95% CI: 0.1, 0.8) lower TTR in the long term.[110] Szummer et al estimated that patients with the composite risk factor of vascular disease which included peripheral arterial disease had 10.5% (95% CI: 2.5, 18.5) lower TTR than those without.[124]

In a randomised controlled population trial comparing acenocoumarol with warfarin for patients with venous thromboembolism, those with prior venous thromboembolism events had higher TTR than those who were experiencing their first event.[132]

2.5.3.5 Medical history: Non-cardiovascular diseases

Diabetic patients have been shown to have lower TTR than non-diabetics in a number of studies.[109-111,115-118,120,122,123,133] For example, Maecdo et al found diabetic patients indicated for oral anticoagulation with atrial fibrillation or venous thromboembolism had 2.01 (95% CI: 0.44, 3.58) and 2.83 (95% CI: 0.31, 5.35) lower TTR, respectively, compared to patients without diabetes.[109]

Nine studies assessed the association between history of bleeding and TTR. Two studies found a significant negative association between bleeding history and INR control. In a cohort of atrial fibrillation patients recruited from primary care centres in Spain, Barrios et al found 11.3% of patients with poor INR control had a history of bleeding, anaemia or a pre-disposition to bleeding compared to 7.1% of patients with good INR control.[128] In a prospective atrial fibrillation registry based in the US, Pokorney et al showed that 10% of patients in the lowest quartile of INR control (TTR 0-53%) had a prior gastrointestinal bleeding event compared with 7% of patients in the highest INR control group (TTR >80%).[118] In univariable analysis, Singer et

al showed that patients with a prior gastrointestinal bleeding event had higher mean TTR compared with patients who did not, however in multivariable analysis this association was not significant.[116]

Ten studies assessed the relationship between renal disease and INR control and five estimated negative associations,[110,118,121,129,133] while the remaining five studies found no significant association.[109,117,128,130,131]

Three studies identified a negative association between history of liver disease and TTR. Rose et al found liver disease was associated poor TTR during long term anticoagulation (more than 6 months), but was not a risk factor during the first 6 months of anticoagulation.[110] Macedo et al found liver disease was associated with lower TTR in venous thromboembolism patients but not atrial fibrillation patients.[109] In univariable analysis, Singer et al estimated that patients with liver disease had a lower mean TTR compared to those without.[116] Chronic lung diseases including chronic obstructive pulmonary disease and asthma have been shown to be associated with worse INR control.[109,110,116-118] In veterans' healthcare data, Rose et al observed that patients with hyperlipidaemia had better TTR than patients without.[110] However, in other studies hyperlipidaemia has been found to have either a negative association[128] or no significant association with TTR.[109,127,129,131] Cancer has been shown to be associated with poor INR control.[109,110] Macedo et al found that for patients undergoing oral anticoagulation for venous thromboembolism those with rheumatoid arthritis had lower TTR than those without.[109] Epilepsy has been shown to be associated with poor INR control.[109,110] Using veterans' healthcare data, Rose et al found a range of mental health conditions, including depression, bipolar disorder and dementia to be associated with poorer INR control, in both the initial 6 months of anticoagulation and longer term treatment.[110] Macedo et al also evaluated the associations of mental health conditions on TTR and found dementia to have a negative impact for venous thromboembolism patients, but not atrial fibrillation patients.[109]

2.5.3.6 Clinical biomarkers

The associations between TTR and clinical biomarkers are inconclusive, mostly owing to varying study designs. In large-scale linked electronic health records from primary care and hospital admissions, overweight (body mass index (BMI): 25-29.9kg/m^2), obese (BMI: 20-39.9kg/m^2) and morbidly obese (BMI: ≥ 40kg/m^2) patients with atrial fibrillation or venous thromboembolism were reported to be associated with higher TTR compared with normal weight patients (BMI: 18.5-24.9 kg/m^2), whilst underweight patients (BMI: <18.5kg/m^2) were associated with lower TTR.[109] Smaller scale studies have found higher BMI to be associated with lower TTR.[117,127]

In patients with renal dysfunction diagnosed with atrial fibrillation, higher estimated glomerular filtration rate (eGFR) was associated with higher TTR.[124] A similar association was found in a more general atrial fibrillation registry cohort. Patients had 8% increased odds of being in the lowest TTR quartile per 5mg/dL decrease in eGFR and eGFR <60mg/dL.[118] Conversely, among high risk AF patients enrolled in the ROCKET-AF clinical trial, in unadjusted analysis patients with eGFR ≥68mL/min had 0.3% (p=0.016) lower TTR than patients with eGFR<68mL/min. In multivariable analysis a quadratic shaped association between eGFR and TTR was observed.[116]

2.5.3.7 Prescribed medication

There appeared to be a null or negative association between prescribed medications such as aspirin and antiplatelets and TTR. This may be due to interactions with warfarin or reverse causality in observational studies. On the other hand, in a few studies patients prescribed lipid lowering drugs[109,116] or beta blockers[116,117] were found to have higher TTR.

2.5.4 Time in therapeutic range as a risk factor for prognosis following oral anticoagulation

Studies of outcomes following oral anticoagulation with INR monitoring are shown in **Table 2.11.** These studies focus on atherothrombotic events, and bleeding – relating to the benefits and harms of oral anticoagulation and also all-cause mortality.

2.5.4.1 All-cause mortality and atherothrombotic events

In large observational cohorts and population-based EHR studies of patients treated with oral anticoagulants, researchers have found that higher INR control, measured by TTR is associated with lower risk of all-cause mortality and atherothrombotic events.

Morgan et al[136] assessed time to mortality amongst atrial fibrillation patients stratified by $CHADS_2$ score, a measure of stroke risk commonly used with atrial fibrillation patients. Amongst high stroke risk patients ($CHADS_2$ score ≥2), warfarin therapy benefits were observed only if TTR was maintained at a sufficiently stable level. Those treated with warfarin had lower risks of stroke if their TTR>70% and mortality if their TTR>40% compared with those not treated with warfarin. Similar results were found among patients with lower stroke risk ($CHADS_2$ score of 1).

In a study of 2504 patients with stable INR (defined as 100% TTR) compared with 3569 patients with <100% TTR the stable patients had lower risk of all-cause mortality (0.4% vs 1.6% at 180 days follow up respectively), but the incidence of thrombosis events did not significantly differ between the groups.[111] In a study with a longer follow up of 1 year, patients with stable INR

had fewer thrombosis events than comparator patients and the mortality benefits of stable INR remained.[112]

Using linked EHRs in Finland, Lehto et al. studied outcomes in 54568 atrial fibrillation patients. Patients with TTR≤40% had increased risk of all-cause mortality and stroke compared to patients with TTR 60-70% (Hazard ratios: 2.4 and 1.8 respectively). Patients with TTR >80% had lower risks of all-cause mortality and stroke compared to patients with TTR 60-70% (Hazard ratios: 0.4 and 0.7 respectively).[137]

In a large scale study from anticoagulation clinic EHR of 42415 patients treated with warfarin for various indications, a U-shaped association between INR and all-cause mortality was identified. That is patients with very low or very high INR were at increased risk of death.[138]

2.5.4.2 Bleeding events

In large observational cohorts and population-based EHR studies of patients treated with oral anticoagulants, researchers have found that higher INR control, measured by TTR, is associated with lower risk of bleeding events.

In a study of 2504 patients with stable INR (100% TTR) compared with 3569 patients with TTR<100%, the stable patients had lower incidence of bleeding at 180 days follow up (0.8% vs. 2.8% respectively).[111] In a similar study with a longer follow up time of 1 year, the association between stable INR and lower bleeding risk remained.[112]

Similarly, Lehto et al found patients with TTR≤40% had increased risk of bleeding compared to patients with TTR 60-70% (Hazard ratio: 1.6). Patients with TTR >80% had lower risk of bleeding compared to patients with TTR 60-70% (Hazard ratio: 0.6).[137]

2.5.5 Novel markers of INR control

A number of studies have investigated measures of INR control as an alternative to TTR, summarised in **Table 2.12**. Examples of these measures of INR control include mean INR,[139] INR variability,[139-144] time above therapeutic range,[144,145] time below therapeutic range,[145] percentage of records within therapeutic range,[139,144] percentage of records outside therapeutic range,[139] percentage of records above therapeutic range,[139] and percentage of records below therapeutic range,[139] first INR measure,[145] last INR measure,[145] linear slope,[145] and area under the curve.[144]

In these studies, INR variability appeared to be a potentially a better measure of INR control than TTR, according to its association with all-cause mortality, atherothrombotic and bleeding events.[139-144]

Proietti et al studied the predictive ability of TTR when it was included in established bleeding risk scores as an additional prognostic factor. The authors used statistical measures such as the net reclassification index and the integrated discrimination improvement, and found that the models had improved predictive performance when accounting for patients previous TTR.[146] It is unknown whether the novel markers of INR control can confer any additional predictive ability to existing prognostic models.

2.5.6 Conclusion and implications for this thesis

Despite numerous studies demonstrating TTR as a marker of INR control and prognostically important for future cardiovascular, bleeding and all-cause mortality events, few guidelines have adopted recommendations relating to TTR and INR control, particularly in non-AF populations.

Previous studies, in a range of settings, have identified numerous risk factors associated with TTR. However the majority of studies populations comprised of atrial fibrillation patients. It is not known whether the findings in atrial fibrillation patients are applicable to or comparable with other disease populations indicated for oral anticoagulation and INR monitoring.

Despite the prognostic importance of TTR being demonstrated, it is rarely adopted in prognostic models for relevant populations e.g. for stroke and bleeding risk. It is not known if TTR or other measures of INR control such as INR variability will contribute to prognostic models in the presence of known and regularly used atherothrombotic and bleeding risk factors.

Within the CALIBER platform, using longitudinal INR records there is the opportunity to assess predictors of TTR and all-cause mortality, atherothrombotic and bleeding outcomes in patients following oral anticoagulation and we can compare across different patients groups indicated for oral anticoagulation, such as atrial fibrillation, venous thromboembolism and heart valve replacement.

2.6 Tables and Figures

Table 2.1: Antithrombotic therapies and their indications

	Antithrombotic therapy classes and agents	Primary CVD prevention for high risk individuals	AT event prevention (following CAD)	AT event prevention (following PAD)	AT event prevention (following AF)	AT event prevention (following ischaemic stroke)	Treatment and prevention of VTE	Heart valve disease or prosthetic heart valves	Surgery with high VTE risk (e.g. orthopaedic)	Haemo dialysis	Other
Antiplatelets	**COX inhibitors**										
	Aspirin	✓	✓	✓	✓	✓	✓	✓	✓		✓[a]
	ADP receptor inhibitors										
	Clopidogrel		✓	✓	✓	✓					
	Ticagrelor		✓								
	Prasugrel		✓								
	Ticlopidine		✓	✓	✓	✓					
	Thromboxane inhibitors										
	Dipyridamole		✓			✓		✓			
	Glycoprotein IIb/ IIIa inhibitors										
	Abciximab		✓								
Oral anticoagulants	**Vitamin K antagonists**										
	Warfarin		✓	✓	✓	✓	✓	✓			✓[b]
	Acenocoumarol		✓	✓	✓	✓	✓	✓			✓[b]
	Phenindione				✓		✓	✓			
	Direct oral anticoagulants										
	Dabigatran				✓				✓		
	Rivaroxaban				✓		✓		✓		
	Apixaban				✓				✓		

	Antithrombotic therapy classes and agents	Primary CVD prevention for high risk individuals	AT event prevention (following CAD)	AT event prevention (following PAD)	AT event prevention (following AF)	AT event prevention (following ischaemic stroke)	Treatment and prevention of VTE	Heart valve disease or prosthetic heart valves	Surgery with high VTE risk (e.g. orthopaedic)	Haemo dialysis	Other
Parenteral anticoagulants	Heparin		✓	✓			✓		✓	✓	✓[c]
	Low molecular weight heparins										
	Bemiparin								✓	✓	
	Certoparin								✓		
	Dalteparin		✓						✓	✓	
	Enoxaparin		✓				✓		✓	✓	
	Fondaparinux		✓				✓		✓	✓	
	Reviparin						✓		✓	✓	
	Tinzaparin						✓		✓		

CVD= cardiovascular disease; AT= atherothrombotic; CAD= coronary artery disease [myocardial infarction, acute coronary syndromes, unstable and stable angina]; PAD= peripheral arterial disease; AF= atrial fibrillation; VTE= venous thromboembolism

[a] other indications for aspirin include pain relief, fever, inflammatory diseases (e.g. rheumatoid arthritics, systemic lupus erythematosus), antiphospholipid syndrome, and pregnancy with high pre-eclampsia risk;

[b] other indications for warfarin and acenocoumarol include treatment of cardiomyopathy;

[c] other indications for heparin include AT event prevention during pregnancy

Table 2.2: Components of bleeding classifications used in randomised clinical trials and prospective observational studies

Component	TIMI[48]	GUSTO[50]	ACUITY[51], HORIZONS[52]	CURE[54]	CURRENT-OASIS 7[56]	STEEPLE[58]	PLATO[59]	GRACE[61]	ISTH[64]	BARC[3]
Fatality	●	●	●	●	●	●	●	●	●	●
Anatomic bleeding site	● Intracranial	● Intracerebral	● Intracranial Intraocular Retro-peritoneal	● Intracranial Intraocular	● Intracranial Intraocular	● Retro-peritoneal Intracranial Intraocular Epistaxis Gastro-intestinal	● Intracranial Intra-pericardial Epistaxis	● Intracerebral	● Intracranial Intraspinal Intraocular Retro-peritoneal Intraarticular Pericardial Intramuscular	● Intracranial Intraocular
Haemoglobin drop	● 3 - <5g/dL ≥5g/dL	○	● ≥3g/dL w. overt bleeding ≥4g/dL w/o overt bleeding	● ≥5g/dL	● ≥5g/dL	● ≥3g/dL 2-3g/dL	● >5g/dL 3-5g/dL	● Haematocrit ≥10%	● ≥2 g/dL	● 3 - <5g/dL ≥5g/dL
Blood transfusion	●	●	●	●	●	●	●	●	●	●
Number of units transfused	○	○	○	●	●	●	●	●	●	●
Haematoma size	○	○	● ≥5cm	○	○	● >5cm	○	○	○	○
Hospitalisation	○	○	○	○	○	○	○	○	●	●
Medical/surgical consultation	●	○	○	○	○	○	○	○	●	●
Medical/ surgical intervention	●	○	●	●	●	●	●	○	●	●
Haemodynamic compromise	○	●	○	○	○	●	○	○	●	○

Component	TIMI[48]	GUSTO[50]	ACUITY[51], HORIZONS[52]	CURE[54]	CURRENT-OASIS 7[56]	STEEPLE[58]	PLATO[59]	GRACE[61]	ISTH[64]	BARC[3]
Severe hypotension	○	○	○	●	●	○	●	○	○	○
Significantly disabling	○	○	○	○	○	○	●	○	○	○
Change in antithrombotic therapy	●	○	○	●	●	○	○	○	●	○

Note: BARC=Bleeding Academic Research Consortium; ISTH= International Society on Thrombosis and Haemostasis; TIMI=Thrombosis In Myocardial Infarction

Table 2.3: Bleeding phenotypes developed using electronic health records

Author	Year	Bleeding endpoint(s) evaluated	No. of Bleeding anatomical sites	Data source(s)	Setting	Study Population	Coding system (n codes)	Supporting EHR data used in case definition	Algorithm figure reported	Assessment of phenotype accuracy
Raiford et al[70]	1996	Upper gastrointestinal bleeding or perforation	1	Saskatchewan Hospital Services Plan	Hospital admissions	Patients hospitalised for upper GI bleeding	ICD-9 (30)	No	No	Site specific codes PPV: 91% Nonspecific codes PPV: 68%
De Abajo et al[71]	1999	Upper gastrointestinal bleeding	1	GPRD (UK)	Primary care	Patients with a record for acute upper GI bleeding	Read (codes not stated)	No	No	PPV: 95/96
Arnason et al[75]	2006	Any bleeding Major bleeding	8	A university hospital, Ottawa	Hospital admissions	Patients with a record for thromboembolism or bleeding	ICD-9 (81)	No (information from patient charts were used to classify severity)	No	Definite bleeding PPV: 91%; NPV: 91% Major bleeding PPV: 87%; NPV: 92%
Wahl et al[72]	2010	Severe upper gastrointestinal bleeding	1	HealthCore Integrated Research Database (USA)	Hospital admissions	Patients with a record for upper GI bleeding	ICD-9 (original:75; refined:33)	Procedure codes	No	PPV: original: 56.5% refined: 87.8%
Cunningham et al[76]	2011	Serious bleeding related to oral anticoagulation	>4	Tennessee Medicaid program	Hospital admissions	Medicaid enrolees >30 years old	ICD-9 (39)	No (information from patient charts were used to classify severity)	No	PPV assessed for individual codes ranged from 71.4% to 100% (>5 charts)

Author	Year	Bleeding endpoint(s) evaluated	No. of Bleeding anatomical sites	Data source(s)	Setting	Study Population	Coding system (n codes)	Supporting EHR data used in case definition	Algorithm figure reported	Assessment of phenotype accuracy
Crooks et al[73]	2012	Upper gastrointestinal bleeding	1	CPRD HES ONS (UK)	Primary care Hospital admissions Death registry	Patients with a record for acute upper GI bleeding	Read (46) ICD-10 (22)	Causes, symptoms, endoscopy, death, transfusion, procedures, alcohol, anaemia, coagulation, collapse, other	Yes	None
Valkhoff et al[74]	2014	Upper gastrointestinal bleeding	1	IPCI (Netherlands) HSD (Italy) ARS (Italy) Aarhus (Denmark)	Primary care Hospital admissions	Patients with a record for upper GI bleeding	IPCI (4) ICD-9 (26) ICD-10 (16)	No	No	IPCI - PPV: 21% HSD - PPV: 78% ARS - PPV: 72% Aarhus - PPV: 77%
Friberg et al[77]	2016	Fatal Non-fatal major Hospitalised Minor	4	Swedish Patient register	Hospital admissions Hospital outpatients Death registry	Atrial fibrillation patients	ICD-10 (38)	Anatomical site (intracranial) Transfusion Hospitalisation Diagnosis position	No	Fatal PPV: 88.1%; NPV: 99.7% Non-fatal major PPV: 90.6%; NPV: 91.5% Hospitalised PPV: 65.1%; NPV: 97.5% Minor PPV: 84.2%; NPV: 98.9%

PPV: positive predictive value; NPV: negative predictive value;

Table 2.4: Population-based estimates of bleeding incidence

Study	Year	Setting	Population	Bleeding endpoint(s)	n	Bleeding incidence and risk factors
Steffensen et al[78]	1997	North Jutland County pharmaco-epidemiologic database linked to regional hospital discharge registry	Patients treated with oral anticoagulants	Major bleeding (fatal or requiring hospitalisation)	682	Major bleeding rate: 6 per 100 treatment years Gastrointestinal bleeding rate: 2.7 per 100 treatment years Intracranial bleeding rate: 1.3 per 100 treatment years Major bleeding incidence increases with age
Hollowell et al[79]	2003	General Practice Research Database	Patients prescribed warfarin	Fatal bleeding (death within 7 days) Hospitalised bleeding Minor bleeding	3958	Gastrointestinal (20%), genitourinary (25%) and epistaxis (20%) most common bleeding events. All intracranial events were fatal/ hospitalised. Bleeding incidence per 100 patient years: Any: 15.2 (13.5, 17.0); Fatal/hospitalised: 3.5 (2.7, 4.6); Minor: 9.1 (7.8, 10.6) Patients with valve disorders had the highest risk of bleeding.
Hallas et al[86]	2006	Funen County patient admin system linked to pharmaco-epidemiological database and Danish central person register	Cases with serious upper GI bleeding and age/sex matched controls	Serious upper gastrointestinal bleeding	1443 cases 57720 controls	Gastrointestinal bleeding adjusted odds ratios with ATT use: Aspirin: 1.8 (1.5, 2.1) Clopidogrel: 1.1 (0.6, 2.1) VKA: 1.8 (2.3, 2.4) Aspirin +clopidogrel: 7.4 (3.5, 15) Aspirin + VKA: 5.3 (2.9, 9.5)
Lovelock et al[84]	2007	Oxford community stroke project & Oxford vascular study	Patients with previous vascular events	Intracerebral haemorrhage	52 cases in OXVASC 55 cases in OCSP	Standardised incidence (per 1000 per year) of intracerebral haemorrhage OXVASC <75yrs: 0.06 (0.03, 0.08) OXVASC ≥75yrs: 1.44 (0.95, 1.92) OCSP <75yrs: 0.10 (0.07, 0.14) OCSP ≥75yrs: 1.55 (0.96, 2.13)

Study	Year	Setting	Population	Bleeding endpoint(s)	n	Bleeding incidence and risk factors
Toyoda et al[80]	2008	19 stroke and CVD centres in Japan	Patients taking ATT for stroke and CVD	Life threatening or major bleeding Any bleeding Minor bleeding (MATCH trial criteria)	4009	Incidence of bleeding events was higher in patients treated with more intense ATT combinations Life threatening or major bleeding incidence (% per year): Single antiplatelet: 1.21 Dual antiplatelet: 2.00 Warfarin: 2.06 Warfarin + antiplatelet: 3.56 Similar trends for any and minor bleeding
Hansen et al[81]	2010	Danish National Patient Registry, Danish Register of Medicinal Product Statistics, National Causes of Death Register	Patients hospitalised with atrial fibrillation	Nonfatal; Fatal Intracranial; Airway Gastrointestinal; Urinary tract	118606	Prescribing of ATT drugs within 90 days of discharge increased from 21.4% in 1997 to 44.8% in 2006 Bleeding within 180 days of discharge increased from 4.7% to 9.0% Increasing bleeding risk with increasing intensity of ATT combination
Lamberts et al[82]	2012	Danish National Patient Registry, Danish Register of Medicinal Product Statistics, National Causes of Death Register	Atrial fibrillation patients admitted with MI/PCI	Fatal or nonfatal bleeding 1 year follow up	11480	Overall, 6.3% had fatal or nonfatal bleeding Triple therapy patients had highest bleeding incidence (14.2 events per 100 person years) Bleeding incidence increased with increasing intensity of ATT 70/75 fatal bleeding events were intracranial or gastrointestinal
Hreinsson et al[85]	2013	National University Hospital Iceland	All patients who underwent upper gastrointestinal endoscopy	Upper gastrointestinal bleeding	1731	Crude upper GI bleeding incidence: 0.87 per 1000 patients per year Incidence increases with age groups Duodenal (20.5%) and gastric ulcer (14.7%) most common bleeding diagnoses Warfarin, NSAIDS, aspirin associated with higher odds of bleeding

Study	Year	Setting	Population	Bleeding endpoint(s)	n	Bleeding incidence and risk factors
Gallagher et al[83]	2014	CPRD, HES, ONS linked data	Atrial fibrillation patients	Any bleeding Fatal Intracranial Extracranial Gastrointestinal	16513	Incidence of bleeding events (per 100 person years):

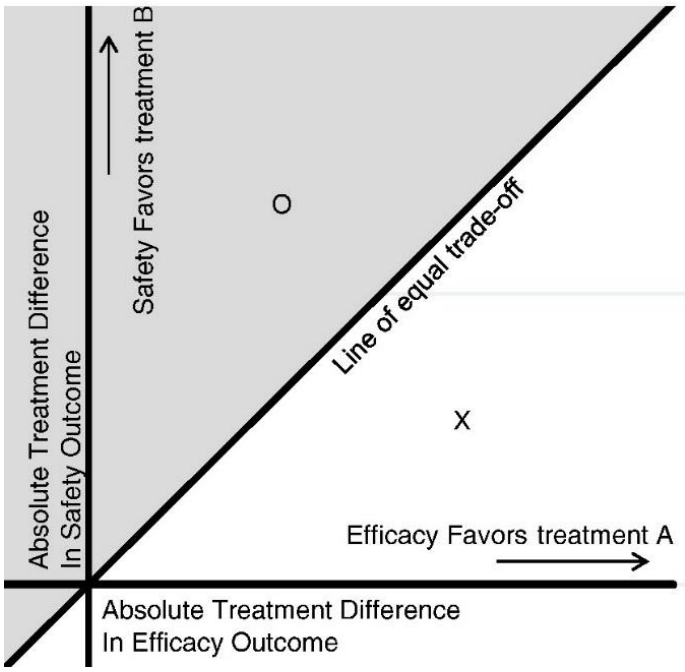
Bleeding event	Current VKA exposure	Past VKA exposure	No VKA use
Any	3.9	2.7	2.9
Fatal	0.3	0.2	0.2
Intracranial	0.4	0.2	0.4
Extracranial	3.4	2.5	2.5
GI	0.9	1.0	1.1

Table 2.5: Comparison of bleeding risk prediction models

	REACH[97]	CRUSADE[27]	ACUITY/ HORIZONS-AMI[28]	HAS-BLED[7]	HAEMORR$_2$HAGES[98]	QBLEED[93]
Year	2010	2009	2010	2010	2006	2014
Population	Stable patients at risk for atherothrombotic events	Treated NSTEMI	ACS	Atrial fibrillation	Atrial fibrillation	Patients aged 21-99 not currently using anticoagulants
n population	68236	71277	17421	3978	0 (model derivation driven by literature alone)	4.4 million
Bleeding outcome	Serious bleed leading to both hospitalization and transfusion within 2 years	In-hospital major bleeding	Non-CABG related major bleeding within 30 days of ACS presentation	Major bleeding within 1 year	Hospitalisation for haemorrhage up to 1000 days	Upper gastrointestinal and intracranial
Follow up t	2 years		30 days	1 year	1000 days	
n events	804	6701	520	53	-	21641 upper GI 9040 intracranial
Prognostic factors						
Demographics & behaviours	2 (age, smoking)	2 (age, sex)	2 (age, sex)	2 (age, drug or alcohol abuse)	2 (age, alcohol abuse)	6 (age, sex, ethnicity, deprivation, smoking, alcohol)
Clinical diagnoses	5 (peripheral arterial disease, congestive heart failure, diabetes, hypercholesterolemia, hypertension)	2 (prior vascular disease, diabetes)	2 (anaemia, ACS presentation)	4 (hypertension, abnormal renal/ liver function, history of stroke, bleeding history or anaemia)	7 (Liver or renal disease, malignancy, prior bleed, uncontrolled hypertension, anaemia, excessive fall risk, prior stroke)	9 (atrial fibrillation, heart failure, high blood pressure, cancer, liver disease, pancreatitis, oesophageal varices, bleeding, VTE)
Biomarkers	0	4 (Haematocrit, creatinine clearance, heart rate, systolic blood pressure)	2 (serum creatinine, white blood cell count)	1 (INR in 2-3 range <60% of time)	2 (reduced platelet count, genetic factors (CYP 2C9 single nucleotide polymorphisms))	2 (reduced platelet count, BMI)
Pharmacological interventions	2 (antiplatelets, oral anticoagulants)	0	1 (ATT: heparin +GPI vs. bivalirudin monotherapy)	1 (antiplatelet or NSAID)	0	5 (antiplatelet drugs, NSAIDs, steroids, antidepressants, anticonvulsants)

	REACH[97]	CRUSADE[27]	ACUITY/ HORIZONS-AMI[28]	HAS-BLED[7]	HAEMORR$_2$HAGES[98]	QBLEED[93]
Included patients with missing data	No	Missing age, sex, ethnicity, haematocrit excluded. Single imputation used on remaining variables.	No	No	-	Yes. Multiple imputation used for missing BMI, SBP, smoking and alcohol status
Externally validated	Yes CHARISMA trial population N=15603	Yes N=4500 ACS patients	Yes N=4500 ACS patients	Yes SPORTIF trial population N=7329 AF patients	Yes SPORTIF trial population N=7329 AF patients	Yes QResearch validation population N=1.4 million patients
Discrimination measure used	c-index	c-index	c-index	c-index	c-index	c-index
Discrimination	0.64	0.8	0.75	0.65	0.62	Upper gastrointestinal bleeding: 0.77 Intracranial bleeding: 0.86
Calibration measure used	Modified Hosmer Lemeshow	Hosmer Lemeshow	Hosmer Lemeshow	Hosmer Lemeshow	Hosmer Lemeshow	Graphically compared mean predicted and observed risks in deciles
Calibration	P=0.31	P=0.5	P=0.3	P>0.05	P>0.05	N/A
Used electronic health records	No	No	No	No	No	Yes

Figure 2.1: A graphical representation of weighing benefits and harms to aid treatment decisions



Adapted from Pocock et al. American Heart Journal 2014 168, 607-610DOI: 10.1016/j.ahj.2014.08.003

Table 2.6: The TRIPOD checklist

Section/Topic	Item		Checklist Item	Page
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	
	5b	D;V	Describe eligibility criteria for participants.	
	5c	D;V	Give details of treatments received, if relevant.	
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	
Sample size	8	D;V	Explain how the study size was arrived at.	
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	
	10c	V	For validation, describe how the predictions were calculated.	
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	
	15b	D	Explain how to the use the prediction model.	
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	

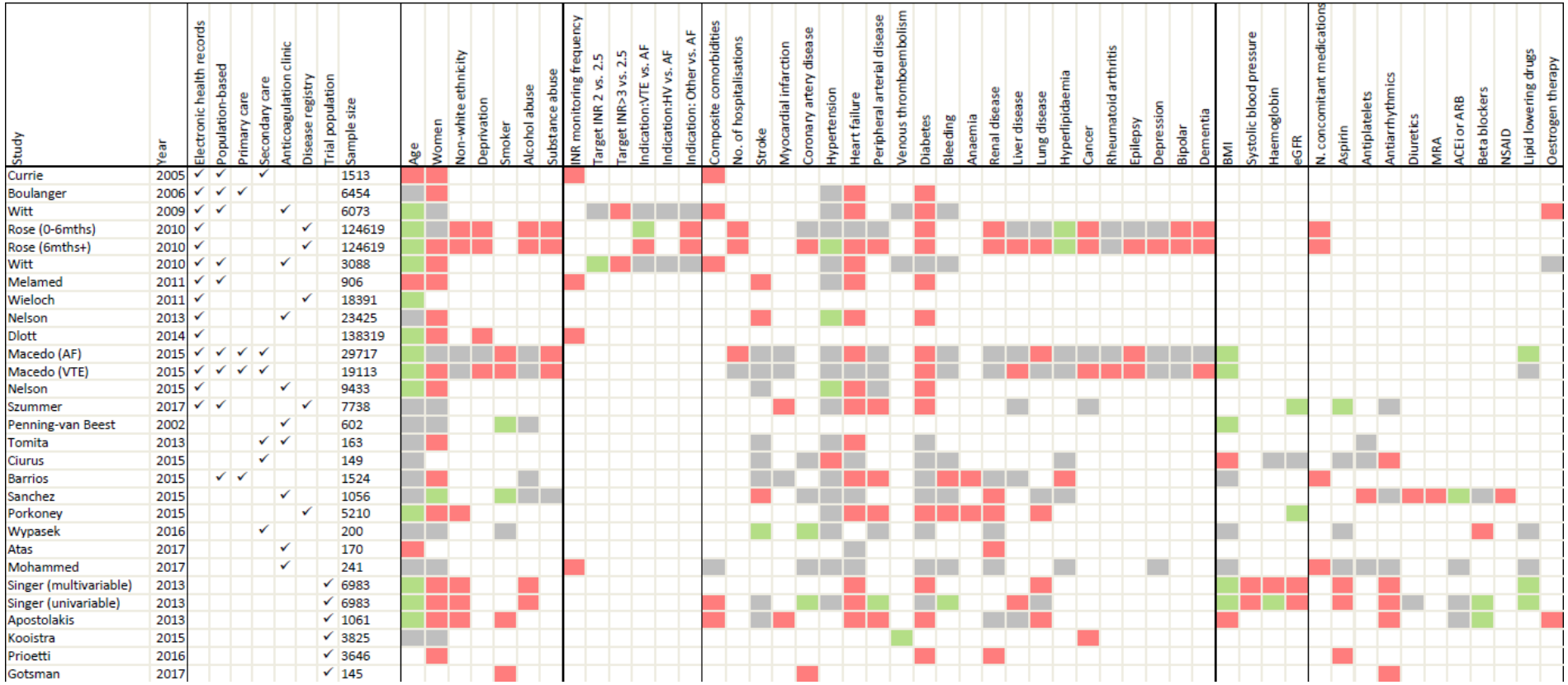
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	

Table 2.7: International normalised ratio (INR) and time in therapeutic range (TTR) in European and US guidelines for atrial fibrillation, venous thromboembolism and heart valve replacement patients

Indication	Guidance	INR therapeutic range	VKA therapy duration	INR test frequency	Recommended TTR threshold %	Other recommended INR or TTR checks
Atrial fibrillation	NICE 2014[4]	2-3	Lifelong	NR	65	INR>5 INR< 1.5
	ESC 2016[5]	2-3	Lifelong	NR	70	SAME-TT_2R_2 score
	AHA/ACC/HRS 2014[147]	2-3	Lifelong	Initiation: weekly Stable: monthly	NR	
Venous thromboembolism	NICE 2012[33]	NR	3 months+	NR	NR	NR
	ESC 2014[34]	2-3	3 months+	NR	NR	Pharmacogenetic testing (CYP2C9) may improve TTR
	CHEST 2016[106]	2-3	3 months+	NR	NR	NR
Heart valve replacement	ESC 2012[108]	Mechanical valve w/o risk factors or bio prosthetic valve: 2-3 Mechanical valve w. risk factors: 2.5-3.5	Mechanical valve: Lifelong Bio prosthetic valve: 3months+	NR	NR	NR
	AHA/ACC 2014[107]	w/o risk factors: 2-3 w. risk factors: 2.5-3.5	Mechanical valve: Lifelong Bio prosthetic valve: 3months+	NR	NR	NR

NR= No recommendation given

Figure 2.2: A summary of risk factors for INR time in therapeutic range



Note: Green boxes indicate risk factors presence was associated with higher TTR, red boxes indicate risk factors presence was associated with worse TTR, grey boxes indicate no association found between risk factors presence and TTR and white boxes indicate the risk factor was not studied.

Table 2.8: Risk factors for INR time in therapeutic range: assessed in electronic health records

Author	Year	Electronic health records	Study Population	n	Endpoint	Overall Mean (SD) TTR	Main findings; (+)= associated with higher TTR, (-) = associated with lower TTR
Currie et al[119]	2005	Hospital inpatient Haematology lab ONS mortality linked data UK	NVAF patients	1513	TTR	Stable group 74.9 Unstable group 44.7	Morbidity (-), monitoring frequency (-), female+age (-)
Boulanger et al[122]	2006	General Electric Electronic Medical Records Logician database USA	NVAF patients 1998-2003	6454	PT/%INR in range	48 (47, 49)	Men (+), congestive heart failure (-), diabetes (-), Northeast vs. Western/Midwest (-)
Witt et al[111]	2009	Kaiser Permanente Colorado, Clinical Pharmacy Anticoagulation Service USA	Patients with >90 days warfarin therapy 2000-2005	6073	TTR=100%	Comparator group: mean proportion of INR values in therapeutic range 49.9% (22)	2504 stable patients, 3569 comparator patients Age>70 (+), INR target >3 vs. 2.5 (-), diabetes (-), heart failure (-), oestrogen therapy (-), chronic disease score (-)

Author	Year	Electronic health records	Study Population	n	Endpoint	Overall Mean (SD) TTR	Main findings; (+)= associated with higher TTR, (-) = associated with lower TTR
Rose et al[110]	2010	Veterans Health Administration (VARIA) USA	Patients receiving oral anticoagulation therapy (valvular heart disease, VTE, AF, other) 2006-2008	124619	TTR during inception and experienced phases	First 6 months: 48 6 months+: 61	**TTR (first 6 months):** Age(+), non-white(-), deprivation (-), distance to nearest VA facility (-), VTE vs. AF (+), other indications vs. AF (-), cancer (-), CKD (-), diabetes (-), hyperlipidaemia (+), alcohol (-), bipolar (-), dementia (-), substance abuse (-), no. of concomitant medications (-), no. of hospitalisations (-) **TTR (6 months+):** female (-), age (+), black ethnicity (-), deprivation (-), VTE and other indications vs. AF (-), cancer (-), CKD (-), liver disease (-), lung disease (-), CAD (-), diabetes (-), epilepsy (-), HF (-), hyperlipidaemia (+), hypertension (+), PAD (-), alcohol (-), bipolar (-), dementia (-), depression (-), substance abuse (-), no. of concomitant medications (-), no. of hospitalisations (-) Developed a prediction model for TTR
Witt et al[112]	2010	Kaiser Permanente Colorado, Clinical Pharmacy Anticoagulation Service USA	Patients with >90 days warfarin therapy 2000-2005	3088	TTR=100%	Comparator group: 42.1 (5.7)	533 patients with 100% TTR, 2555 comparator patients Age (+), male (+), INR target 2 vs 2.5 (+), INR target>3 vs. 2.5 (-),heart failure (-), chronic disease score (-)
Melamed et al[120]	2011	Clalit Health Services database (Israel)	AF patients within a large managed care organisation 2006-2007	906	TTR<60% TTR 60-75% TTR>75%	48.6 (23.1)	TTR<60% (n=611, 67.4%) TTR 60-75 (n= 187, 20.6%) TTR>75% (n=108, 11.9%) Age(-), female (-), >16 INR records (-), diabetes (-), heart failure (-), prior stroke (-), non-board certified physician (-)
Walker et al[148]	2011	Veterans Health Administration	AF/atrial flutter patients 2002-2003	296	TTR	MHC 56.8 (16.9) No MHC 65.9 (18.2)	Mental health conditions (MHC) (-)

Author	Year	Electronic health records	Study Population	n	Endpoint	Overall Mean (SD) TTR	Main findings; (+)= associated with higher TTR, (-) = associated with lower TTR
Wieloch et al[113]	2011	AuriculA Sweden	Atrial fibrillation 2008	18391	TTR	76.2	Age (+)
Nelson et al[123]	2013	CoagClinic USA (49 states)	NVAF 2006-2010	23425	TTR<55	67.3 (14.4)	Men (+),US region: West vs Northeast (-), US region: South vs Northeast (-), heart failure (-), Diabetes (-), previous stroke (-), hypertension (+)
Rodriguez et al[149]	2013	Massachusetts Gen. Hosp. Anticoagulation Management Service, Research Patient Data Repository Linked data, USA	Indicated for oral anticoagulation (2-3) 2009-2010	3770	TTR	73.8	Limited English proficiency (-)
Dlott et al[114]	2014	Quest Diagnostics Informatics Data USA	AF patients 2007-2008	138319	TTR	53.7 (23.3)	INR test frequency (-quadratic), study length<6months (-), age (+), women (-), median income (+), US region
Razouki et al[150]	2014	Veterans Health Administration (VARIA) USA	Patients receiving warfarin, target INR:2-3 2006-2008	103897	TAR (time above range) TBR (Time below range)	Not stated	(Note: Associations with TAR and TBR not TTR.) **TAR:** Women (+), age (-), non-white (+), deprivation (+), AF vs. VTE (+), cancer (+), CKD (+), lung disease (+), diabetes (+), HF (+), alcohol (+), bipolar (+), dementia (+), depression (+), substance abuse (+), no. of concomitant meds (+), no. of hospitalisations (+) **TBR:** Women (+), age (-), deprivation (+), distance to nearest VA (-), AF vs. VTE (-), cancer (+), CKD (+), liver disease (+), lung disease (+), diabetes (-), epilepsy (+), HF (+), alcohol (+), depression (+), alcohol (-), no. of concomitant meds (+), no. of hospitalisations (+)

Author	Year	Electronic health records	Study Population	n	Endpoint	Overall Mean (SD) TTR	Main findings; (+)= associated with higher TTR, (-) = associated with lower TTR
Macedo et al [109]	2015	CPRD, HES Linked data (UK)	AF + VTE patients with warfarin rx in first 12 months of anticoagulation 2000-2013	AF: 29717 VTE: 19113	TTR TTR<70% TBR>30% TAR>30%	44% AF and 36% VTE had TTR>70%	**AF:** age (+), BMI, current smoking (-), acute respiratory infection (-), chronic lung disease (-), pain medication (-), diabetes (-), epilepsy (-), lipid lowering drugs (+), 5-10 hospitalisations per year (-) **VTE:** female (-), age (+), BMI (+), current smoking (-), substance abuse (-), cancer (-), chronic lung disease (-), pain medication (-), dementia (-), diabetes (-), epilepsy (-), rheumatoid arthritis (-)
Nelson et al[115]	2015	CoagClinic USA (49 states)	NVAF 2006-2010	9433	Quartile with highest % INR out of range	66.8	Age (+), male (+), heart failure (-), diabetes (-), hypertension (+)
Yong et al[151]	2016	TREAT-AF Veterans Health Administration	Newly diagnosed AF 2003-1012	184161	TTR	Ranged from 59(18) to 52(20) amongst ethnicity groups	Black ethnicity (-)
Gryzmala-Lubanski et al[152]	2017	AURICULA, Swedish National Patient Registry, SWEDEHEART, Cause of Death Registry Linked data	Mechanical heart valve patients 2006-2011	3831	TTR		Mitral valve vs. aortic valve (-)
Hellyer et al[153]	2017	Veterans Health Administration USA	AF patients 2003-2012	167190	TTR	Not stated	Similar rates of INR monitoring across risk groups CHA_2DS_2-VASc score (-) HAS-BLED score (-)

Author	Year	Electronic health records	Study Population	n	Endpoint	Overall Mean (SD) TTR	Main findings; (+)= associated with higher TTR, (-) = associated with lower TTR
Szummer et al[124]	2017	Stockholm CREAtinine Measurements (SCREAM), Swedish Renal Register Linked data	Newly diagnosed AF patients initiating warfarin 2006-2011	7738	TTR	Median (IQR) 83 (71, 92)	Poor renal function (low eGFR) (-), diabetes (-), vascular disease (-), heart failure (-), aspirin (+)
Williams et al[154] (ACC abstract only)	2017	Electronic medical records (exact resource not specified)	Newly diagnosed NVAF patients	8867	TTR	Median (IQR) 55 (34, 68)	19 factors included in final prediction model (85 evaluated in total) Strongest predictors of low TTR: antiarrhythmic drug use, anaemia, lung disease, aspirin, low red blood cell count Better predictive performance than SAMEe-TT_2R_2 score in validation cohort

Table 2.9: Risk factors for INR time in therapeutic range: assessed in observational cohorts and disease registries

Author	Year	Setting	Study Population	n	Endpoint	Overall Mean (SD) TTR	Main findings; (+)= associated with higher TTR, (-) = associated with lower TTR
Penning-van Beest et al[125]	2002	Red Cross anticoagulation clinic Netherlands	All patients treated with oral anticoagulants 1997-1999	602	INR>6	Median INR Cases: 6.8 Controls: 3.2 (2-4)	BMI<20 (-), weight loss (-), below average physical activity (-), decrease in drinking (-), smoking (+)
Veeger et al[155]	2005	Anticoagulation clinic Netherlands	Patients with VTE 1995-1998	2304	TTR (2-3.5)	63%	30 day TTR(+)
Burton et al[156]	2006	27 general practices Scotland	Atrial fibrillation patients 1998	601	TTR	68%	
Ansell et al[134]	2007	Multicentre USA, Canada, France, Italy, Spain	NVAF patients taking OAC for at least 60 days	1511	TTR	USA 57 (24.1) Canada 61 (25.2) France 58.1 (25.1) Italy 68.9 (17.0) Spain 64.4 (19.4)	USA vs. Italy (-), France vs. Italy (-), France vs. Spain (-), USA vs. Spain (-)
Cavallari et al[157]	2009	Antithrombosis Clinic University of Illinois Medical Centre	African American patients (and Caucasian controls) at the ATC	118	Unstable (% INR out of range > median)	Median proportion out of range INR 44% (14-82%)	**African American:** No. of clinic visits (-), warfarin adherence (+), ≥3 episodes of vomiting or diarrhoea (-), ≥1 anti-infective rx (-) **Caucasian:** No. of clinic visits (-), warfarin adherence (+)

Author	Year	Setting	Study Population	n	Endpoint	Overall Mean (SD) TTR	Main findings; (+)= associated with higher TTR, (-) = associated with lower TTR
Okumura et al[158]	2011	Multicentre study 6 institutions, 1 clinic in 5 prefectures in Japan	NVAF undergoing warfarin ≥2 years 2008-2010	501	TTR	64 (25)	Age (+), Warfarin dose (-)
Costa et al[159]	2012	Single anticoagulation clinic Brazil	Patients taking OAC for at least 90 days from enrolment 2006-2008	134	TTR	64.7 (17.1)	Variability of daily vitamin K intake (-), male (+), warfarin treatment duration >2months (+), presence of family support (+), good medication management capacity *(cognitive ability to take meds as prescribed)* (+)
Han et al[160]	2013	Multicentre 10 cardiology practices and 30 primary care practices USA	NVAF patients	392	TTR	56.7	Cardiology vs. primary care practices(+) (Mean TTR: 60.8 vs. 55.3)
Tomita et al[126]	2013	Outpatients from 4 institutes Japan	NVAF patients	163	TTR	69.7 (25.1)	Women (-), congestive heart failure (-)
Ciurus et al[127]	2015	Department of Cardiology of the Medical University of Lodz Poland	NVAF or VTE patients indicated for OAC	149	TTR>80 (stable) TTR<80 (unstable)	76 (21)	Arterial hypertension (-), amiodarone (-), BMI (-)
Barrios et al[128]	2015	Multicentre Spain	NVAF patients receiving VKA >1 year	1524	TTR	69.0 (17.7)	Women (-), dyslipidaemia (-), heart failure (-), PAD (-), dietary habits (-), bleeding/anaemia (-), history of TTR<60 (-), no. of concomitant meds (-)
Sanchez et al[129]	2015	120 cardiology clinics (CALIFA) Spain	NVAF patients	1056	TTR<65	63.8 (25.9)	Kidney disease (-), no ARB treatment (-), antiplatelets (-), regular NSAID use (-) male (-), unemployed vs. employed (-), homemaker vs. employed (-), institutional resident vs. lives with partner (-), previous stroke (-), MRA (-), diuretics (-)

Author	Year	Setting	Study Population	n	Endpoint	Overall Mean (SD) TTR	Main findings; (+)= associated with higher TTR, (-) = associated with lower TTR
Porkorney et al[118]	2015	ORBIT-AF registry USA	AF patients treated with warfarin	5210	TTR in the lowest quartile (TTR<53)	65 (20)	Age (+), Diabetes(-), eGFR (+), INR at clinic (+), weight (+), haematocrit (+), COPD (-), prior valve surgery (-), NYHA class (HF) (-), AF diagnosis >1 year prior to enrolment (+), frailty (-)
Bishop et al[161]	2016	John Hopkins Hospital USA	Adult patients whose warfarin managed by face-to-face visits at clinic	249	TTR	50.1	Primary AC provider vs. no primary AC provider (+)
Wypasek et al[130]	2016	Poland	Patients with aortic valve replacement (mechanical or bioprosthesis) with genetic guided warfarin dosing	200	TTR<60	Median 59.6 (38.7, 82.7)	CAD (+), previous stroke (+), CYP2C9*2 (-)
Alyousif et al[162]	2016	KAMC Saudi Arabia	AF patients treated with warfarin >3months 2012-2013	110	TTR	59 (24.2)	$CHADS_2$ (-)
Atas et al[121]	2017	Tertiary care, university hospital Turkey	AF patients treated with warfarin 2014-2016	170	TTR<55	54.2 (21.4)	Elderly [age>75] (-), renal dysfunction [eGFR<60] (-)
Mohammed et al[131]	2017	Heart Hospital outpatient clinic Qatar	NVAF patients with >6 months warfarin	241	TTR<65	Median (range) 70 (19, 100)	Polypharmacy (-), no. of clinic visits (-)

Table 2.10: Risk factors for INR time in therapeutic range: assessed in randomised controlled trial populations

Author	Year	Trial(s)	Study Population	n	Endpoint	Overall Mean (SD) TTR	Main findings; (+)= associated with higher TTR, (-) = associated with lower TTR
Rombouts et al[163]	2007	Leiden anticoagulation clinic Netherlands	Patients with an indication for long-term OAC randomised to Vitamin K or placebo	182	TTR	Vitamin K: 89.5 (86.4, 92.5) Placebo: 85.5 (82.3, 88.6)	Vitamin K supplementation (+)
Sconce et al[164]	2007	Anticoagulation clinics Freeman Hospital and Royal Victoria Infirmary UK	Atrial fibrillation patients treated with warfarin (2-3) for at least 9 months with unstable control randomised to Vitamin K or placebo	68	TTR	Baseline Vitamin K: 59 (20) Control: 63 (18)	Vitamin K supplementation (+)
Singer et al[116]	2013	ROCKET-AF trial (45 countries, 7 regions)	AF patients randomised to warfarin	6983	TTR	55.2 (21.3)	Age (+), VKA/warfarin naïve (-), Latin America vs. USA/Canada (-), South Africa vs. USA/Canada (-), Eastern Europe vs. USA/Canada (-), East Asia vs. USA/Canada (-), India vs. USA/Canada (-), Congestive heart failure (-), women (-), COPD (-), eGFR (-quadratic), haemoglobin (- quadratic), SBP (-), BMI (+quadratic), diabetes (-), heavy alcohol consumption (-), amiodarone (-), statin (+), aspirin (-), non-white ethnicity(-)
Apostolakis et al[117]	2013	Model derivation: AFFIRM Model validation: prospective cohort from Sandwell + West Birmingham Hospitals	AF patients treated with OAC	1061 286	TTR	Derivation 64 (18) Validation 66 (16)	Women (-), minority (-), age (+), BMI>30 (-), MI (-), heart failure (-), PAD (-), diabetes (-), pulmonary disease (-), smoking (-), rhythm control (-), beta blockers (+), verapamil (+), amiodarone (-), >2 comorbidities (-), >1 comorbidity (-)

Author	Year	Trial(s)	Study Population	n	Endpoint	Overall Mean (SD) TTR	Main findings; (+)= associated with higher TTR, (-) = associated with lower TTR
Smith et al[165]	2013	TREAT (intensive educational intervention vs. usual care) UK	AF patients newly referred for warfarin therapy	97	TTR	Intervention group 76.2 (64.1, 97.3) Control group 71.3 (51.2, 84.7)	Intensive educational therapy vs. usual care (+)
Kooistra et al[132]	2015	EINSTEIN-DVT EINSTEIN-PE	VTE patients randomised to VKA therapy Warfarin or acenocoumarol	3825	TTR<64.7 (also TTR variability)	Median 64.7	**Warfarin:** Eastern Europe vs. Western Europe/Israel/South Africa (-), Australia/NZ vs. Western Europe/Israel/South Africa (+), Asia/South America vs. Western Europe/Israel/South Africa (-), weight <50kg (-), active cancer (-), secondary VTE (-), previous VTE (+) **Acenocoumarol:** Eastern Europe vs. Western Europe/Israel/South Africa (+), secondary VTE (-), age (+)
Proietti et al[133]	2016	SPORTIF III and V	Atrial fibrillation patients randomised to warfarin with data on renal function	3646	TTR TTR>70	Median 68.6 (56.5, 80.6)	**All Patients:** Chronic kidney disease (-), chronic AF (+), Men (+), diabetes (-), aspirin (-) **CKD subgroup:** weight (+), chronic AF (+), congestive heart failure (+)
Gotsman et al[135]	2017	standard care vs. one time intervention assessing potential risk factors for labile INR and giving advice	Patients with heart failure on warfarin therapy	145	TTR	Median 61 (42, 85) Intervention: 80 (62, 93) Control: 44 (29, 61)	**Pre intervention TTR:** Education>10years (+), ischaemic heart disease (-), smoker (-), amiodarone (-) **Post intervention TTR:** Patient tailored AC advice (+), amiodarone(-)

Table 2.11: Studies of prognosis following monitoring for oral anticoagulation

Author	Year	Setting	Population	n	Exposure(s)	Endpoint(s)	Main Findings
Hylek et al[166]	2000	Anticoagulation clinic; prospective cohort USA	Patients treated with warfarin with therapeutic range 2-3	114+ 268	INR>6 vs. INR within target range (1.7 – 3.3)	Major haemorrhage (fatal, intracranial or requiring hospitalisation and transfusion of at least 2U of blood) within 2 weeks INR decay to <4 in the INR>6 group	4.4% of patients with INR>6 had major haemorrhage (vs 0 in target range group) in the 2 week follow up Patients with INR>6 had an observed INR of <4 at a median time of 4 days
Oden at al[138]	2002	EHR: Anticoagulation clinic Sweden	Patients treated with warfarin (various indication)	42451	INR (continuous values)	All-cause mortality Death caused by intracranial haemorrhage	Lowest risk of death seen at INR 2.2 to 2.3 for all indications for warfarin and therapeutic ranges U-shaped hazard for all-cause mortality Increasing risk of bleeding deaths with increasing INR
Freixa et al[167]	2003	Prospective observational study of outpatients	Patients treated with acenocoumarol (various indications)	104	High INR (INR >5) Controlled INR (patients with all INR within target range for 3 months prior to inclusion)	Haemorrhage (major if required transfusion otherwise classed as minor)	2% of all INR tests were >5 and corresponded to 55 patients Factors associated with the high INR group: weight, female, number of drugs, concomitant medication, prosthesis, compliance, intercurrent disease High INR had more minor (20% vs 4.8%) and major (1.8% vs 0%) bleeding episodes than the control group.
Wittkowsky et al[168]	2004	Retrospective medical record review Anticoagulation clinic	Warfarin treated patients (various indications)	1020	Over-anticoagulation (INR >4) Under-anticoagulation (INR<2) INR within range	Thromboembolic events Major bleed Causes for out of range INR also examined	Major bleed rate 4%/patient-yr – most likely to occur in over-anticoagulated patients Thromboembolic event rate 2%/patient-yr – most likely to occur in under-anticoagulated patients INR<2 (n=2881), INR>4 (n=603) Causes: initiation, treatment intentionally held prior to procedure, response to dosage change, non-compliance, dose error, food/drug interactions. Majority unexplained.

Author	Year	Setting	Population	n	Exposure(s)	Endpoint(s)	Main Findings
Clark et al[169]	2008	Retrospective matched cohort; centralised anticoagulation service	Patients treated with warfarin	2597	Stable INR (2 INR values within or above INR therapeutic range. 3rd INR within therapeutic range) Low INR (As above but 3rd INR value ≥0.5 units below therapeutic range)	Anticoagulation related thromboembolic events within 90 days of index INR	Characteristics: Patients in the Low INR group had higher target ranges, poorer compliance, and less likely to have had previous thrombosis compared with the therapeutic INR group. Outcomes: No significant difference in thrombosis events between the groups at 90 days (0.4% vs. 0.1%) No significant difference found for other AC related complications (bleeding and death). Patients in the Low INR group were more likely to receive a dose boost (65% vs. 0.3%)
Morgan et al[136]	2009	Retrospective cohort analysis of linked inpatient, haematology and mortality data	Patients with AF and >5 INR tests	2235	TTR $CHADS_2$ score≥2	Time to stroke Time to mortality	Patients with a $CHADS_2$ risk score≥2 were less likely to have INR TTR >70%, more likely to have further comorbidities, were younger and more likely to be female than those with a lower risk score. $CHADS_2$ score≥2: Compared with patients not on warfarin therapy, treated patients only have a significantly lower risk of stroke when their TTR% >70 [HR: 0.2 [0.05-0.82]] but have a lower risk of mortality when their TTR% >40. $CHADS_2$ score=1: Compared with patients not on warfarin therapy, treated patients have a significantly lower risk of stroke when their TTR% >60 but have a lower risk of mortality when their TTR% >40. Overall patients with TTR<40% had reduced time to stroke and patients with TTR<30% had reduced time to death

Author	Year	Setting	Population	n	Exposure(s)	Endpoint(s)	Main Findings
Witt et al[111]	2009	Kaiser Permanente Colorado, Clinical Pharmacy Anticoagulation Service USA	Patients with >90 days warfarin therapy 2000-2005	2504 stable 3569 comparators	Stable INR (100% TTR) vs comparators	All-cause mortality Thrombosis Bleeding 180 day follow up	Stable patients **were older**, less likely to have comorbidities such as diabetes and heart failure, less likely to have a target INR ≥3 The stable group had fewer incidences of receiving heparin (0.3% vs 3.2%), mortality (0.4% vs 1.6%) and bleeding (0.8% vs 2.8%). There was no difference in thrombosis events between the groups (0.4% vs 0.7%)
Wallentin et al[170]	2010	RE-LY randomised control trial 906 international centres	AF patients randomised to warfarin or dabigatran	18024	Centre mean INR control (cTTR)	Stroke and systemic embolism (SE) Major bleeding Intracranial bleeding Death Composite of stroke, SE, MI, PE, death and major bleeding Composite of non-haemorrhagic stroke, SE, MI, PE, and death	Centres with higher cTTR had older mean ages, lower mean CHADS2 scores, fewer cases of previous stroke The benefit of dabigatran to reduce bleeding adverse events compared with warfarin was consistent across centres irrespective of cTTR quartile For vascular events, non-haemorrhagic events and mortality the benefits of dabigatran more pronounced at centres with poorer INR control.

Author	Year	Setting	Population	n	Exposure(s)	Endpoint(s)	Main Findings
Witt et al[112]	2010	EHR: Kaiser Permanente Colorado, Clinical Pharmacy Anticoagulation Service USA	Patients with >90 days warfarin therapy 2000-2005	533 stable 2555 comparator	Stable INR (100% TTR)	All-cause mortality Thrombosis Bleeding 365 day follow up	Characteristics: The stable group was older, more likely to be male, more likely to have AF, less likely to have heart valve replacement, less likely to have previous venous thrombosis, had a higher (not sig) median duration of warfarin therapy and had a lower chronic disease score. Outcomes: During the 365 follow up period, stable group were less likely to receive heparin, die, have an AC-related thrombosis event or have an AC-related bleeding event
Gallagher et al[171]	2011	GPRD	AF patients	37907	No warfarin use vs. warfarin use stratified by TTR groups	Stroke and transient ischaemic attack	Patients with INR<40% had increased stroke risk compared with patients not on warfarin. Adjusted RR, TTR<30 vs. non users: 1.44 (1.28-1.59) Patients with TTR>40%had lower stroke risk Adjusted RR, TTR ≥70 vs. non users: 0.33 (0.30, 0.36) These associations remained consistent in subgroup analysis by CHA_2DS_2-VASc score
Lehto et al[137]	2017	Linkages between: FinWAF; Finnish Care Register; Finnish Cancer Registry; National prescription Registry; National Cause of Death Register; 6 regional laboratory databases; Population Register	AF patients	54568	TTR	Bleeding, stroke, cardiovascular death, all-cause mortality	Adjusted hazard ratios: TTR≤40% vs. TTR 60-70 Bleeding: 1.6 (1.5, 1.8) Stroke: 1.8 (1.7, 2.0) CV death: 2.0 (1.8, 2.2) All-cause mortality: 2.4 (2.2, 2.5) TTR>80% vs. TTR 60-70 Bleeding: 0.6 (0.5, 0.7) Stroke: 0.7 (0.6, 0.8) CV death: 0.5 (0.4, 0.5) All-cause mortality: 0.4 (0.4, 0.5)

Table 2.12: Studies of novel measures of INR control: risk factors, prognosis and predictive ability

Name	Year	Study population	n	Objectives	Exposures	Endpoints	Main findings
Jones et al[139]	2005	Record linkage study in secondary care setting, patients with NVAF, no heart valve replacement and ≥5 INR readings	2223	To assess the association between measures of INR control with outcomes	Mean INR INR standard deviation % in target % INR<2 % INR>3 % out target	Mortality, ischaemic stroke, thromboembolic events, bleeding, hospitalisation, INR patterns monitoring (ICD-10 codes used are given in appendix)	All-cause mortality was associated with all measures of INR Increased bleed odds associated with SD INR Thromboembolic events associated with %INR<2,% in target range and mean INR Ischaemic stroke similarly associated with %INR<2 Increased hospitalisation when out of INR range
van Leeuwen et al[140]	2008	LAVA cohort Patients with mechanical heart valve prosthesis (1985-1993)	630	To assess 3 INR variability measures and their association with thrombotic and bleeding events	3 measures of INR variance growth rate (Fihn and Cannegieter) A: captures deviation from target range B1 and B2: captures deviation from previous record (measure to measure)	Thrombotic events Haemorrhagic events Combination of all adverse events	Method A was best associated with events. B1 and B2 had equally good performance when combined with TTR.
Lind et al[141]	2012	AF patients Anticoagulation registries linked with, hospital admissions and death registry	19180	To determine if INR standard deviation is more prognostically important than TTR	TTR INR standard deviation	All-cause mortality Stroke Bleeding	Low correlation between TTR and INR SD. INR SD had better predictive ability (measured by hazard ratios). With both in the model TTR was no longer sig.

van Den Ham et al[145]	2013	AF patients aged >40 excl. heart valve disease/replacement patients Linked CPRD-HES records (1987-2010)	27381	To evaluate if TTR can be improved by considering patterns of INR over time	%TTR Combination of simple measures split into 6 clusters: %Time below range, %Time above range, mean INR above normal, number of INR measures within 6 month period, first measure, last measure, linear slope	All-cause mortality (CPRD record) Ischaemic or haemorrhagic stroke and TIA (CPRD-HES) Major bleed (CPRD) Minor bleed (CPRD) (ISTH criteria) Hospitalised bleed (HES)	Clusters of INR patterns and measures resulted in improved prediction of events
Ibrahim et al[142]	2013	AF, DVT/PE, heart valve and other patients from the European Action on Anticoagulation study	819	To assess 3 INR variability measures and their association with thrombotic and bleeding events and compare with TTR	3 measures of INR variance growth rate (Fihn and Cannegieter) A: captures deviation from target range B1 and B2: captures deviation from previous record (measure to measure)	All-cause mortality Thromboembolism Bleeding	Method A was the strongest predictor at 3 months prior to events INR variability was a better predictor than TTR
Razouki et al[143]	2014	AF patients Veterans Health Administration	40404	To determine whether %TTR and INR variability associated with risks of events	%TTR and INR variability	Ischaemic stroke Major bleeding Fatal bleeding	High INR variability was associated with higher risks of events. High TTR alone may not be enough to reduce risks of events - variability should be considered too.
Rose et al[144]	2015	Admin database anticoagulation clinic data	676	To determine which INR summary measures are best correlated with bleeding	Proportion of INR in range INR variability % TTR TTR and INR variability % time above range % time with INR >4 Area under curve above range Area under curve above range squared	Clinically relevant bleeding	INR variability had strongest association with bleeding, followed by TTR Simple measures were also shown to have some value (proportion)

Proietti et al[146]	2016	Atrial fibrillation SPORTIF trial population	3551	Investigating if predictions of bleeding events improve with TTR included in ATRIA, $HEMORR_2HAGES$ and ORBIT scores	labile INR (TTR <65%)	Major bleeding 1) all reported 2) adjudicated	Net reclassification index and Integrated discrimination improvement estimates showed improved predictive performance in modified models that included labile TTR as a prognostic factor

3 Approaches to curating and phenotyping national linked electronic health records for research in CALIBER

Chapter Summary

Linked electronic health records contain a wealth of information that may be useful for researchers. However some data manipulation is required to transform the raw health records into a research ready datasets.

In this chapter I provide an introduction to electronic health records, and to the CALIBER linked electronic health record platform. I describe the data available within CALIBER and the approaches taken to transform raw data into research ready datasets.

I then describe the datasets used within each of the studies in this thesis, including the study populations, the risk factors and endpoints and how these components were formulated and generated.

Finally, I identified the limitations of the CALIBER platform and make recommendations for future improvements.

3.1 Electronic health records

3.1.1 General background to EHRs

Electronic health records (EHRs) are computerised collections of health data, from health services. Health data can include patient demographics, prescriptions, medical history and diagnoses, treatment and interventions, patient monitoring, drug safety monitoring, public health surveillance, costings and hospital level information. These data may be collected for clinical care, administration or auditing and quality assurance purposes.

EHR data can be input at the point of care (e.g. at general practices), or input retrospectively using medical notes by specialist coders.

3.1.2 Potential strengths of EHRs for clinical research

The increasing use and availability of EHRs has provided opportunities for clinical research. EHRs have the potential to contain large amounts of longitudinal data, which may be costly to procure in prospective studies and therefore allows studies of epidemiological trends over time, and studies with long follow up. Administrative records are useful for investigating

quality of care within patient groups. Population-based EHR cohorts allow for large scale studies on cohorts that are representative of the real world which is ideal for epidemiological, public health, drug uptake and safety studies.

In this chapter I describe the CALIBER linked electronic health records platform used in this thesis and the principles and approaches (**Figure 3.1)** I used to prepare the data for statistical analysis.

3.2 CALIBER: a linked electronic health records platform

3.2.1 Introduction

The CALIBER[8] (Clinical research using Linked Bespoke studies and Electronic health Records) platform is a series of linkages between UK electronic healthcare databases: Clinical Practice Research Datalink (CPRD) a longitudinal primary care database, Hospital Episode Statistics (HES) a database of hospital admissions and procedures, Myocardial Ischaemia National Audit Project (MINAP) a national acute coronary disease registry and the Office of National Statistics (ONS) for cause-specific mortality and social deprivation data. The CALIBER dataset used in this PhD holds data comprising of approximately 2 million patients from 225 general practices in England that have consented to linkage between 1997 and 2010.

3.2.2 Data linkage

CALIBER provides linkages between the anonymised datasets using encrypted CPRD patient identifiers allowing a more complete insight into patient's medical journeys than if any data resource was used individually **(Figure 3.2)**. Individual data sources alone may not accurately capture incidence of events, and different data sources collect different data.

Data linkage of patient level data from CPRD to other anonymised data sources are performed by a Trusted Third Party using NHS numbers, gender and date of birth. 58% of UK CPRD general practices currently consent to data linkage and it has previously been demonstrated that CPRD-HES linked data are representative of the general UK population.[172-174]

3.2.3 High resolution phenotypes

While in traditional cohorts diseases and events are confirmed at the time of occurrence or retrospectively by checking hospital notes, within CALIBER high quality disease phenotypes are developed using the linked data through collaborations between clinicians, epidemiologists and statisticians. The phenotype development process is outlined in **Figure 3.3.**

- In a test cohort descriptive and explorative analyses are performed on the potential phenotype components to determine which are suitable for use in the algorithm

- A preliminary phenotype algorithm is formed which is then tested and revised iteratively
- The revised phenotype algorithm is implemented and tested and revised accordingly to reach the final version of the algorithm
- The final phenotype is added to the CALIBER portal, including the codelists, metadata, code and programming scripts required to implement the algorithm given raw data
- The phenotype is made available for the EHR community who provide feedback to help further enhance the algorithm

Phenotypes for diseases and medical conditions have been developed for primary care records using Read codes in CPRD and secondary care records using ICD-10 codes in HES. The high resolution disease phenotypes usually include categories which are determined by wording and clinical usage of Read and ICD terms. Phenotype categories can denote:

- Subtype (e.g. MI: STEMI, NSTEMI, unspecified; Cancer: metastases, anatomical sites)
- Severity (e.g. Renal disease: mild, moderate, severe)
- Status (e.g. History of; monitoring; possible diagnosis; confirmed diagnosis)

For many diseases, CALIBER researchers have developed composite phenotypes which fully harness the data across the linked data sources. The composite phenotypes often comprise of disease diagnoses in both primary and secondary care, procedures, test results and prescribed medications relevant to the disease. Such composite phenotypes ensure maximal case ascertainment of a disease within the linked electronic health records.

Examples of phenotype development and validation In CALIBER have been published for myocardial infarction[36] and atrial fibrillation.[38] These studies demonstrate which codes are used from the various sources or how diagnoses may be inferred (e.g. from relevant biomarkers, prescriptions or procedures) and compare patient characteristics, risk factors or outcomes with those from traditional cohorts to confirm the validity of the phenotype.

3.2.4 CALIBER study approval

Studies of anonymised UK primary care data and linked data, such as CALIBER, are subject to approval from the Independent Scientific Advisory Committee (ISAC). ISAC approval is gained through submitting a protocol which outlines the study background and objectives, the data required including rationale and definitions (e.g. Read and ICD-10 codelists) for the study population, exposures and endpoints and a statistical analysis plan. The ISAC committee members (a multidisciplinary group of clinicians, statisticians, epidemiologists, health

informaticians, data scientists and lay members) provide detailed feedback and advise whether the study protocol is approved or requires revisions and resubmission.

Lay summaries of studies approved by ISAC are available online. Published research articles which use UK primary care and linked data are required to report their approved ISAC protocol number. Any minor or major changes to an ISAC approved study protocol are subject to re-review by the ISAC committee.

3.2.5 CALIBER user tools

The CALIBER data portal [https://www.caliberresearch.org/portal] contains a comprehensive collection of all phenotypes and their code lists developed in the CALIBER, spanning across numerous cardiovascular and non-cardiovascular disease areas. **(Figure 3.4)**

A series of R packages to support the use of CALIBER data have been developed by Dr Anoop Shah, *CALIBERlookups*, *CALIBERcodelists*, *CALIBERdatamanage*. These packages include dictionaries for ICD-10, Read and ONS codes, functions to look up and generate codelists, map codelists between dictionaries and to generally aid management of large datasets.

3.2.6 CALIBER data management

The CALIBER data platform is managed by the Data Lab: a team of data scientists who manage and maintain the catalogue of disease phenotypes in the CALIBER portal, perform data extraction and assist cohort formation.

3.2.7 Strengths of the CALIBER platform and approach to phenotyping

The CALIBER platform has a number of advantages for researchers. There is a wide range of linked data available, a representative sample of the English general population (2 million people registered at GP practices 1997-2010). The linkages between primary care, hospital admissions, disease registry and cause of death provide a comprehensive overview of these patients journey through the healthcare system.

For researchers using EHRs there is potential for disease definitions to vary from study to study. This can be problematic for interpreting and comparing results and study replicability. Under the CALIBER programme, there is a standardised approach for developing disease phenotypes. The portal contains a large and growing number of validated and reproducible disease phenotypes available to researchers, therefore encouraging scientific replicability.

3.3 CALIBER data sources

In the following subsections I describe each of the data sources that are linked within the CALIBER platform and their key data fields. A full CALIBER data dictionary is available in **supplementary appendix 11.1.1.**

3.3.1 Clinical Practice Research Datalink

The General Practice Research Database (GPRD) was founded in 1987 as a resource used to collect longitudinal primary care data on a large and representative sample covering 7% of the UK population.[174] In 2012 the GPRD was relaunched as the Clinical Practice Research Datalink (CPRD) signalling an emphasis towards data linkage with other anonymised data sources.

General practitioners can enter data on patients including demographics, blood test results and clinical biomarkers, disease diagnoses, prescribed medications and date of patient death.

Diagnoses, clinical test results, and procedures are coded using Read terms, a hierarchical coding system developed by Dr James Read and eventually acquired by the NHS. Read terms are very granular and there are usually multiple terms for single conditions. CPRD developed a more refined coding system, medcodes, which are mapped to Read terms for easy transition between the coding systems.

Prescriptions are recorded according to chapters of the British National Formulary (BNF), using BNF codes. Each chapter of the BNF refers to different disease areas, for example chapter 2 contains all drugs used to treat cardiovascular diseases. The subsections within chapters refer to broad drug classes for example chapter 2.9 Antiplatelet drugs, and all drugs within this chapter have the BNF code 2090000. All prescriptions also have a prodcode which is a more granular coding system than BNF codes as they vary by drug substance, dosage and manufacturer.

<u>CPRD patient and general practice data fields:</u>

patid: Encrypted CPRD patient identifier

pracid: Unique CPRD general practice identifier number

prac_region: the UK region in which the general practice is situated; The NHS administrative regions in England are North East, North West, Yorkshire and Humberside, West Midlands, East Midlands, East of England, London, South East, South Central and South East.

prac_uts: date at which the general practice data was considered to be of up-to-standard research quality. The date is determined using an algorithm which examines gaps in the data and the consistency of death recording in the general practice.

gender: Patients gender

dob: Patients date of birth. CPRD data includes year and month of birth only.

tod: date of transfer out of CPRD general practice

toreason: Reason for patients transfer out of CPRD practice. Reasons include patient death, the patient moving away from CPRD practice, or the last data collection date for the general practice.

deathdate: patient date of death as recorded by their general practitioner

date_entry: The date the patient enters the CPRD cohort, the latest date of patient registration at the CPRD practice and the date the practice reaches up-to-standard quality status

date_exit: The date at which the patient exits the CPRD cohort. The earliest date of patient date of death, transfer out of general practice date, and general practice last data collection date

CPRD diagnosis data fields:

patid: Encrypted CPRD patient identifier

eventdate: date of diagnosis record

medcode: unique CPRD diagnosis code

CALIBER phenotype category: The diagnosis category of the medcode as determined by CALIBER researchers

CPRD prescription data fields:

patid: Encrypted CPRD patient identifier

eventdate: date of prescription record

prodcode: unique CPRD product code

bnfcode: code referring to the BNF chapter and section

qty: The quantity of the product prescribed

ndd: The numeric daily dose prescribed

numdays: Length of prescription/ therapy (days)

numpacks: The number of product packs prescribed

packtype: The size or type of pack

issueseq: The sequence number for repeat prescriptions

CPRD clinical biomarkers data fields:

patid: Encrypted CPRD patient identifier

eventdate: date of record

data1: Depending on data type - usually biomarker measurement

data2: Depending on data type - usually biomarker measurement units

3.3.2 Hospital Episode Statistics

Hospital Episode Statistics (HES) captures data for all admissions to NHS hospitals in England. The primary diagnosis for the admission, admission and discharge dates and procedures performed while hospitalised can all be recorded. Diagnoses are coded using International Classification of Diseases (ICD-10) codes and procedures are recorded using Office of Population, Censuses and Surveys Classification of Surgical Operations and Procedures (OPCS) codes.

ICD-10 codes are grouped into 22 chapters which each refer to distinct disease types or diseases of particular anatomical sites. ICD-10 codes are less granular than Read terms (16,000 ICD-10 codes vs >200,000 Read codes).

HES admission data fields:

patid: Encrypted CPRD patient identifier

date_admission: date of hospital admission

date_discharged: date of hospital discharge

icd: ICD-10 code for the hospital admission

hosp_primary: Indicates if the ICD-10 code is the primary diagnosis in the hospitalisation

CALIBER phenotype category: The diagnosis category of the ICD-10 code as determined by CALIBER researchers

HES procedure data fields:

patid: Encrypted CPRD patient identifier

date_admission: date of hospital admission

date_procedure: date of procedure

date_discharged: date of hospital discharge

opcs: OPCS-4 code for the procedure

CALIBER phenotype category: The diagnosis category of the OPCS-4 code as determined by CALIBER researchers

3.3.3 Myocardial Ischaemia National Audit Project

The Myocardial Ischaemia National Audit Project (MINAP)[175] is a national registry for acute coronary events resulting in hospitalisation covering 230 hospitals in England and Wales from 2000 onwards. Data is collected on patient demographics, the MI subtype (ST-elevation MI (STEMI) or non ST-elevation MI (NSTEMI)), electrocardiogram test results, cardiac biomarkers, complications, comorbidities, procedures and treatment prior to, during and following hospitalisation.

3.3.4 Office for National Statistics

The Office for National Statistics (ONS) independently collects and produces statistics regarding the population and economy of the UK. CALIBER has linkages to the social deprivation and mortality data collected by the ONS. The social deprivation data includes the Index of Multiple Deprivation (IMD) score, which is determined by the deprivation statistics in the local geographical area of general practices and is used as a proxy for patients' socio-economic status. The mortality data consists of the date of death as recorded on death certificates and the underlying cause and up to 15 secondary causes of death which is coded using ICD-9 codes (1997-2000) and ICD-10 codes (2000 onwards).

ONS mortality data fields:

patid: Encrypted CPRD patient identifier

dod: Patient date of death as recorded on their death certificate

cod: ICD-10 code recorded as the underlying cause of death

cod1-cod15: ICD-10 codes for up to 15 secondary causes of death

3.4 Approach to data curation – principle and application

In this section I describe the approaches I took for planning and preparing CALIBER data for analysis to answer my research questions, and provide examples of from the analysis chapters 4-9, including excerpts of R programming scripts. Each of the steps described in this section are shown in **Figure 3.1.**

I provide an estimated time required to complete each step and an example timeline based on these estimates is shown in **Table 3.1**.

3.4.1 Develop a study protocol

For a given research question it is important to develop a detailed study protocol that describes the background to the research question, how you intend to answer the question and the potential impact of the findings.

A study protocol is required in the ISAC application and helps researchers to clarify their objectives and the methods and materials required to achieve the objectives. Having a study protocol also encourages scientific reproducibility. The contents of a study protocol will usually include a study background, a list of primary and secondary objectives, a statistical analysis plan and potential limitations.

See my ISAC for the prognostic modelling work which includes a study protocol in **supplementary appendix 11.1.2.**

Estimated duration: 4 weeks

3.4.2 Exploratory/feasibility analysis

Exploratory or feasibility analyses are required to determine if the available CALIBER data is sufficiently powered to answer the research question. These analyses are usually simple, for example, initial counts of the study population to estimate the sample size which in turn can inform power calculations. Such counts can be performed by the Data Lab before being granted access to the data which is subject to ISAC approval and the ISAC application requires an estimated sample size.

For example, in the ISAC for the prognostic modelling study **(Chapter 7)** I provided the following sample size estimate:

Between 2000 and 2010 there were 35,858 patients with MI in CALIBER (linked electronic health records of CPRD, HES, MINAP and ONS), of which approximately 16,000 had continuous registration for >12 prior to and after their index MI, and survived with no recurrent MI for the 12 month period following their index MI. These 16,000 post-MI 1-year survivors experienced approximately 2,400 CVD events and 1,300 non-CVD deaths over a mean follow-up of 3 (SD=2) years. The sample size and number of events will be considerably higher when data up to July 2014 is used.

Estimated duration: 1 week

3.4.3 Define study population inclusion and exclusion criteria

The study population required to answer the research question must be defined. This may be based on patient characteristics (e.g. all patients with a certain disease history), availability of data (e.g. patients with non-missing records for a certain biomarker), or date range etc.

For the prognostic modelling study, I required all patients stable following a myocardial infarction. This was defined as patients who had an MI event and survived a year without a recurrent MI. I required all patients aged above 18 years and the study period was 2000-2010. I used MI events that were captured in CPRD, HES or MINAP.

Estimated duration: 2 weeks

3.4.4 Define the variables required to fulfil study objectives

This usually includes required covariate data such as patient characteristics, medical history or biomarkers, exposures and study endpoints.

First I checked whether the phenotypes I required were published on the CALIBER portal. A large number of analysis-ready and validated phenotypes are available on the CALIBER portal [https://www.caliberresearch.org/portal], and can simply be referred to in the ISAC application and data request stage. However, for required variables without an entry in the portal it is necessary to generate codelists as a starting point to developing a new phenotype. Codelists can be generated using the *CALIBERcodelists* and *CALIBERlookups* packages in R. Codelists can be formed using search terms and keywords with Boolean operators.

For the bleeding phenotype development study I required a comprehensive list of bleeding events relevant to the harms of antithrombotic therapies. The code below is the initial search that I performed for bleeding codes in the Read dictionary in primary care. Initial search terms were informed by literature and included general bleeding related terms such as ‘bleed’ and ‘haemorrhage’ whilst initial exclusion terms aimed to remove codes that may not be relevant such as historical events and trauma. This search produced a codelist containing 502 codes

which was eventually reduced to 215 bleeding codes relevant to the study, through further refinement of the searching algorithm, manual review and clinician input. An example of refinement included removing all codes with the prefixed with the letters 'L' or 'Q' as they pertain to childbirth related bleeding (pre-/peri-/post-natal bleeding).

```
> setdictionary('read')

CALIBERcodelists package, version 0.2.5
Clearing categories in master dictionary.
Using read dictionary.
CPRD Pegasus Medical Dictionary, March 2010
>
> codes<-termhas("haemorrhage"
+           %OR% "bleed"
+           %OR% "blood in stool"
+           %OR% "blood in faeces"
+           %OR% "peptic ulcer"
+           %OR% "gastric ulcer"
+           %OR% "duodenal ulcer"
+           %OR% "gastrojejunal ulcer"
+           %OR% "haematuria"
+           %OR% "haemoptysis"
+           %OR% "haemoperitoneum"
+           %OR% "melaena"
+           %OR% "haematemesis"
+           %OR% "haemorrhagic"
+           %OR% "epistaxis"
+           %OR% "haematoma"
+           %AND% NOT("partum")
+           %AND% NOT("fever")
+           %AND% NOT("history")
+           %AND% NOT("H/O")
+           %AND% NOT("FH")
+           %AND% NOT("F/H")
+           %AND% NOT("disorder")
+           %AND% NOT("screen")
+           %AND% NOT("symptom")
+           %AND% NOT("traumatic")
+           %AND% NOT("operation")
+           )
> bleed_cprd_initial <- data.frame(codelist(codes))
```

Once the bleeding Read codes were finalised, I ran an initial search of bleeding ICD-10 codes by converting the Read codes to ICD-10 using the mappings built into the *CALIBERcodelists* package.

```
> # Convert bleeding Read codelist to ICD-10
> # bleed_cprd is the final list of bleeding Read codes
>
> bleed_cprd <- codelist(bleed_cprd)
>
> bleed_icd10 <- data.frame(convert(bleed_cprd,
+                                   toDictionary="icd10",
+                                   fromDictionary="read"))
```

The result was not perfect and required further refinement, but provided a good starting point for developing the ICD-10 codelist. The refinement measures included removing ICD-10 codes picked up in the mapping that were not related to bleeding and searching the ICD-10 dictionary for further bleeding codes that may have been missed.

To a produce a OPCS codelist for endoscopic examinations relevant to the development of bleeding phenotypes, I performed an initial search using the term 'endoscop' to allow for codes with the terms endoscopy or endoscopic and 'exam' or 'diag' to limit the search result to include examination or diagnostic procedures. I also included a range of NOT statements to exclude OPCS terms not relevant to the study.

```
> setdictionary('opcs')

CALIBERcodelists package, version 0.2.9
Clearing categories in master dictionary.
Using opcs dictionary.
Office of Population Censuses and Surveys Codes 4.6, Jan 2011 V1.0, NH
S Classifications Service
>
> codes<-termhas("endoscop"
+                %AND% ("exam"%OR% "diag")
+                %AND% NOT("biopsy")
+                %AND% NOT("lavage")
+                %AND% NOT("organ")
+                %AND% NOT("fetus")
+                %AND% NOT("insert")
+                %AND% NOT("brush")
+                %AND% NOT("stain")
+                %AND% NOT("sampling")
+                %AND% NOT("joint"))
>
> endoscopy_opcs <- data.frame(codelist(codes))
```

To develop codelists for prescription data, we can perform searches of the CALIBER product dictionary. We may search for prescriptions by applying filters to columns containing BNF chapter names, BNF codes (derived from BNF chapter numbers) and drug substances. For example, I performed a search of the dictionary for all products in the BNF chapter 2.11: Antifibrinolytic drugs and Haemostatics using the BNF code '02110000' as shown in the code excerpt below.

```
> # load the CALIBER product dictionary
> data(CALIBER_PRODDICT)
>
> # search for all products in BNF chapter 2.11
> # Antifibrinolytic drugs and haemostatics
>
> antifib_drugs <- CALIBER_PRODDICT[which(grepl("02110000",bnfcode)),]
```

Estimated duration: 2 weeks

3.4.5 Apply for ISAC approval

Next an ISAC application must be submitted and approved. The application requires an overall study protocol including study background and objectives, defined study population, exposures, and endpoints which includes codelists or described phenotypes as appropriate and a detailed statistical analysis plan. The application is peer-reviewed by the ISAC committee and we received feedback approximately within 4 weeks of submission. Usually minor revisions and resubmission is required prior to acceptance.

The complete ISAC application form for the prognostic modelling study is shown in **supplementary material 11.1.2.**

Estimated duration: 4 weeks

3.4.6 Request the data

Once the study has received approval from ISAC, details of the study population and variables required can be compiled into a simple spreadsheet which is then provided to the Data Lab team member who will extract the data. The data request for the bleeding phenotype study is shown in **Figure 3.5.**

Estimated duration: 3 days

3.4.7 Receive the linked data

We receive the data in the format of multiple '.csv' files **(Figure 3.6).** Each file corresponds to longitudinal records for a single variable from a single data source. The format of the file name is 'variable_datasource' for example the file name af_hes contains all atrial fibrillation admission records in HES for patients in the cohort.

We also receive a general cohort file with each patient's CPRD general practice identifier, date of birth, gender, entry and exit dates to their CPRD general practice, ONS cause of death ICD-10 codes and IMD score.

Estimated duration: 4 weeks

3.4.8 Construct cohort: applying inclusion and exclusion criteria

The data is usually extracted for a broader cohort and we apply the inclusion and exclusion criteria ourselves.

For example, for the prognostic modelling study, I was provided with a table of all patients who had an MI record in CPRD, HES or MINAP, and the dates of all of their MI records. With this data, I had to determine which patients could be classed as stable post-MI, i.e. the patients that survived 1 year following an MI without a recurrent event.

Estimated duration: 1 week

3.4.9 Develop new phenotypes

For the data required that does not have a phenotype defined in the CALIBER portal, it is necessary to develop a new phenotype. Once the phenotype has been tested and finalised it can be uploaded to the CALIBER portal where other researchers can access it and provide feedback.

In my thesis I demonstrated the process of developing a bleeding phenotype algorithm relating to the side effects of antithrombotic therapy use **(Chapter 5).** Within the phenotype I defined 3 distinct levels of bleeding event severity.

The development of the bleeding phenotype is fully described in **Chapter 5.** In brief, I developed bleeding Read and ICD-10 codelists and performed a descriptive analysis of bleeding events captured in CPRD, HES, MINAP and ONS. I assessed the prevalence of procedures, diseases diagnoses, biomarkers and prescriptions recorded within different time windows (e.g. same day, within 30 days) with respect to the bleeding records. I identified events and records that may be associated with bleeding severity, by using short term all-cause mortality used as a proxy for severity. Therefore I was able to determine which factors were suitable for use in the phenotype and formulated an algorithm **(Figure 5.18)**. The bleeding phenotype algorithm was validated through application to the acute coronary disease and atrial fibrillation population and assessing the incidence and outcomes associated with bleeding events of differing severity.

Estimated duration: 2 weeks

3.4.10 Generate study exposures

We generally need to transform the longitudinal records into a single variable per patient to describe their characteristics and disease status at baseline.

Below I show the R scripts I used to apply the alcohol phenotype to the bleeding phenotype cohort, to determine which patients have a history of excess alcohol consumption. This phenotype uses patient gender, alcohol drinking status, alcohol units consumed (continuous and categorical) primary care records to determine patients alcohol consumption status.

The excess alcohol consumption phenotype algorithm is shown as a visualisation in **Figure 3.7.**

```
> setwd("R:/Pop_Health/EPH_CEG_CALIBER/_KEEP/14_133_Bleeding")
>
> # read in cohort patient ID's
> cohort <- fread("Processed cohort/cohort_patid.csv")
>
```

```
> # read in alcohol drinker status  phenotype components
>
> # CPRD alcohol status (categorical)
> alcohol_drinker <- fread("cal_14_133_cohort_alcohol_drinker_gprd.csv")
>
> # CPRD alcohol units (continuous) records
> alcohol_units <- fread("cal_14_133_cohort_alcohol_units.csv")
>
> # CPRD alcohol units (categorical) records
> alcohol_units_cat <- fread("cal_14_133_cohort_alcohol_units_cat_gprd.csv")
>
> # read in gender data
> gender <- fread("Processed cohort/cohort_entry_events_demog_death.csv")
> gender <- gender[,c("patid","gender"),with=F]
>
> # merge cohort and gender data
> cohort1 <- merge(cohort, gender, by="patid", all.x=T)
>
> # apply category labels to gender column
> cohort1$gender <- factor(cohort1$gender)
> levels(cohort1$gender) <- c("Men", "Women")
>
> # CALIBER alcohol composite phenoptype algorithm
>
> # If Read code in alcohol drinker code list,
> # THEN alcohol_drinker_comp = relevant category
>
> # IF sex = 0 AND (alcohol_units > 21 per week OR alcohol_units_cat = 4 or 5),
> # THEN alcohol_drinker_comp = 4 (Excess drinker)
> # IF sex = 1 AND (alcohol_units > 14 per week OR alcohol_units_cat = 3, 4 or 5),
> # THEN alcohol_drinker_comp = 4 (Excess drinker)
>
> # IF sex = 0 AND (alcohol_units >= and <= 21 per week OR alcohol_units_cat = 2 or 3),
> # THEN alcohol_drinker_comp = 3 (Current drinker)
> # IF sex = 1 AND (alcohol_units >= 1 and <= 14 per week OR alcohol_units_cat = 2),
> # THEN alcohol_drinker_comp = 3 (Current drinker)
>
> # IF sex = 0 AND (alcohol_units <1 on single day OR alcohol_units_cat = 1),
> # THEN alcohol_drinker_comp = 2 (Occasional drinker)
> # IF sex = 1 AND (alcohol_units < 1 on single day OR alcohol_units_cat = 1),
> # THEN alcohol_drinker_comp = 2 (Occasional drinker)
>
>
> # apply category labels to CPRD alcohol status
> alcohol_drinker[ ,alcohol_drinker_gprd:=as.factor(alcohol_drinker_gprd)]
> levels(alcohol_drinker$alcohol_drinker_gprd)<- c("Non-drinker", "Ex-drinker",
+                                                  "Occasional drinker", "Current drinker",
+                                                  "Excess drinker", "Excess drinker",
+                                                  "Drink status NOS")
>
> # merge cohort and alcohol status data
> cohort_alc <- merge(cohort1, alcohol_drinker, by="patid", all.x=T)
>
> # calculate time between baseline date and alcohol record
> cohort_alc[, alc_t:=as.numeric(difftime(eventdate.y, eventdate.x, units="days"))]
>
> # create binary exceess alcohol variable
> # if alcohol category is excess drinker and
> # the record preceeds cohort entry
> cohort_alc[,alcohol_history:= as.numeric(alc_t<0 & alcohol_drinker_gprd=="Excess drink
er")]
>
> # remove duplicate records
```

```
> cohort_alc<-unique(cohort_alc[,c("patid","eventdate.x","gender","alcohol_history"),wit
h=F])
>
> # merge with numeric units data
> cohort_alc1 <- merge(cohort_alc, alcohol_units, by="patid", all.x=T)
>
> # calculate time between baseline date and alcohol units record
> cohort_alc1[, alc_t:=as.numeric(difftime(eventdate, eventdate.x, units="days"))]
>
> # determine which patients have excess alcohol history
> # according to gender and units consumption (continuous) prior to baseline date
> cohort_alc1[which(alc_t<0 & alcohol_units>21 & gender=="Men"),alcohol_history:=1]
> cohort_alc1[which(alc_t<0 & alcohol_units>14 & gender=="Women"),alcohol_history:= 1]
>
> # remove duplicate records
> cohort_alc1<-unique(cohort_alc1[,c("patid","eventdate.x","gender","alcohol_history"),w
ith=F])
>
> # merge with categorical units data
> cohort_alc2 <- merge(cohort_alc1, alcohol_units_cat, by="patid", all.x=T)
>
> # calculate time between baseline date and alcohol units record
> cohort_alc2[, alc_t:=as.numeric(difftime(eventdate, eventdate.x, units="days"))]
>
> # determine which patients have excess alcohol history
> # according to gender and units consumption (categorical) prior to baseline date
> cohort_alc2[which(alc_t<0 & alcohol_units_cat_gprd %in% 4:5 & gender=="Men"), alcohol_
history:= 1]
> cohort_alc2[which(alc_t<0 & alcohol_units_cat_gprd %in% 3:5 & gender=="Women"), alcoho
l_history:= 1]
>
> # remove duplicate records
> cohort_alc2 <- unique(na.omit(cohort_alc2[,c("patid","eventdate.x","gender","alcohol_h
istory"), with=F]))
>
> cohort_excess_alc_hist <- merge(cohort, cohort_alc2, all.x=T, by="patid")
>
> # if patients have no past alcohol records, assume no excess alcohol history
> cohort_excess_alc_hist[which(is.na(alcohol_history)), alcohol_history:=0]
```

For patients biomarker data we generally select the nearest record to their baseline date within an appropriate time window. For example, we could select the nearest biomarker measure to baseline within a year prior to baseline. For patients with no record in the specified time window, their data for that biomarker is recorded as missing. Other approaches may be taken to generate baseline biomarker data, for example averaging multiple records instead of selecting a single value.

Below I show an example of how I selected baseline blood pressure records for the bleeding phenotype cohort. For this study I defined baseline blood pressure as the record nearest to patients baseline date, but not more than a year prior.

```
> library(data.table)
>
> setwd("R:/Pop_Health/EPH_CEG_CALIBER/_KEEP/14_133_Bleeding")
>
> # read in cohort patient ID's
```

```
> cohort <- fread("Processed cohort/cohort_patid.csv")
>
> # set format of cohort baseline dates
> cohort$eventdate <- as.Date(cohort$eventdate)
>
> # read in longitudinal blood pressure records
> bp <- fread("cal_14_133_cohort_bp_gprd.csv")
Read 11363905 rows and 9 (of 9) columns from 0.475 GB file in 00:00:22
>
> # set format of bp record dates
> bp$eventdate <- as.Date(bp$eventdate)
>
> # merge cohort patient ids and bp records
> # retains only patients with at leastone bp record
> cohort_bp <- merge(cohort, bp, by="patid")
>
> # calculate time difference (days) between
> # cohort entry and blood pressure records
> cohort_bp[, t_diff:=eventdate.y-eventdate.x]
>
> # select only blood pressure records within a year prior to baseline
> cohort_bp <- cohort_bp[which(t_diff<=0 & t_diff>-366),]
>
> cohort_bp <- cohort_bp[,c("patid","sys_bp","t_diff"),with=F]
>
> setkeyv(cohort_bp,c("patid","t_diff"))
>
> # choose the most recent record relative to baseline dates
> cohort_bp1 <- cohort_bp[,.SD[.N], by="patid"]
>
> # merge bp data back with cohort patient ID's
> # patients with no bp record will have 'NA' entry
> cohort_bp_all <- merge(cohort, cohort_bp1, by="patid", all.x=T)
```

Estimated duration: 1 week

3.4.11 Generate endpoints

Next, we need to generate endpoints for patients. I mostly used time-to-event analysis in this thesis, therefore the endpoint information required is patient follow up time (i.e. the time between cohort entry and the event of interest or loss to follow up) and an indicator variable for whether the patient had an event or was censored.

For cause-specific death, researchers have the option use events where the cause of interest was recorded as the underlying cause of death and in any of the 15 secondary cause of death fields.

Below is an example of generating all-cause mortality time to event variables for the bleeding phenotype cohort.

```
> # create all-cause mortality
> # time to event variables
>
> # Initially set death binary indicator as 0 for all patients
> cohort[, dead:=0]
>
```

```
> # format ONS date of death
> cohort[ ,dod:=as.Date(dod)]
>
> # If patients have an ONs date of death
> # set the death indicator to 1
> cohort[which(!is.na(dod)), dead:=1]
>
> # create a new death date variable
> # date of death for patients who died
> # end of follow up for patients who were censored
> cohort[ ,dead_date:=as.Date(NA)]
> cohort[which(dead==1), dead_date:=dod]
> cohort[which(dead==0), dead_date:=date_exit]
>
> # patient follow up time from cohort entry to death or censoring
> cohort[,dead_t:=as.numeric(difftime(dead_date, eventdate,units="days
"))]
```

Estimated duration: 3 days

3.5 CALIBER study populations and phenotypes used in this thesis

Following the processes described in the previous section, here I describe and summarise the study populations and variables used in the analyses chapters of this thesis.

3.5.1 Study populations

In this section I describe the inclusion and exclusion criteria that were used to define the 4 distinct study populations used in this thesis. The study populations described in this section are summarised in **Table 3.3**.

3.5.1.1 Chapter 4: Indications for antithrombotic therapies in linked electronic health records

All patients in CALIBER with at least one antithrombotic therapy prescription [BNF chapters 2.8: Anticoagulants and protamine and 2.9: Antiplatelet drugs] between 1997 and 2010 were entered into the study population.

Patients were excluded if their first antithrombotic therapy prescription was within six months of entry to the CPRD cohort (i.e. excluding those on continuing prescriptions). Patient's index date was the date of their first ATT prescription.

3.5.1.2 Chapter 5: Development and validation of prognostic models for atherothrombotic and bleeding events in stable myocardial infarction survivors

For this study, patients who had an MI and survived 1 year without any recurrent MI's were entered into the cohort between 2000 and 2010. Their cohort entry date was their 1 year post the date of their index MI. The algorithm to determine MI patients is as follows:

Myocardial Infarction **(**https://www.caliberresearch.org/portal/show/phenotype_mi)

Patients had a diagnosis for myocardial infarction if they had a STEMI, NSTEMI or unspecified acute MI record in CPRD, an unspecified acute MI record in HES, a STEMI or NSTEMI recorded in MINAP, or a transluminal coronary thrombolysis procedure in OPCS

Patients were split into model development and model validation cohorts depending on the region of their general practice using a North-South divide. Patients registered at general practices in East Midlands, East of England, Greater London, South East, South Central and South West were used for model development and patients registered at general practices in the West Midlands, Yorkshire and Humberside, North East and North West were used for model validation.

3.5.1.3 Chapter 6: Developing bleeding phenotypes in CALIBER; Chapter 7: Incidence of bleeding events in four common cardiovascular diseases

The cohort used for this study was patients with a diagnosis of atrial fibrillation, myocardial infarction, stable angina or unstable angina, excluding those whose first events were fatal between 1997 and 2010. Patients were entered into the cohort at their first diagnosis of one of the four diseases. The complete codelists and phenotype descriptions can be found on the CALIBER portal:

Atrial fibrillation **(**https://www.caliberresearch.org/portal/show/phenotype_af)

Patients had a diagnosis for atrial fibrillation if they had an atrial fibrillation record in HES or CPRD, excluding those with solely historical or monitoring records. Further cases were inferred for patients with no atrial fibrillation record in either HES or CPRD but have a prescription for warfarin and no history of heart valve replacement or prior deep vein thrombosis or pulmonary embolism or if they have a prescription for digoxin and no history of heart failure.

Myocardial Infarction **(**https://www.caliberresearch.org/portal/show/phenotype_mi)

Patients had a diagnosis for myocardial infarction if they had a STEMI, NSTEMI or unspecified acute MI record in CPRD, an unspecified acute MI record in HES, a STEMI or NSTEMI recorded in MINAP, or a transluminal coronary thrombolysis procedure in OPCS

Unstable angina (https://www.caliberresearch.org/portal/show/phenotype_ua)

Patients had a diagnosis of unstable angina if they had unstable angina diagnosed in primary care, secondary care (HES or MINAP), or a diagnosis of acute coronary syndrome in primary care or an admission for acute ischaemic heart disease in secondary care.

Stable angina (https://www.caliberresearch.org/portal/show/phenotype_sa)

Stable angina was diagnosed based on chest pain attributed to coronary causes recorded in CPRD, a stable angina record in CPRD, a stable angina admission in HES, prescribed anti-anginal medication, primary or secondary care records of PCI or CABG performed or abnormal test results for any of stress echocardiogram, invasive coronary angiogram, CT coronary angiogram, MR coronary angiogram, exercise ECG, resting ECG or myocardial perfusion scan.

3.5.1.4 Chapter 8: Predictors and outcomes of INR time in therapeutic range; Chapter 9: The predictive value of measures of INR control for atherothrombotic and bleeding outcomes

All patients with an oral anticoagulant prescription [BNF chapter 2.8.2] and at least 5 consecutive INR records were included in the study population. INR records were considered

to be consecutive with they occurred within 90 days of each other. For patients with multiple valid spells of INR monitoring, I chose the earliest spell as the index INR spell used in analysis.

3.5.2 Risk factor and covariate phenotypes

Risk factors and exposures available to use within CALIBER include disease history, procedures, longitudinal records of clinical biomarkers and records of prescribed medications. All of the risk factor and covariate phenotypes described in this section are summarised in **Table 3.3**

3.5.2.1 Demographics

The majority of demographics data used in this thesis are available from CPRD. Age is calculated using the date of birth recorded in CPRD (only birth month and year available) and the dates of cohort entry. Sex, ethnicity and the geographical area of CPRD practices are also available. The index of multiple deprivation (IMD) is available from ONS, which is used as a measure of patients socioeconomic status.

3.5.2.2 Behaviours

The two behaviours used in analysis were smoking status and alcohol consumption. Both of these variables have composite CALIBER phenotypes. The smoking phenotype uses smoking status records in CPRD and HES as well as prescription records for smoking cessation products to define patients smoking status. The alcohol phenotype uses records of drinker status and alcohol units consumed captured in CPRD to define patients drinking status.

3.5.2.3 Cardiovascular disease history

All of the cardiovascular disease history variables were defined using data from more than one source. For example, to define a history of myocardial infarction, diagnosis records from CPRD, HES and MINAP as well as thrombolysis procedures OPCS-4 codes are used. Other phenotypes, such as hypertension and atrial fibrillation use prescription records in their definitions, for blood pressure lowering medication and warfarin or digoxin respectively. The hypertension phenotype also uses also uses blood pressure records.

3.5.2.4 Non-cardiovascular disease history

All of the non-cardiovascular disease history variables were defined using data from more than one source. However at this stage, few of the non-cardiovascular disease variables have definitions that use data beyond diagnosis records in CPRD and HES.

3.5.2.5 Procedures

Procedures phenotypes, such as those for transfusion and revascularisation were defined using OPCS-4 and Read codes.

3.5.2.6 Clinical biomarkers

Clinical biomarker records were obtained from CPRD only. The biomarkers had varying levels of completeness. For example heart rate had a high rate of missingness whereas blood pressure was relatively well recorded.

3.5.2.7 Prescribed medication

Prescription records were obtained from CPRD only. A range of issued prescriptions for cardiovascular and non-cardiovascular therapy were used as covariates in analysis. In general they were included in analysis using a binary indicator for prescriptions issued prior to baseline, or as an estimated duration of therapy prior to baseline.

3.5.3 Endpoint phenotypes

The study endpoints used in this thesis are described in this section are summarised in **Table 3.4**.

All-cause mortality

All-cause mortality is defined using date of death recorded in ONS, and CPRD for those without a date of death recorded in ONS. Patients without a date of death in either source are considered not to have died.

Cardiovascular death

Cardiovascular death is defined using ONS data as one of the following ICD-10 codes: I00-I02, I05-I09, I10-I15, I20-I25, I26-I28, I30-I52, I60-I69, I70-I79, I80-I89 and I95-I99 or for deaths prior to 2001, ICD-9 codes: 390, 4010, 4011, 4019, 4200, 4211, 4220, 42490, 42491, 42499, 4258, 4299, 449, 452, 7854, recorded as the underlying cause of death.

Cardiovascular death, stroke or myocardial infarction

Cardiovascular death, stroke or myocardial infarction is a composite of cardiovascular death, ischaemic or unspecified stroke and myocardial infarction and is defined as the first occurrence of the aforementioned events in follow up.

Bleeding

The definition of bleeding events evolves throughout this thesis.

For the prognostic modelling chapter **[chapter 5]**, three bleeding endpoints of varying incidence and severity were used. The definitions and codes used for the three bleeding endpoints in this study are shown in **Table 3.5.**

The first, fatal or hospitalised bleeding was intended to be a broader, more common less severe bleeding endpoint, followed by CALIBER major bleeding which was constructed emulate TIMI major bleeding, the primary safety endpoint of the PEGASUS-TIMI trial,[17] as closely as possible by including intracranial bleeding, fatal bleeding, bleeding requiring transfusion and bleeding requiring an extended hospitalisation. Finally the fatal or intracranial bleeding was the rarest but most severe bleeding endpoint.

In chapter 5, bleeding ICD-10 and Read codes were reviewed and a more comprehensive bleeding phenotype algorithm was developed **(Figure 5.18).**

3.6 Limitations of CALIBER

CALIBER is not without its limitations. The version of CALIBER used in this thesis covers records from 1997 to 2010, which may be considered out of date, especially for research in medical areas with newly approved treatments and changing guidelines. However, data updates are subject to the data licenses held by the Institute and can be a time consuming process.

Linked data is only available for patients enrolled at CPRD general practices that consent to linkage. HES and ONS have national coverage and therefore the data for patients who are not enrolled at consenting practices is not available in CALIBER.

Missing data is an issue for observational cohorts; in particular for electronic health records it is difficult to know what the missingness mechanism of data may be therefore not all methods for handling missing data may be appropriate. For example, a missing entry for a disease diagnosis may imply a patient does not have the disease, or that the patient has the disease but it has not been coded. Not all general practices receive electronic records for laboratory tests. Therefore missing data is common for clinical biomarkers. There are also cases where the data is genuinely missing for example missing discharge dates in HES records or missing procedure dates in OPCS-4 records.

Administrative health data is not designed for research and therefore requires a great deal of cleaning and manipulation to create research ready datasets which is relatively time consuming.

There are occasionally conflicts or discrepancies in data between the sources, for example the date or type of a single event may be recorded differently in CPRD and HES. It is up to researchers how to handle such issues.

The gold-standard method for validating electronic health records and disease phenotypes is to compare them with manual charts and calculate statistics such as the positive predictive

value, sensitivity and specificity. However this process is beyond the scope of the CALIBER remit, and would require separate approvals to contact general practitioners, patients and to gain access to and review medical charts - a time consuming and expensive task.

Recommendations

CALIBER could benefit from widening available linkages to include HES outpatient data, HES accident and emergency data, prescriptions issued in hospitals, biomarker records captured during hospitalisation. These records could provide marginal improvements to case ascertainment and also widen the field of research available to CALIBER users.

An automated data request and delivery process implemented via the CALIBER portal could potentially be more efficient than current 'manual' methods. Given that disease phenotypes are available on the portal, researchers could select the appropriate phenotypes to form their study population and the phenotypes required as study variables. The request could be sent via the data lab for their approval and an automated generation of the data can be carried out.

A graphical user interface for generating top-level data summaries and visualisations from the CALIBER data, for example, to aid feasibility and exploratory analysis, may be useful. This would allow an overview of the data to be more accessible to researchers without a data science or statistical background.

3.7 Conclusion

In this chapter I described the approaches taken to prepare data from the CALIBER platform ready for analysis to answer the research questions in chapters 4 to 9.

3.8 Tables and figures

Figure 3.1: The steps I took to plan and curate CALIBER linked electronic health records for statistical analysis

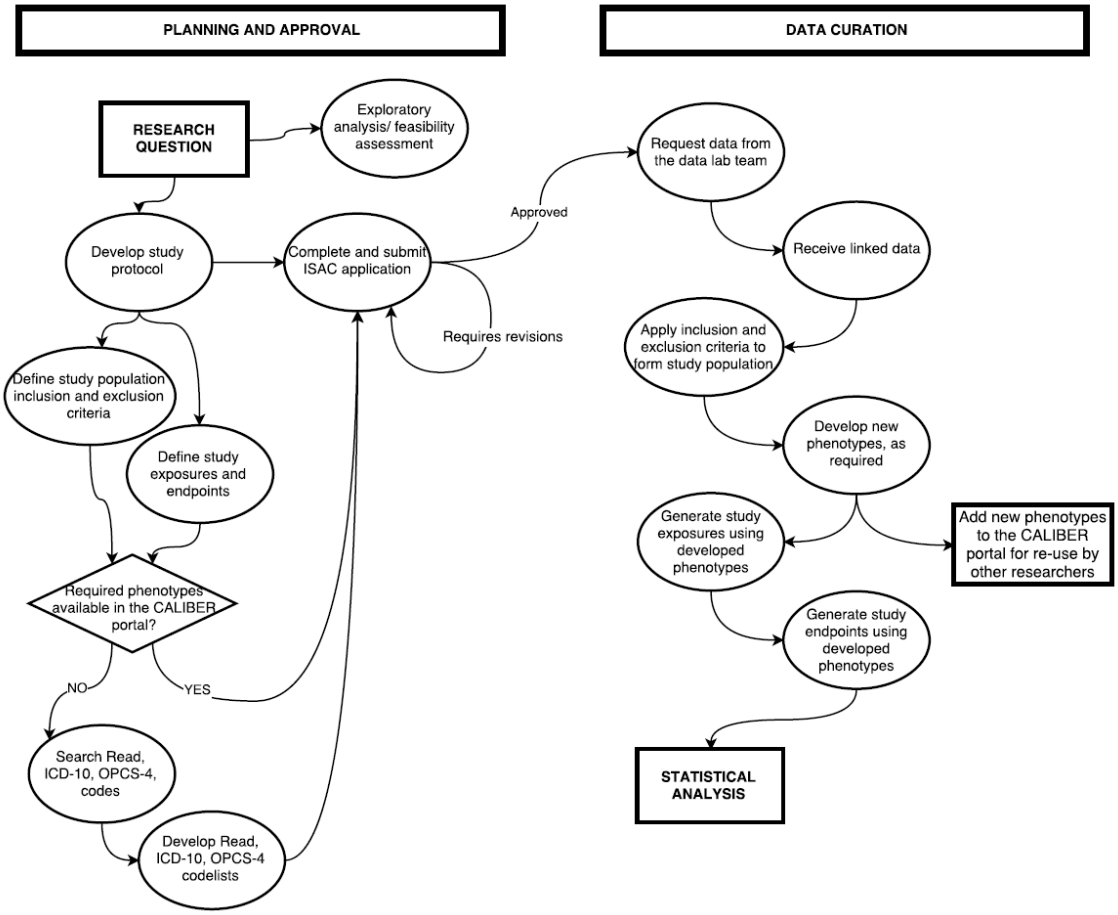
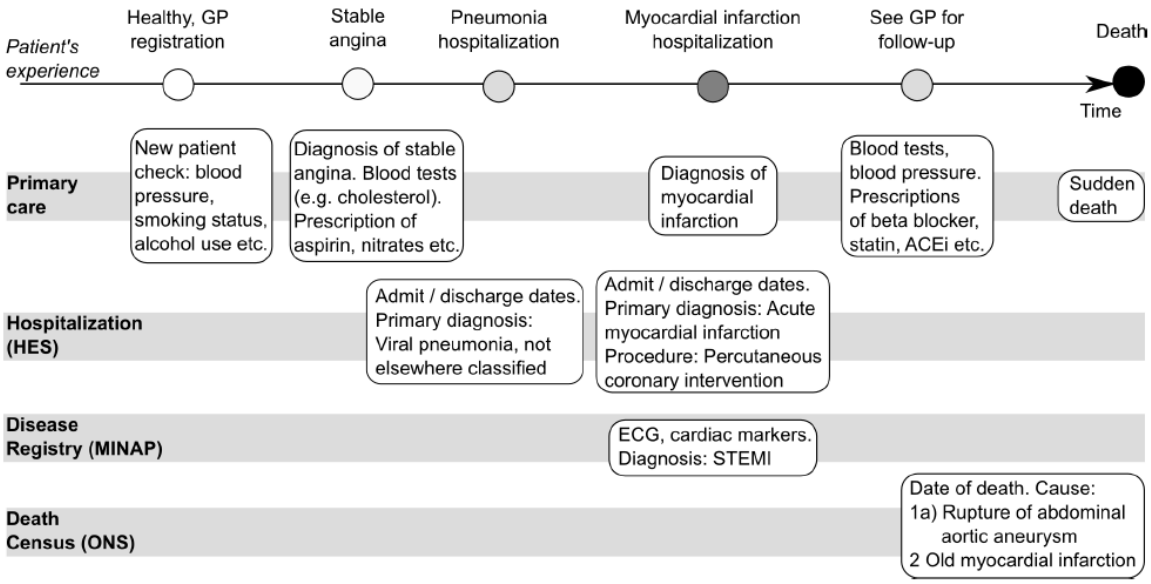
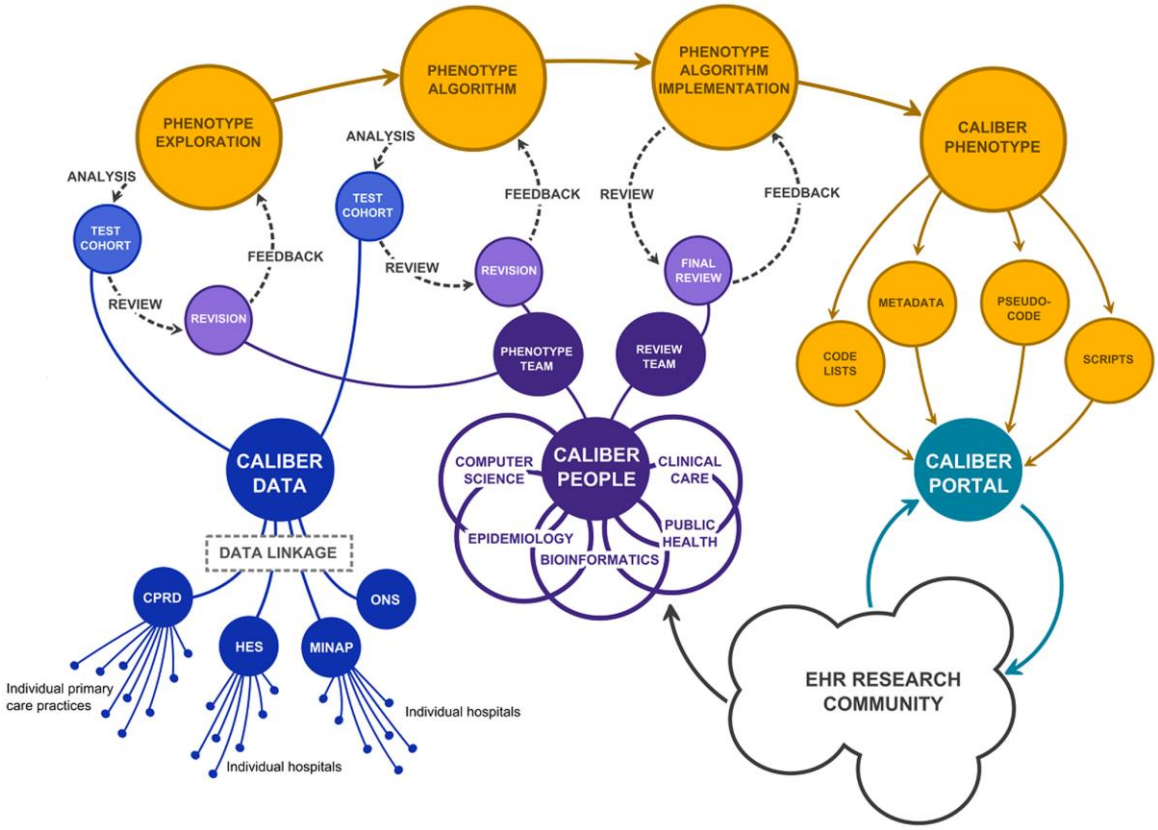


Figure 3.2: Demonstration of how linked data provides information across a patient's medical journey



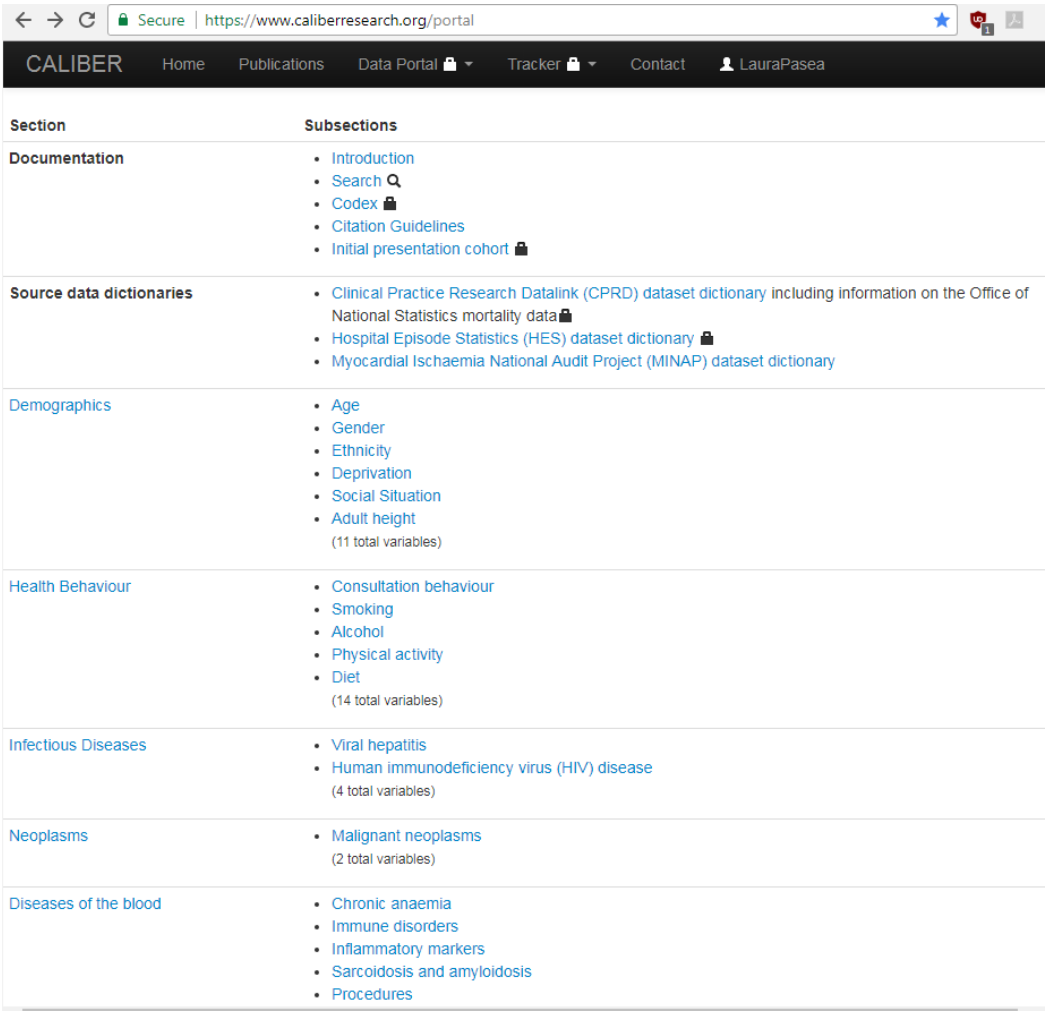
Denaxas et al, Data Resource Profile: CALIBER, *Intl Journal Epidemiology*, 2012

Figure 3.3: The process of phenotype development in the CALIBER platform



Sourced from: Morley et al, 2014, Defining Disease Phenotypes Using National Linked Electronic Health Records: A Case Study of Atrial Fibrillation. PLOS ONE 9(11): e110900.

Figure 3.4: CALIBER data portal front page

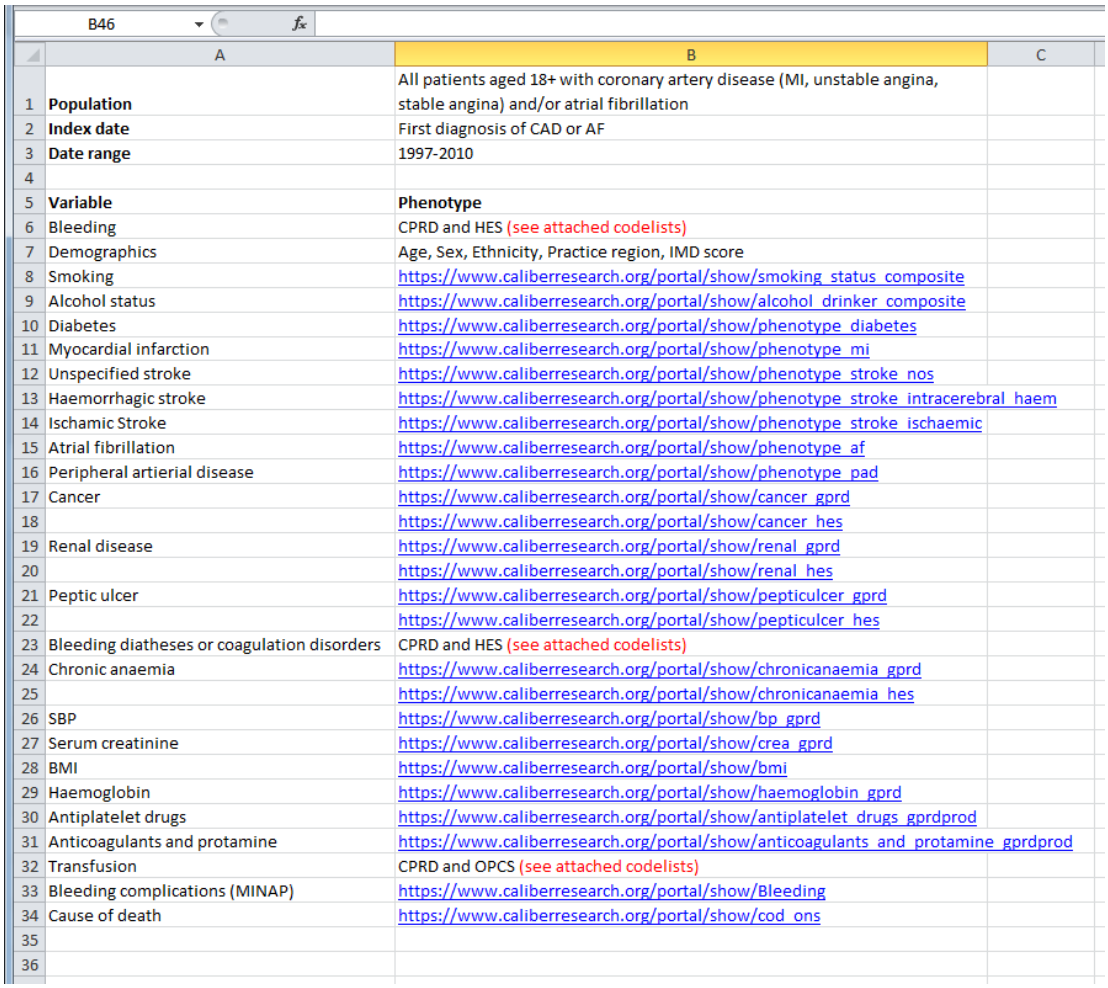


Section	Subsections
Documentation	• Introduction • Search • Codex • Citation Guidelines • Initial presentation cohort
Source data dictionaries	• Clinical Practice Research Datalink (CPRD) dataset dictionary including information on the Office of National Statistics mortality data • Hospital Episode Statistics (HES) dataset dictionary • Myocardial Ischaemia National Audit Project (MINAP) dataset dictionary
Demographics	• Age • Gender • Ethnicity • Deprivation • Social Situation • Adult height (11 total variables)
Health Behaviour	• Consultation behaviour • Smoking • Alcohol • Physical activity • Diet (14 total variables)
Infectious Diseases	• Viral hepatitis • Human immunodeficiency virus (HIV) disease (4 total variables)
Neoplasms	• Malignant neoplasms (2 total variables)
Diseases of the blood	• Chronic anaemia • Immune disorders • Inflammatory markers • Sarcoidosis and amyloidosis • Procedures

Table 3.1: Estimated timeline for CALIBER data preparation

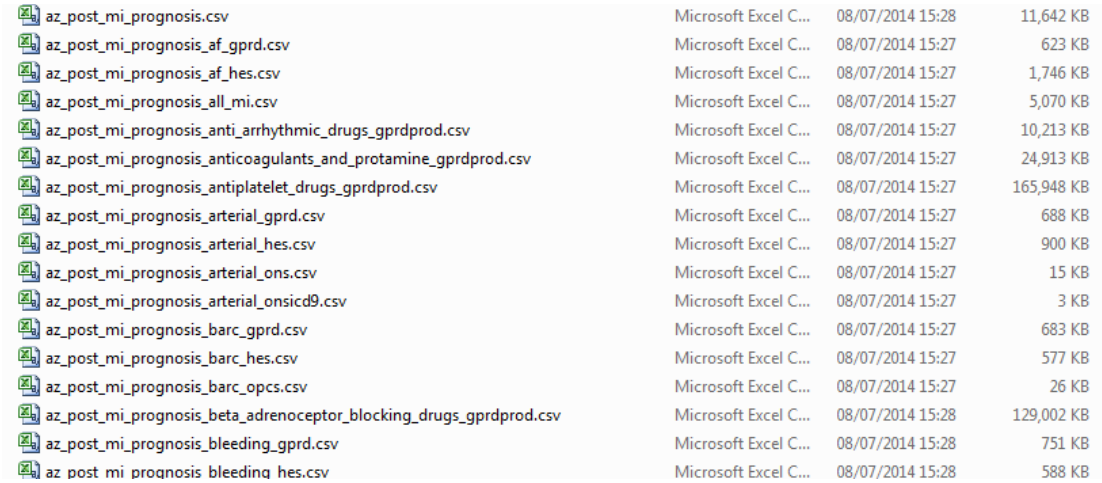
Step	Week															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Develop study protocol																
Feasibility analysis																
Define study population																
Define study variables																
ISAC approval																
Request data										Data request being processed and fulfilled						
Receive data																
Construct cohort																
Develop phenotypes																
Generate study exposures																
Generate study endpoints																

Figure 3.5: Data request provided to the data manager for the bleeding phenotype study



	A	B	C
1	**Population**	All patients aged 18+ with coronary artery disease (MI, unstable angina, stable angina) and/or atrial fibrillation	
2	**Index date**	First diagnosis of CAD or AF	
3	**Date range**	1997-2010	
4			
5	**Variable**	**Phenotype**	
6	Bleeding	CPRD and HES (see attached codelists)	
7	Demographics	Age, Sex, Ethnicity, Practice region, IMD score	
8	Smoking	https://www.caliberresearch.org/portal/show/smoking_status_composite	
9	Alcohol status	https://www.caliberresearch.org/portal/show/alcohol_drinker_composite	
10	Diabetes	https://www.caliberresearch.org/portal/show/phenotype_diabetes	
11	Myocardial infarction	https://www.caliberresearch.org/portal/show/phenotype_mi	
12	Unspecified stroke	https://www.caliberresearch.org/portal/show/phenotype_stroke_nos	
13	Haemorrhagic stroke	https://www.caliberresearch.org/portal/show/phenotype_stroke_intracerebral_haem	
14	Ischamic Stroke	https://www.caliberresearch.org/portal/show/phenotype_stroke_ischaemic	
15	Atrial fibrillation	https://www.caliberresearch.org/portal/show/phenotype_af	
16	Peripheral artierial disease	https://www.caliberresearch.org/portal/show/phenotype_pad	
17	Cancer	https://www.caliberresearch.org/portal/show/cancer_gprd	
18		https://www.caliberresearch.org/portal/show/cancer_hes	
19	Renal disease	https://www.caliberresearch.org/portal/show/renal_gprd	
20		https://www.caliberresearch.org/portal/show/renal_hes	
21	Peptic ulcer	https://www.caliberresearch.org/portal/show/pepticulcer_gprd	
22		https://www.caliberresearch.org/portal/show/pepticulcer_hes	
23	Bleeding diatheses or coagulation disorders	CPRD and HES (see attached codelists)	
24	Chronic anaemia	https://www.caliberresearch.org/portal/show/chronicanaemia_gprd	
25		https://www.caliberresearch.org/portal/show/chronicanaemia_hes	
26	SBP	https://www.caliberresearch.org/portal/show/bp_gprd	
27	Serum creatinine	https://www.caliberresearch.org/portal/show/crea_gprd	
28	BMI	https://www.caliberresearch.org/portal/show/bmi	
29	Haemoglobin	https://www.caliberresearch.org/portal/show/haemoglobin_gprd	
30	Antiplatelet drugs	https://www.caliberresearch.org/portal/show/antiplatelet_drugs_gprdprod	
31	Anticoagulants and protamine	https://www.caliberresearch.org/portal/show/anticoagulants_and_protamine_gprdprod	
32	Transfusion	CPRD and OPCS (see attached codelists)	
33	Bleeding complications (MINAP)	https://www.caliberresearch.org/portal/show/Bleeding	
34	Cause of death	https://www.caliberresearch.org/portal/show/cod_ons	
35			
36			

Figure 3.6: Example of CALIBER cohort and variable files extracted



az_post_mi_prognosis.csv	Microsoft Excel C...	08/07/2014 15:28	11,642 KB
az_post_mi_prognosis_af_gprd.csv	Microsoft Excel C...	08/07/2014 15:27	623 KB
az_post_mi_prognosis_af_hes.csv	Microsoft Excel C...	08/07/2014 15:27	1,746 KB
az_post_mi_prognosis_all_mi.csv	Microsoft Excel C...	08/07/2014 15:27	5,070 KB
az_post_mi_prognosis_anti_arrhythmic_drugs_gprdprod.csv	Microsoft Excel C...	08/07/2014 15:27	10,213 KB
az_post_mi_prognosis_anticoagulants_and_protamine_gprdprod.csv	Microsoft Excel C...	08/07/2014 15:27	24,913 KB
az_post_mi_prognosis_antiplatelet_drugs_gprdprod.csv	Microsoft Excel C...	08/07/2014 15:27	165,948 KB
az_post_mi_prognosis_arterial_gprd.csv	Microsoft Excel C...	08/07/2014 15:27	688 KB
az_post_mi_prognosis_arterial_hes.csv	Microsoft Excel C...	08/07/2014 15:27	900 KB
az_post_mi_prognosis_arterial_ons.csv	Microsoft Excel C...	08/07/2014 15:27	15 KB
az_post_mi_prognosis_arterial_onsicd9.csv	Microsoft Excel C...	08/07/2014 15:27	3 KB
az_post_mi_prognosis_barc_gprd.csv	Microsoft Excel C...	08/07/2014 15:27	683 KB
az_post_mi_prognosis_barc_hes.csv	Microsoft Excel C...	08/07/2014 15:27	577 KB
az_post_mi_prognosis_barc_opcs.csv	Microsoft Excel C...	08/07/2014 15:27	26 KB
az_post_mi_prognosis_beta_adrenoceptor_blocking_drugs_gprdprod.csv	Microsoft Excel C...	08/07/2014 15:28	129,002 KB
az_post_mi_prognosis_bleeding_gprd.csv	Microsoft Excel C...	08/07/2014 15:28	751 KB
az_post_mi_prognosis_bleeding_hes.csv	Microsoft Excel C...	08/07/2014 15:28	588 KB

Table 3.2: Summary of study populations

Chapter	Study date range	Entry criteria	Cohort entry date
4) Indications for antithrombotic therapies	1997-2010	All patients with at least one prescription for an antithrombotic therapy	The date of the first antithrombotic therapy prescription
5) bleeding phenotypes; 6) bleeding incidence	1997-2010	Patients with a diagnosis for atrial fibrillation, myocardial infarction, unstable angina or stable angina	The date of the first diagnosis for any of the entry criteria diseases
7) Development and validation of prognostic models	1997-2010	Patients who survived event free for one year following a myocardial infarction	The date one year following their index myocardial infarction record
8) TTR; 9) Predictive value of INR control measures	1997-2010	All patients with at least one oral anticoagulant prescription and at least 5 consecutive INR records	The date of the first INR record in the first valid INR spell

Table 3.3: Summary of data sources for CALIBER risk factor phenotypes

Variable	Phenotype	CPRD diagnosis	CPRD biomarker	CPRD prescription	CPRD procedure	HES admission	HES procedure	MINAP	ONS
Demographics									
Age	CPRD								
Sex	CPRD								
Ethnicity	CPRD								
Index of multiple deprivation	ONS								
Practice region	CPRD								
Behaviours									
Smoking status	CALIBER composite	x		x		x			
Alcohol consumption	CALIBER composite	x	x						
Cardiovascular disease history									
Myocardial infarction	CALIBER composite	x				x	x	x	
Unstable angina	CALIBER composite	x				x		x	
Hypertension	CALIBER composite	x	x	x		x	x		
Ischaemic Stroke	CALIBER composite	x				x			
Unspecified stroke	CALIBER composite	x				x			
Haemorrhagic stroke	CALIBER composite	x				x			
Atrial fibrillation	CALIBER composite	x		x		x			
Heart failure	CPRD; HES	x				x			
Peripheral arterial disease	CALIBER composite	x			x	x	x		
Venous thromboembolism	CPRD; HES	x				x			
Non-cardiovascular disease history									
Diabetes	CALIBER composite	x		x		x			
Renal disease	CPRD; HES	x				x			
Liver disease	CPRD; HES	x				x			
Cancer	CPRD; HES	x				x			
Chronic obstructive pulmonary disease	CPRD; HES	x				x			
Peptic ulcer	CPRD; HES	x				x			
Bleeding diatheses or coagulation disorders	CPRD; HES	x				x			
Chronic anaemia	CPRD; HES	x				x			
Procedures									
Revascularisation	CPRD; OPCS				x		x		
Heart valve replacement	CPRD; OPCS				x		x		
Transfusion	CPRD; OPCS				x		x		
Surgical arrest of bleeding	OPCS						x		
Bleeding cessation	OPCS						x		
Haematoma evacuation	CPRD; OPCS				x		x		
Endoscopy	OPCS						x		
Clinical biomarkers									

Variable	Phenotype	CPRD diagnosis	CPRD biomarker	CPRD prescription	CPRD procedure	HES admission	HES procedure	MINAP	ONS
Systolic blood pressure	CPRD		x						
Diastolic blood pressure	CPRD		x						
Serum creatinine	CPRD		x						
Body mass index	CPRD		x						
Heart rate	CPRD		x						
White blood cell count	CPRD		x						
Total cholesterol	CPRD		x						
HDL cholesterol	CPRD		x						
eGFR	CPRD		x						
Body mass index	CPRD		x						
International normalised ratio	CPRD		x						
Platelet count	CPRD		x						
Hba1c	CPRD		x						
Haemoglobin	CPRD		x						
Prescribed medications									
Antiplatelet drugs	CPRD			x					
Anticoagulants and protamine	CPRD			x					
Statins	CPRD			x					
Thiazides	CPRD			x					
K-sparing diuretics and aldosterone	CPRD			x					
K-sparing diuretics with other diuretics	CPRD			x					
Beta blockers	CPRD			x					
Ace inhibitors	CPRD			x					
Angiotensin receptor blockers	CPRD			x					
Calcium channel blockers	CPRD			x					
Statins	CPRD			x					
Insulin	CPRD			x					
Anti-diabetics	CPRD			x					
Digoxin	CPRD			x					
Amiodarone	CPRD			x					
Non-steroidal anti-inflammatory drugs	CPRD			x					
Antidepressants	CPRD			x					

Table 3.4: Defining all-cause mortality and atherothrombotic endpoints in CALIBER linked electronic health records

Endpoint	CPRD	HES	MINAP	ONS
All-cause mortality*	Date of death			Date of death
Cardiovascular death				I00-I02, I05-I09, I10-I15, I20-I25, I26-I28, I30-I52, I60-I69, I70-I79, I80-I89, I95-I99
Cardiovascular death, stroke or MI		Ischaemic stroke Unspecified stroke Myocardial infarction	STEMI NSTEMI	I00-I02, I05-I09, I10-I15, I20-I25, I26-I28, I30-I52, I60-I69, I70-I79, I80-I89, I95-I99

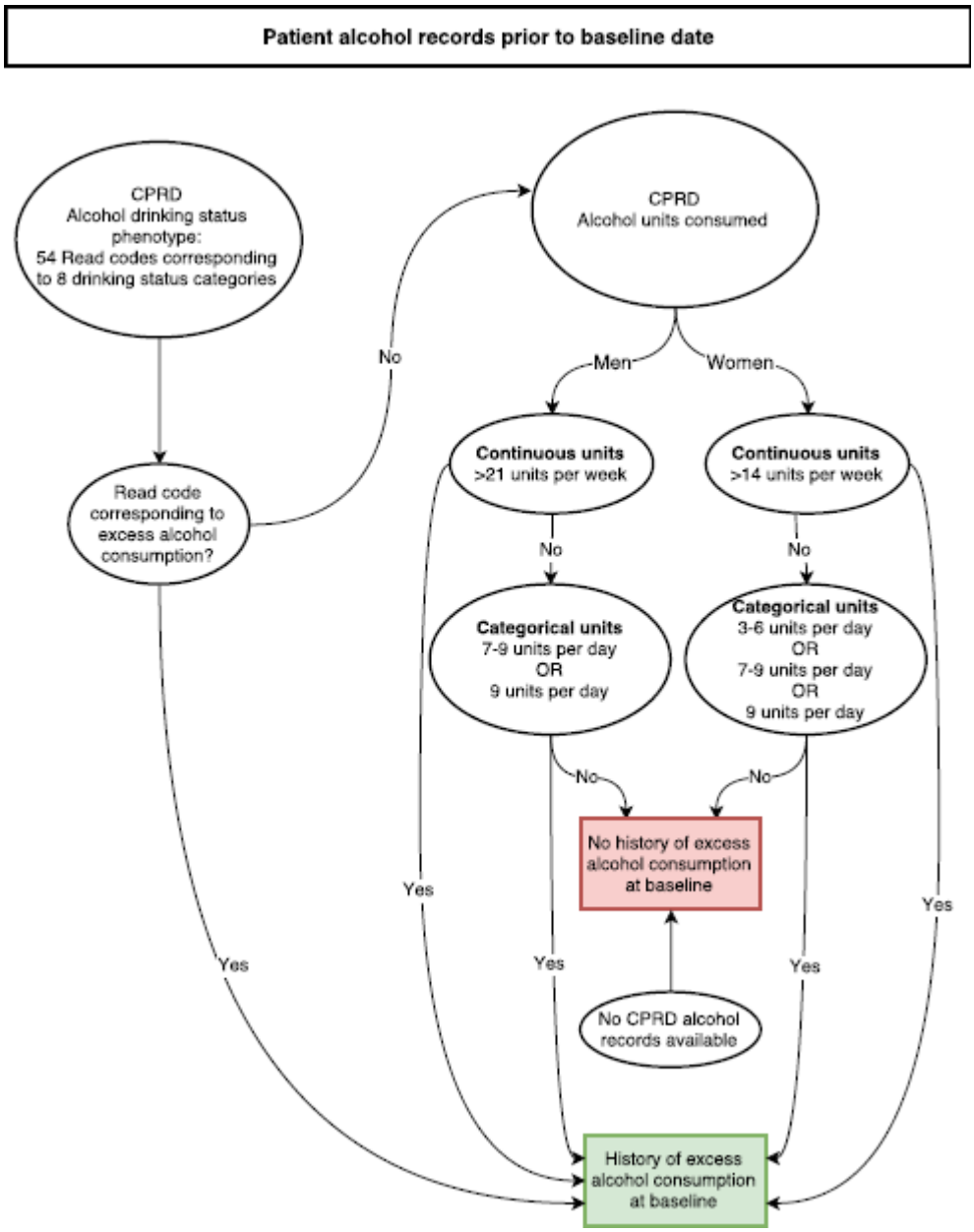
*if date of death appears in both CPRD and ONS with conflicting dates, the ONS date is used

Table 3.5: Defining three bleeding endpoints using codes in linked electronic health records

Endpoint	ICD-10 codes
Fatal or hospitalised bleeding	In hospital admissions (HES) OR death registry(ONS): I60 I61 I62 K250 K252 K254 K256 K260 K262 K264 K266 K270 K272 K274 K276 K280 K282 K284 K286 K290 K625 K920 K921 K922 P261 R040 R041 R048 R049 H356 H431 H450
CALIBER major bleeding	In hospital admissions (HES): I60 I61 I62 In death registry (ONS): I60 I61 I62 K250 K252 K254 K256 K260 K262 K264 K266 K270 K272 K274 K276 K280 K282 K284 K286 K290 K625 K920 K921 K922 P261 R040 R041 R048 R049 H356 H431 H450 **Or** all-cause mortality within 7 days of a hospital admission for any of the above codes **Or** any of the above codes in hospital admissions with primary admission and hospitalisation >14days **Or** any of the above codes in hospital admissions with a transfusion code in primary care (Read codes: 7L14.00, 7L14000, 7L14100, 7L14300, 7L14311, 7L14y00, 7L14z00, TAy0.00, TB1y000, ZV58200) or hospital (OPCS codes X33, X331, X332, X333, X338, X339) within 30 days
Fatal or intracranial bleeding	In hospital admissions (HES): I60 I61 I62 In death registry(ONS): I60 I61 I62 K250 K252 K254 K256 K260 K262 K264 K266 K270 K272 K274 K276 K280 K282 K284 K286 K290 K625 K920 K921 K922 P261 R040 R041 R048 R049 H356 H431 H450 **Or** all-cause mortality within 7 days of a hospital admission for any of the above codes

Note: the prefix of the code can identify the bleeding location: I=Intracranial, K=Gastrointestinal, P=pulmonary, R=respiratory, H= Eye; Fatal or hospitalised bleeding included fatal bleeding or hospitalisation of any duration with bleeding as a primary or secondary reason for admission; CALIBER major bleeding included fatal or intracranial bleeding, bleeding as a primary cause of hospitalisation with length of stay > 14 days or bleedings requiring transfusion. Bleeding requiring transfusion were identified as bleedings with a relevant transfusion record in either primary or secondary care within 30 days following the bleeding; Fatal or intracranial bleeding included fatal bleeding or intracranial bleeding only. Bleedings were considered fatal if there was a bleeding code recorded as the underlying cause of death or if a patient died from any cause within 7 days of hospitalised bleeding;

Figure 3.7: An algorithm to determine patients' history of excess alcohol consumption at baseline



4 Preliminary analysis of the CALIBER linked electronic health records: Antithrombotic drug use and inferring drug indications

Chapter Summary

Background

Antithrombotic therapies are prescribed to prevent and treat a wide range of cardiovascular and non-cardiovascular conditions. In primary care electronic health records, the indication or reasons for prescribing drugs are not explicitly stated. The aim of this descriptive analysis was to use linked electronic health records to investigate the prevalence of antithrombotic prescribing and to infer indications for antithrombotic prescriptions.

Methods

Using CALIBER linked electronic health records I studied a population of patients prescribed at least one antithrombotic therapy between 1997 and 2010. I estimated the prevalence of antithrombotic prescribing over the duration of the study period. To infer indications for patients first antithrombotic prescription, I searched for 20 cardiovascular and non-cardiovascular diseases in primary care and hospital admissions diagnosis records in a [-180, 30] day window around the date of the prescription.

Results

In CALIBER, 277,598 patients were prescribed at least one antithrombotic therapy prescription between 1997 and 2010. The prevalence of antithrombotic prescribing increased over the course of the study period. I identified an indication for 45.7% of patients' first antithrombotic therapy prescription. Overall, the most common indication was coronary heart disease however this differed when stratified by different antithrombotic agents.

Conclusion

Using linked electronic health records, I demonstrated a method to infer indications for antithrombotic therapies. More than half of patients, mostly those prescribed aspirin, did not have one of the investigated indications recorded in a time window around their prescription. Further work should be done to investigate the bleeding risks of these patients.

4.1 Background

As described in chapter 2, various guidelines recommend antithrombotic drugs for atherothrombotic event prevention for a number of cardiovascular and non-cardiovascular diseases **(Table 2.1).**

To date, studies of antithrombotic therapies in usual care settings have focused on single disease populations such as atrial fibrillation[176] or acute coronary syndromes[177] to analyse the uptake of the treatment within these particular patient populations or have focused on a subset of antithrombotic therapies such as oral anticoagulants[178]. No previous study has looked at the longitudinal prescribing trends of all types of antithrombotic therapies in an unselected national cohort.

Clinical Practice Research Datalink (CPRD) prescription records do not explicitly state the indication for which they were prescribed, however it may be possible to infer the indications from diagnoses made around the time of the prescription. Attempts to determine indications in national electronic health records for other classes of drugs such as oral corticosteroids,[179] antibacterial,[180] antidepressant[181,182] and antipsychotic drugs[183] have previously been made, to inform monitoring of drug use or over use and to identify the distribution of diseases amongst individuals prescribed the drugs.

Antithrombotic therapies are associated with bleeding side effects which range from minor nosebleeds to serious gastrointestinal or intracranial haemorrhage. Therefore it is of interest which patients are prescribed these drugs and for what reason.

The aims of this descriptive analysis were:

- to investigate the prevalence of the antithrombotic therapy prescribing in England
- to investigate the heterogeneity and distribution of cardiovascular and non-cardiovascular diseases in patients prescribed antithrombotic therapies.

4.2 Methods

4.2.1 Study population

I used the CALIBER (CArdiovascular research using LInked Bespoke and Electronic health Records) research platform, consisting of EHR linkages between primary care data (Clinical Practice Research Datalink (CPRD)), secondary care data (Hospital Episode Statistics), disease registry data (Myocardial Ischaemia National Audit Project) and cause-specific mortality (Office for National Statistics) in England.[8] The 4% sample of England's population in CPRD available for linkage is unselected, representative in terms of age, sex and overall mortality.[172-174]

The study population for this study included patients with at least one antithrombotic therapy prescription recorded in CPRD between 1997 and 2010. Patients were eligible for inclusion in the cohort if their first antithrombotic therapy prescription was recorded at least 180 days following registration at their general practice.

Antithrombotic therapies include any of the substances in the BNF chapters: 2.8.1 parenteral anticoagulants, 2.8.2: oral anticoagulants or 2.9: antiplatelet drugs.

4.2.2 Antithrombotic prescribing prevalence

The prevalence of antithrombotic prescribing in CALIBER was examined graphically. The most prolific antithrombotic drugs aspirin, clopidogrel and warfarin were analysed individually. The remaining drugs were examined as groups: antiplatelet agents (dipyridamole, prasugrel, ticlopidine, tirofiban and abciximab), oral anticoagulants (acenocoumarol and phenindione), direct oral anticoagulants (dabigatran and rivaroxaban) and parenteral anticoagulants (heparin and low molecular weight heparin).

Prevalence was calculated at monthly intervals from January 1998 to December 2009. The denominator was the total number of patients registered with their GP practices in CALIBER that month. The numerator was the monthly number of the patients with prescription records in primary care within each of the previously described antithrombotic therapy drug groups.

4.2.3 Determining indications of antithrombotic therapy prescriptions

Indications for cardiovascular and non-cardiovascular conditions were investigated and were based on those suggested for the various antithrombotic agents recorded in the BNF. The cardiovascular indications were atrial fibrillation, myocardial infarction, unstable angina, stable angina, heart valve disease or replacement, heart failure, unspecified coronary disease, transient ischemic attack, ischaemic or unspecified stroke, venous thromboembolism, peripheral arterial disease, cardiomyopathy, and a recorded international normalised ratio<2.

Non-cardiovascular indications included hip or knee replacement surgery, cancer, dialysis, systemic lupus erythematosus and rheumatoid arthritis.

Phenotyping algorithms and codelists for the investigated indications were procured from the CALIBER portal [https://www.caliberresearch.org/portal] and have been validated in previous published studies.[38-45,184-187]

The time window for indications around antithrombotic therapy prescriptions was defined as up to 180 days prior to the prescription to 30 days following the indication (to allow for potential time lag in recording), illustrated in **Figure 4.1.**

4.3 Results

4.3.1 Prevalence of antithrombotic therapy prescriptions

In CALIBER there were 277,598 patients with at least one antithrombotic therapy prescription between 1997 and 2010, with their first prescription recorded at least 180 days following their general practice registration. The flow of patients into the study population is displayed in **Figure 4.2**. The mean (SD) age of these patients was 66.3 (14.8) and 138,573 (49.9%) were women **(Table 4.1).** The study cohort had a total of 8.6 million antithrombotic therapy prescription records between 1997 and 2010.

Figure 4.3 (top panel) shows the changing prevalence of antithrombotic therapy prescriptions in CALIBER over time. Between 1997 and 2010 the prevalence of aspirin prescribing rose from about 1.5% to 4.5%, warfarin prescribing rose from 0.3% to 0.8% and clopidogrel which was approved for use in the UK in late 1998, rose to 0.5% by the start of 2010.

A log transformation of the percentages shown in **Figure 4.3 (bottom panel)** enables us to visualise the changing trends over time particularly for the less commonly used drugs. We can see the introduction and approval of clopidogrel in 1998 and its overtaking of other antiplatelet agents as the platelet inhibitor of choice. Whilst warfarin prescriptions have risen, the prescribing of alternative oral anticoagulants has remained fairly constant. Finally, we can see the introduction of the novel oral anticoagulants, dabigatran and rivaroxaban which were first approved for prevention of thromboembolic disease following hip or knee replacement in 2008 and subsequently approved for treating thromboembolic disease and preventing stroke in patients with non-valvular atrial fibrillation.

4.3.2 Indication for first antithrombotic therapy prescription

Index antithrombotic therapies in the study population comprised of 54 unique drug combinations. The distribution of the index antithrombotic therapies is shown in **Figure 4.4**.

266901 (96.1%) patients were prescribed a single antithrombotic drug: 222962 (80.31%) patients were prescribed aspirin, 27436 (9.88%) patients were prescribed warfarin, 7772 (2.80%) patients were prescribed clopidogrel, 6196 (2.23%) patients were prescribed a parenteral anticoagulant, 2413 (0.87%) patients were prescribed an antiplatelet other than aspirin or clopidogrel, 93 (0.03%) patients were prescribed a non-warfarin oral anticoagulant and 29 (0.01%) patients were prescribed a direct oral anticoagulant.

The most common index antithrombotic therapy combination was two antiplatelets for 9631 (3.47%) patients including 7136 (2.6%) prescribed aspirin and clopidogrel dual therapy. Furthermore, 683 (0.25%) patients were prescribed warfarin and one antiplatelet agent, 233 (0.08%) patients were prescribed warfarin and one parenteral anticoagulant, 31 (0.01%) patients were prescribed warfarin and two antiplatelet agents and 119 (0.04%) patients were prescribed other combinations of antithrombotic therapies.

The indications identified for the antithrombotic therapy groups are summarised in **Table 4.2**. Overall, we found an indication for 126923 (45.7%) index antithrombotic therapy prescriptions and 49635 (17.9%) patients had more than one indication recorded in the time window of 180 days prior to up to 30 days following the date of their index antithrombotic therapy.

Across all antithrombotic therapy agents the most common indications were stable angina, unspecified coronary heart disease, atrial fibrillation and myocardial infarction. Less than 50% of patients prescribed aspirin had a record of the investigated conditions in primary care or hospital admissions records and stable angina, unspecified coronary heart disease and myocardial infarction were the most common indications. The majority of patients prescribed clopidogrel (82.1%) were allocated an indication, with coronary disease (myocardial infarction, unspecified coronary disease, unstable angina, stable angina) being the most common. Patients prescribed other antiplatelet agents mainly had records for transient ischaemic attack, and ischaemic or unspecified stroke.

89.1% of patients prescribed warfarin had one of the investigated conditions recorded within the time window. Venous thromboembolism, atrial fibrillation, INR<2, heart failure, heart valve disease or replacement and cancer were among the most common indications. Indications were found for 74% of patients treated with other oral anticoagulants with venous thromboembolism, atrial fibrillation and INR<2 the most common. Only 29 patients were prescribed direct oral anticoagulants due to their approval for use for venous thromboembolism prevention occurring near the end of the study period. Most of these patients had a record for a hip or knee replacement procedure. 59.3% of patients prescribed

parenteral anticoagulants were allocated an indication. The most common indications for this group were venous thromboembolism, cancer, INR<2 and hip or knee replacement surgery.

4.3.3 Characteristics of patients with and without an associated indication for their index antithrombotic therapy prescription

Table 4.3 displays the characteristics of patients with and without an associated indication recorded between 180 days before and 30 days following their index ATT prescription. Patients without an indication had a lower mean age, 61.2 years (SD: 15.6), than patients with an indication, 64.6 years (SD: 14.1). The proportion of women was higher in the no indication group (52.4% vs. 46.9% in the indication group). There were also fewer smokers and ex-smokers amongst the patients with no indication compared to those with an indication. Type 2 diabetes was more common amongst patients with no indication (18.4% vs. 8.7% in the indication group). Patients did not appear to differ in terms of common cardiovascular clinical biomarkers such as body mass index, systolic blood pressure and cholesterol.

4.4 Discussion

Using linked electronic health records, we studied a cohort of 277,598 patients prescribed antithrombotic therapies. We identified a rising prevalence of antithrombotic prescribing and inferred cardiovascular and non-cardiovascular indications for 126923 (45.7%)patient's first prescription in primary care.

Antithrombotic therapy prescribing prevalence

With the antithrombotic prescription records in captured in primary care records we identified rising prevalence amongst most of the antithrombotic subclasses, including aspirin, clopidogrel, other antiplatelet agents warfarin and parenteral anticoagulants. We also observed the rapid uptake of clopidogrel following its approval for use in 1998 for treating myocardial infarction and the introduction of direct oral anticoagulants in 2009 for venous thromboembolism prevention. This demonstrates that primary care prescription records may provide evidence of trends and changes in treatment practices over time. When linked to clinical outcome primary care and hospital admissions records, such data can be useful to evaluate the effectiveness of new treatments in a real-world setting.

Index antithrombotic therapy indications

Within each of the groups of antithrombotic therapies the most prevalent conditions identified within 180 days prior to and 30 days following index prescriptions were generally as expected, according to the guidelines (**Table 2.1**). For example, antiplatelet therapy is recommended for atherothrombotic prevention following coronary disease and almost half of patients (43.4%)

with a clopidogrel prescription had a myocardial infarction record in the time window. Atrial fibrillation and venous thromboembolism prevention are among the leading indications for warfarin and of the patients prescribed warfarin, 44.1% had a venous thromboembolism record and 31.8% had an atrial fibrillation record. This data suggests that antithrombotic prescribing in usual clinical care largely follows treatment guidelines.

Patients with 'no recorded indication'

150675 (54.3%)patients had no record of the investigated indications in a time window around their index antithrombotic therapy. This was mostly attributable to patients with aspirin prescriptions, of which more 60% did not have a record of the investigated indications. However aspirin is associated with the lowest risk of bleeding and the broadest range of possible indications, for example general pain relief is a plausible reason that may not be recorded.

Of the agents associated with higher bleeding risk and not recommended for general usage, 1 in 5 patients prescribed clopidogrel and 1 in 10 patients prescribed warfarin did not have a recorded indication.

The higher prevalence of diabetes in the 'no indication' group could indicate a group of high-risk patient's prescribed antithrombotic therapy, usually aspirin, for primary cardiovascular disease prevention. However some cardiovascular risk factors were less prevalent in the 'no indication' group, such as age, men and smokers. This may coincide with risk profiles of patients with lower bleeding risk, a consideration clinicians and patients take when prescribing antithrombotic therapies.

Strengths and limitations

To my knowledge, the first study of its kind for antithrombotic drugs. CALIBER allows a more comprehensive search of indications for prescriptions than if primary care data was used alone, with linkages to hospital admissions and myocardial infarction disease registry data. It has previously been demonstrated that no single EHR data source has complete coverage of diagnosis records.[36]

There are further indications for antithrombotic therapies that may be investigated such as rare and/or genetic blood disorders that increase risk of thrombosis. Such conditions are not well coded in primary care or hospital admissions data at this time but may be explored in future with linkages to more specialised data e.g. disease registries. In this study, despite having linked records across primary and secondary care, only primary care prescriptions were

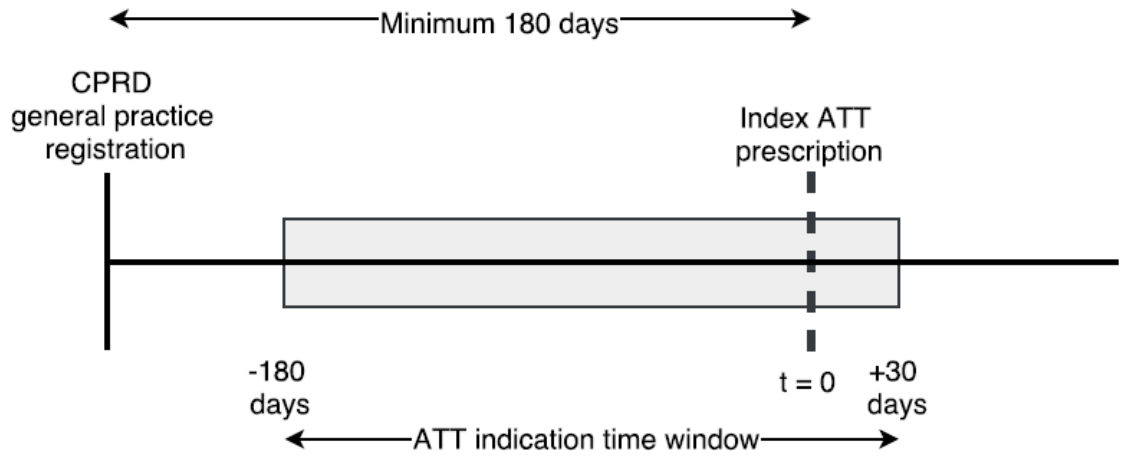
analysed. Aspirin is a common over the counter drug and whilst it was by far the most prescribed drug of the antithrombotic its use is likely under-represented in this study. Furthermore, in assessing aspirin prescriptions I did not differentiate between high-dose and low-dose aspirin, of which the former is not used as an antithrombotic therapy. High dose aspirin is indicated for general aches and has not been shown to be any more effective as an antithrombotic therapy than low-dose aspirin.

Research implications

Further potential areas of research include: 1) investigating duration of antithrombotic treatment within disease groups, 2) assessing the prevalence of contraindications in patients prescribed antithrombotic therapies, 3) estimating bleeding risks in patients prescribed antithrombotic therapy and the balance between benefits and harms in an unselected real-world setting and 4) estimating interactions between antithrombotic therapies and blood pressure control in hypertensive patients.

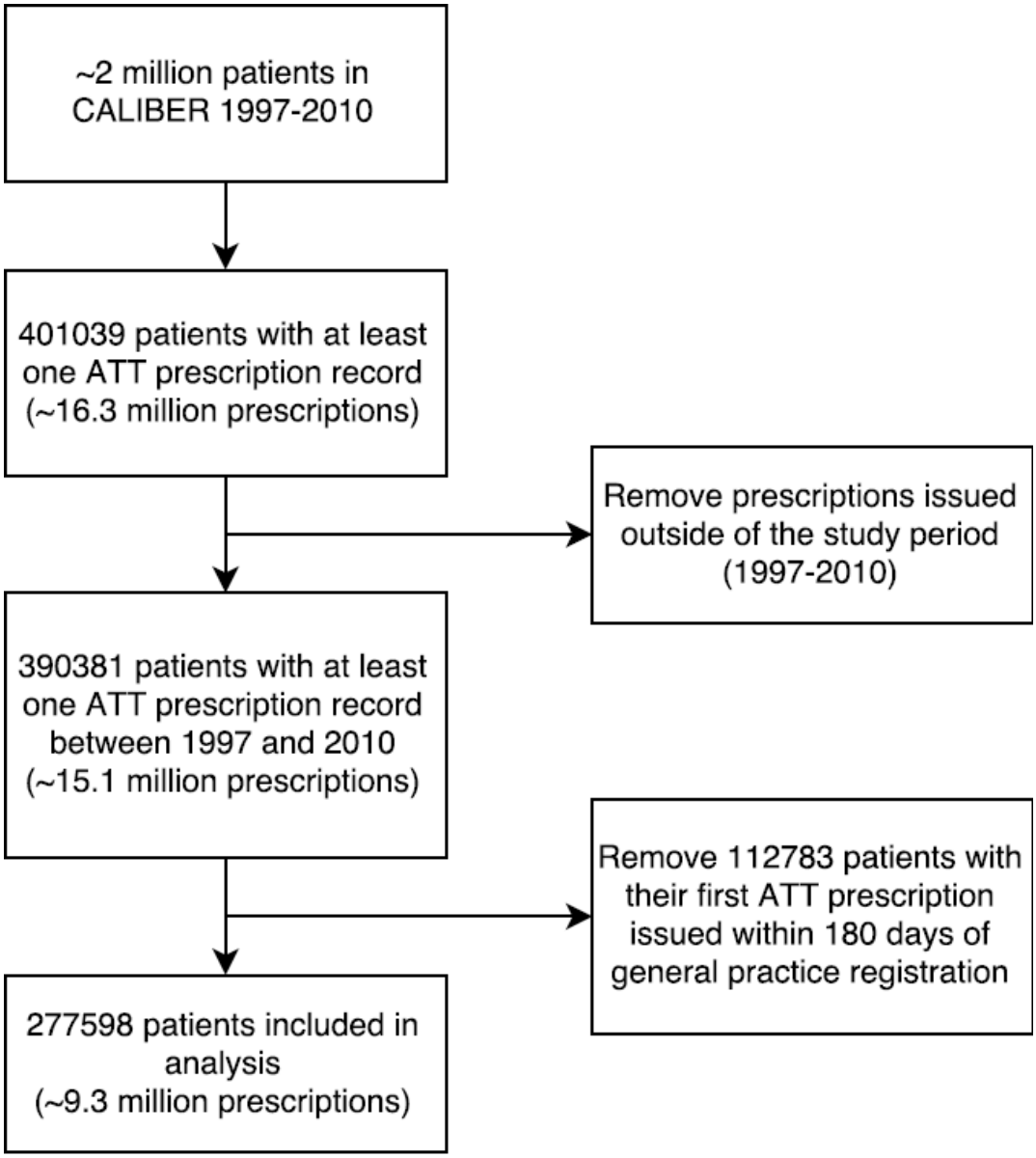
4.5 Tables and Figures

Figure 4.1: Illustrating the time window used to allocate indications for antithrombotic therapy prescriptions



ATT= antithrombotic therapy; CPRD= Clinical Practice Research Datalink

Figure 4.2: Study population flowchart



ATT= antithrombotic therapy

Table 4.1: Demographics of the study population

	N=277,598
Age (years)	66.3 (14.8)
Women	138573 (49.9)
Number of antithrombotic therapy prescriptions in follow up, median (IQR)	19 (4, 45)
Follow up time (years), median (IQR)	4.1 (1.8, 7.1)

IQR= interquartile range

Figure 4.3: Prevalence of antithrombotic therapy prescribing captured in primary care electronic health records 1998-2009

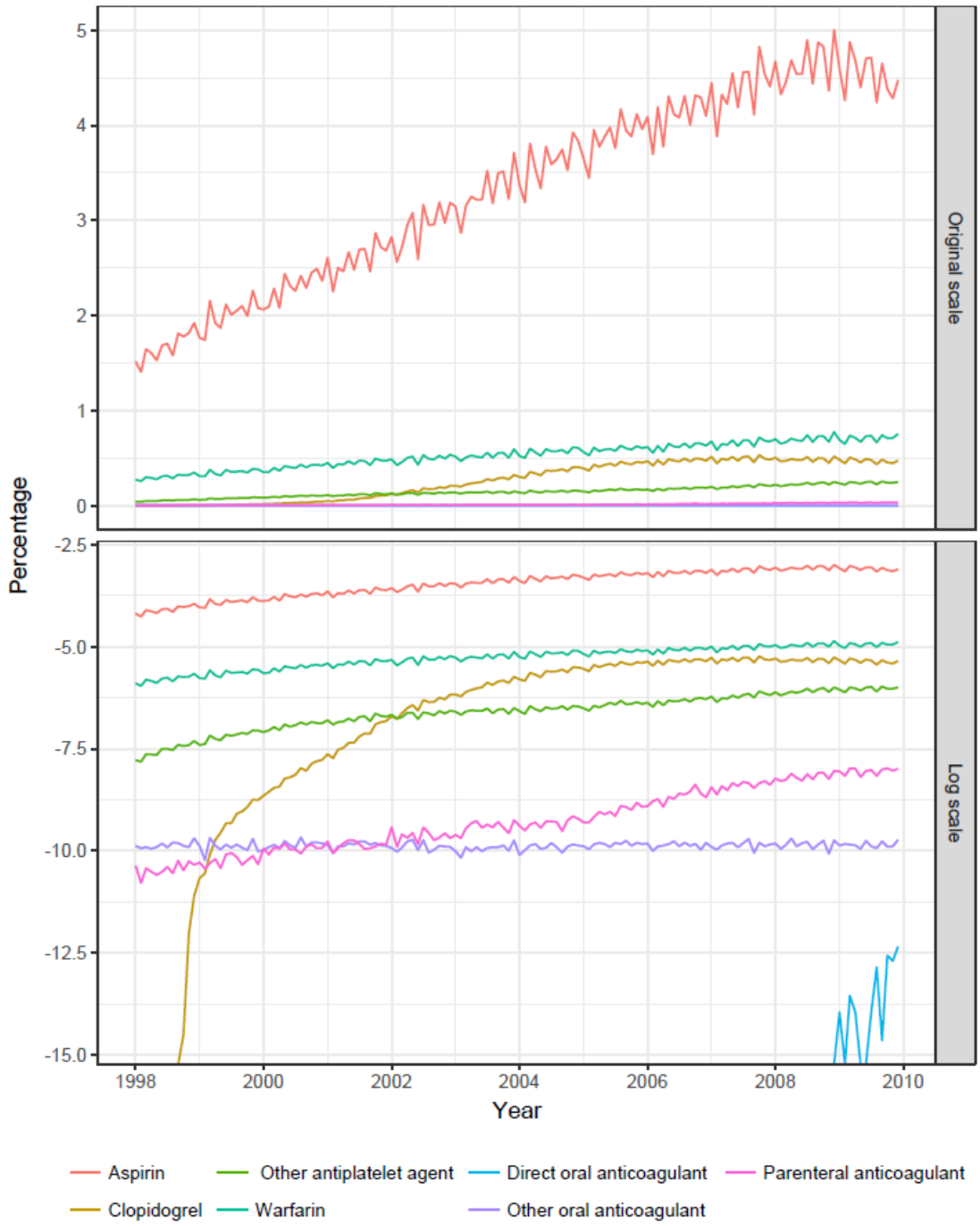
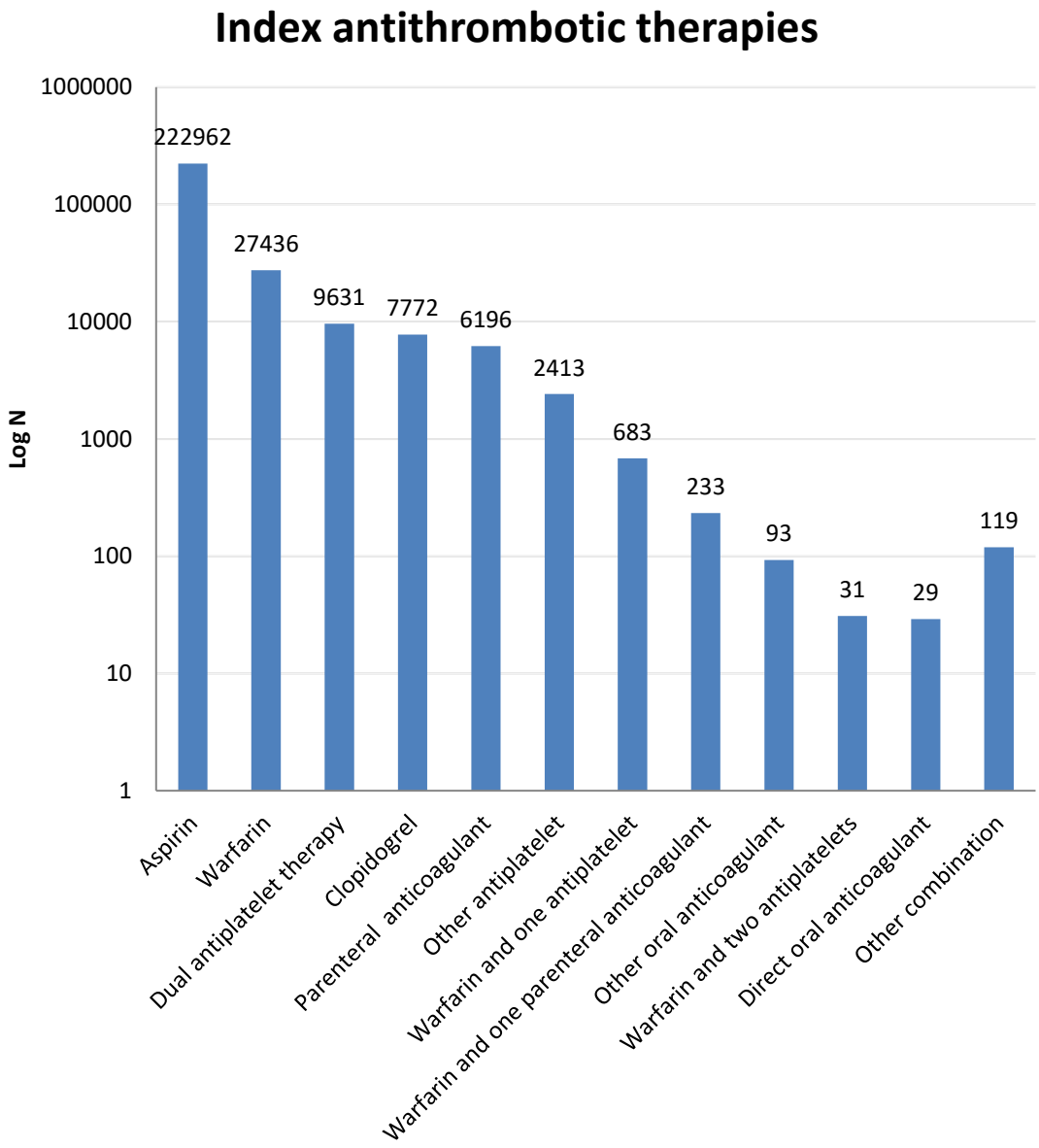


Figure 4.4: The distribution of index antithrombotic therapy combinations



Note: log scale (actual n displayed in chart above the bars)

Table 4.2: Distribution of the indications for 277,598 patient's index antithrombotic prescription

		Index antithrombotic therapy including any of:						
Indication[1]	**Any ATT**	**Aspirin**	**Clopidogrel**	**Other antiplatelet**	**Warfarin**	**Other oral anticoagulant**	**Direct oral anticoagulant**	**Parenteral anticoagulant**
N[2]	**277598**	**233316**	**15027**	**4930**	**28384**	**96**	**29**	**6540**
Myocardial Infarction	18387 (6.6)	16633 (7.1)	6518 (43.4)	87 (1.8)	805 (2.8)	1 (1)	0 (0)	13 (0.2)
Unspecified coronary heart disease	25211 (9.1)	22173 (9.5)	6339 (42.2)	214 (4.3)	1308 (4.6)	0 (0)	1 (3.4)	37 (0.6)
Unstable angina	7524 (2.7)	6650 (2.9)	2628 (17.5)	42 (0.9)	272 (1)	0 (0)	0 (0)	10 (0.2)
Stable angina	32596 (11.7)	29454 (12.6)	4912 (32.7)	272 (5.5)	1142 (4)	3 (3.1)	0 (0)	46 (0.7)
Atrial Fibrillation	24318 (8.8)	14772 (6.3)	1037 (6.9)	212 (4.3)	9034 (31.8)	17 (17.7)	0 (0)	110 (1.7)
Ischaemic or unspecified stroke	12122 (4.4)	10245 (4.4)	681 (4.5)	1723 (34.9)	922 (3.2)	5 (5.2)	0 (0)	27 (0.4)
Transient ischaemic attack	11757 (4.2)	10299 (4.4)	627 (4.2)	1192 (24.2)	375 (1.3)	0 (0)	0 (0)	14 (0.2)
Haemorrhagic stroke	688 (0.2)	546 (0.2)	48 (0.3)	69 (1.4)	86 (0.3)	1 (1)	0 (0)	8 (0.1)
Heart failure	10814 (3.9)	7867 (3.4)	1109 (7.4)	94 (1.9)	2604 (9.2)	1 (1)	0 (0)	45 (0.7)
Heart valve disease or replacement	5146 (1.9)	3334 (1.4)	497 (3.3)	61 (1.2)	1632 (5.7)	3 (3.1)	0 (0)	40 (0.6)
Peripheral arterial disease	7453 (2.7)	6231 (2.7)	652 (4.3)	166 (3.4)	612 (2.2)	2 (2.1)	0 (0)	62 (0.9)
Cardiomyopathy	97 (0)	59 (0)	15 (0.1)	0 (0)	31 (0.1)	0 (0)	0 (0)	2 (0)
Venous thromboembolism	15041 (5.4)	749 (0.3)	62 (0.4)	19 (0.4)	12505 (44.1)	39 (40.6)	0 (0)	1965 (30)
INR<2	10018 (3.6)	2208 (0.9)	281 (1.9)	98 (2)	7126 (25.1)	15 (15.6)	0 (0)	750 (11.5)
Hip or knee replacement	5356 (1.9)	3248 (1.4)	122 (0.8)	24 (0.5)	1422 (5)	1 (1)	24 (82.8)	589 (9)
Cancer	10123 (3.6)	5330 (2.3)	422 (2.8)	152 (3.1)	2729 (9.6)	6 (6.2)	0 (0)	1816 (27.8)
Dialysis	517 (0.2)	394 (0.2)	38 (0.3)	8 (0.2)	92 (0.3)	0 (0)	0 (0)	10 (0.2)
Systemic lupus erythematosus	215 (0.1)	148 (0.1)	10 (0.1)	6 (0.1)	44 (0.2)	0 (0)	0 (0)	12 (0.2)
Rheumatoid arthritis	1657 (0.6)	1186 (0.5)	186 (1.2)	46 (0.9)	309 (1.1)	0 (0)	0 (0)	60 (0.9)
None of the above	150675 (54.3)	141102 (60.5)	3129 (20.8)	1670 (33.9)	3095 (10.9)	25 (26)	5 (17.2)	2665 (40.7)

[1] 49635 (17.9%) patients had more than one indication recorded in the time window of [-180, 30] days around their index antithrombotic prescription(s)

[2] 55026 (19.7%) patients were prescribed more than one antithrombotic drug on their index date;

Table 4.3: Characteristics of patients with and without an allocated indication for their index antithrombotic therapy prescription

	Indication	No indication
N	**126923**	**150675**
Age; Mean (SD)	64.6 (14.1)	61.2 (15.6)
Women	59545 (46.9)	79028 (52.4)
Highest IMD Quintile	18734 (14.8)	22668 (15)
Missing %	0.3	0.3
Non-Smoker	55835 (44)	71230 (47.3)
Ex-Smoker	27931 (22)	31551 (20.9)
Smoker	21978 (17.3)	23872 (15.8)
Missing %	16.7	15.9
Type 1 Diabetes	794 (0.6)	2210 (1.5)
Type 2 Diabetes	11022 (8.7)	27720 (18.4)
Unspecified diabetes	1190 (0.9)	1813 (1.2)
Body mass index; Mean (SD)	27.7 (5.72)	28.5 (5.90)
Missing %	68	57.2
Systolic blood pressure; Mean (SD)	142 (21.4)	143 (20.4)
Missing %	25	19.7
Low density lipoproteins; Mean (SD)	3.19 (1.11)	3.24 (1.10)
Missing %	78.6	67.2
High density lipoproteins; Mean (SD)	1.39 (0.438)	1.37 (0.419)
Missing %	76.3	63.9
Total cholesterol; Mean (SD)	5.36 (1.24)	5.43 (1.23)
Missing %	63.2	50.9
Creatinine; Mean (SD)	95.6 (29.9)	90.9 (24.8)
Missing %	55.3	47.9
Haemoglobin; Mean (SD)	13.6 (1.76)	14.0 (1.54)
Missing %	61.2	61.2

SD= standard deviation; IMD= Index of multiple deprivation;

5 Developing bleeding phenotypes in CALIBER

Chapter Summary

Background

The main safety concern associated with the use of antithrombotic therapies is increased risk of bleeding. Bleeding harms can vary in presentation and severity. With the potential for such devastating treatment side effects, it is important to consider both the benefits and harms when studying antithrombotic therapies. To do this, standardised bleeding definitions for use in electronic health records are required. The objectives of this analysis were to determine whether bleeding endpoints used in clinical trials could be replicated using linked electronic health records and whether bleeding severity can be defined using supporting records such as transfusion.

Methods

First a review of bleeding diagnosis terms in Read and ICD-10 was performed, and bleeding code lists were developed. In the study population of 128815 patients with atrial fibrillation or coronary disease in CALIBER 1997-2010, the characteristics of the bleeding events were described. Potential markers of severity were assessed through their crude association with short term all-cause mortality using Kaplan-Meier plots.

Results

A bleeding phenotype algorithm was developed that identifies fatal bleeding, and bleeding with and without markers of severity in primary care and hospital admissions records. The markers of severity identified were bleeding anatomical site, transfusion recorded in hospital records, bleeding recorded as the primary reason for hospitalisation and duration of hospitalisation and the number of bleeding diagnosis codes recorded in one day.

Conclusion

Using linked electronic health records I constructed a scalable bleeding phenotype which identifies bleeding of differing severity, which can be used in studies of bleeding incidence and outcomes. No single source of data was sufficient to fully capture bleeding events in the study population.

5.1 Background

5.1.1 Antithrombotic therapies and bleeding risks

In common cardiac disorders, including atrial fibrillation (AF), acute coronary syndromes and stable coronary disease, antithrombotic drugs have been shown in randomised trials to be effective in reducing the risk of subsequent atherothrombotic events.[188-190] There are increasing numbers of people with such common heart diseases on one, two or three antithrombotic agents because of better implementation of longstanding evidence trial evidence (e.g. aspirin for the secondary prevention of non-myocardial infarction (MI) coronary disease), the introduction of new antithrombotic drugs (e.g. novel oral anticoagulant's [NOAC's], ticagrelor) and the recognition of the role of prolonging dual antiplatelet therapy for MI (e.g. ticagrelor).[17] Indeed prolonged (lifelong) treatment is now recommended not only for anticoagulation in AF but also in dual antiplatelet therapy following acute MI.[17,191]

The main safety concern associated with the use of antithrombotic therapies is increased risk of bleeding. Bleeding harms can vary in their manifestation, ranging from minor nosebleeds to major or fatal intracranial haemorrhage. With the potential for such devastating treatment side effects, it is important for clinicians and patients to consider both the benefits and harms.

5.1.2 Definitions for bleeding events in electronic health records

Electronic health records (EHRs) have increasingly been used for medical research; however there has been a lack of standardised definitions to measure bleeding occurrence and severity using national population based primary care and hospital admission records. This limits the comparability and interpretation of EHR studies investigating harms associated with the use of antithrombotic therapies.

Previous EHR studies of bleeding endpoints have been confined to hospitalised patients,[75-77] or one anatomical site (e.g. upper gastrointestinal bleeding[70,71,73]), or based on commercially insured or administrative claims data[72,74] (**Table 5.1**).

5.1.3 Disease phenotypes in CALIBER linked electronic health records

The CALIBER platform provides linkages between primary care, hospital admissions, myocardial infarction registry, and cause of death sources.

Within CALIBER high resolution disease phenotypes are developed using the linked data through collaborations between clinicians, epidemiologists and statisticians. The disease phenotypes can be granular enough to include categories, or sub-phenotypes which classify for example, disease subtype (e.g. ST-elevation or non-ST elevation myocardial infarction),

severity of the condition (e.g. mild, moderate or severe renal disease) or disease status (e.g. history of, possible diagnosis, confirmed diagnosis).

An example of the approach to phenotyping in CALIBER has been demonstrated for atrial fibrillation.[38] Atrial fibrillation cases were defined in primary care and hospitalisations. Additional cases were inferred for those without a diagnosis code in primary care or hospital admissions, but with prescriptions for drugs commonly used by atrial fibrillation patients (warfarin and digoxin) and no diagnosis for any of the other indications for the prescriptions.

Within CALIBER there is information on bleeding diagnosis, and other relevant data that might be used to assess severity across primary and hospital care (e.g. site, fatality, length of hospital stay, haemoglobin, transfusion, endoscopy, surgical interventions).

Using the CALIBER data I sought to address the following questions:

First to what extent can population-based electronic health records (EHR) replicate the definitions of bleeding used in trials? **Secondly,** can bleeding severity be defined using information on bleeding diagnosis, site, fatality, length of stay, haemoglobin, transfusion, endoscopy, surgical interventions using linked primary care and hospital admissions EHR.

5.2 Methods

5.2.1 Reviewing code lists for bleeding and related procedures

Bleeding diagnosis terms in Read and ICD-10 were reviewed for defining bleeding in primary care and hospital admissions and determine which are relevant to antithrombotic use. I identified 96 ICD-10 codes **(Supplementary appendix 11.2.1**) and 215 Read codes **(Supplementary appendix 11.2.2**) for bleeding events which I categorised into a total of 18 anatomical sites:

- Intracranial (Intracerebral; Subarachnoid; Extradural; Subdural; Unspecified)
- Gastrointestinal (Upper; Lower; Unspecified)
- Respiratory (Upper; Lower; Unspecified)
- Ruptured aortic aneurysm or haemopericardium
- Genitourinary
- Bleeding disorders
- Ocular
- Ear
- Renal
- Unspecified

MINAP has one field related to bleeding complications which consists of 6 categories pertaining to bleeding location and haemoglobin drop. **(Supplementary appendix 11.2.3)** Furthermore I reviewed primary care Read codes and OPCS codes for hospital procedures related to bleeding that may provide supporting evidence for bleeding events:

- Transfusion **(Read codes: Supplementary appendix 11.2.5; OPCS codes: Supplementary appendix 11.2.4)**
- Haematoma evacuation or aspiration **(OPCS codes: Supplementary appendix 11.2.6)**
- Surgical arrest for bleeding **(OPCS codes: Supplementary appendix 11.2.7)**
- Endoscopy **(OPCS code: Supplementary appendix 11.2.8)**

5.2.2 Study Population

The study population consisted of patients with either coronary disease or atrial fibrillation, i.e. those who were potential candidates for antithrombotic therapy, in CALIBER during 1997-2010. To define this population I used pre-existing validated disease phenotypes in CALIBER described in **Chapter 3**. Patients were eligible if they were aged 18 years and above and entered the cohort at their first diagnosis of atrial fibrillation, myocardial infarction, unstable angina or stable angina, and were followed up until death or transfer out of their primary care practice (i.e. loss to follow-up).

5.2.2.1 Patient characteristics

I examined the characteristics, (demographics, behaviours, medical history) of the population in stratified by their disease diagnosis at cohort entry. Using prescribing data I summarised duration (median and interquartile range days) of antithrombotic use between cohort entry and first bleeding event. Prescriptions were assumed to last a maximum of 90 days as this is the longest allowed duration of prescriptions in the UK. Antithrombotic therapies were grouped as aspirin monotherapy, adenosine diphosphate (ADP) receptor monotherapy, dual antiplatelet therapy, vitamin K antagonist (VKA) monotherapy, VKA and one antiplatelet (aspirin or ADP receptor inhibitor) and VKA and two antiplatelets (aspirin and ADP receptor inhibitor).

5.2.3 Developing the phenotype

Based on a combination of the exploratory analyses of bleeding codes, their characteristics, severity markers and consensus amongst the study team I iteratively developed an algorithm to define bleeding in linked electronic health records. Bleeding was grouped as with or without markers of severity within primary care or hospital admissions or as inferred.

I assessed how well bleeding events are captured amongst the data sources using a Venn diagram. I allowed a minimum of 30 days between records from different data sources to consider events to be unique, to allow for time lag in events being recorded.

5.2.3.1 Bleeding characteristics

I performed preliminary analysis of the bleeding records in the study population to determine which data are viable for use in the final algorithm.

I examined the characteristics of bleeding records stratified by anatomical bleeding site. For hospitalised bleeding I calculated the length of hospitalisation and the proportion of records with primary diagnosis position. I investigated the presence of procedures records, transfusion (recorded in primary care and hospital admissions), bleeding surgical arrest, haematoma evacuation and endoscopy, at different time intervals (on the same day, within 7 days and within 30 days) of bleeding records in primary care and hospital admissions.

I examined the availability of haemoglobin values captured in primary care at various time intervals (acute: same day, +/-1 day, +/7 days, and chronic: in the year prior and in the year post) to determine whether it is possible identify haemoglobin levels dropping due to a bleeding event. Haemoglobin drop related to a bleeding event was then calculated as the peak haemoglobin value within 365 to 7 days prior to bleeding minus the lowest haemoglobin value recorded within 7 days of bleeding for patients with a minimum of 2 haemoglobin values recorded in primary care.

5.2.3.2 Indicators of bleeding severity

To assess the suitability of indicators for bleeding severity I used Kaplan Meier plots to examine short term (30 day) all-cause and bleeding specific mortality following bleeding with and without the indicators. Guided by current clinical bleeding definitions that are used to assess bleeding at the point of care **(Table 5.2)** and availability of data within CALIBER the following severity indicators were considered:

- Anatomical site
- Presence of a transfusion record
- Presence of surgical interventions
- Haemoglobin drop
- Presence of multiple bleeding records on a single date.
- For hospitalised bleeding records I considered primary versus secondary diagnosis position and length of hospitalisation.

5.2.3.3 Inferring bleeding cases

Using the information gained through analysing the characteristics of bleeding events, I attempted to capture potential bleeding cases in patients with no bleeding code in primary care or hospital admissions through the following pathways:

- Surgical procedures (surgical arrest, haematoma evacuation) recorded in primary care and hospital admissions
- Transfusion with
    - Iron deficiency anaemia record in primary care or hospital admissions within 30 days
    - Low haemoglobin (<10g/dL) and endoscopy within 30 days and without cancer, liver disease, renal disease diagnosis 1 year prior
- Low haemoglobin (<10g/dL) with
    - Iron deficiency anaemia record in primary care or hospital admissions and endoscopy within 30 days and without cancer, liver disease, renal disease diagnosis 1 year prior

5.3 Results

5.3.1 Study population

The study population consisted of 128815 patients in 224 general practices with a new diagnosis of at least one of diagnosis of atrial fibrillation, acute myocardial infarction, unstable angina or stable angina from 1997-2010. They were followed up for a total of 559,161 person-years, a median of 3.7 years (IQR: 5.4 1.5, 6.9), and were 48.5% women and had a mean age of 71.5 years old at cohort entry.

Patient characteristics stratified by cardiac disease are shown in **Table 5.3**. Atrial fibrillation patients were older than the coronary disease patients and were majority women, whereas the coronary disease patients were mostly men. The atrial fibrillation patients also had higher prevalence of history of stroke, renal disease, cancer and chronic anaemia. The majority of patients in all four disease groups were prescribed at least 1 antithrombotic therapy during follow up.

Around 70% of patients with coronary diseases (acute myocardial infarction, stable or unstable angina) were prescribed aspirin in follow up for median durations near to 2 years suggesting it was generally prescribed for long term treatment and/or secondary prevention. Dual antiplatelet therapy was most common in myocardial infarction patients (34.6%), and was prescribed for a median duration of just under a year (349 days) which corroborates with

current guidelines for the treatment of acute myocardial infarction. The most aggressive form of antithrombotic therapy, triple therapy (VKA + 2 antiplatelets) was uncommon. Around 1% of patients in each cardiac disease group were prescribed triple therapy and the median durations of around 60-70 days suggest it was generally for short term treatment.

5.3.2 Phenotype Development

5.3.2.1 Characteristics of bleeding events in HES, CPRD and MINAP

During follow up 16087 (12.5%) patients had a hospitalisation bleeding record in HES (23719 bleeding records in total), 17716 (13.8%) patients had a primary care bleeding record in CPRD (30107 bleeding records in total), 1494 (1.2%) patients had bleeding recorded as underlying cause of death in ONS, 2405 (1.9%) patients had bleeding recorded as a secondary cause of death in ONS and 797 (0.6%) patients had a bleeding complication recorded in MINAP (823 bleeding complications in total).

5.3.2.2 Bleeding anatomical site

Distribution across data sources

The distribution of bleeding anatomical sites is presented in **Table 5.4**. Within each data source, the most common anatomical bleeding sites were genitourinary (31.1%), lower gastrointestinal (31.1%), ruptured aortic aneurysm or haemopericardium (36.6%), unspecified gastrointestinal (30.6%) and unspecified site (93.8%) in HES, CPRD, ONS (underlying cause of death), ONS (other cause of death) respectively and MINAP. Gastrointestinal bleeding (including upper, lower and unspecified) was the most common bleeding site accounting for 40.9% of bleeding events across HES, CPRD and ONS. Intracranial bleeding and ruptured aortic aneurysm or haemopericardium were relatively uncommon outside of the death registry. Bleeding respiratory sites was most commonly found in primary care records, in particular upper respiratory (15.1%). Bleeding related to the eyes, ears, renal system or bleeding disorders were uncommon in primary care and hospital admissions and were never recorded as an underlying or other cause of death in ONS.

Anatomical site and short term mortality

In both HES (**Figure 5.1**) and CPRD (**Figure 5.2**) ruptured aortic aneurysm or haemopericardium were related with the highest 30 day mortality: 53.2% in HES and 22.8% in CPRD. Intracranial bleeding was also associated with high short term mortality. In MINAP (**Figure 5.3**) intracranial bleeding was associated with 60.7% 30 day mortality, the highest amongst the available categories. Bleeding requiring hospitalisation in general had higher 30 day mortality than bleeding events captured in primary care.

5.3.2.3 Length of hospitalisation and primary reason for admission

Distribution across bleeding anatomical sites

Descriptive statistics for length of stay and reason for admission for hospitalised bleeding is shown in **Table 5.5**.

The median length of stay for bleeding events that required hospitalisation was 3 days (IQR: 0, 10 days). Intracranial bleeding was associated with longest length of stay, with median hospitalisation days ranging from 9 (3, 21) to 13.5 (5.8, 25.5) amongst the subgroups. Genitourinary and ocular bleeding had median hospitalisations of 0 days.

Overall, 14846 (62.6%) hospitalised bleeding events were recorded as the primary reason for hospitalisation. The vast majority of intracranial bleeding events were the primary reason for hospital admission; in particular 85.2% of intracerebral, 84.6% of subarachnoid and 80.8% of subdural were coded as such. Ruptured aortic aneurysm and haemopericardium also were mostly primary admission events (81.1%). Bleeding disorders and unspecified bleeding events were rarely coded as the primary reason for hospitalisation, 33% and 25.2% respectively.

Length of hospitalisation, reason for admission and short term mortality

Patients hospitalised for longer durations had higher mortality at discharge (**Figure 5.4**). At discharge the risk of all-cause mortality was 4.5%, 11.6%, 15.1%, 20.6%, 23.3% for 0-3, 3-7, 7-14, 14-30 and >30 days hospitalisation respectively. At 30 days following discharge the risk of all-cause mortality was 6.3%, 15.1%, 20.0%, 28.5% and 32.3% for 0-3, 3-7, 7-14, 14-30 and >30 days hospitalisation respectively.

Hospital admissions primarily for bleeding had lower 30 day all-cause mortality than admissions where bleeding was not the primary event (8.7% [8.2, 9.1%] vs. 13.5% [12.8, 14.2%]). However it was considered that this comparison may have been confounded by there being more severe non-bleeding primary events in the cases that bleeding is secondary. Therefore I considered the same comparison but with a bleeding specific cause of mortality outcome. Here I saw that when bleeding is the primary reason for admission the risk of 30 day bleeding mortality is higher than when it was not the primary reason for admission (4.5% [4.2, 4.8%] vs. 3.7% [3.3, 4.1%]) (**Figure 5.5**).

5.3.2.4 Transfusion

Distribution of transfusion events around HES and CPRD bleeding events

Of the HES bleeding events, 0.4% had a transfusion record in CPRD and 2.3% had a transfusion record in OPCS on the same day. Expanding the time window to +/-30 days around the

bleeding event increased the capture of potentially related transfusions to 1.6% in CPRD and 7.5% in OPCS **(Table 5.6).** Transfusion records appeared to mostly coincide with gastrointestinal bleeding events – 13% upper GI, 10.4% lower GI and 15.1% unspecified GI bleeding events had a transfusion record in OPCS within 30 days.

CPRD bleeding events had fewer related transfusion events in CPRD and OPCS (**Table 5.7).** 0.4% had a transfusion record in CPRD and 0.4% had a transfusion record in OPCS on the same day. Expanding the time window to +/-30 days around the bleeding event increased the capture of potentially related transfusions to 1% in CPRD and 2.8% in OPCS**.** Transfusion records appeared to mostly coincide with gastrointestinal bleeding events – 13% upper GI, 10.4% lower GI and 15.1% unspecified GI bleeding events had a transfusion record in OPCS within 30 days. Similarly to HES, transfusion records were most commonly associated with gastrointestinal bleeding events.

Transfusion and short term mortality

Amongst HES bleeding events (**Figure 5.6**), 90 day mortality was highest in the group with a transfusion record in OPCS within 30 days of their bleeding event (30.1%), followed by bleeding events with transfusion recorded in both CPRD and OPCS within 30 days (16.7%), bleeding events with no transfusion (16.5%) and bleeding events with transfusion captured in CPRD within 30 days (10.0%).

Amongst CPRD bleeding events **(Figure 5.7)**, 90 day mortality was highest in the group with transfusion record in OPCS within 30 days of a bleeding event (18.7%), followed by bleeding events with transfusion recorded in both OPCS and CPRD within 30 days (18.3%), bleeding events with transfusion captured in CPRD within 30 days (10.9%) and bleeding events with no transfusion record within 30 days (4.8%).

5.3.2.5 Haemoglobin

Recording of haemoglobin in CPRD general practices

The total number of haemoglobin records per year in each of the 224 general practices over the course of the study period (1997-2010) is shown in **Supplementary appendix 11.2.9**. All 224 general practices had at least one haemoglobin record and in total, 811218 haemoglobin values were recorded. Under half (42.4%) of the general practices received at least one haemoglobin lab result per year for the entire study period and many practices increased recording haemoglobin values over time. On average, general practices had 293 haemoglobin records per year (median 249, IQR: 87, 434)

Capture of acute and chronic haemoglobin values around bleeding events

891(3.8%) bleeding events in HES had a haemoglobin value recorded in CPRD within +/- 1 day **(Table 5.8)**. The mean (SD) haemoglobin in these cases was 10.2 (2.9) g/dL. Increasing the time window to +/- 7 days improved the capture of acute haemoglobin to 9.2%, with a mean (SD) of 11.0 (2.7) g/dL. Gastrointestinal and upper respiratory bleeding events and bleeding disorders had the highest relative availability of haemoglobin within +/- 7 days (**Supplementary appendix 11.2.10**). Chronic haemoglobin data availability was better. In the year prior, 61.3% HES bleeding events had a haemoglobin record, mean (SD): 12.7 (2.2) g/dL and the year following, 47.3%, 12.3 (2.0) g/dL.

2542 (8.4%) bleeding events in CPRD had a haemoglobin value recorded in CPRD within +/- 1 day (**Table 5.9**). The mean (SD) haemoglobin in these cases was 12.3 (2.4) g/dL. Increasing the time window to +/- 7 days improved the capture of acute haemoglobin to 18.1%, with a mean (SD) of 12.4 (2.3) g/dL. Gastrointestinal, genitourinary and upper respiratory bleeding events and bleeding disorders had the highest availability of haemoglobin within +/- 7 days. In the year prior, 60.9% CPRD bleeding events had a haemoglobin record, mean (SD): 13.1 (1.9) g/dL and the year following, 59.0%, 12.8 (2.0) g/dL.

Estimating haemoglobin drop due to bleeding

The average chronic haemoglobin values were consistently higher than the acute values in both CPRD and HES, suggesting some degree of a haemoglobin drop can be detected but only in a limited number of patients. I was able to calculate haemoglobin drop for 1636 (6.9%) of bleeding events in HES and the average drop was 1.3 (2.3) g/dL. I was able to calculate haemoglobin drop for 3720 (12.3%) of bleeding events in CPRD for which the average drop was 1.1 (1.8) g/dL. The average haemoglobin drop was highest amongst unspecified gastrointestinal bleeding events in both HES and CPRD, 2.7 (2.5) g/dL and 2.4 (2.7) g/dL respectively. The distribution of estimated haemoglobin drops are shown in **Figure 5.8.**

Haemoglobin drop and short term mortality

Following HES bleeding 30 day all-cause mortality was similar amongst events with haemoglobin drop ≤3d/dL (9.95%),>3g/dL (10.8%) and no data (10.5%) (**Figure 5.9)**. Differences in 30 day mortality between the groups were more apparent following CPRD bleeding (**Figure 5.10)**. Those with a higher haemoglobin drop (>3 g/dL) had 30 day mortality risk of 5.14%, while those with ≤3g/dL and no data had risks of 2.07% and 2.28% respectively. In MINAP a small number of bleeding events were coded with haemoglobin drop information

(**Figure 5.3**). 30 day all-cause mortality risk was 17.2%, 15.9% and 13.3% following bleeding events with >5g/dL, 3-5 g/dL and >3 g/dL respectively.

5.3.2.6 Examinations and interventions

For hospitalised bleeding events, endoscopic examinations **(Table 5.10)** were recorded in OPCS on the same day for 22.6%, within 7 days for 33.2% and within 30 days for 39.7%. Endoscopies were found to be mostly associated with gastrointestinal (present for 36.1%, 51.1%, 40.5% of upper, lower and unspecified gastrointestinal bleeding events respectively), lower respiratory (31.9%) and genitourinary (55.1%) within 30 days of hospitalised bleeding. Surgical arrest of bleeding **(Table 5.10)** was observed to be associated with upper respiratory bleeding events. 11.2% and 22.7% upper respiratory bleeding events had a surgical arrest procedure record on the same day or within 30 days respectively. Surgical arrest procedures were otherwise rare to be recorded within 30 days of bleeding from any other anatomical site. Similarly, haematoma evacuation or aspiration records appeared almost exclusively around subdural intracranial bleeding events. 6.6% and 41% of subdural intracranial bleeding events had a haematoma evacuation or aspiration procedure on the same day or within 30 days respectively. Prescribed medication that may be related to bleeding cessation or bleeding recovery (antifibrinolytics, haemostatics and vitamin K) **(Table 5.10)** were uncommon. Overall, prescription records for antifibrinolytic or haemostatic drugs were present for 0.7% bleeding events within 30 days and vitamin K prescriptions were present for 0.07% bleeding events within 30 days.

Examinations, interventions and short-term mortality

The examinations and interventions investigated appeared to have a protective effect. Gastrointestinal, genitourinary and lower respiratory bleeding events that had a record of endoscopic examination within 30 days were associated with lower 90 day all-cause mortality than those that were not (9.0% vs. 22.9%) **(Figure 5.11)**. Upper respiratory bleeding that required surgical arrest was associated with 5.3% 90 day all-cause mortality compared with 11.3% for those that had no record of surgical arrest procedures within 30 days **(Figure 5.12)**. Subdural intracranial bleeding that had a record of haematoma evacuation or aspiration within 30 days had 18.3% 90 day all-cause mortality compared with 48.0% for those that didn't **(Figure 5.13)**.

5.3.2.7 Number of bleeding codes recorded on a single date

There were 1323 occurrences in HES of more than one bleeding ICD-10 code being recorded on the same date for a single patient. In 1192 cases 2 codes were recorded, in 113 cases 3 codes were recorded and in 18 cases 4 codes were recorded. The combinations of anatomical

sites in each of these cases are shown in **Figure 5.14.** Combinations of codes including upper, lower and unspecified gastrointestinal and genitourinary bleeding were the most common.

There were 1063 occurrences in CPRD of more than one bleeding Read code being recorded on the same date for a single patient. In 1007 cases 2 codes were recorded, in 53 cases 3 codes were recorded and in 3 cases 4 codes were recorded. The combinations of anatomical sites in each of these cases are shown in **Figure 5.15**. In the majority of these cases the codes recorded belonged to the same anatomical site category. There was much less variation in the anatomical sites captured in these cases in CPRD compared with HES.

Multiple bleeding codes and short-term mortality

Cases in HES with >1 ICD-10 bleeding code recorded on the same date had higher 30 day mortality than cases with a single code, 15.9% vs. 9.8% (**Figure 5.16**). Cases in CPRD with > 1 Read bleeding code recorded on the same date also had higher 30 day mortality than cases with a single code, 3.0% vs. 2.2% **(Figure 5.17**)

5.3.3 Inferring bleeding events

Of the 128815 patients in the study cohort, 27249 had at least one bleeding record in HES, CPRD or ONS. I examined the records of the remaining 101566 patients to determine whether any patients had evidence of a bleeding event that had not been captured in the bleeding codes. Through the pathways described in **Section 5.2.3.3**, I identified 477 surgical arrest or haematoma evacuation procedures in 451 patients; 689 cases in 514 patients of a transfusion code in OPCS accompanied by an iron deficiency anaemia diagnosis in HES or CPRD within 30 days; 77 cases in 62 patients of a transfusion code in OPCS accompanied by a haemoglobin value of <10g/dL in CPRD within 30 days, an endoscopic examination within 30 days and no history of cancer, liver or renal disease in the year prior to transfusion; and 249 cases in 182 patients of haemoglobin <10g/dL in CPRD, an endoscopic examination within 30 days and no history of cancer, liver or renal disease in the year prior to the haemoglobin record. That is, overall 1492 potential bleeding events identified in 1144/101566 patients with no bleeding record in HES or CPRD.

5.3.4 The CALIBER bleeding phenotype

Based on the results in **Section 5.3.2** I determined that the following characteristics were appropriate to incorporate into the algorithm to define non-fatal bleeding severity:

- Anatomical site: Intracranial or ruptured aortic aneurysm or haemopericardium
- Transfusion record in hospital admissions (OPCS) within 30 days
- Primary reason for hospitalisation AND >14 days stay

- >1 bleeding code recorded on a single date in a single data source

Fatal bleeding events were determined by bleeding records in ONS (underlying or other cause of death) or bleeding records in HES or CPRD followed by all-cause mortality within 7 days.

The CALIBER bleeding phenotype algorithm is presented in **Figure 5.18.** When identifying bleeding events first search we for fatal bleeding events in ONS, followed by bleeding in hospital admissions and determine whether it was fatal or non-fatal event and check for markers of bleeding severity. Then we search for bleeding in primary care and determine fatality or severity markers as applicable. Otherwise we search for evidence of bleeding outside of the bleeding codes. The algorithm is scalable and may be adapted to researchers' needs, for example if they are interested in non-fatal bleeding events only, or if they only have access to a single data source e.g. hospital admission records in HES.

Applying the algorithm to the cohort (n=128815) there were 39804 bleeding events in 27259 patients (21.2%) from primary care, hospital admissions and death registry records. 1492 bleeding events were inferred in 1144 (0.9%) patients with no bleeding code in primary care or hospital admissions. 59.4% of coded bleeding events were captured in primary care, 50.2% in hospital admissions, and 3.8% events in the death registry. However, allowing a 30 day window, only 13.2% of coded bleeding events were captured in 2 or more data sources. The overlap of bleeding events between the data sources used is shown in **Figure 5.19**.

5.4 Discussion

Based on the electronic health records of 128815 patients with newly diagnosed atrial fibrillation, myocardial infarction, unstable angina or stable angina, I developed a comprehensive bleeding phenotype algorithm **(Figure 5.18)**. The established phenotype identified bleeding events and markers of bleeding severity using combinations of diagnoses, procedures, transfusion and haemoglobin in primary care and hospital admissions records.

5.4.1 Developing a bleeding phenotype in linked electronic health records

In developing the bleeding phenotype I carried out a comprehensive assessment of bleeding record distribution in primary care and hospital admission data and concurrent records of relevant events such as transfusion, haemoglobin and drug prescriptions various time windows which to my knowledge has not been carried out previously in linked electronic health records. Unlike previous EHR studies which used simple code lists for bleeding, I demonstrated the depth of information readily available within linked electronic health records and the capability to achieve a more granular case definition. The results highlighted the importance of using multiple linked data sources for defining bleeding phenotype in electronic health records. No

individual data source used in this study had complete coverage of coded bleeding diagnoses, transfusions, causes of death and other bleeding relevant data, and only 13.2% of bleeding cases were captured in multiple data sources (**Figure 5.19**). However, although the phenotype was developed in linked electronic health records; it may be adapted for use in non-linked sources, but with careful interpretation given the differences in incidence and outcomes associated with bleeding in primary care and hospital admissions. CALIBER includes linkages to MINAP, a national myocardial infarction registry. However, bleeding records in these myocardial infarction patients were sparse and details of bleeding anatomical site were largely absent therefore the decision was made not to include these records in the overall phenotype.

5.4.2 Markers of bleeding severity

I identified markers of bleeding severity available within the linked EHRs, which corroborate with bleeding definitions used in clinical trials **(Table 5.2)**. In addition to fatality, hospitalisation, transfusion and anatomical bleeding site which are commonly adopted within bleeding definitions, I found evidence to incorporate mode of hospital admission, length of hospitalisation and the presence of multiple bleeding records as markers of severity. It is important that bleeding severity is accurately captured within phenotype algorithms due to the heterogeneous nature of bleeding and the outcomes following bleeding. The prognostic importance of haemoglobin falling below the normal range in stable coronary disease and myocardial infarction patients has previously been demonstrated.[192] With inevitable future availability of data such as hospital haemoglobin records, and amount of blood transfused I envision the definition of severity may be further refined.

5.4.3 Inferring bleeding events

A previous study showed that it is appropriate to infer disease cases in electronic health records where diagnosis codes are absent.[38] I identified 1144 patients with no coded bleeding diagnosis present but exhibited signs or symptoms of bleeding, such as low haemoglobin, iron deficiency anaemia or had a bleeding related procedure, with careful scrutiny to rule out cases where bleeding may not be the cause of these signs, symptoms and procedures, such as cancer, liver and renal diseases. This suggests the importance of looking beyond diagnosis codes in electronic health records for case finding. Identifying further bleeding events may ameliorate the underestimation of bleeding risk and improve statistical power in current safety studies of antithrombotic use.

5.4.4 Addition to previous implementations of bleeding phenotypes in EHRs

While some previous bleeding definitions have been limited to single disease areas in particular gastrointestinal bleeding,[70,72-74] I was fully inclusive with regards to anatomical site.

Indeed I found gastrointestinal bleeding records were among the most common in CPRD and HES. However clinically relevant bleeding from upper respiratory and genitourinary sites were also common within a population that is at relatively high risk of bleeding due to their eligibility for antithrombotic therapy.

5.4.5 Limitations of EHRs

Our EHR lacked sufficient data for acute haemoglobin or the number of units transfused support the classification of bleeding severity. These are markers commonly used to assess bleeding severity in practice and will likely made a significant contribution in providing greater refinement to bleeding definitions in EHRs when the data inevitably becomes available. Due to the fully anonymised data in CALIBER, I was unable to validate bleeding cases using a gold-standard measure, such as manual review of hospital charts. However individual components of the phenotype, such as subgroups of the bleeding codes, have been validated in previous studies in CPRD, HES[73] and other electronic health record systems[70,72,74-76] and the analysis of outcomes following bleeding adequately reflected expected results across levels of bleeding severity.

5.4.6 Conclusion

In linked electronic health records I have developed bleeding phenotypes with distinct levels of severity which may be useful and informative for studies of bleeding incidence and outcomes. No single source of data was sufficient to fully capture bleeding events in the study population.

5.5 Tables and Figures

Table 5.1: Bleeding phenotypes developed using electronic health records

Author	Year	Bleeding endpoint(s) evaluated	No. of anatomical sites	Data source(s)	Setting	Study Population	Coding system (n codes)	Supporting EHR data used in case definition	Algorithm figure reported	Assessment of phenotype accuracy
Raiford et al[70]	1996	Upper GI bleeding or perforation	1	Saskatchewan Hospital Services Plan	Hospital admissions	Patients hospitalised for upper GI bleeding	ICD-9 (30)	No	No	Site specific codes PPV: 91% Nonspecific codes PPV: 68%
De Abajo et al[71]	1999	Upper GI bleeding	1	GPRD (UK)	Primary care	Patients with a record for acute upper GI bleeding	Read (codes not stated)	No	No	PPV: 95/96
Arnason et al[75]	2006	Any bleeding Major bleeding	8	A university hospital, Ottawa	Hospital admissions	Patients with a record for thromboembolism or bleeding	ICD-9 (81)	No (information from patient charts were used to classify severity)	No	Definite bleeding PPV: 91%; NPV: 91% Major bleeding PPV: 87%; NPV: 92%
Wahl et al[72]	2010	Severe upper GI bleeding	1	HealthCore Integrated Research Database (USA)	Hospital admissions	Patients with a record for upper GI bleeding	ICD-9 (original:75; refined:33)	Procedure codes	No	PPV: original: 56.5% refined: 87.8%
Cunningham et al[76]	2011	Serious bleeding related to oral anticoagulation	>4	Tennessee Medicaid program	Hospital admissions	Medicaid enrolees >30 years old	ICD-9 (39)	No (information from patient charts were used to classify severity)	No	PPV assessed for individual codes ranged from 71.4% to 100% (>5 charts)

Author	Year	Bleeding endpoint(s) evaluated	No. of anatomical sites	Data source(s)	Setting	Study Population	Coding system (n codes)	Supporting EHR data used in case definition	Algorithm figure reported	Assessment of phenotype accuracy
Crooks et al[73]	2012	Upper GI bleeding	1	CPRD HES ONS (UK)	Primary care Hospital admissions Death registry	Patients with a record for acute upper GI bleeding	Read (46) ICD-10 (22)	Causes, symptoms, endoscopy, death, transfusion, procedures, alcohol, anaemia, coagulation, collapse, other	Yes	None
Valkhoff et al[74]	2014	Upper GI bleeding	1	IPCI (Netherlands) HSD (Italy) ARS (Italy) Aarhus (Denmark)	Primary care Hospital admissions	Patients with a record for upper GI bleeding	IPCI (4) ICD-9 (26) ICD-10 (16)	No	No	IPCI - PPV: 21% HSD - PPV: 78% ARS - PPV: 72% Aarhus - PPV: 77%
Friberg et al[77]	2016	Fatal Non-fatal major Hospitalised Minor	4	Swedish Patient register	Hospital admissions Hospital outpatients Death registry	Atrial fibrillation patients	ICD-10 (38)	Anatomical site (intracranial) Transfusion Hospitalisation Diagnosis position	No	Fatal PPV: 88.1%; NPV: 99.7% Non-fatal major PPV: 90.6%; NPV: 91.5% Hospitalised PPV: 65.1%; NPV: 97.5% Minor PPV: 84.2%; NPV: 98.9%

Author	Year	Bleeding endpoint(s) evaluated	No. of anatomical sites	Data source(s)	Setting	Study Population	Coding system (n codes)	Supporting EHR data used in case definition	Algorithm figure reported	Assessment of phenotype accuracy
CALIBER – present study	2017	Fatal Hospitalised with markers of severity Hospitalised Primary care with markers of severity Primary care Inferred	18	CPRD HES ONS	Primary care Hospital admissions Death registry	Coronary disease and atrial fibrillation patients	Read (201) ICD-10(96)	Yes: Transfusion, anatomical site; procedures;	Yes	Validated bleeding severity through assessment of prognosis following bleeding

GI=Gastrointestinal; PPV= positive predictive value; NPV= negative predictive value

Table 5.2: Components of bleeding definitions used in randomised trials of antithrombotic therapies and the CALIBER bleeding phenotype

Component	Bleeding Academic Research Consortium (BARC)	International Society on Thrombosis and Haemostasis (ISTH)	Thrombosis In Myocardial Infarction (TIMI)	CALIBER bleeding phenotype
Fatality	●	●	●	●
Anatomic bleeding site	● Intracranial Intraocular	● Intracranial Intraspinal Intraocular Retroperitoneal Intraarticular Pericardial Intramuscular w. compartment syndrome	● Intracranial	● Intracranial Gastrointestinal Respiratory Ruptured aortic aneurysm or haemopericardium Genitourinary Bleeding disorders Ocular Ear Renal Unspecified
Haemoglobin drop	● 3 - <5g/dL ≥5g/dL	● ≥2 g/dL	● 3 - <5g/dL ≥5g/dL	○
Hospitalisation	●	●	○	●
Mode of hospital admission	○	○	○	●
Length of hospitalisation	○	○	○	●
Blood transfusion	●	●	●	●
Number of units transfused	●	●	○	○
Medical/surgical consultation	●	●	●	●
Medical or surgical intervention	●	●	●	●
Multiple bleeding codes	○	○	○	●
Haemodynamic compromise	○	●	○	○
Change in antithrombotic therapy	○	●	●	○

Table 5.3: Baseline patient characteristics in 4 common cardiac diseases

	Atrial Fibrillation (n= 27061)	Acute Myocardial Infarction (n=25031)	Unstable Angina (n=9500)	Stable Angina (n= 67223)
Demographics and behaviours at cohort entry				
Age (years), mean (SD)	76.6 (12.8)	69.9 (13.5)	69.1 (13.2)	70.4 (12.3)
Women, n (%)	14266 (52.7)	9206 (36.8)	4169 (43.9)	31365 (46.7)
Highest quintile of deprivation (most deprived)	5137 (19)	4758 (19.1)	1947 (20.5)	13837 (20.6)
Smoking status, n (%)				
Non-Smoker	11938 (44.1)	8895 (35.5)	3851 (40.5)	28721 (42.7)
Smoker	2277 (8.4)	3691 (14.7)	1058 (11.1)	6229 (9.3)
Ex-smoker	7531 (2.8)	6355 (25.4)	2634 (27.7)	19743 (29.4)
Missing, (%)	19.6	24.3	20.6	18.6
History of alcohol abuse, n (%)	2627 (9.7)	2430 (9.7)	908 (9.6)	6459 (9.6)
Medical history ***(Any record ever prior to entry)***				
Diabetes, n (%)				
Type 1	153 (0.6)	277 (1.1)	99 (1)	609 (0.9)
Type 2	2695 (10)	2922 (11.7)	1222 (12.9)	8728 (13)
Unspecified type	486 (1.8)	448 (1.8)	220 (2.3)	1352 (2)
Ischaemic or unspecified stroke, n (%)	2169 (8)	1462 (5.8)	558 (5.9)	3435 (5.1)
Peripheral arterial disease, n (%)	2276 (8.4)	2147 (8.6)	865 (9.1)	6199 (9.2)
Renal disease, n (%)	2570 (9.5)	1731 (6.9)	694 (7.3)	4351 (6.5)
Non-metastatic cancer, n (%)	5427 (20.1)	3158 (12.6)	1155 (12.2)	8701 (12.9)
Metastatic cancer, n (%)	526 (1.9)	209 (0.8)	74 (0.8)	520 (0.8)
Peptic ulcer, n (%)	1814 (6.7)	1713 (6.8)	753 (7.9)	5074 (7.5)
Bleeding diatheses and coagulation disorders, n (%)	312 (1.2)	175 (0.7)	77 (0.8)	534 (0.8)
Chronic anaemia, n (%)	4982 (18.4)	2808 (11.2)	1198 (12.6)	8125 (12.1)
Biomarkers ***(Nearest record to entry within 1 year prior to entry)***				
SBP (mmHg), mean (SD)	140 (21.8)	143 (21.2)	142 (21.2)	142 (20.5)
Missing, (%)	29.7	33.2	25.8	21.6
BMI, mean (SD)	27.2 (6.19)	27.7 (5.44)	28.4 (5.73)	28.5 (5.60)
BMI, n (%)				

	Atrial Fibrillation (n= 27061)	Acute Myocardial Infarction (n=25031)	Unstable Angina (n=9500)	Stable Angina (n= 67223)
Underweight	378 (1.4)	194 (7.8)	67 (0.7)	420 (0.6)
Normal	2609 (9.6)	2145 (8.6)	832 (8.8)	6411 (9.5)
Overweight	2556 (9.4)	2909 (11.6)	1219 (12.8)	9997 (14.9)
Obese	2024 (7.5)	2078 (8.3)	1053 (11.1)	8440 (12.6)
Missing, (%)	72	70.7	66.6	62.4
Haemoglobin (g/dL), mean (SD)	12.9 (1.97)	13.5 (1.91)	13.5 (1.75)	13.6 (1.69)
Missing, (%)	56.2	65.9	62	59.1
Creatinine (mol/l), mean (SD)	107 (59.1)	108 (56.1)	105 (55.9)	102 (46.1)
Missing, (%)	49.5	57.6	54.1	49.7
Antithrombotic therapies (n, %) and duration (median, IQR) ***(Between cohort entry and 1st bleeding event or end of follow-up)***				
Any antithrombotic therapy	16868 (62.3)	19950 (79.7)	7947 (83.7)	55619 (82.7)
Aspirin monotherapy	10787 (39.9)	16511 (66)	6695 (70.5)	48262 (71.8)
Duration (days)	382 (114, 908)	791 (267, 1742)	765 (268, 1691)	842 (305, 1752)
ADP receptor inhibitor monotherapy	1264 (4.7)	3683 (14.7)	1425 (15)	7351 (10.9)
Duration (days)	150 (46, 486)	94 (30, 376)	121 (42, 495)	181 (52, 652)
Dual antiplatelet therapy	1594 (5.9)	8673 (34.6)	2417 (25.4)	9539 (14.2)
Duration (days)	186 (90, 426)	349 (143, 478)	272 (98, 488)	261 (90, 476)
VKA monotherapy	7149 (26.4)	1666 (6.7)	853 (9)	6287 (9.4)
Duration (days)	427 (146, 1083)	216 (82, 626)	318 (110, 844)	344 (113, 938)
VKA + 1 antiplatelet	3003 (11.1)	1426 (5.7)	637 (6.7)	3892 (5.8)
Duration (days)	85 (51, 163)	106 (55, 262)	90 (54, 228)	90 (51, 214)
VKA + 2 antiplatelets	266 (1)	321 (1.3)	102 (1.1)	430 (0.6)
Duration (days)	68.5 (39, 93.2)	68.0 (43, 116.0)	62.5 (39, 90.0)	57.0 (35, 84.0)

Note: SD=standard deviation, SBP=systolic blood pressure, BMI= body mass index, IQR= interquartile range, ADP=adenosine diphosphate, VKA=vitamin K antagonist

Table 5.4: Distribution of records of bleeding in various anatomical sites in each data source

	CPRD (Primary care) n=30107	HES (Hospital admissions) n=23719	ONS - underlying cause (Death registry) n= 1494	ONS - other cause (Death registry) n= 2405	MINAP (MI registry) n=823
Intracranial					
Intracerebral	226 (0.75)	715 (3.01)	365 (24.43)	445 (18.5)	(Any intracranial) 28 (3.40)
Subarachnoid	137 (0.46)	201 (0.85)	92 (6.16)	94 (3.91)	
Extradural	1 (0.00)	22 (0.09)	0 (0)	1 (0.04)	
Subdural	83 (0.28)	442 (1.86)	56 (3.75)	92 (3.83)	
Unspecified	5 (0.02)	55 (0.23)	42 (2.81)	54 (2.25)	
Gastrointestinal					
Upper	1455 (4.83)	3142 (13.25)	192 (12.85)	133 (5.53)	
Lower	9365 (31.11)	5806 (24.48)	7 (0.47)	25 (1.04)	
Unspecified	1062 (3.53)	1530 (6.45)	182 (12.18)	735 (30.56)	
Respiratory					
Upper	4543 (15.09)	2207 (9.30)	0 (0)	4 (0.17)	(Retro-peritoneal) 23 (2.79)
Lower	1579 (5.24)	1281 (5.40)	6 (0.40)	42 (1.75)	
Unspecified	0 (0)	26 (0.11)	1 (0.07)	31 (1.29)	
Ruptured aortic aneurysm or haemopericardium	57 (0.19)	280 (1.18)	546 (36.55)	596 (24.78)	
Genitourinary	9032 (30.0)	7378 (31.11)	1 (0.07)	18 (0.75)	
Ocular	703 (2.34)	263 (1.11)	0 (0)	0 (0)	N/A
Ear	0 (0)	5 (0.02)	0 (0)	0 (0)	N/A
Renal	2 (0.01)	0 (0)	0 (0)	0 (0)	N/A
Bleeding disorders	33 (0.11)	227 (0.96)	0 (0)	0 (0)	N/A
Unspecified	1824 (6.06)	139 (0.59)	4 (0.27)	135 (5.61)	772 (93.80)

Figure 5.1: 30 day mortality following bleeding captured in HES stratified by anatomical site

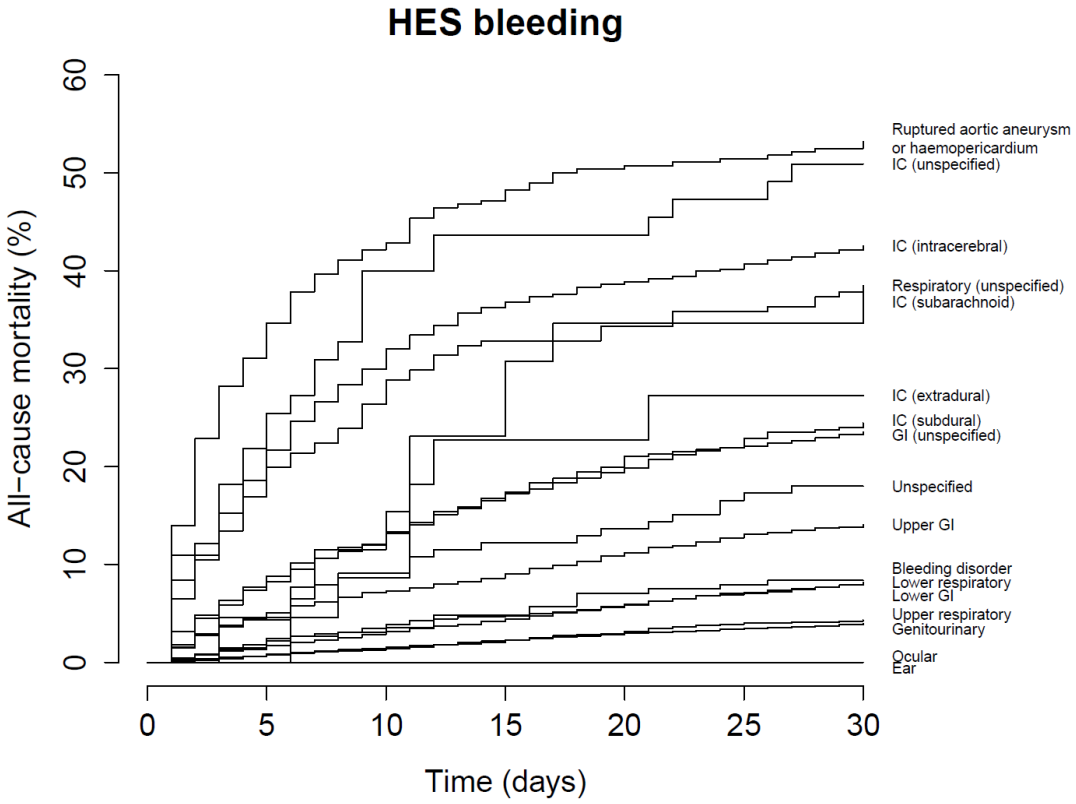


Figure 5.2: 30 day mortality following bleeding captured in CPRD stratified by anatomical site

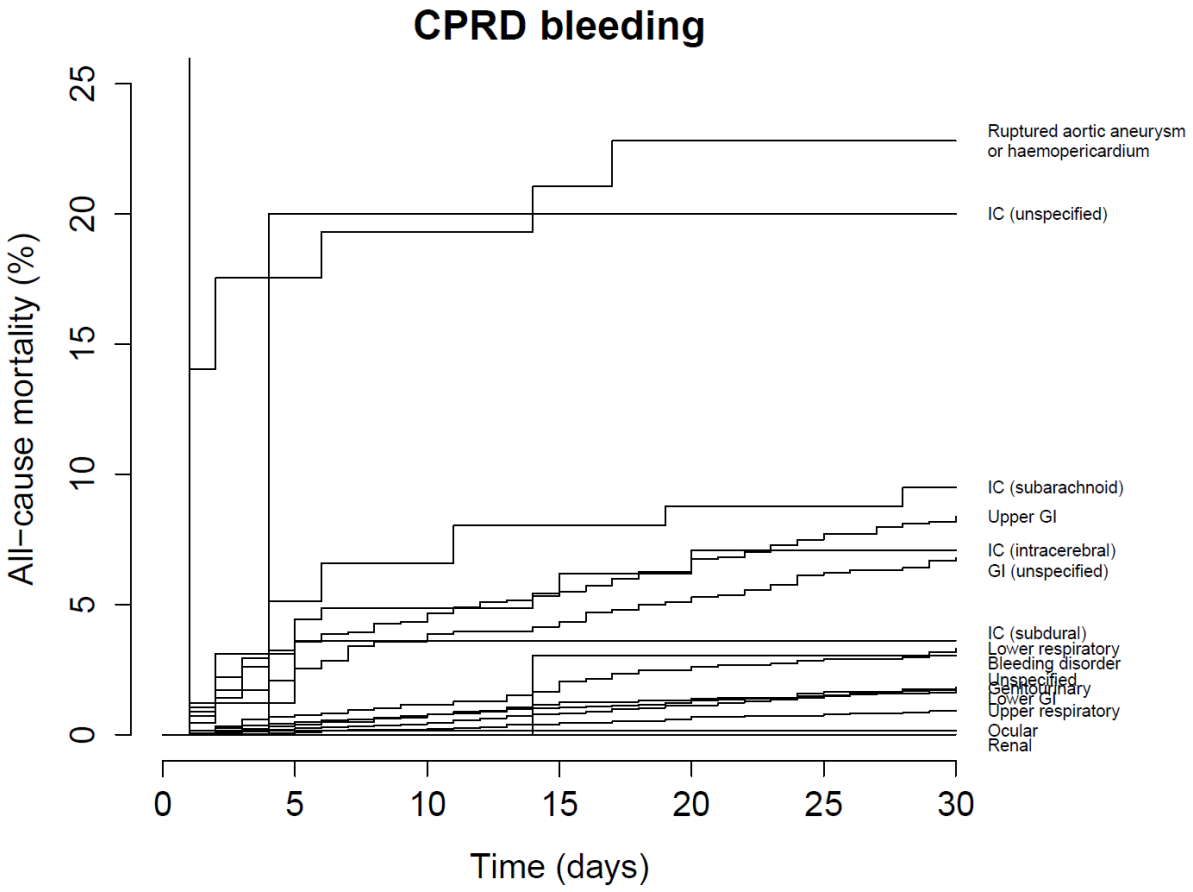
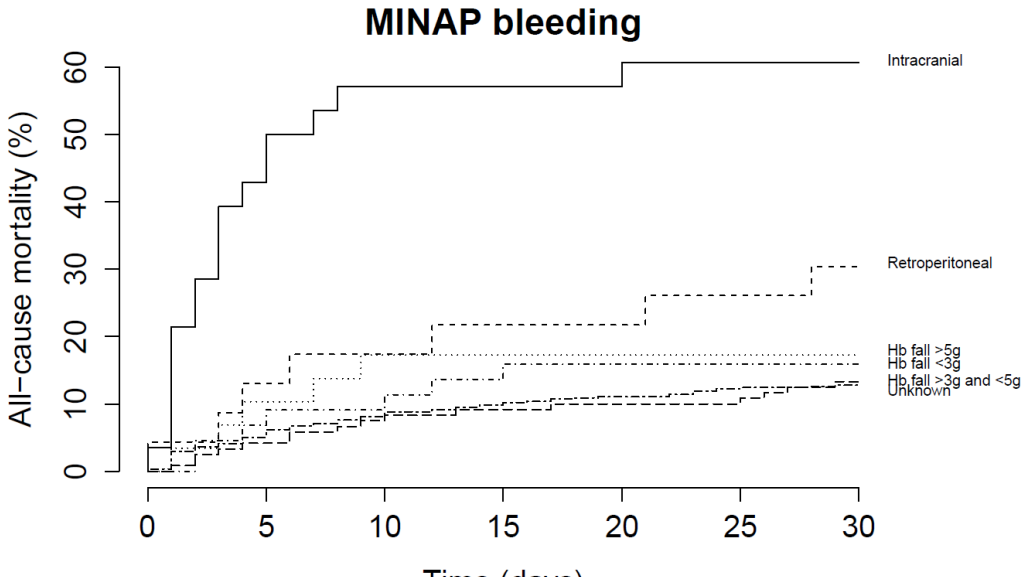


Figure 5.3: 30 day mortality following bleeding captured in MINAP stratified by MINAP bleeding category



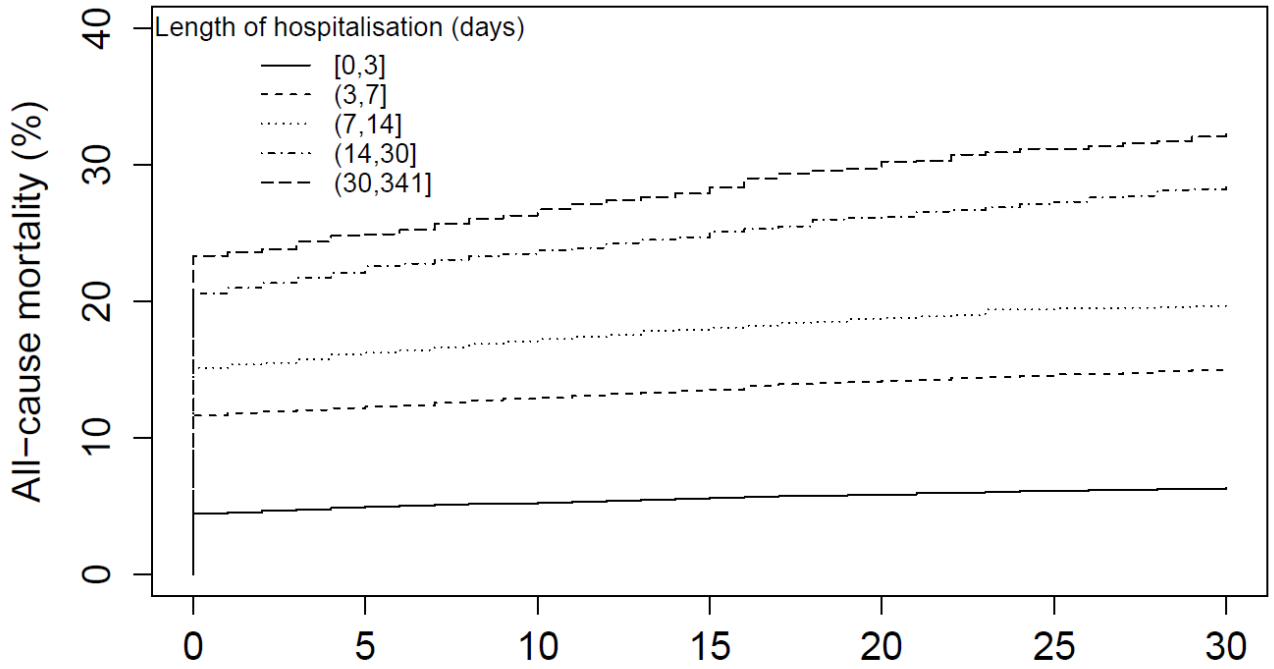
MINAP category		Number at risk		
Intracranial	28	12	12	11
Retroperitoneal	23	19	18	16
Hb fall >5g	29	24	24	24
Hb fall >3g and <5g	44	40	37	37
Hb fall <3g	120	111	108	104
Unknown	579	532	513	502

Table 5.5: Length of hospitalisation and reason for admission for 23719 HES bleeding codes in 16087 patients by bleed location

	IC (intracerebral)	IC (subarachnoid)	IC (subdural)	IC (extradural)	IC (NOS)	Upper GI	Lower GI	GI (NOS)	Upper resp.	Lower resp.	Resp. (NOS)	Rupt. AAA/ HP*	Genito-urinary	Ocular	Ear	Bleeding disorder	NOS	Overall
n	715	201	442	22	55	3142	5806	1530	2207	1281	26	280	7378	263	5	227	139	23719
Length of hospitalisation (days)																		
Mean (SD)	24.82 (32.49)	21.45 (33.13)	16.58 (21.93)	27.82 (37.28)	16.20 (18.55)	13.64 (19.71)	9.68 (18.10)	13.46 (21.08)	6.11 (13.95)	7.46 (14.05)	14.69 (16.79)	14.33 (22.20)	5.38 (13.59)	2.55 (8.41)	11.20 (20.61)	10.14 (16.28)	10.88 (16.04)	9.36 (18.16)
Median (IQR)	12.0 (5.00, 34.0)	10.0 (3.00, 26.0)	9.0 (3.00, 21.0)	13.5 (5.75, 25.5)	9.0 (3.50, 24.0)	7.0 (2.00, 16.8)	3.0 (0.00, 11.0)	7.0 (2.00, 16.0)	2.0 (1.00, 5.0)	2.0 (0.00, 9.0)	9.5 (4.50, 18.5)	6.5 (2.00, 17.0)	0.0 (0.00, 5.0)	0.0 (0.00, 2.0)	2.0 (1.00, 4.0)	5.0 (1.00, 12.8)	6.0 (1.00, 15.0)	3 (0, 10)
Min, Max	0, 299	0, 269	0, 196	0, 164	0, 90	0, 197	0, 341	0, 243	0, 299	0, 192	0, 67	0, 180	0, 214	0, 72	1, 48	0, 151	0, 130	0, 341
Missing %	0.1	0.5	0	0	0	0.6	0.4	0.1	0.1	0.2	0	0	0.1	0	0	0.4	0.7	
Primary reason for hospitalisation, n (%)	609 (85.2)	170 (84.6)	357 (80.8)	14 (63.6)	38 (69.1)	1939 (61.7)	3238 (55.8)	886 (57.9)	1678 (76)	725 (56.6)	12 (46.2)	227 (81.1)	4687 (63.5)	154 (58.6)	2 (40)	75 (33)	35 (25.2)	14846 (62.6)

*ruptured abdominal or thoracic aortic aneurysm or haemopericardium

Figure 5.4: The association between length of hospitalisation and 30 day mortality



Length of hospitalisation (days)	Number at risk			
(0,3]	12620	11919	11790	11686
(3,7]	3625	3136	3085	3045
(7,14]	2992	2471	2413	2379
(14,30]	2548	1930	1859	1797
(30,341]	1876	1334	1261	1214

Figure 5.5: The association between the reason for hospitalisation and 30 day mortality

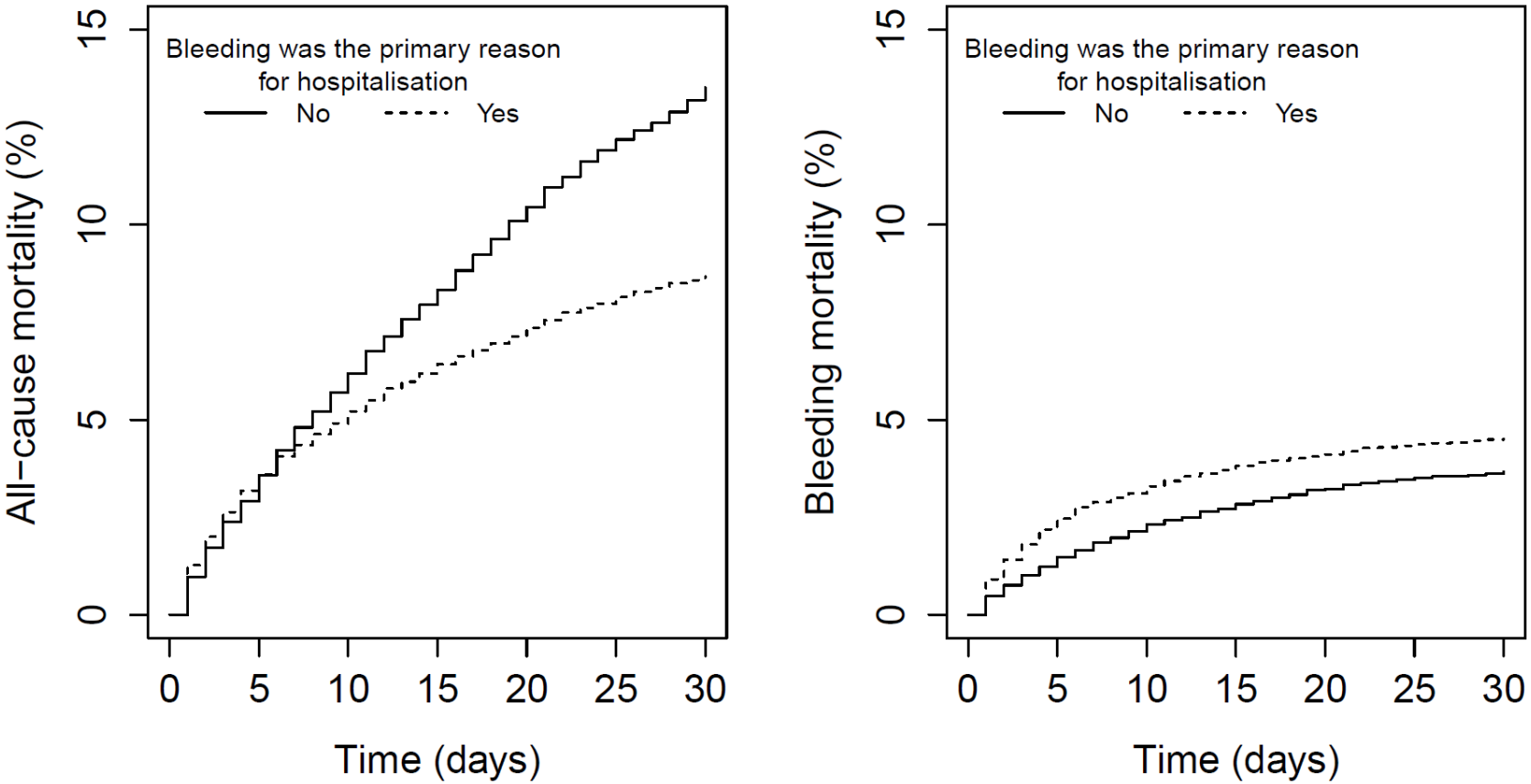


Table 5.6: HES bleeding (n= 23719 bleeding events in 16087 patients) and transfusions

	IC (intracerebral)	IC (subarachnoid)	IC (subdural)	IC (extradural)	IC (NOS)	Upper GI	Lower GI	GI (NOS)	Upper resp.	Lower resp.	Resp. (NOS)	Rupt. AAA/ HP*	Genito-urinary	Ocular	Ear	Bleeding disorder	NOS	Overall
n	715	201	442	22	55	3142	5806	1530	2207	1281	26	280	7378	263	5	227	139	23719
Transfusion																		
Transfusion (CPRD) within 0 days of bleed, n (%)	0 (0)	0 (0)	2 (0.5)	0 (0)	0 (0)	26 (0.8)	46 (0.8)	14 (0.9)	5 (0.2)	2 (0.2)	0 (0)	0 (0)	10 (0.1)	0 (0)	0 (0)	0 (0)	0 (0)	105 (0.4)
Transfusion (CPRD) within +/-7 days of bleed, n (%)	2 (0.3)	0 (0)	3 (0.7)	0 (0)	0 (0)	53 (1.7)	89 (1.5)	31 (2)	14 (0.6)	4 (0.3)	0 (0)	0 (0)	25 (0.3)	0 (0)	0 (0)	0 (0)	2 (1.4)	223 (0.9)
Transfusion (CPRD) within +/-14 days of bleed, n (%)	3 (0.4)	0 (0)	3 (0.7)	0 (0)	0 (0)	64 (2)	117 (2)	43 (2.8)	19 (0.9)	6 (0.5)	0 (0)	0 (0)	33 (0.4)	0 (0)	0 (0)	0 (0)	2 (1.4)	290 (1.2)
Transfusion (CPRD) within +/-30 days of bleed, n (%)	4 (0.6)	0 (0)	3 (0.7)	0 (0)	0 (0)	75 (2.4)	141 (2.4)	51 (3.3)	27 (1.2)	8 (0.6)	0 (0)	1 (0.4)	58 (0.8)	0 (0)	0 (0)	0 (0)	2 (1.4)	370 (1.6)
Transfusion (OPCS) within 0 days of bleed, n (%)	2 (0.3)	1 (0.5)	2 (0.5)	0 (0)	0 (0)	131 (4.2)	215 (3.7)	70 (4.6)	26 (1.2)	2 (0.2)	0 (0)	5 (1.8)	83 (1.1)	0 (0)	1 (20)	6 (2.6)	8 (5.8)	552 (2.3)
Transfusion (OPCS) within +/-7 days of bleed, n (%)	7 (1)	1 (0.5)	4 (0.9)	0 (0)	1 (1.8)	317 (10.1)	485 (8.4)	182 (11.9)	78 (3.5)	19 (1.5)	3 (11.5)	11 (3.9)	182 (2.5)	0 (0)	2 (40)	20 (8.8)	16 (11.5)	1328 (5.6)
Transfusion (OPCS) within +/-14 days of bleed, n (%)	8 (1.1)	1 (0.5)	7 (1.6)	0 (0)	1 (1.8)	355 (11.3)	534 (9.2)	207 (13.5)	101 (4.6)	28 (2.2)	4 (15.4)	11 (3.9)	220 (3)	0 (0)	2 (40)	21 (9.3)	19 (13.7)	1519 (6.4)
Transfusion (OPCS) within +/-30 days of bleed, n (%)	9 (1.3)	3 (1.5)	8 (1.8)	0 (0)	1 (1.8)	410 (13)	601 (10.4)	231 (15.1)	136 (6.2)	36 (2.8)	4 (15.4)	14 (5)	266 (3.6)	0 (0)	2 (40)	26 (11.5)	21 (15.1)	1768 (7.5)

*ruptured abdominal or thoracic aortic aneurysm or haemopericardium

Table 5.7: CPRD bleeding (n=30107 bleeding events in 17716 patients) and transfusions

	IC (intracerebral)	IC (subarachnoid)	IC (subdural)	IC (extradural)	IC (NOS)	Upper GI	Lower GI	GI (NOS)	Upper resp.	Lower resp.	Rupt. AAA/ HP*	Genito-urinary	Ocular	Renal	Bleeding disorder	NOS	Overall
n	226	137	83	1	5	1455	9365	1062	4543	1579	57	9032	703	2	33	1824	30107
Transfusion																	
Transfusion (CPRD) within 0 days of bleed, n (%)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	19 (1.3)	35 (0.4)	33 (3.1)	4 (0.1)	1 (0.1)	0 (0)	9 (0.1)	0 (0)	0 (0)	0 (0)	5 (0.3)	106 (0.4)
Transfusion (CPRD) within +/-7 days of bleed, n (%)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	29 (2)	62 (0.7)	45 (4.2)	11 (0.2)	2 (0.1)	0 (0)	27 (0.3)	0 (0)	0 (0)	0 (0)	5 (0.3)	181 (0.6)
Transfusion (CPRD) within +/-14 days of bleed, n (%)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	31 (2.1)	84 (0.9)	50 (4.7)	20 (0.4)	3 (0.2)	0 (0)	38 (0.4)	0 (0)	0 (0)	1 (3)	7 (0.4)	234 (0.8)
Transfusion (CPRD) within +/-30 days of bleed, n (%)	0 (0)	0 (0)	1 (1.2)	0 (0)	0 (0)	35 (2.4)	104 (1.1)	61 (5.7)	24 (0.5)	5 (0.3)	0 (0)	61 (0.7)	2 (0.3)	0 (0)	1 (3)	8 (0.4)	302 (1.0)
Transfusion (OPCS) within 0 days of bleed, n (%)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	24 (1.6)	55 (0.6)	24 (2.3)	4 (0.1)	1 (0.1)	1 (1.8)	15 (0.2)	0 (0)	0 (0)	1 (3)	0 (0)	125 (0.4)
Transfusion (OPCS) within +/-7 days of bleed, n (%)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	71 (4.9)	202 (2.2)	100 (9.4)	35 (0.8)	11 (0.7)	1 (1.8)	58 (0.6)	0 (0)	0 (0)	1 (3)	11 (0.6)	490 (1.6)
Transfusion (OPCS) within +/-14 days of bleed, n (%)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	88 (6)	270 (2.9)	129 (12.1)	45 (1)	15 (0.9)	1 (1.8)	90 (1)	0 (0)	0 (0)	1 (3)	19 (1)	659 (2.2)
Transfusion (OPCS) within +/-30 days of bleed, n (%)	0 (0)	0 (0)	1 (1.2)	0 (0)	0 (0)	105 (7.2)	324 (3.5)	148 (13.9)	83 (1.8)	17 (1.1)	1 (1.8)	119 (1.3)	1 (0.1)	0 (0)	1 (3)	26 (1.4)	831 (2.8)

*ruptured abdominal or thoracic aortic aneurysm or haemopericardium

Figure 5.6: The association between HES bleeding requiring transfusion and short term all-cause mortality

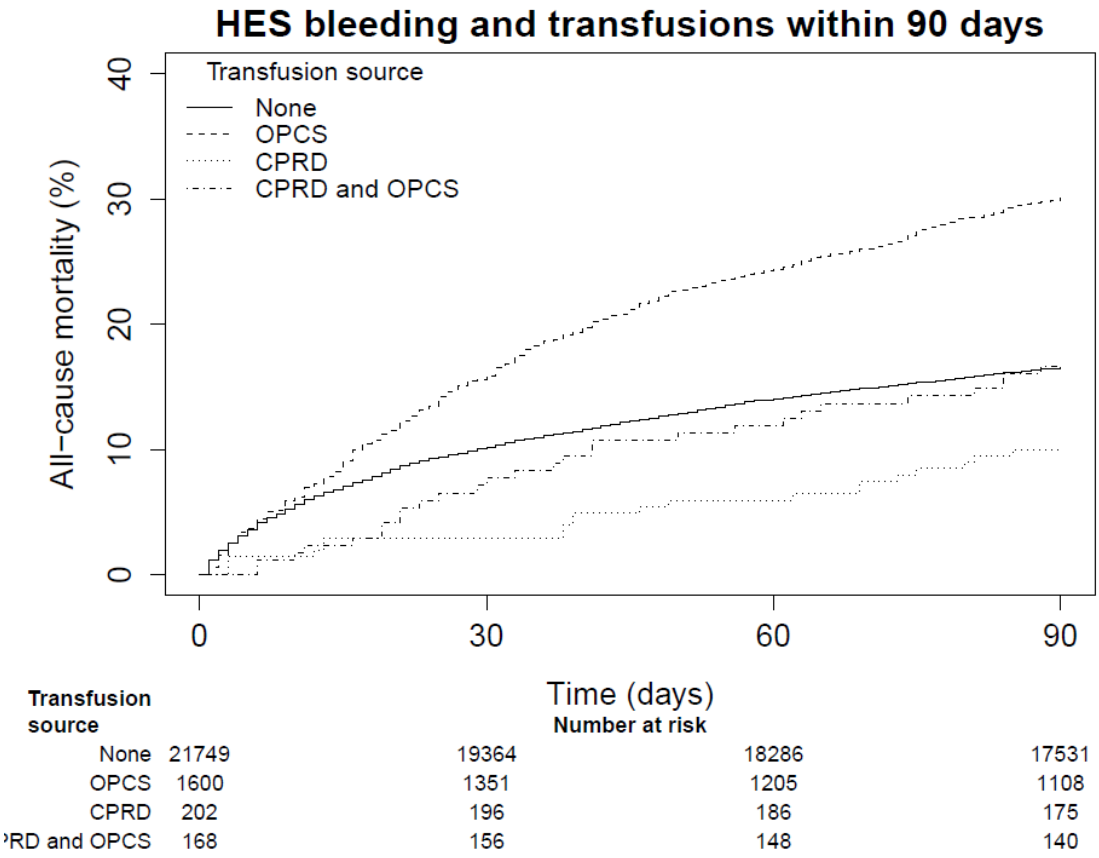
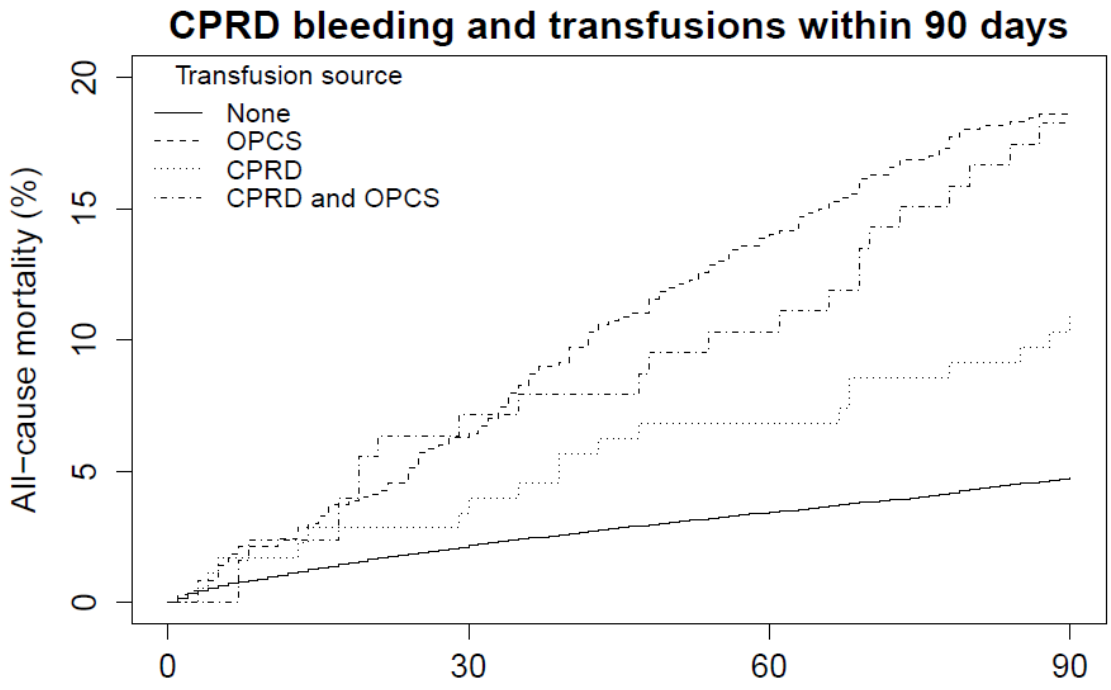


Figure 5.7: The association between CPRD bleeding requiring a transfusion and short term mortality



Transfusion source	Number at risk			
None	29105	28267	27642	27027
OPCS	700	656	603	568
CPRD	176	170	161	155
CPRD+OPCS	126	117	113	103

.

Table 5.8: HES bleeding (n= 23719 bleeding events in 16087 patients) and haemoglobin

	IC (intracerebral)	IC (subarachnoid)	IC (subdural)	IC (extradural)	IC (NOS)	Upper GI	Lower GI	GI (NOS)	Upper resp.	Lower resp.	Resp. (NOS)	Rupt. AAA/ HP*	Genito-urinary	Ocular	Ear	Bleeding disorder	NOS	Overall
n	715	201	442	22	55	3142	5806	1530	2207	1281	26	280	7378	263	5	227	139	23719
Acute haemoglobin (g/dL)																		
Haemoglobin record (+/- 1 day of bleed), n (%)	8 (1.1)	4 (2)	9 (2)	0 (0)	0 (0)	160 (5.1)	282 (4.9)	95 (6.2)	97 (4.4)	35 (2.7)	1 (3.8)	8 (2.9)	172 (2.3)	1 (0.4)	1 (20)	11 (4.8)	7 (5)	891 (3.8)
Haemoglobin (+/- 1 day of bleed), mean (SD)	12.51 (2.048)	12.07 (0.655)	10.88 (3.688)	-	-	9.67 (2.929)	9.56 (2.727)	8.52 (2.310)	12.00 (2.517)	11.89 (1.836)	12.10 (-)	11.91 (2.574)	11.28 (2.835)	12.60 (-)	8.40 (-)	11.07 (3.192)	10.38 (1.881)	10.23 (2.90)
Haemoglobin record (+/- 7 days of bleed), n (%)	17 (2.4)	11 (5.5)	26 (5.9)	3 (13.6)	2 (3.6)	339 (10.8)	667 (11.5)	191 (12.5)	241 (10.9)	94 (7.3)	2 (7.7)	12 (4.3)	523 (7.1)	11 (4.2)	1 (20)	28 (12.3)	22 (15.8)	2190 (9.2)
Haemoglobin (+/- 7 days of bleed), mean (SD)	12.61 (2.120)	11.68 (1.280)	12.01 (2.919)	12.00 (3.041)	13.30 (2.404)	10.52 (2.801)	10.45 (2.591)	9.50 (2.554)	11.77 (2.284)	12.18 (2.253)	13.00 (1.273)	12.07 (2.292)	11.65 (2.490)	11.00 (0.894)	8.40 (-)	11.11 (2.617)	9.88 (2.035)	10.95 (2.65)
Chronic haemoglobin (g/dL)																		
Haemoglobin record (in year prior to bleed), n (%)	378 (52.9)	113 (56.2)	284 (64.3)	13 (59.1)	24 (43.6)	1824 (58.1)	3681 (63.4)	983 (64.2)	1338 (60.6)	788 (61.5)	14 (53.8)	150 (53.6)	4579 (62.1)	145 (55.1)	2 (40)	135 (59.5)	96 (69.1)	14547 (61.3)
Haemoglobin (in the year prior to bleed), mean (SD)	13.0 (1.817)	13.1 (1.889)	13.0 (2.029)	14.2 (0.961)	12.9 (1.832)	12.3 (2.150)	12.4 (2.104)	11.8 (2.148)	12.8 (2.083)	13.1 (1.858)	12.6 (1.561)	12.9 (1.858)	13.0 (1.911)	12.8 (1.701)	12.4 (1.061)	12.9 (1.991)	11.6 (2.262)	12.66 (2.06)
Haemoglobin record (in year post bleed), n (%)	142 (19.9)	51 (25.4)	124 (28.1)	5 (22.7)	8 (14.5)	1435 (45.7)	2955 (50.9)	688 (45)	1200 (54.4)	567 (44.3)	8 (30.8)	76 (27.1)	3648 (49.4)	134 (51)	3 (60)	104 (45.8)	71 (51.1)	11219 (47.3)
Haemoglobin (in the year post bleed), mean (SD)	13.3 (1.60)	13.5 (1.76)	13.2 (1.99)	12.1 (2.27)	14.2 (2.19)	12.0 (1.95)	12.1 (1.97)	11.5 (2.02)	12.2 (2.03)	12.8 (1.93)	12.3 (1.73)	11.8 (1.87)	12.7 (1.96)	12.8 (1.80)	12.3 (2.25)	12.2 (2.18)	11.4 (2.08)	12.32 (2.01)
Haemoglobin drop (g/dL)**																		
Haemoglobin drop, n (%)	10 (1.4)	6 (3)	16 (3.6)	3 (13.6)	2 (3.6)	236 (7.5)	512 (8.8)	138 (9)	182 (8.2)	76 (5.9)	2 (7.7)	8 (2.9)	396 (5.4)	10 (3.8)	0 (0)	20 (8.8)	19 (13.7)	1636 (6.9)
Haemoglobin drop, mean (SD)	0.700 (1.63)	1.067 (1.46)	0.637 (0.77)	1.600 (3.39)	1.900 (0.99)	2.206 (2.67)	2.317 (2.39)	2.664 (2.45)	1.521 (2.02)	0.776 (1.48)	1.950 (0.07)	1.013 (1.11)	1.350 (1.87)	0.740 (0.70)	-	2.220 (1.58)	2.247 (2.11)	1.88 (2.27)

*ruptured abdominal or thoracic aortic aneurysm or haemopericardium

Haemoglobin drop defined as **max haemoglobin in year prior to bleed minus **min** haemoglobin recorded within 7 days of bleed

Table 5.9: CPRD bleeding (n=30107 bleeding events in 17716 patients) and haemoglobin

	IC (intracerebral)	IC (subarachnoid)	IC (subdural)	IC (extradural)	IC (NOS)	Upper GI	Lower GI	GI (NOS)	Upper resp.	Lower resp.	Rupt. AAA/ HP*	Genito-urinary	Ocular	Renal	Bleeding disorder	NOS	Overall
n	226	137	83	1	5	1455	9365	1062	4543	1579	57	9032	703	2	33	1824	30107
Acute haemoglobin																	
Haemoglobin record (+/- 1 day of bleed), n (%)	6 (2.7)	1 (0.7)	1 (1.2)	0 (0)	0 (0)	125 (8.6)	1018 (10.9)	110 (10.4)	301 (6.6)	102 (6.5)	2 (3.5)	794 (8.8)	15 (2.1)	0 (0)	8 (24.2)	59 (3.2)	2542 (8.4)
Haemoglobin (+/- 1 day of bleed), mean (SD)	13.88 (0.691)	11.50 (-)	13.50 (-)	-	-	11.81 (2.670)	12.07 (2.346)	9.48 (2.662)	12.48 (2.033)	13.18 (1.787)	10.80 (1.131)	13.01 (2.063)	13.00 (2.427)	-	13.51 (2.018)	11.71 (2.199)	12.3 (2.36)
Haemoglobin record (+/- 7 days of bleed), n (%)	12 (5.3)	3 (2.2)	1 (1.2)	0 (0)	0 (0)	237 (16.3)	2094 (22.4)	229 (21.6)	729 (16)	220 (13.9)	4 (7)	1689 (18.7)	49 (7)	0 (0)	14 (42.4)	155 (8.5)	5436 (18.1)
Haemoglobin (+/- 7 days of bleed), mean (SD)	13.23 (1.061)	12.23 (0.702)	13.50 (-)	-	-	11.73 (2.453)	12.24 (2.252)	9.99 (2.586)	12.46 (1.958)	12.88 (2.052)	11.10 (1.573)	12.91 (2.078)	12.79 (2.055)	-	13.31 (1.560)	12.10 (2.036)	12.4 (2.25)
Chronic haemoglobin																	
Haemoglobin record (in year prior bleed), n (%)	122 (54)	75 (54.7)	55 (66.3)	0 (0)	0 (0)	899 (61.8)	5833 (62.3)	690 (65)	2759 (60.7)	936 (59.3)	30 (52.6)	5466 (60.5)	383 (54.5)	0 (0)	23 (69.7)	1068 (58.6)	18339 (60.9)
Haemoglobin (in the year prior to bleed), mean (SD)	13.2 (1.57)	13.6 (1.80)	12.6 (2.08)	-	-	12.4 (2.03)	13.0 (1.88)	11.8 (2.40)	13.1 (1.81)	13.3 (1.75)	13.4 (1.64)	13.3 (1.86)	12.9 (1.66)	-	12.9 (1.46)	13.1 (1.67)	13.1 (1.89)
Haemoglobin record (in year post-bleed), n (%)	120 (53.1)	56 (40.9)	35 (42.2)	0 (0)	2 (40)	813 (55.9)	5749 (61.4)	690 (65)	2787 (61.3)	844 (53.5)	24 (42.1)	5181 (57.4)	390 (55.5)	1 (50)	24 (72.7)	1045 (57.3)	17761 (59.0)
Haemoglobin (in the year post-bleed), mean (SD)	13.1 (1.61)	13.9 (1.58)	13.1 (1.80)	-	13.2 (0.00)	12.2 (1.91)	12.7 (1.98)	11.8 (2.01)	12.8 (1.91)	13.1 (1.85)	11.8 (1.93)	12.9 (1.98)	12.9 (1.87)	13.9 (-)	13.0 (1.79)	12.8 (1.82)	12.8 (1.96)
Haemoglobin drop (g/dL)**																	
Haemoglobin drop, n (%)	9 (4)	2 (1.5)	1 (1.2)	0 (0)	0 (0)	164 (11.3)	1475 (15.8)	167 (15.7)	517 (11.4)	147 (9.3)	3 (5.3)	1075 (11.9)	36 (5.1)	0 (0)	13 (39.4)	111 (6.1)	3720 (12.3)
Haemoglobin drop, mean (SD)	0.44 (1.59)	0.00 (0.28)	1.10 (-)	-	-	1.32 (1.99)	1.15 (1.85)	2.35 (2.66)	0.86 (1.59)	0.72 (1.37)	2.37 (0.23)	0.85 (1.46)	0.19 (1.53)	-	0.38 (0.94)	1.46 (1.70)	1.06 (1.77)

IC= intracranial; GI= gastrointestinal; NOS= not otherwise specified; SD= standard deviation

*ruptured abdominal or thoracic aortic aneurysm or haemopericardium

Haemoglobin drop defined as **max haemoglobin in year prior to bleed minus **min** haemoglobin recorded within 7 days of bleed

Figure 5.8: Distribution of estimated haemoglobin drop around bleeding events in HES and CPRD

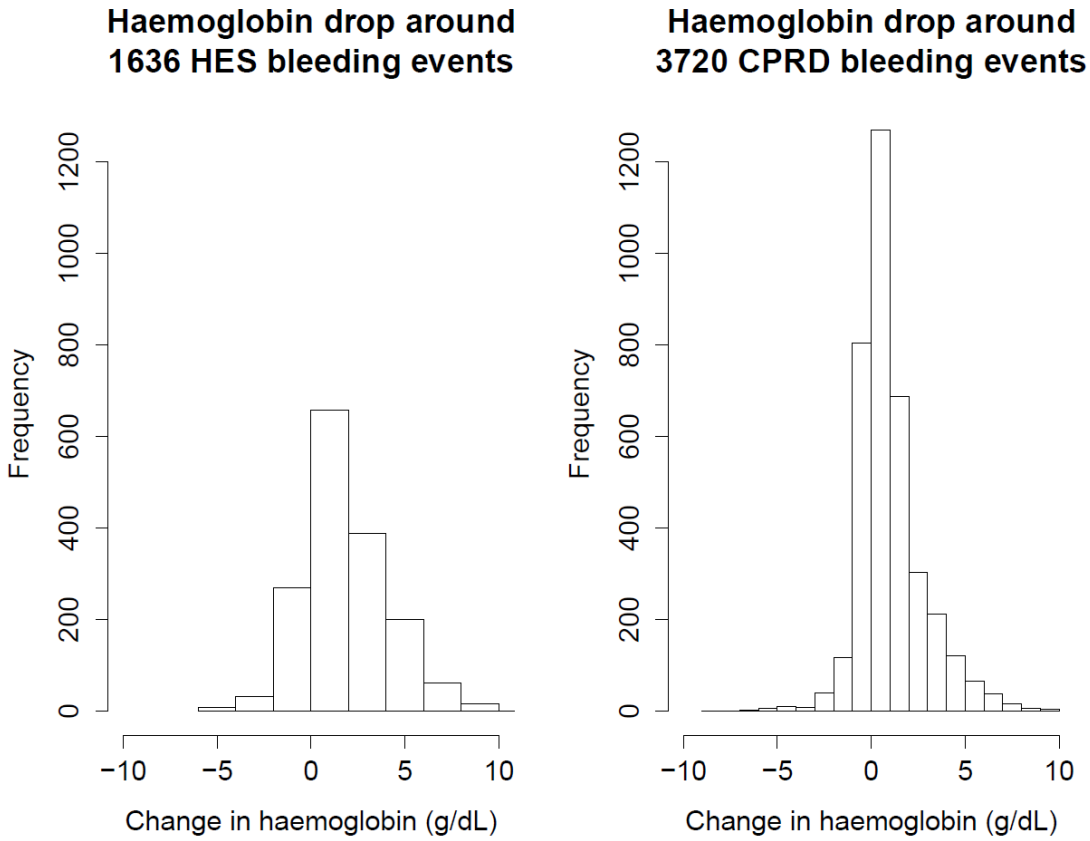
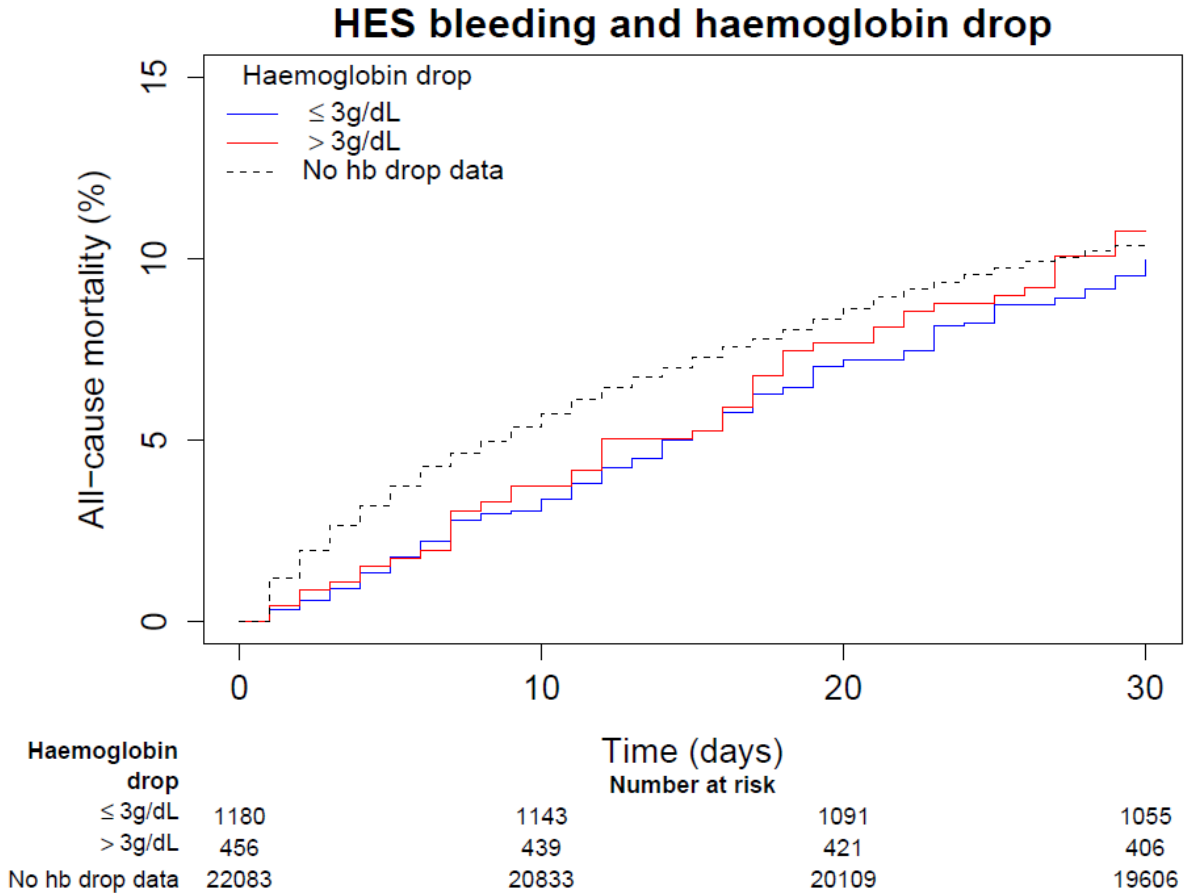
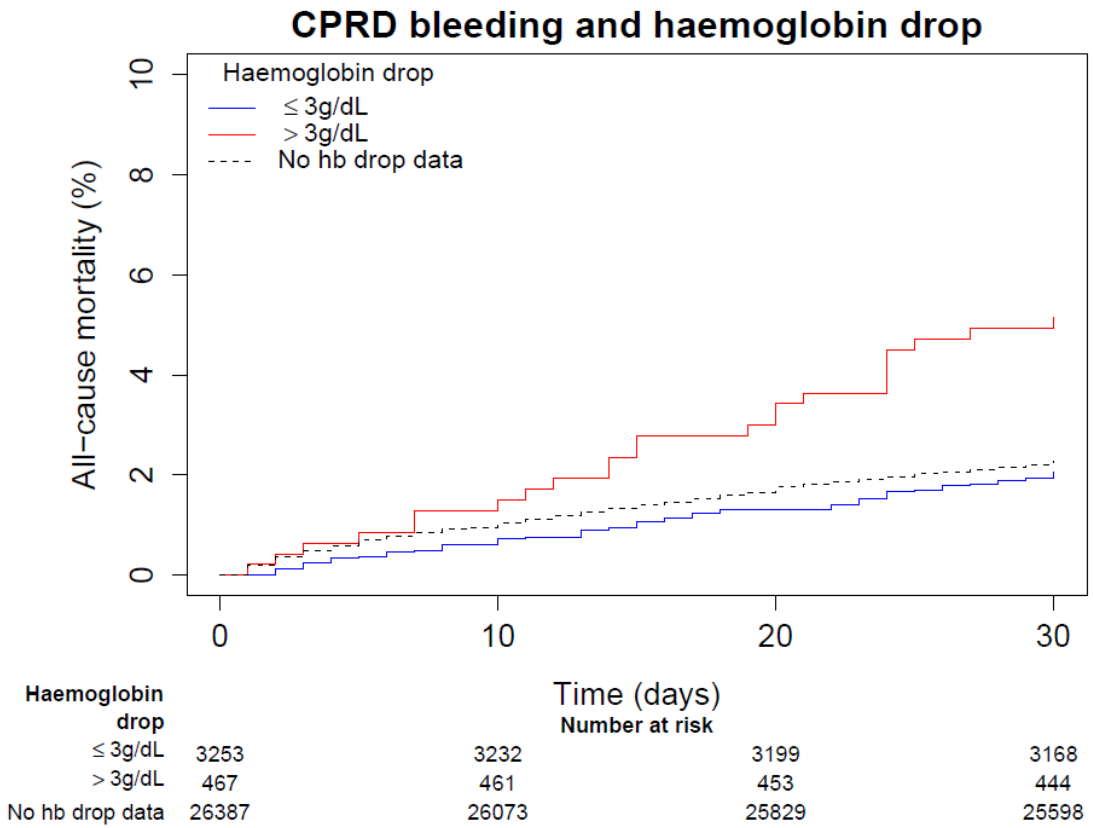


Figure 5.9: Association between haemoglobin drop and short term mortality following HES bleeding



hb= haemoglobin

Figure 5.10: Association between haemoglobin drop and short term mortality following CPRD bleeding



hb= haemoglobin

Table 5.10: Examinations and interventions and HES bleeding (n= 23719 bleeding events in 16087 patients)

	IC (intracerebral)	IC (subarachnoid)	IC (subdural)	IC (extradural)	IC (NOS)	Upper GI	Lower GI	GI (NOS)	Upper resp.	Lower resp.	Resp. (NOS)	Rupt. AA/ HP*	Genito-urinary	Ocular	Ear	Bleeding disorder	NOS	Overall
N bleeds:	715	201	442	22	55	3142	5806	1530	2207	1281	26	280	7378	263	5	227	139	23719
Endoscopy																		
On same day of bleed, n (%)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	294 (9.4)	1406 (24.2)	184 (12)	14 (0.6)	267 (20.8)	3 (11.5)	1 (0.4)	3170 (43)	1 (0.4)	0 (0)	5 (2.2)	5 (3.6)	5350 (22.6)
Within +/- 7 days of bleed, n (%)	8 (1.1)	1 (0.5)	9 (2)	0 (0)	1 (1.8)	839 (26.7)	2461 (42.4)	489 (32)	77 (3.5)	329 (25.7)	5 (19.2)	7 (2.5)	3611 (48.9)	3 (1.1)	0 (0)	18 (7.9)	9 (6.5)	7867 (33.2)
Within +/- 30 days of bleed, n (%)	18 (2.5)	1 (0.5)	14 (3.2)	2 (9.1)	4 (7.3)	1133 (36.1)	2966 (51.1)	619 (40.5)	122 (5.5)	408 (31.9)	7 (26.9)	14 (5)	4063 (55.1)	8 (3)	0 (0)	27 (11.9)	15 (10.8)	9421 (39.7)
Surgical arrest of bleeding																		
On same day of bleed, n (%)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0)	0 (0)	0 (0)	248 (11.2)	2 (0.2)	0 (0)	2 (0.7)	0 (0)	0 (0)	1 (20)	1 (0.4)	0 (0)	255 (1.1)
Within +/- 7 days of bleed, n (%)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	6 (0.2)	1 (0)	0 (0)	453 (20.5)	3 (0.2)	0 (0)	5 (1.8)	4 (0.1)	0 (0)	1 (20)	1 (0.4)	1 (0.7)	475 (2.0)
Within +/- 30 days of bleed, n (%)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	12 (0.4)	6 (0.1)	2 (0.1)	502 (22.7)	3 (0.2)	0 (0)	5 (1.8)	5 (0.1)	0 (0)	1 (20)	1 (0.4)	1 (0.7)	538 (2.3)
Haematoma evacuation/aspiration																		
On same day of bleed, n (%)	7 (1)	1 (0.5)	29 (6.6)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0.1)	0 (0)	1 (0.1)	0 (0)	0 (0)	1 (0)	0 (0)	0 (0)	0 (0)	1 (0.7)	41 (0.2)
Within +/- 7 days of bleed, n (%)	13 (1.8)	3 (1.5)	146 (33)	2 (9.1)	1 (1.8)	3 (0.1)	3 (0.1)	1 (0.1)	0 (0)	3 (0.2)	0 (0)	0 (0)	2 (0)	1 (0.4)	0 (0)	1 (0.4)	2 (1.4)	181 (0.8)
Within +/- 30 days of bleed, n (%)	18 (2.5)	5 (2.5)	181 (41)	3 (13.6)	1 (1.8)	7 (0.2)	6 (0.1)	1 (0.1)	0 (0)	3 (0.2)	0 (0)	1 (0.4)	4 (0.1)	1 (0.4)	0 (0)	2 (0.9)	2 (1.4)	235 (1.0)
Antifibrinolytics or haemostatic rx																		
On same day of bleed, n (%)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0.1)	0 (0)	0 (0)	0 (0)	0 (0)	4 (0.1)	0 (0)	0 (0)	0 (0)	0 (0)	5 (0.02)
Within +/- 7 days of bleed, n (%)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	11 (0.2)	6 (0.4)	6 (0.3)	2 (0.2)	0 (0)	0 (0)	27 (0.4)	0 (0)	0 (0)	0 (0)	0 (0)	52 (0.2)
Within +/- 30 days of bleed, n (%)	1 (0.1)	0 (0)	2 (0.5)	0 (0)	0 (0)	3 (0.1)	35 (0.6)	15 (1)	14 (0.6)	13 (1)	0 (0)	0 (0)	71 (1)	0 (0)	0 (0)	0 (0)	6 (4.3)	160 (0.7)
Vitamin K rx																		
On same day of bleed, n (%)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0.004)
Within +/- 7 days of bleed, n (%)	0 (0)	0 (0)	1 (0.2)	0 (0)	0 (0)	1 (0)	2 (0)	0 (0)	0 (0)	1 (0.1)	0 (0)	0 (0)	1 (0)	0 (0)	0 (0)	2 (0.9)	0 (0)	8 (0.03)
Within +/- 30 days of bleed, n (%)	0 (0)	0 (0)	1 (0.2)	0 (0)	0 (0)	3 (0.1)	4 (0.1)	1 (0.1)	1 (0)	2 (0.2)	0 (0)	0 (0)	3 (0)	0 (0)	0 (0)	2 (0.9)	0 (0)	17 (0.07)

IC= intracranial; GI= gastrointestinal; NOS= not otherwise specified; rx= prescription

*ruptured abdominal or thoracic aortic aneurysm or haemopericardium;

Figure 5.11: The association between bleeding requiring endoscopic examination and short term all-cause mortality

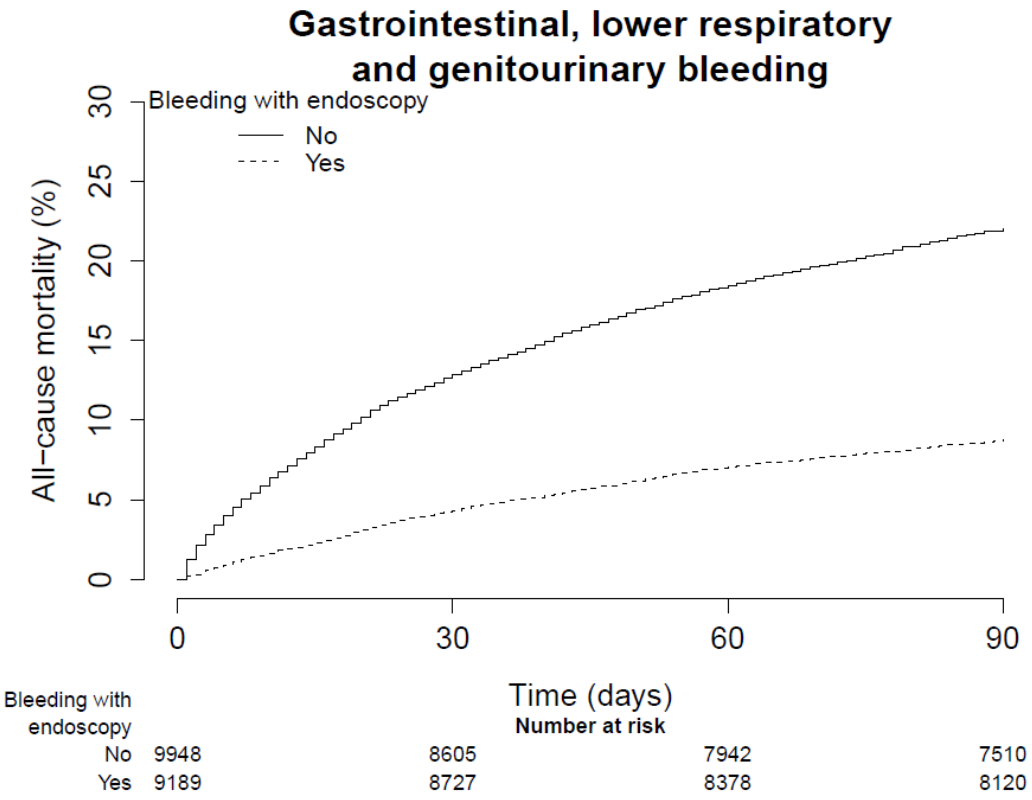


Figure 5.12: The association between bleeding requiring surgical arrest and short term all-cause mortality

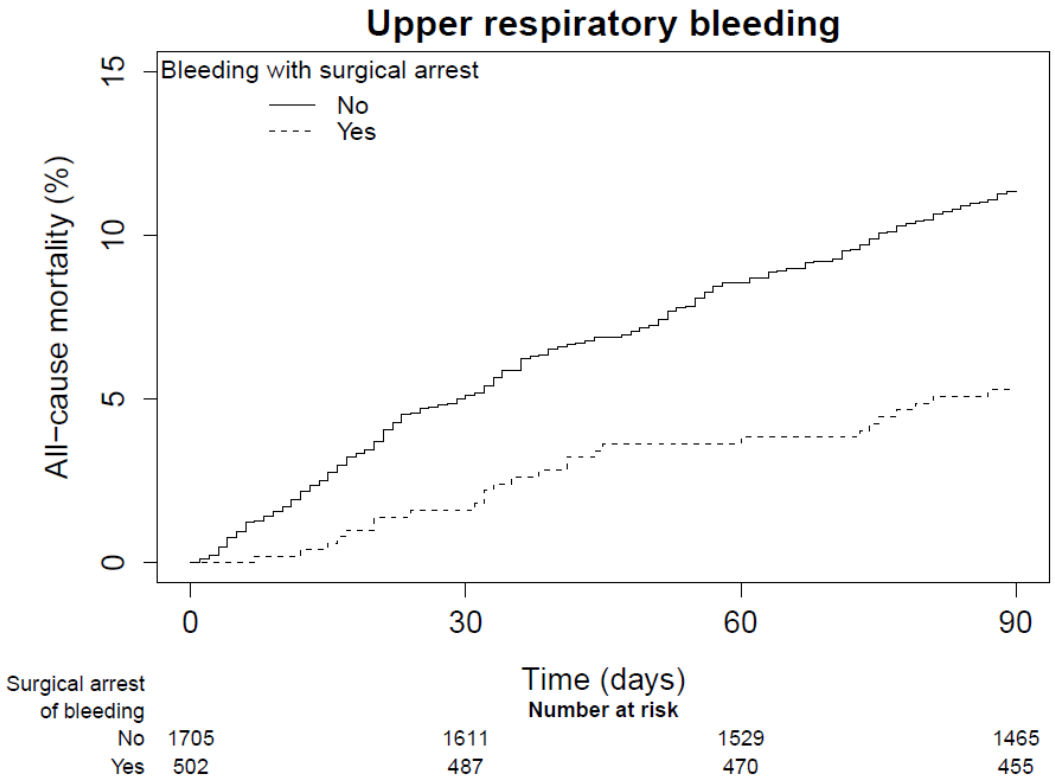


Figure 5.13: The association between bleeding requiring haematoma evacuation or aspiration and short term all-cause mortality

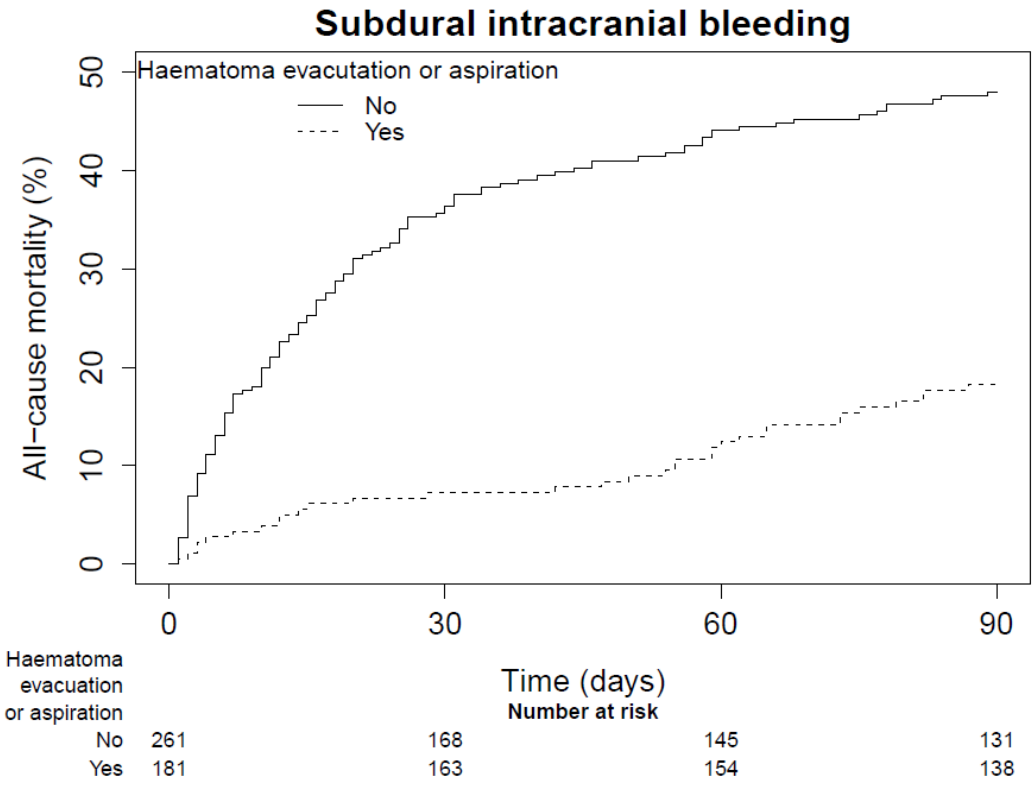
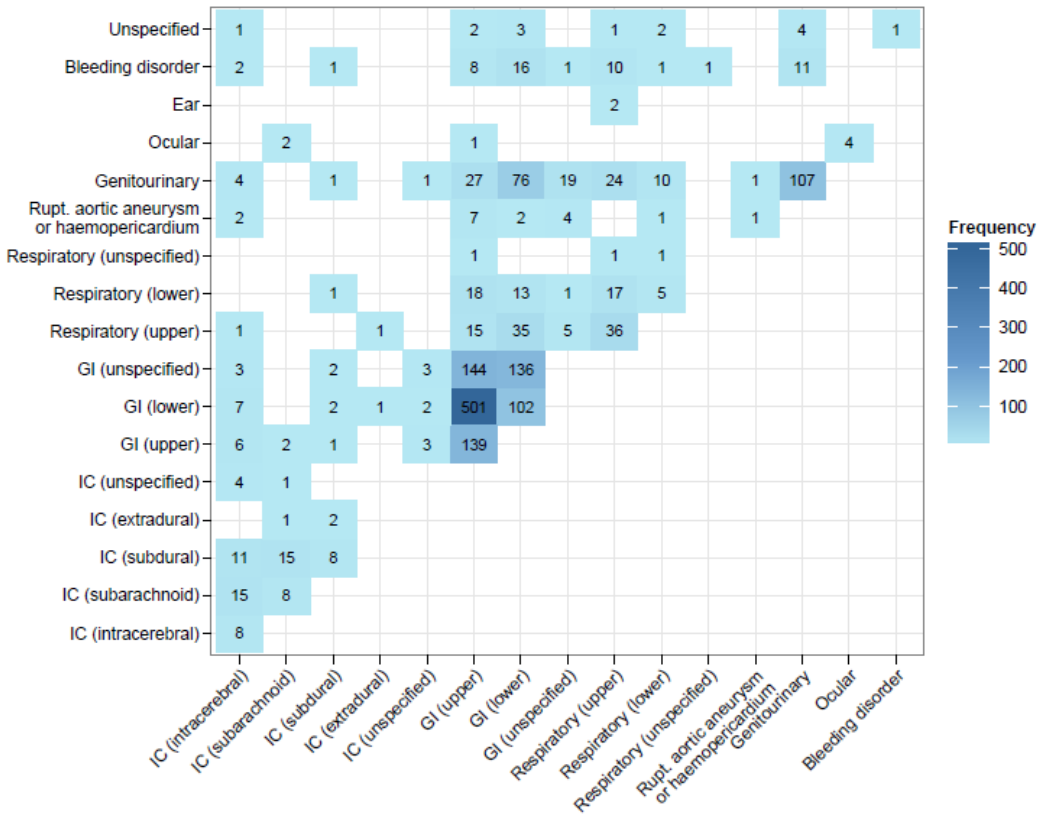
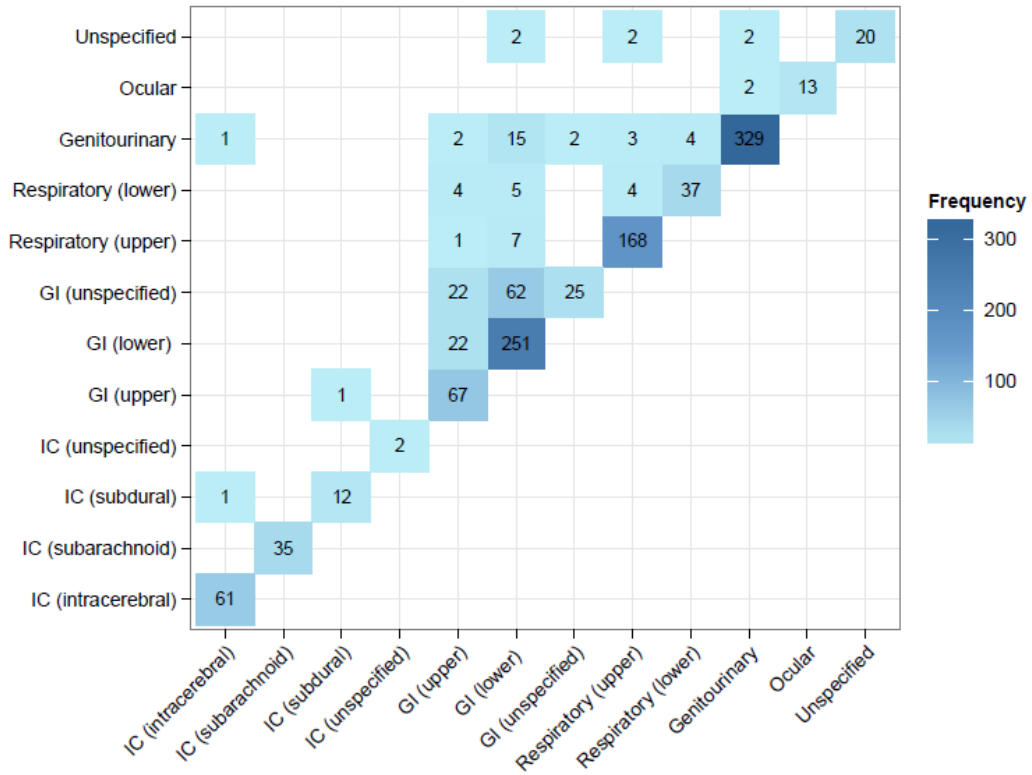


Figure 5.14: The anatomical site combinations for cases with multiple bleeding codes recorded in HES on a single date (1323 cases: 1192 with 2 codes, 113 with 3 codes, 18 with 4 codes recorded)



GI= gastrointestinal; IC= intracranial;

Figure 5.15: The distribution of anatomical site combinations for cases with multiple bleeding codes recorded in CPRD on a single date (1063 cases: 1007 with 2 codes, 53 with 3 codes, 3 with 4 codes recorded)



GI= gastrointestinal; IC= intracranial;

Figure 5.16: The association between number of ICD-10 bleeding codes recorded and 30 day mortality

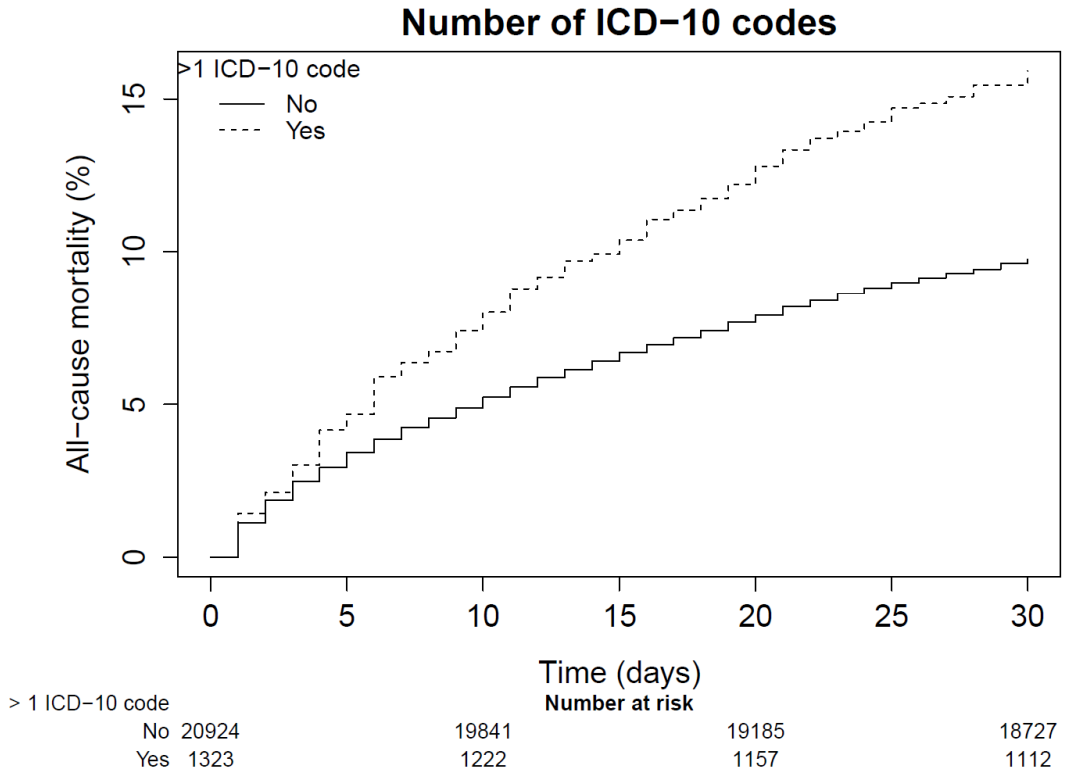


Figure 5.17: The association between number of Read bleeding codes recorded and 30 day all-cause mortality

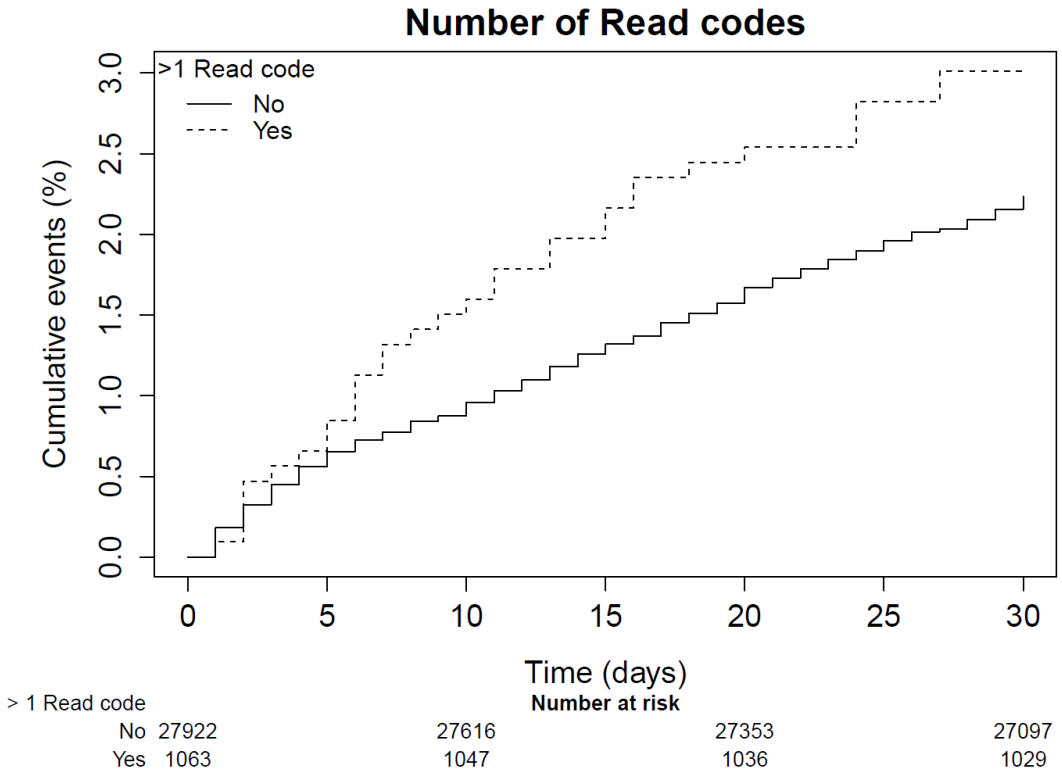


Figure 5.18: Phenotype algorithm showing the use of CPRD (primary care), HES (hospital admissions) and ONS (death registry) to define major and minor bleeding in primary care and hospital admissions and infer additional bleeding events

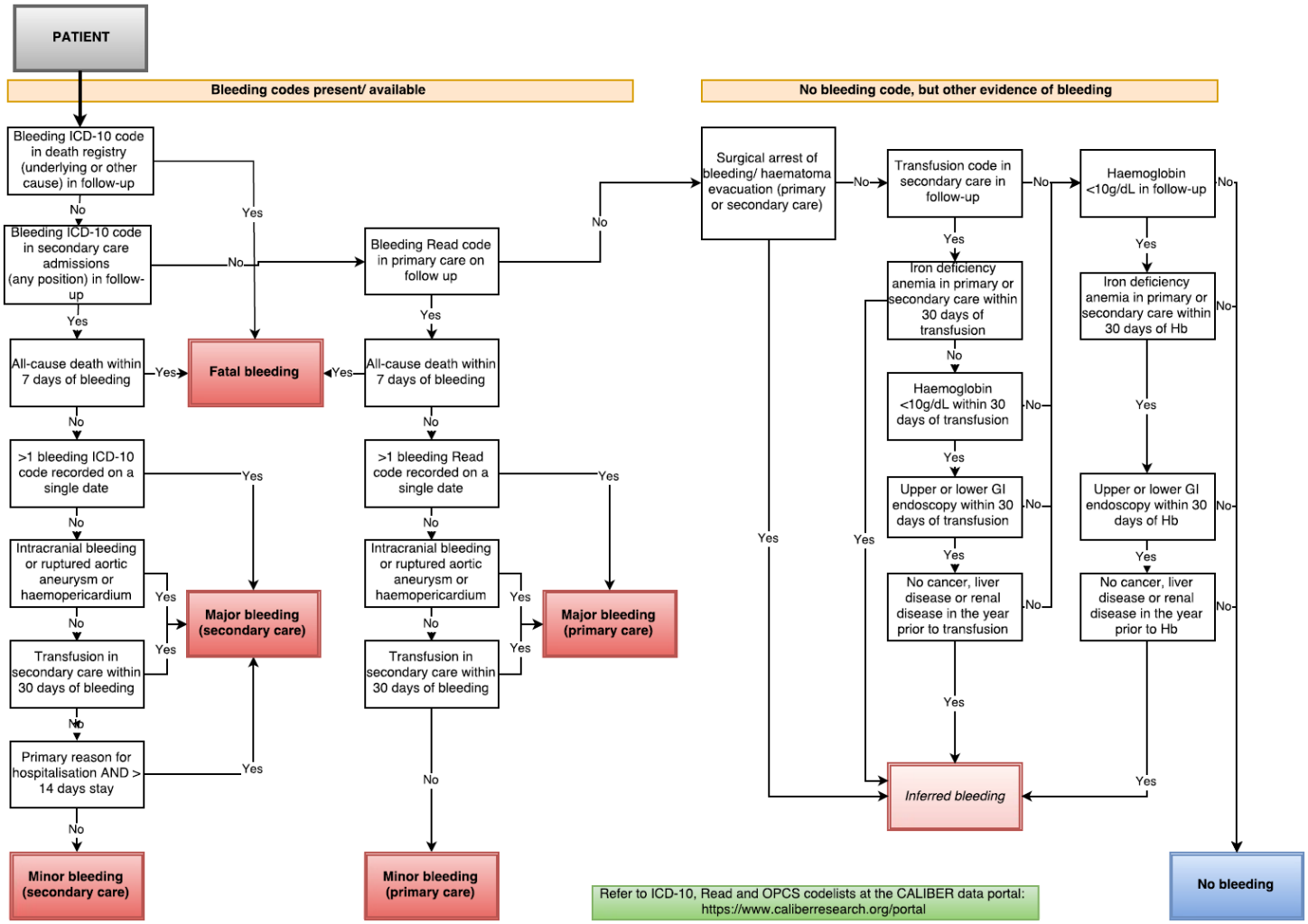
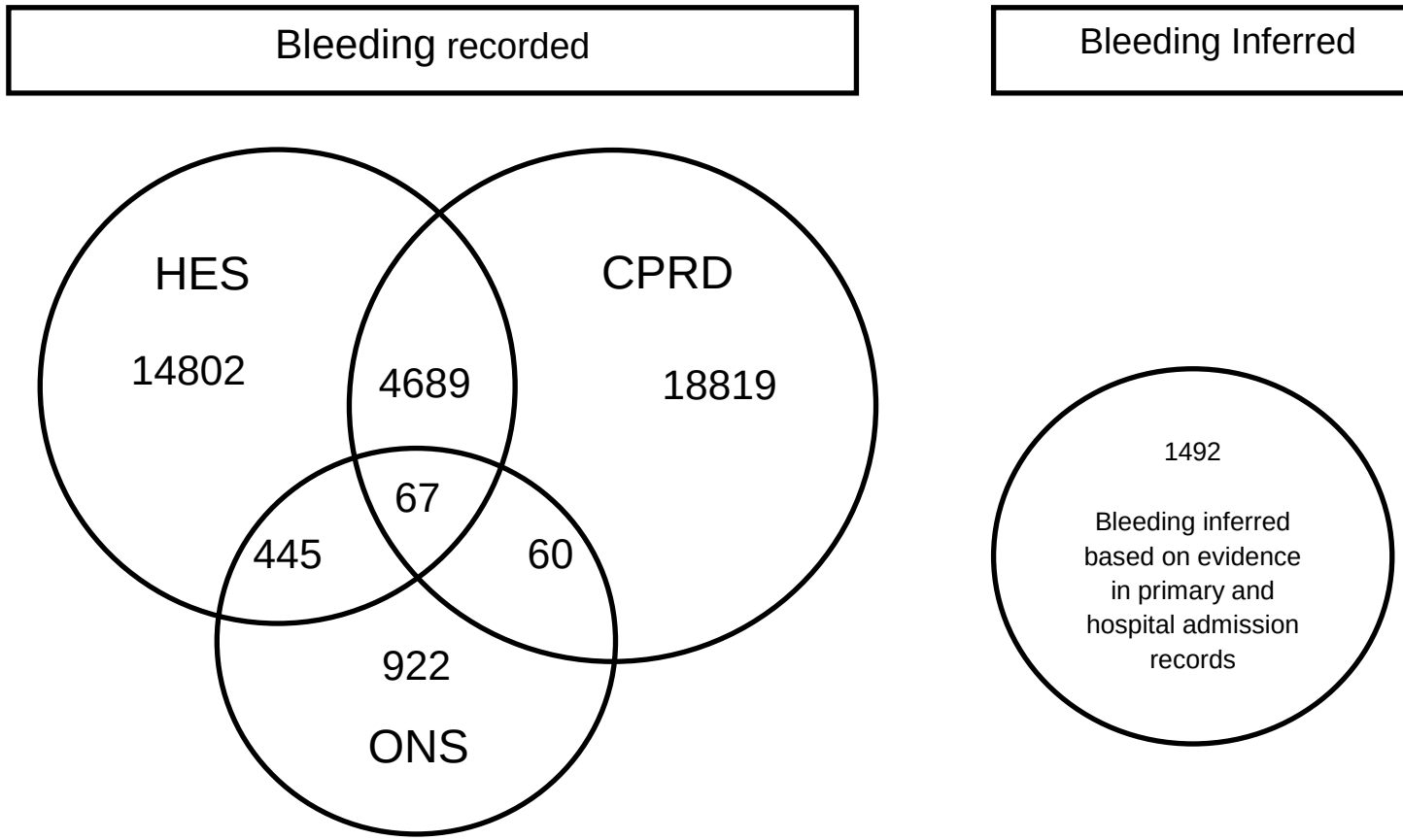


Figure 5.19: Overlap of 39,804 bleeding recorded in CPRD (primary care), HES (hospital care) and ONS (death registry) and the number of inferred bleeding cases in patients without a bleeding record in primary or hospital care (n= 128,815 patients)



6 Incidence and prognosis of bleeding events in four common cardiovascular diseases

Chapter Summary

Background

There are increasing numbers of patients being treated with antithrombotic therapies, new agents being approved and guidelines recommending lifetime treatment to reduce risk of subsequent atherothrombotic events. In the previous chapter I developed a bleeding phenotype algorithm. In this chapter my objectives were to determine the incidence of major and any bleeding in atrial fibrillation and coronary disease patients, to investigate the incidence of antithrombotic prescribing and bleeding over time and to investigate the effect of bleeding on long term prognosis.

Methods

In a study population of 128815 patients with atrial fibrillation or coronary disease in CALIBER from 1997-2010, 5 year incidence of bleeding events was assessed using Kapan-Meier plots. Cox proportional hazard models were used to estimate hazard ratios for the association between antithrombotic therapies and the risk of bleeding. Bleeding and antithrombotic prescribing prevalence was by counting the number of events and prescriptions per patient in CALIBER at monthly intervals between 1997 and 2010. Cox proportional hazard models were used to estimate hazard ratios for the associations between bleeding severity and the risk of all-cause mortality and atherothrombotic events.

Results

The overall 5 year risk of any bleeding was 24.2% (95% CI: 23.9, 24.5%) and fatal or bleeding with markers of severity was 6.5% (5.4, 6.7%). Bleeding risk was higher in atrial fibrillation patients compared with coronary disease patients. More intense antithrombotic regimens were associated with increased bleeding risk; patients prescribed triple therapy versus aspirin had more than 3 times the risk of any bleeding event. The rate of hospitalised bleeding increased over the duration of the study period. Bleeding recorded in both primary care and hospitalised settings were associated with increased risk of all-cause mortality and atherothrombotic events.

Conclusion

Bleeding is a major public health problem; it is common, increasing hospitalisations and associated with high mortality. With linked electronic health records I have developed and validated a comprehensive and reproducible bleeding phenotype with 3 distinct levels of severity that is can be used and further developed.

6.1 Background

In common cardiac disorders, including atrial fibrillation, acute coronary syndromes and stable coronary disease, antithrombotic drugs have been shown to be effective in reducing the risk of subsequent atherothrombotic events in randomised trials.[188-190] Yet, major and minor bleeding is the main safety concern of different antithrombotic drug classes in clinical trials. Clinical bleeding scores[7] have been incorporated into guidelines and bleeding prognostic models have been developed[186] to aid assessment of patients bleeding risk prior to making treatment decisions. However such tools are limited by miscalibration when applied to new populations.[193]

There are increasing numbers of people with such common heart diseases on one, two or three antithrombotic agents because of better implementation of longstanding trial evidence, (e.g. aspirin in the secondary prevention of non-myocardial infarction coronary disease), the introduction of new antithrombotic drugs (e.g. direct oral anticoagulants (DOACs), ticagrelor) and the recognition of the benefit of prolonging dual antiplatelet therapy post-MI (e.g. ticagrelor).[17] Indeed prolonged (lifelong) treatment is now recommended not only for anticoagulation in atrial fibrillation but also in dual antiplatelet therapy following acute myocardial infarction.[17,191,194]

For the bleeding risk associated with antithrombotic therapy, it remains unclear the modern population burden of bleeding in people with common cardiac disorders. While previous studies reported bleeding risk of atrial fibrillation[81], acute coronary syndromes[195], and stable coronary disease[66], there are no studies which compare across common cardiac disorders; and the bleeding definitions used within single disease studies differ. While small regional studies of intracerebral haemorrhage suggest increased incidence[84,196] there has been a lack of understanding of time trend in bleeding among other cardiac disorders.

In the previous chapter I developed a bleeding phenotype using linked electronic heath records (CALIBER). The bleeding phenotype was developed with a view to encourage standardised bleeding definitions within electronic health record studies with bleeding endpoints and enable monitoring patients at risk of bleeding (e.g. those treated with

antithrombotic therapies). The assessment of the incidence and prognosis of bleeding in the study population can contribute towards demonstrating the validity of the phenotypes whilst estimating the burden of bleeding in populations indicated for antithrombotic therapies in England.

There has been a lack of research in large scale population based samples of patients with common cardiac disorders evaluating the incidence of bleeding, time trends and prognosis. I sought to address the following questions:

First, what is the long term incidence of major bleeding and any bleeding across patients with atrial fibrillation, acute myocardial infarction, unstable and stable angina who are on different antiplatelet and anticoagulation regimens? There are no studies which compare across common cardiovascular disorders; and bleeding definitions differ.

Second to what extent has the incidence of bleeding increased over time with the changes in antithrombotic management? Small regional studies of intracerebral haemorrhage suggest an increase incidence.

Third how does major, minor bleeding impact on long term prognosis in terms of all-cause mortality, a composite of cardiovascular mortality, MI and stroke and recurrent bleeding?

6.2 Methods

6.2.1 Study population

The study population consisted of patients with either coronary disease or atrial fibrillation, i.e. those who were potential candidates for antithrombotic therapy, in CALIBER during 1997-2010. To define this population I used pre-existing validated disease phenotypes in CALIBER. Patients were eligible if they were aged 18 years and above and entered the cohort at their first diagnosis of atrial fibrillation, acute myocardial infarction, unstable angina or stable angina. If patients were diagnosed with more than one of the aforementioned diseases on their index date they were placed into a group using the ranking 1) atrial fibrillation, 2) myocardial infarction, 3) unstable angina, and 4) stable angina. Patients were followed up until death, transfer out of their primary care practice (i.e. loss to follow-up), or the date of administrative censoring (March 2010).

I analysed baseline characteristics of patients stratified baseline bleeding risk. Patients bleeding risk was calculated using a previously validated prognostic model which uses baseline characteristics to determine long term risks[186]. Multiple imputation[197] was used to impute missing systolic blood pressure, creatinine and smoking status data that was required in order

to apply the prognostic model and calculate individual predictions. Ten imputed data sets were generated. The prognostic model was applied to each imputed data set and the calculated linear predictors were averaged for each patient. Using the 16%, 50% and 84% cut points from the prognostic model development,[186] patients were grouped as low, medium low, medium high and high bleeding risk.

Using prescribing data I summarised duration (median and interquartile range days) of antithrombotic use between cohort entry and first bleeding event. To calculate duration, a patient's prescription was assumed to be continuous if issued within 90 days of the previous one (90 days is the longest allowed duration of prescriptions in the UK). Antithrombotic therapies were grouped as aspirin monotherapy, adenosine diphosphate (ADP) receptor inhibitor monotherapy, dual antiplatelet therapy (aspirin and ADP receptor inhibitor), vitamin K antagonist (VKA) monotherapy, VKA and one antiplatelet (aspirin or ADP receptor inhibitor) and triple therapy (VKA, aspirin and ADP receptor inhibitor).

6.2.2 Classification of bleeding events

Bleeding events were defined using the algorithm defined in the previous chapter. Non-fatal bleeding events were classified as hospitalised or primary care with markers of severity (referred to as hospitalised+ and primary care+), and hospitalised or primary care without markers of severity (referred to as hospitalised and primary care). For patients with no bleeding code in either primary care or hospital records, bleeding events may be inferred where there are records that provide evidence suggesting bleeding, for example transfusions and low haemoglobin.

6.2.3 Bleeding incidence

The incidence of any bleeding and fatal, hospitalised+ and primary care+ bleeding events was assessed using Kaplan-Meier plots stratified by cardiovascular disease group (atrial fibrillation, acute myocardial infarction, unstable angina and stable angina) and by bleeding risk group (high, medium high, medium low and low). Follow-up was capped at 5 years.

6.2.4 The association between antithrombotic therapy prescribing and the risk of bleeding

Cox proportional hazard models were used to estimate hazard ratios for the association between antithrombotic therapies and first bleeding event of any severity and fatal or bleeding+ event. Antithrombotic therapy prescriptions were included in the models as a time-dependent variable due to the dynamic nature of such treatment within individuals. Possible states were no antithrombotic therapy (reference group), aspirin, ADP receptor inhibitor, dual antiplatelet therapy, VKA monotherapy, VKA and one antiplatelet (aspirin or ADP receptor inhibitor) and triple therapy. Patients were followed up until their first bleeding event of any

severity and until their first fatal or bleeding+ event. Models were also adjusted for age, sex, bleeding risk group, and initial cardiovascular disease.

6.2.5 Time trends in bleeding and antithrombotic therapy prescribing

I estimated the change in prevalence of primary care, hospitalised and fatal bleeding within the study population at monthly intervals between 1997 and 2010. To do this I calculated the number of bleeding events in primary care, secondary care and death registry per active patients in each month and fitted local regression (LOESS) smoothed lines to assess the trends. I also investigated the trends in antithrombotic therapy prescribing (aspirin, ADP receptor inhibitor and VKAs) using the same methods.

Bleeding and antithrombotic therapy prescribing rates were also assessed in the initial cardiovascular disease subgroups: atrial fibrillation, myocardial infarction, unstable angina and stable angina.

6.2.6 Prognosis following non-fatal bleeding events

Cox proportional hazard models were used to estimate hazard ratios for the association between first bleeding events and all-cause mortality and atherothrombotic events (composite of cardiovascular death, ischaemic or unspecified stroke, or myocardial infarction). Bleeding severity (hospitalised +, primary care +, hospitalised, primary care and inferred) was treated as a time-dependent variable in the models to prevent immortal time bias. The bleeding variable could have the states no bleeding, primary care, primary care+, hospitalised or hospitalised+. Models were also adjusted for age, sex and baseline disease history (diabetes, stroke, peripheral arterial disease, cancer, renal disease, peptic ulcer, bleeding diatheses, chronic anaemia). The risk of recurrent bleeding events (at least 30 days following initial bleeding event to ensure the recurrent bleeding is a new event and not a continuation of the initial bleeding) following non-fatal bleeding events of any severity and primary care+ or hospitalised+ bleeding was examined using Kaplan-Meier plots.

6.3 Results

6.3.1 Study Population baseline characteristics

The baseline characteristics of the 128815 patients in the study population, stratified by bleeding risk group are presented in **Table 6.1.** By nature of the risk prediction model, patients were on average older in the higher risk groups. Large proportions of the high risk patients had comorbidities such as atrial fibrillation (31.9% vs. 12.8% low risk patients), ischaemic or unspecified stroke (18.9% vs. 0.4%), type 2 diabetes (22.1% vs. 3.5%), cancer (36.4% vs 1.4%) and chronic anaemia (34.2% vs. 4.7%). Patients in the low risk group had a higher average BMI.

14.9% of patients in the low risk group were classified as obese compared with 6.7% of the patients in high risk group. Average systolic blood pressure and creatinine was higher and haemoglobin was lower in the higher risk groups.

The percentage of patients prescribed antithrombotic therapy during follow-up was lowest in the high bleeding risk group (65.9%) compared with the medium high risk group (78.4%), the medium low risk group (85.5%) and the low risk group (72.0%). Aspirin monotherapy was less common in the high risk group compared with the other groups. The percentage of patients prescribed dual antiplatelet therapy increased across risk groups (6.6% in the low risk group vs. 14.6% in the high risk group). The median duration of dual antiplatelet therapy was similar across the risk groups, approximately a year, which aligns with guidelines that were in place during the study period. VKA monotherapy prescribing was highest in the medium low (21.3%) and low (18.4%) risk groups. Prescribing VKAs with 1 or 2 antiplatelets concurrently was most common in the medium high and medium low risk groups.

6.3.2 Long term risk of bleeding

At 5 years the overall risk of any bleeding event in the study population was 24.2% (95% CI: 23.9, 24.5%) and the 5 year risk of fatal, hospitalised+ or primary care+ bleeding events was 6.5% (95% CI: 6.4, 6.7%).

The Kaplan-Meier curves of CALIBER bleeding events, stratified by initial cardiovascular disease are shown in **Figure 6.1.** At 5 years 29.1% (95% CI: 28.2, 29.9%) of atrial fibrillation patients, 21.9% (95% CI: 21.2, 22.5%) of myocardial infarction patients, 25.3% (95% CI: 24.2, 26.3%) of unstable angina patients and 23.4% (95% CI: 23.0, 23.8%) of stable angina had bleeding of any kind. At 5 years fatal, hospital+ or primary care+ bleeding risks were 9.9% (95% CI: 9.3, 10.4%) for atrial fibrillation patients, 6.1% (95% CI: 5.8, 6.5%) for myocardial infarction patients, 6.8% (95% CI: 6.0, 7.2%) for unstable angina patients and 5.7% (95% CI: 5.5, 5.9%) for stable angina.

At 5 years any bleeding risks were 16.8% (95% CI: 16.0, 17.5%), 19.5% (95% CI: 19.0, 19.9%), 27.2% (95% CI: 26.7, 27.7%) and 37.1% (95% CI: 36.0, 18.1%) for patients in the low, medium, medium high and high bleeding risk groups respectively. Fatal, hospital+ or primary care+ bleeding event risks were 2.0% (95% CI: 1.8, 2.3%), 3.6% (95% CI: 3.4, 3.8%), 8.2%, (95% CI: 7.9, 8.5%) and 15.1% (95% CI: 14.3, 15.9%) for patients in the low, medium, medium high and high bleeding risk groups respectively.

The increasing incidence of bleeding across bleeding risk groups was consistent within each cardiovascular disease group **Table 6.2.** The 5 year risk of fatal, hospitalised+ or primary care+

bleeding in patients with atrial fibrillation and classified as high bleeding risk was 17.5% (15.6, 19.3%), and for patients with stable angina classified as low bleeding risk 1.85% (1.00, 2.70%).

6.3.3 The association between antithrombotic prescriptions and bleeding

Patients prescribed with more aggressive antithrombotic therapies (dual antiplatelet therapy, vitamin K antagonists, and triple therapy) had significantly higher risk of bleeding events compared with to those not prescribed antithrombotic therapies **(Figure 6.2)**. Patients who were prescribed triple therapy had 3.3 (95% CI: 2.6, 4.4) times increased risk of any bleeding and 5.2 (95% CI: 3.4, 9.0) times increased risk of fatal, hospitalised+ or primary care+ bleeding events compared with those not prescribed antithrombotic therapies.

6.3.4 Time trends in bleeding incidence and antithrombotic prescribing

In **Figure 6.3** estimated time trends in bleeding incidence and antithrombotic prescribing are shown. The estimated number of hospitalised+ bleeding events per 1000 active patients increased from 0.32 (95% CI: 0.24, 0.40) in January 1998 to 0.54 (95% CI: 0.45, 0.62) in December 2009. Primary care+ bleeding events per 1000 active patients decreased from 0.80 (95% CI: 0.70, 0.91) in January 1998 to 0.34 (95% CI: 0.23, 0.45) in December 2009, and fatal bleeding remained steady. There were increases in hospitalised and primary care bleeding events without markers of severity. The estimated number of hospitalised bleeding events per 1000 active patients increased from 1.02 (95% CI: 0.83, 1.22) in January 1998 to 2.68 (95% CI: 2.49, 2.88) in December 2009. The estimated number of primary care bleeding events per 1000 active patients increased from 1.70 (95% CI: 1.44, 1.95) in January 1998 to 3.31 (95% CI: 3.06, 3.57) in December 2009. Rates of prescribed antithrombotic therapies also rose over the study period. The number of aspirin prescriptions issued per 1000 active patients rose from 147.9 (95% CI: 127.4, 168.3) in January 1998 to 465.1 (95% CI: 444.6, 485.6) in December 2009. ADP receptor inhibitor prescriptions per 1000 active patients rose from 2.8 (95% CI: 0.2, 5.4) in January 1998 to 94.8 (95% CI: 92.2, 97.4) in December 2009. VKA prescriptions per 1000 active patients rose from 22.7 (95% CI: 19.2, 26.1) in January 1998 to 83.7 (95% CI: 80.2, 87.1) in December 2009.

The rates of fatal, hospitalised+ and primary care+ stratified by cardiovascular disease are displayed in **Supplementary appendix 11.3.1.** The increasing rates of hospitalised and primary care bleeding events appeared to be similar amongst the cardiovascular diseases **(Supplementary appendix 11.3.2).**

Antithrombotic therapy prescribing rates differed amongst the cardiovascular diseases, in particular between atrial fibrillation and the coronary diseases **(Supplementary appendix 11.3.3).** The rates of prescribing of each drug class increased in all four diseases, and the

prescribing of aspirin was consistently higher than those of vitamin K antagonists and ADP receptor inhibitors. However the prescribing of vitamin K antagonists was highest for atrial fibrillation patients and the prescribing of ADP receptor inhibitors was highest for myocardial infarction, unstable angina and stable angina patients.

6.3.5 Death, atherothrombotic events and recurrent bleeding following first bleeding event

Patients were at increased risk of all-cause mortality and cardiovascular death, stroke or myocardial infarction following their first bleeding event and this association was observed across all bleeding severities **(Figure 6.4**). Compared with patients with no bleeding, risk of all-cause mortality was highest in patients following hospitalised+ bleeding (adjusted HR: 2.97; 95% CI: 2.84, 3.12). Similarly, risk of cardiovascular death stroke or MI events increased the most following hospitalised+ bleeding (adjusted HR: 2.55; 95% CI: 2.38, 2.74). Following primary care bleeding, patients were also at significantly increased risk of all-cause mortality (adjusted HR: 1.23; 95% CI: 1.19, 1.27) and cardiovascular death, stroke or MI events (adjusted HR: 1.08; 95% CI: 1.04, 1.13) although to a lesser extent.

The estimates of prognosis were largely consistent between atrial fibrillation **(Supplementary appendix 11.3.4)**, myocardial infarction **(Supplementary appendix 11.3.5)**, unstable angina **(Supplementary appendix 11.3.6)** and stable angina **(Supplementary appendix 11.3.7)**. All bleeding events were associated with increased risk of all-cause mortality within each of the cardiovascular diseases. For atrial fibrillation patients the risk of cardiovascular death, stroke or MI was not significantly increased following primary care bleeding events (adjusted HR: 1.03; 95% CI: 0.93, 1.15)**.** Similarly, unstable angina patients were not at increased risk of cardiovascular death, stroke or MI following primary care (adjusted HR: 0.98; 95% CI: 0.84, 1.14) and inferred bleeding events (adjusted HR: 1.41; 95% CI: 0.93, 2.13).

Following an initial bleeding event patients were at high risk of further bleeding in the long term **(Figure 6.5)**. The risks were greater if the initial bleeding event was major. Patients who had a bleeding event of any kind were at 32.4% (95% CI: 31.8, 33.0) risk of 5 year any recurrent bleeding and 8.3% (95% CI: 7.9, 8.6) risk of 5 year fatal or major bleeding. Patients who experienced major bleeding had 37.4% (95% CI: 36.0, 38.8) 5 year risk of any further bleeding and 23.1% (95% CI: 21.9, 24.3) 5 year risk of fatal or major bleeding.

6.3.6 Classification of bleeding severity

Through the assessment of prognosis following bleeding **(Figure 6.4)** I identified 3 distinct levels of severity amongst the bleeding types defined in the phenotype. I therefore classify non-fatal bleeding severity as:

- **I:** Hospitalised + bleeding
- **II:** Hospitalised bleeding; primary care + bleeding; inferred bleeding
- **III:** Primary care bleeding

6.4 Discussion

6.4.1 Bleeding incidence in cardiovascular disease populations

In this chapter I demonstrated a direct comparison of bleeding within four common cardiovascular diseases with varying degrees of antithrombotic use. Atrial fibrillation had highest bleeding 5-year rates both for any bleeding (29.1%) and for fatal or major bleeding (9.9%). This may be indicative of higher prevalence and longer durations of prescribed aggressive ATT treatment (VKA, dual, triple therapy) in atrial fibrillation patients**.** However the incidence of bleeding in myocardial infarction, unstable angina and stable angina patients was still relatively high.

6.4.2 Time trends in bleeding rates over the study period (1997-2010)

So far as I am aware there have been no previous studies which have evaluated the time trends in bleeding incidence. It was hypothesised that the increased use of antithrombotic therapies during this period would be associated with an increased incidence of bleeding. Increases in major and minor hospitalised bleeding and minor primary care bleeding events over time were identified. It is unknown whether the increase in bleeding incidence is directly attributable to increasing range of available antithrombotic therapies and widening indications for their use over time, but there was an increase in antithrombotic prescribing in the coinciding timeframe. These bleeding events are associated with poor outcomes indicating an increasing burden bleeding may have on healthcare systems and costs in England.

6.4.3 Prognosis following bleeding

The analysis of prognosis following a non-fatal bleeding event identified 3 distinct levels of severity; I: Major bleeding requiring hospitalisation, II: Minor bleeding requiring hospitalisation, major bleeding in primary care, or inferred bleeding and III: Minor bleeding in primary care**.** This contrasts with the usual dichotomised major/minor bleeding used in outcome studies. Increased bleeding severity was strongly associated with increased risks of all-cause mortality and atherothrombotic events. 'Type III' bleeding though the least severe, significantly increased risk of death suggesting all types of bleeding captured by the phenotype are clinically important. The association between bleeding and subsequent atherothrombotic events could be due to cessation of ATT following a bleeding event and thus exposing patients to increased risk of atherothrombotic events. Patients at highest risk of atherothrombotic

events may be treated more aggressively with ATT and therefore at a high risk of bleeding. Furthermore there are a number of common risk factors, such as age and comorbidities, for both the propensity to bleeding and the progression from one cardiac to disease to fatal or non-fatal atherothrombotic events.

6.4.4 Phenotype validity

The gold standard method for validating phenotypes in electronic health records is to compare manually with patient charts. Such information is not available within the CALIBER platform. However the analyses performed in this study demonstrate the characteristics of the bleeding events defined by the phenotype and previous studies have demonstrated the good validity of many of the bleeding codes used in the present study.[74,77]

6.4.5 Clinical implications

This study provides an analysis of bleeding incidence, risks and prognosis from a public health perspective. The rise of bleeding and adverse prognosis following bleeding of any kind suggests an iatrogenic epidemic. A potential action to stem this problem is more stringent checking of bleeding risks prior to prescribing antithrombotic therapies. For example risk scores for cardiovascular and bleeding events integrated into general practice systems can help determine suitability for antithrombotic therapy and the phenotype can be used for continual monitoring and automated flagging-up of bleeding events. With growing ability replicating trial populations within real world data in linked electronic health records to estimate real-world impact of interventions phenotypes such as ours allow us to closely match the endpoints used in trials.

6.4.6 Future research

The phenotype should be validated in further electronic health record populations to ensure generalisability. The phenotype should also be applied to more recent populations including patients prescribed newer antithrombotic agents such as novel oral anticoagulants and ticagrelor. It is not known if the wider range of available antithrombotic agents has had an impact on incidence and severity of bleeding events.

6.4.7 Conclusion

Bleeding is a major public health problem; it is common, increasing hospitalisations and associated with high mortality. With linked electronic health records I have developed and validated a comprehensive and reproducible bleeding phenotype with 3 distinct levels of severity that is can be used and further developed for use in electronic health record studies of bleeding outcomes or antithrombotic safety.

6.5 Tables and Figures

Table 6.1: Baseline characteristics stratified by bleeding risk group

		High (n= 21101)	Medium high (n=53362)	Medium low (n= 40950)	Low (n=13402)
Demographics and behaviours at cohort entry					
Age (years), mean (SD)		83.8 (7.95)	77.2 (8.17)	64.9 (8.12)	49.6 (8.85)
Women, n (%)		9184 (43.5)	25697 (48.2)	17892 (43.7)	6233 (46.5)
Highest quintile of deprivation, n (%)		4059 (19.3)	10422 (19.6)	8109 (19.9)	3089 (23.1)
	Missing %	0.3	0.3	0.3	0.3
Smoking status, n (%)	Smoker	2207 (10.5)	4716 (8.8)	4697 (11.5)	1635 (12.2)
	Ex-Smoker	5760 (27.3)	14842 (27.8)	12204 (29.8)	3457 (25.8)
	Non-Smoker	9524 (45.1)	23379 (43.8)	15580 (38.0)	4922 (36.7)
	Missing %	17.1	19.5	20.7	25.3
History of alcohol abuse, n (%)		2928 (13.9)	5197 (9.7)	3727 (9.1)	572 (4.3)
Medical history prior to cohort entry [a]					
Atrial fibrillation, n (%)		6622 (31.4)	13007 (24.4)	5711 (13.9)	1721 (12.8)
Myocardial infarction, n (%)		4938 (23.4)	9970 (18.7)	7790 (19)	2333 (17.4)
Unstable angina, n (%)		1236 (5.9)	3536 (6.6)	3311 (8.1)	1417 (10.6)
Stable angina, n (%)		8305 (39.4)	26849 (50.3)	24138 (58.9)	7931 (59.2)
Ischaemic or unspecified stroke, n (%)		3988 (18.9)	3073 (5.8)	514 (1.3)	49 (0.4)
Peripheral arterial disease, n (%)		5048 (23.9)	4983 (9.3)	1348 (3.3)	108 (0.8)
Diabetes, n (%)	Type 2	4671 (22.1)	7148 (13.4)	3273 (8)	475 (3.5)
	Type 1	58 (0.3)	263 (0.5)	430 (1.1)	387 (2.9)
	Unspecified	533 (2.5)	1119 (2.1)	706 (1.7)	148 (1.1)
Renal disease, n (%)		3862 (18.3)	3885 (7.3)	1314 (3.2)	285 (2.1)
Non-metastatic cancer, n (%)		7691 (36.4)	8506 (15.9)	2062 (5)	182 (1.4)
Metastatic cancer, n (%)		984 (4.7)	308 (0.6)	36 (0.1)	1 (0)
Peptic ulcer, n (%)		2174 (10.3)	4131 (7.7)	2526 (6.2)	523 (3.9)
Bleeding diatheses and coagulation disorders, n (%)		662 (3.1)	330 (0.6)	87 (0.2)	19 (0.1)
Chronic anaemia, n (%)		7217 (34.2)	7088 (13.3)	2183 (5.3)	625 (4.7)
Biomarkers at cohort entry [b]					
BMI, Mean (SD)		26.3 (4.93)	27.5 (5.37)	29.0 (5.74)	30.7 (6.78)
Underweight, n (%)		264 (1.3)	539 (1.0)	221 (0.5)	35 (0.2)
Normal, n (%)		2708 (12.8)	5465 (10.2)	3091 (7.5)	733 (5.5)
Overweight, n (%)		2637 (12.5)	7132 (13.4)	5533 (13.5)	1379 (10.3)
Obese, n (%)		1413 (6.7)	4921 (9.2)	5259 (12.8)	2002 (14.9)
	Missing %	66.7	66.2	65.6	69
SBP (mmHg), mean (SD)		146 (26.2)	143 (21.1)	139 (17.7)	133 (15.8)
	Missing %	23.8	24.5	26	34.1

	High (n= 21101)	Medium high (n=53362)	Medium low (n= 40950)	Low (n=13402)
Haemoglobin (g/dL), mean (SD)	12.4 (1.98)	13.3 (1.75)	14.0 (1.51)	14.1 (1.49)
Missing %	50.2	59.2	63.7	67.8
Creatinine (mmol/l), mean (SD)	139.8 (91.1)	101.9 (34.8)	89.4 (21.8)	81.4 (19.6)
Missing %	42	49.6	54.5	65.1
Antithrombotic therapies (n, %) and duration (median, IQR) during follow-up[c]				
Any antithrombotic therapy, n (%)	13898 (65.9)	41835 (78.4)	34999 (85.5)	9652 (72.0)
Aspirin monotherapy, n (%)	9991 (47.3)	33921 (63.6)	30103 (73.5)	8240 (61.5)
Duration (days)	416 (137, 944)	718 (253, 1532)	965 (357, 1949)	714 (218, 1698)
ADP receptor inhibitor monotherapy, n (%)	1779 (8.4)	5409 (10.1)	5057 (12.3)	1478 (11)
Duration (days)	151 (41, 516)	145 (42, 561)	135 (42, 551)	142 (47, 463)
Dual antiplatelet therapy, n (%)	3086 (14.6)	7531 (14.1)	4455 (10.9)	883 (6.6)
Duration (days)	343 (112, 856)	372 (124, 990)	362 (120, 999)	331 (126, 954)
VKA monotherapy, n (%)	2537 (12)	8502 (15.9)	8715 (21.3)	2469 (18.4)
Duration (days)	229 (90, 448)	292 (101, 482)	316 (118, 477)	305 (114, 471)
VKA + 1 antiplatelet, n (%)	1253 (5.9)	4328 (8.1)	2914 (7.1)	463 (3.5)
Duration (days)	90 (53, 236)	90 (51, 205)	90 (54, 198)	90 (49, 189)
VKA + 2 antiplatelets, n (%)	159 (0.8)	547 (1)	363 (0.9)	50 (0.4)
Duration (days)	65.0 (35.0, 90)	61.0 (38.5, 90)	63.0 (34.0, 90)	74.5 (55.5, 108)

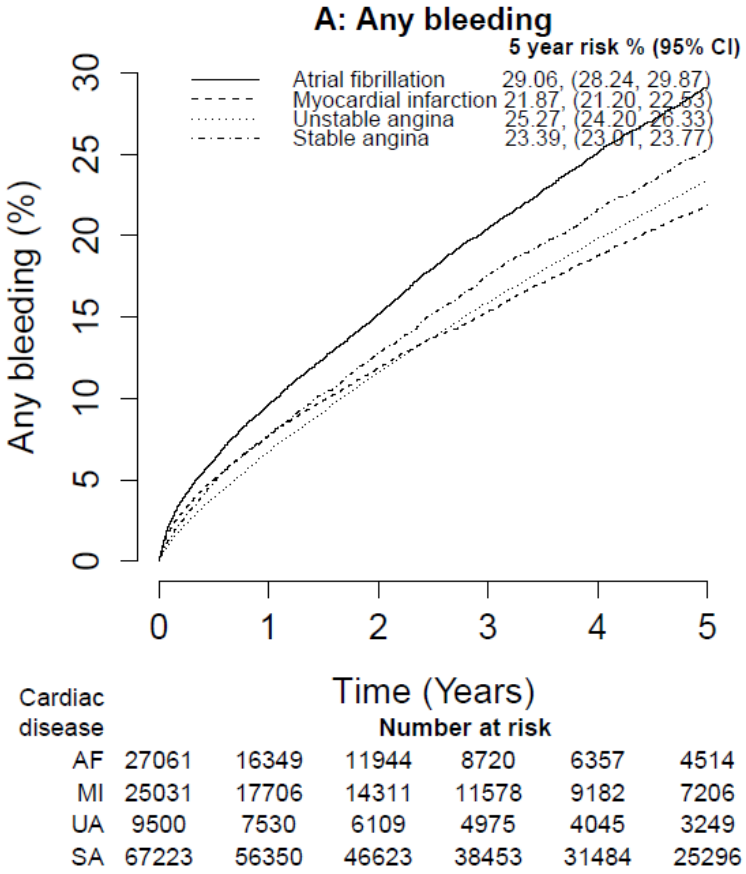
Note: SD=standard deviation, SBP=systolic blood pressure, BMI= body mass index, IQR= interquartile range, ADP=adenosine diphosphate, VKA=vitamin K antagonist

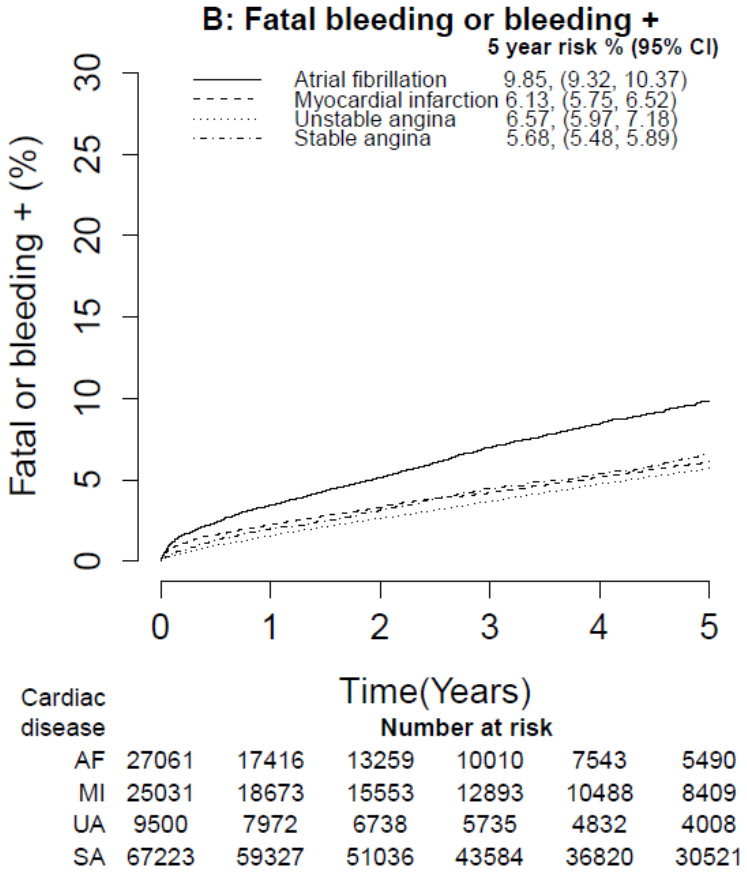
[a] Any record prior to cohort entry

[b] Nearest record to entry within 1 year prior to entry

[c] Between cohort entry and 1st bleeding event or end of follow-up

Figure 6.1: Five year risk of CALIBER bleeding from time of initial atrial fibrillation, acute myocardial infarction, unstable angina or stable angina (n= 128,815 patients); A: any bleeding, B: Fatal bleeding or bleeding with further markers of severity)





Note: In panel A 'Any bleeding' includes fatal, hospitalised +, hospitalised, primary care + and primary care bleeding events; in panel B 'Fatal bleeding or bleeding +' includes fatal, hospitalised + and primary care+ bleeding events only

Table 6.2: Kaplan-Meier estimates (95% confidence interval) for 5 year any bleeding (top row) and fatal bleeding, primary care+ bleeding or hospitalised+ bleeding (bottom row) stratified by baseline bleeding risk group and initial cardiovascular disease

Bleeding risk group	Atrial fibrillation	Myocardial infarction	Unstable angina	Stable angina	Overall
Low	16.05 (13.90 18.14) 3.17 (2.20 4.12)	12.14 (10.55 13.70) 2.10 (1.41 2.78)	16.87 (14.51 19.17) 1.83 (1.50 2.16)	18.05 (17.09 18.99) 1.85 (1.00 2.70)	16.8 (16.0 17.5) 2.0 (1.8 2.3)
Medium low	23.53 (22.11 24.93) 5.96 (5.17 6.74)	16.70 (15.71 17.68) 3.05 (2.59 3.50)	21.53 (19.90 23.11) 4.02 (3.24 4.78)	19.15 (18.58 19.71) 3.21 (2.96 3.46)	19.5 (19.0 19.9) 3.6 (3.4 3.8)
Medium high	31.40 (30.16 32.62) 10.92 (10.10 11.73)	24.47 (23.35 25.57) 7.32 (6.64 7.98)	29.46 (27.55 31.31) 8.90 (7.72 10.06)	26.39 (25.75 27.03) 7.51 (7.12 7.89)	27.2 (26.7 27.7) 8.2 (7.9 8.5)
High	40.08 (37.56 42.50) 17.48 (15.62 19.31)	37.41 (34.98 39.75) 15.11 (13.33 16.86)	38.00 (33.88 41.86) 16.17 (13.06 19.16)	35.62 (34.14 37.06) 13.80 (12.73 14.85)	37.1 (36.0 18.1) 15.1 (14.3 15.9)
Overall	29.1 (28.2 29.9) 9.9 (9.3 10.4)	21.9 (21.2 22.5) 6.1 (5.8 6.5)	25.3 (24.2 26.3) 6.8 (6.0 7.2)	23.4 (23.0 23.8) 5.7 (5.5 5.9)	24.2 (23.9 24.5) 6.5 (6.4 6.7)

Figure 6.2: The association between antithrombotic therapy prescribing and any bleeding and fatal or bleeding + events

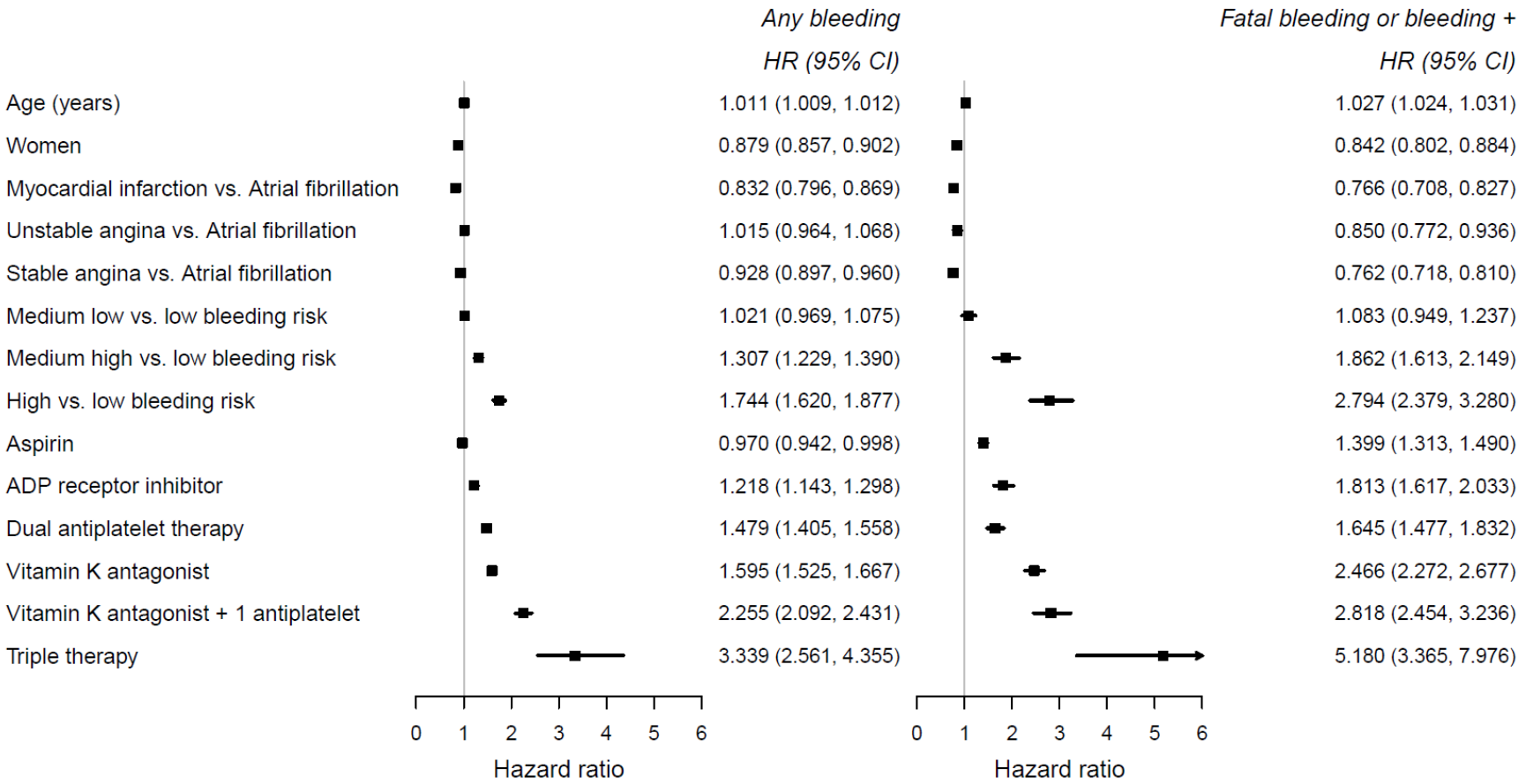
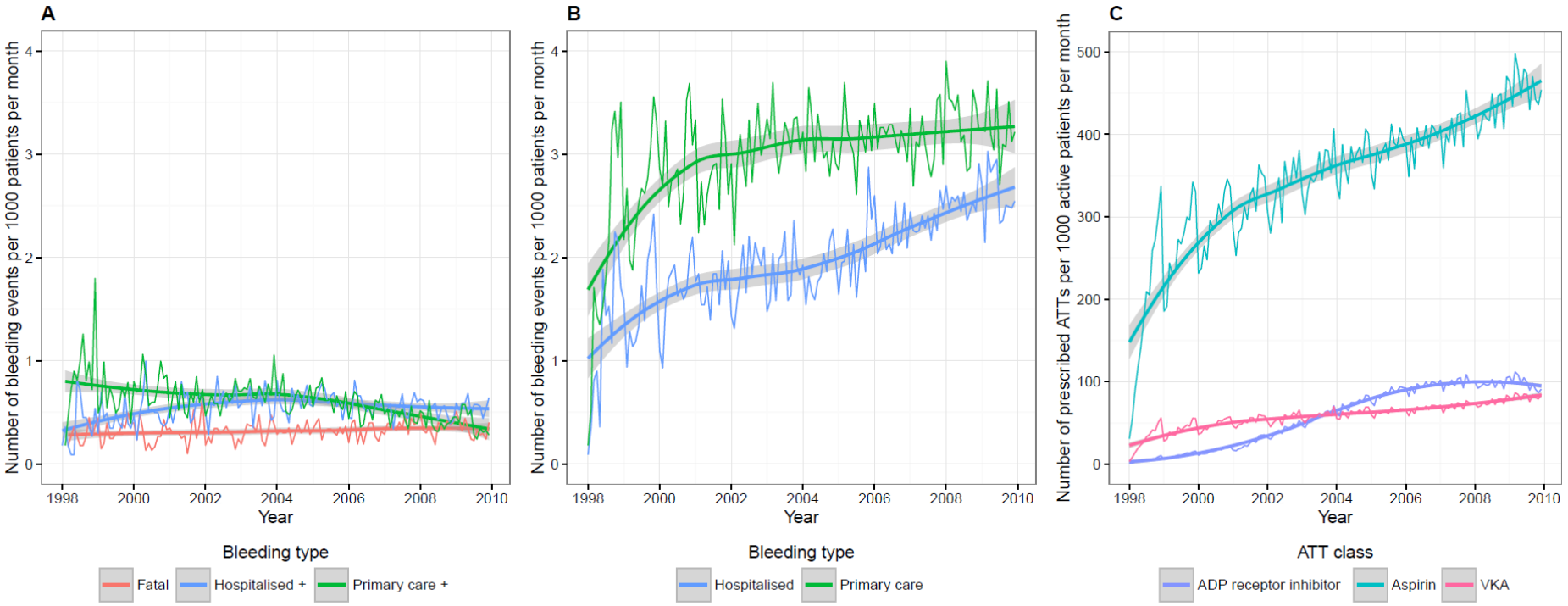


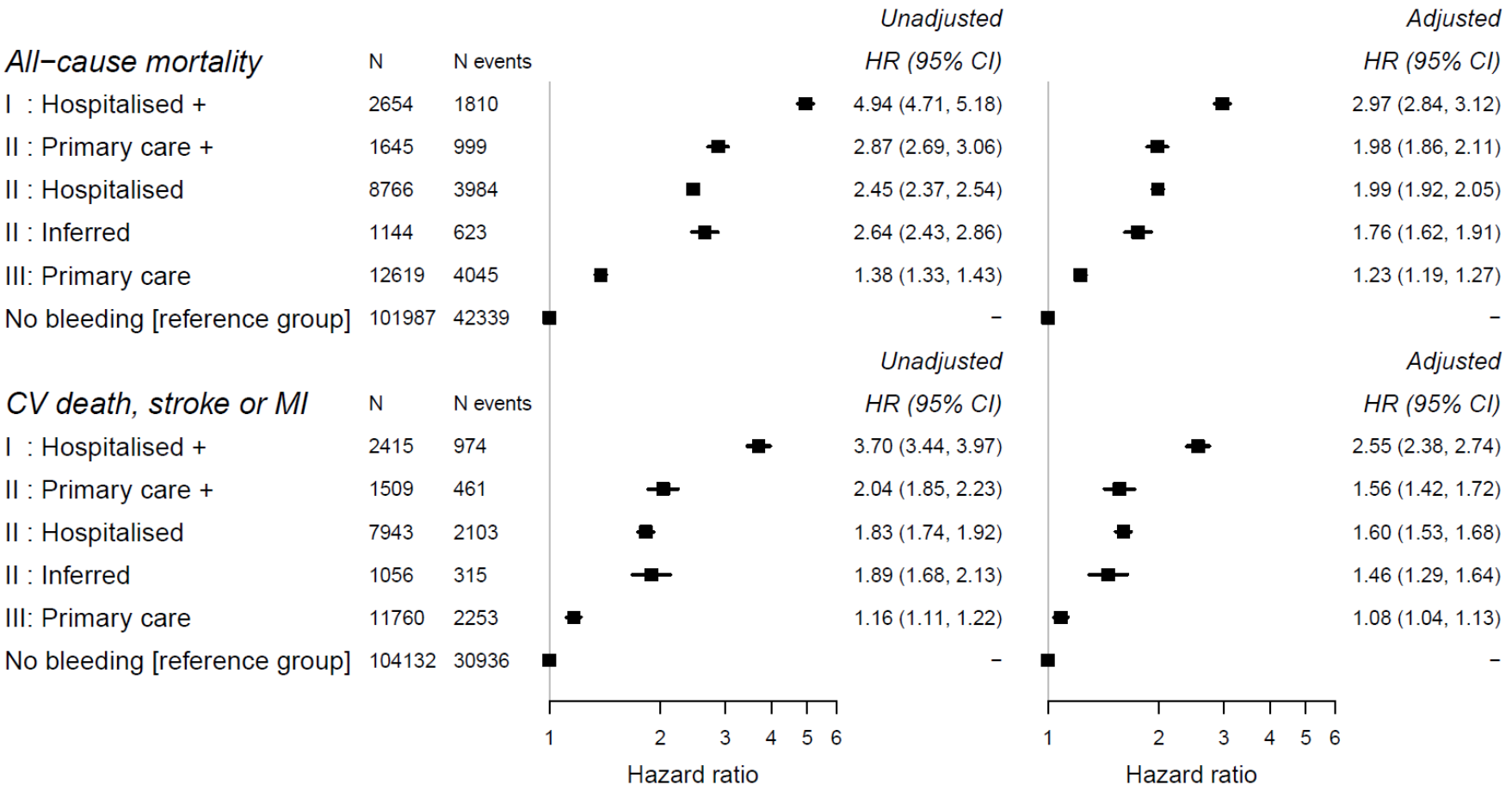
Figure 6.3: Time trends of fatal, hospitalised and primary care bleeding events and antithrombotic prescribing 1998-2010 in CALIBER.



A: Fatal, hospitalised + and primary care+ bleeding events; B: Hospitalised and primary care bleeding events; C: Prescriptions for ADP receptor inhibitors, aspirin and vitamin K antagonists.

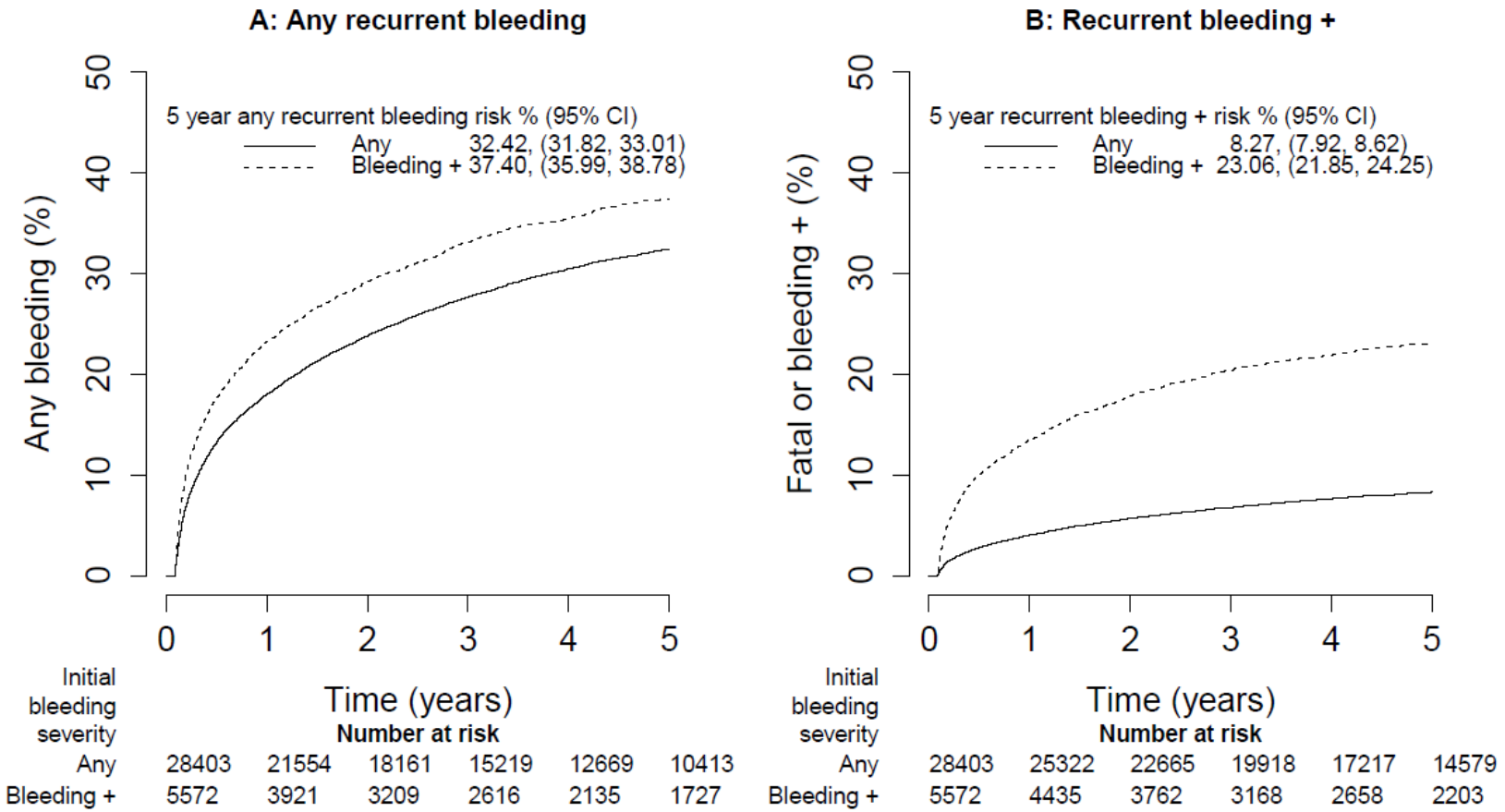
Note: Fitted lines are loess smoothed curved with shaded 95% confidence intervals

Figure 6.4: The association between non-fatal bleeding severity classes and time to all-cause mortality and cardiovascular death, stroke or myocardial infarction



Note: Adjusted estimates are adjusted for age, sex and comorbidities; CV= cardiovascular; MI= myocardial infarction

Figure 6.5: Five year risk of recurrent bleeding stratified by initial bleeding type: any bleeding or bleeding with further markers of severity (bleeding +).



Note: 'Any bleeding' includes hospitalised, hospitalised +, primary care, primary care + and inferred bleeding. 'Bleeding +' includes hospitalised + or primary care + bleeding.

7 Personalising the decision for prolonged dual antiplatelet therapy: Development and validation of prognostic models for atherothrombotic and bleeding events in stable myocardial infarction survivors

This work in this chapter has been published:

Pasea, L., Chung, S.C., Pujades-Rodriguez, M., Moayyeri, A., Denaxas, S., Fox, K.A., Wallentin, L., Pocock, S.J., Timmis, A., Banerjee, A., Patel, R. and Hemingway, H., 2017. Personalising the decision for prolonged dual antiplatelet therapy: development, validation and potential impact of prognostic models for cardiovascular events and bleeding in myocardial infarction survivors. *European heart journal*, *38*(14), pp.1048-1055. https://doi.org/10.1093/eurheartj/ehw683

Chapter Summary

Background

Recent trials have examined the efficacy of prolonging dual antiplatelet therapy in myocardial infarctions survivors for continued reduction of atherothrombotic risk

Methods

Using CALIBER linked electronic health records prognostic models for patients who survived one year following a myocardial infarction were developed (n=12694) and validated (n=5613) for all-cause mortality, atherothrombotic and bleeding events. Trial effect estimates for prolonged dual antiplatelet therapy with ticagrelor were applied to predicted risks of atherothrombotic and bleeding events in the validation cohort to estimate the net clinical benefit of treatment. A web application was developed to produce predicted risks of atherothrombotic and bleeding events given a set of baseline characteristics.

Results

Prognostic models were developed including covariates for patient demographics, behaviours, cardiovascular and non-cardiovascular medical history, clinical biomarkers and prescribed medication. In external validation the models performed well in terms of discrimination (c indexes ranged from 0.67 to 0.81 across the studied endpoints) and calibration. In the highest risk groups we estimated, for every 10,000 patients treated per year 249 (95% CI: 228, 269) cardiovascular events may be prevented and 134 (95% CI: 87,181) major bleeding events may be caused, whereas in the lowest risk groups 28 (95% CI: 19, 37) cardiovascular events may be prevented and 9 (95% CI: 0, 20) major bleeding events may be caused. Depending on the

weighting of benefits versus harms there was a net clinical benefit in favour of dual antiplatelet therapy for 63 to 99% of patients in the validation cohort.

Conclusion

I developed and validated prognostic models for benefits and different bleeding harms, relevant to the decision to prolong DAPT in patients stable 1 year after an acute MI. Personalised treatment decisions, based on individual patient risk profiles, can inform decision-making.

7.1 Introduction

Among patients who survived a year since their last acute myocardial infarction (MI), subsequent major cardiovascular events, all-cause mortality and major bleeding risks are high.[198,199] In unselected populations in USA, Sweden, England and France 20% of such patients experienced subsequent MI, stroke or died during the following 3 years.[199] Prolonged secondary prevention therapy in such patients is already recommended for 4 classes of drugs (statins, beta-blockers, ACE inhibitors and aspirin). Recent trials[17,200] examined addition of an extra antiplatelet. The PEGASUS-TIMI 54 trial[17] found prolonged dual antiplatelet therapy (DAPT) using aspirin and ticagrelor compared with aspirin alone in patients 1-3 years since their last acute MI reduced the risk of cardiovascular death, stroke or MI by 16% but increased major bleeding twofold. In light of this evidence, 2015 European Society of Cardiology guidelines recommend prolonged DAPT may be 'considered after careful assessment of ischaemic and bleeding risks'.[191,201]

How to make this 'careful assessment' is unclear. Risk prediction modelling has proven invaluable in conditions such as atrial fibrillation where similar decisions on benefits and harm need to be weighed.[7] Indeed, such models for bleeding and subsequent MI have been developed to guide use of bivalirudin[202], the choice of $P2Y_{12}$ inhibitor[203] and duration of DAPT.[204,205] An existing model intended to guide the duration of DAPT is based on patients undergoing drug eluting stent placement and selected into a trial and uses patient characteristics at the time of percutaneous coronary intervention.[205] Currently, no prognostic models evaluate the long-term risks of bleeding and cardiovascular events using updated clinical information one year after acute MI, to support the key clinical decision on prolonging DAPT.

In this context, unselected populations are important to provide realistic estimates of long-term cardiovascular and bleeding risks; these estimates are often substantially higher than those observed in the placebo arms of trials.[187,206,207] Linked electronic health records (EHRs)

are ideal sources of data to derive such 'real life' estimates of risks and harms, at scale. The CALIBER dataset,[8] validated for cardiovascular prognostic research,[40,41,184] is a 2 million person resource of linked primary-secondary and mortality data in England including 18,307 MI survivors.

Among these stable MI survivors, I sought to first, develop and validate prognostic models for major cardiovascular and bleeding endpoints. I used prognostic factors present one year following acute MI and widely recorded as part of guideline recommended care. Secondly, to demonstrate how predicted benefits and harms may aid treatment and clinical decisions I applied PEGASUS-TIMI 54 trial[17] relative risks of efficacy and safety to estimate potential numbers of cardiovascular events prevented and harms caused by prolonged DAPT, and net clinical benefits for individuals.

7.2 Methods

The models were developed and validated in line with TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) guidelines.[102] (See **supplementary appendix 11.4.1** for the completed TRIPOD checklist).

7.2.1 Linked electronic health records

I used the CALIBER (CArdiovascular research using LInked Bespoke and Electronic health Records) research platform, consisting of EHR linkages between primary care data (Clinical Practice Research Datalink (CPRD)), secondary care data (Hospital Episode Statistics), disease registry data (Myocardial Ischaemia National Audit Project) and cause-specific mortality (Office for National Statistics) in England.[8] The 4% sample of England's population in CPRD available for linkage is unselected, representative in terms of age, sex and overall mortality.[172-174] Furthermore, there is extensive evidence of risk factor and cardiovascular and non-cardiovascular disease endpoint validity in CALIBER.[36,38-42,184,185] The study was approved by the Independent Scientific Advisory Committee of the Medicines and Health care products Regulatory Agency in the UK and the MINAP Academic Group.

The full details of the data used for this work is described in **Chapter 3**.

7.2.2 Study Population

Patients in CALIBER alive 1 year after their last acute MI (i.e. their index acute MI) were included in the study. I studied patients from 2000-2010, before the introduction of ticagrelor, and when prolonged DAPT was rare. Follow-up started at 1 year after index acute MI, and patients were censored at the earliest date of the endpoints of interest, primary care practice transfer, death, or 5 years of follow-up. I evaluated prescriptions indicating prolonged clopidogrel use in follow up in the cohort to ensure they were untreated with long-term dual

antiplatelet therapy i.e. the decision the models aimed to aid. Patients were split into model development and validation cohorts using a pre-specified geographical divide.[208] The North of England has well documented higher rates of cardiovascular mortality compared to the South.[209] Based on the 10 administrative areas in the National Health Service I chose 6 in the South for model development cohort and 4 in the North for model validation cohort.

7.2.3 Potential prognostic factors

In model development I considered a priori prognostic factors including demographics, behaviours, cardiovascular and non-cardiovascular medical history, medications and clinical biomarkers. Each patient's most recent biomarker records in the year from index acute MI to follow-up start were used. Medications were defined as having ≥1 prescription of a drug in the year prior to follow-up start. As a patient's risk profile evolves with time (risk at index event and at one year may differ) I also analysed risk characteristics at the time of hospital discharge from index acute MI. Patients characteristics at hospital discharge and at one year post-MI were summarised as mean and standard deviation (SD) or median and interquartile range (IQR) for continuous variables and percentages for categorical variables.

7.2.4 Endpoints

The primary endpoint relating to potential benefits of prolonged DAPT, a composite of cardiovascular death, MI or ischemic or unspecified stroke, have validated phenotypes in CALIBER.[36,40,184] For examining potential harms of prolonged DAPT, I evaluated three severe bleeding endpoints with differing incidence: 1) fatal or hospitalised bleeding; 2) CALIBER major bleeding (a composite of fatal bleeding, intracranial bleeding, hospitalised bleeding with length of stay exceeding 14 days and bleeding requiring transfusion); and 3) fatal or intracranial bleeding. Models were also evaluated for all-cause mortality and cardiovascular death.

7.2.5 Statistical analysis

7.2.5.1 Model development

I evaluated associations between endpoints and prognostic factors using proportional hazards models with Weibull baseline hazards. Univariable models were used to detect nonlinear trends and inform inclusion of prognostic factors in multivariable models. Nonlinear continuous prognostic factors were included in the model using restricted cubic splines with three knot points. I chose the number of knots based on descriptive plots in the univariable analysis to provide a sufficient balance between capturing accurate shape and without overfitting.[210] I used 3 knots (a sensitivity analysis using 5 knots resulted in very similar predictions). Proportional hazards assumptions were checked using residual and log(-log) plots. For non-normally distributed continuous prognostic factors, sensible functional form was

obtained by log transformation. For parsimony, backwards selection methods were employed for constructing multivariable models, in which all prognostic factors were initially included and removed stepwise if p>0.1. Variance inflation factors were calculated to detect evidence of multicollinearity problems in the model selection process. If variables were deemed to be collinear (variance inflation factor>3) then a decision was made to remove one of the offending variables. Known important prognostic factors (age, gender, smoking status, index acute MI subtype, diabetes, history of MI prior to index acute MI and stroke) were retained in all models.

Missing values for clinical biomarkers and patient behaviours (smoking status and alcohol consumptions) were multiply imputed using MICE (Multiple Imputation by Chained Equations).[211] I used multiple imputation to allow the use of all available patient data in the analyses and potentially increase the precision of estimates. If the analyses was restricted to complete cases, the sample size would be reduced and would likely not be representative of the study population. Multiple imputation assumes data is missing at random – i.e. the missing data is not associated with the value of the missing data itself and violation of this assumption may result in biased imputations and model estimates. Prognostic factors with more than 50% data missing were not considered for multiple imputation or inclusion in the models. 30 imputed sets were generated using all baseline covariates (demographics and behaviours, disease history, clinical biomarkers and prescribed drugs) and study endpoints as predictors in the imputation models. Coefficient estimates from models fitted to imputed data sets were combined using Rubin's rules; which uses average coefficients and variances account for between and within imputation variability.

7.2.5.2 Risk groups

Unequal sized groups can minimise information loss when categorising compared with equal groupings,[212] I therefore grouped patients according into four risk groups (highest, high, low, lowest) using cut-points at the 16th, 50th and 84th percentiles of the model weighted sum of covariates in the development cohort.[89] These cut points split the validation cohort into risk groups for each outcome Therefore in validation, a well calibrated model would place 16%, 34%, 34% and 16% of patients into the high, medium high, medium low and low risk groups respectively.

7.2.5.3 Model validation

The developed models were initially validated internally in the development population. Bootstrap estimates of Harrell c-indexes were calculated and the observed events in the development population were compared with model predictions stratified by risk group. Models were then validated using the geographically external North of England population.

Harrell's C-indexes and 95% confidence intervals were used to estimate the discriminative ability of each model. Hazard ratios for each of the endpoints between risk groups, using the lowest risk groups being as reference, were calculated as a further measure of discrimination. Model calibration was assessed visually by comparing plots of model expected events with validation cohort observed events, stratified by risk group.

7.2.5.4 Model application

I applied the relative risks for efficacy (cardiovascular death, stroke or MI, cardiovascular death and all-cause mortality) and safety (TIMI major bleeding and fatal or intracranial bleeding) of ticagrelor 60mg versus placebo estimated in the PEGASUS-TIMI 54 trial[17] to the validation cohort stratified by predicted risk groups to calculate the potential events prevented and harms caused per 10,000 patients treated per year **(Equation 1)**. I assumed constant treatment effect across risk groups.

Equation 1: Calculating potential events prevented and harms caused with prolonged dual antiplatelet therapy

$$\text{〚Potential events prevented per 10,000 patients treated per year〛}\_ij = 10000 \times [\text{〚model predicted 3 year absolute risk〛}\_ij - (\text{〚PEGASUS TIMI 54 relative risk〛}\_(i) \times \text{〚model predicted 3 year absolute risk〛}\_ij)]/3$$

$$i \in \{CV\ death, stroke\ or\ MI\}$$

$$j \in \{highest\ risk; high\ risk; low\ risk; lowest\ risk\}$$

$$\text{Potential harms caused per 10,000 patients treated per year}_{ij} = 10000 \times [(\text{〚PEGASUS TIMI 54 relative risk〛}\_(i) \times \text{〚model predicted 3 year absolute risk〛}\_ij) - \text{〚model predicted 3 year absolute risk〛}\_ij]/3$$

$$i \in \{major\ bleeding; fatal\ or\ intracranial\ bleeding\}$$

$$j \in \{highest\ risk; high\ risk; low\ risk; lowest\ risk\}$$

I calculated net cardiovascular death, stroke or MI and net CALIBER major bleeding risks with treatment for each individual in the validation cohort using the trial relative risk estimates **(Equation 2).** I evaluated predicted net benefit in patients (i.e. estimated cardiovascular risk decrease exceeds bleeding risk increase) under three different benefit and harm weighting scenarios:[99,203] 1) patients view benefits and harms equally, 2) patients are more concerned about preventing atherothrombotic events than potential harms and 3) patients are more concerned about potential harms than preventing atherothrombotic events.

Equation 2: Calculating individual net benefits for prolonged dual antiplatelet therapy

$$Net\ benefit_i = [absolute\ CV\ risk_i - (PEGASUS\ TIMI\ 54\ relative\ CV\ risk \times absolute\ CV\ risk_i)] - [absolute\ bleeding\ risk_i - (PEGASUS\ TIMI\ 54\ relative\ bleeding\ risk \times absolute\ bleeding\ risk)]$$

$$i \in \{1 \dots n\}\ patients$$

7.2.6 Web app development

I used the RShiny package [https://shiny.rstudio.com/] to develop a user friendly web app to calculate risks of atherothrombotic and bleeding risks given patient baseline characteristics. I used the effect estimates of prolonged dual antiplatelet therapy from the PEGASUS TIMI 54 trial to show potential changes of risk with further treatment (**Equation 1**).

7.2.7 Assessment of simplified approaches for risk stratification

I assessed simple methods of risk stratification. I compared the model predicted 3 years risks of cardiovascular death, stroke or MI and CALIBER major bleeding amongst patients in the validation cohort with and without binary factors determined by the PEGASUS-TIMI 54 trial to indicate patients at the highest risk: age≥ 65, diabetes (any of types 1, 2 or unspecified type), history of MI prior to their index MI and history of renal disease. I applied commonly used point-base risk scores, CHA_2DS_2-VASc[6] for atherothrombotic risk and HASBLED[7] for bleeding risk to the validation cohort. The CHA_2DS_2-VASc scheme scores patients 1 point each for age 65-74 years, female gender, histories of heart failure, hypertension, diabetes and vascular disease and 2 points for age≥75 years and history of stroke. Patients risk increases with higher values of the CHA_2DS_2-VASc score and the scores range is 0-9. The HASBLED scheme scores patients 1 point each for hypertension, abnormal liver function, abnormal renal function, stroke, bleeding, labile international normalised ratio (time in therapeutic range <60), age>65 years, prescriptions for antiplatelets or non-steroidal anti-inflammatory drugs and history of alcohol abuse. Patients risk increases with higher values of the HASBLED score and the scores range is 0-9. I assessed the performance of CHA_2DS_2-VASc and HASBLED risk stratification schemes using c-indexes and 5 year Kaplan-Meier event estimates for patients with each score value.

7.3 Results

7.3.1 Baseline characteristics and overall event rates

I identified 18307 patients in CALIBER who had survived 1 year following their index MI (**Figure 7.1**). The model development cohort consisted of 12,694 patients (mean age 70.1 years, 66.1% male) from 159 general practices, median follow-up of 3.1 years (range: 0-9.8) and 27%

followed-up for at least 5 years **(Table 7.1)**. The validation cohort consisted of 5,613 patients (mean age 69 years, 64.4% male) from 61 general practices. Histories of stroke, heart failure, peripheral arterial disease, renal disease, chronic obstructive pulmonary disease, dementia, chronic anaemia, peptic ulcer and hospitalised bleeding were more prevalent in the model validation cohort compared with the development cohort. Approximately 3% of patients had prolonged clopidogrel use in follow-up.

Patient characteristics of the cohorts changed from acute MI discharge to the study baseline date of 1 year post-MI (**Supplementary appendix 11.4.2**) including increased heart failure (18.2% to 23.5% in the development cohort and 21.4% to 28.0% in the validation cohort) and renal disease (8.7% to 13.6% in the development cohort and 9.5% to 14.8% in the validation cohort) prevalence and changed smoking statuses. There was also a reduction in mean blood pressure of 12mmHg in both groups (145 (SD: 16.3) to 133(SD: 18.6) in the development cohort and 144(SD: 17.0) to 132 (SD: 18.4) in the validation cohort.

Using Kaplan-Meier curved I observed higher 5 year event rates in the validation cohort than in the development cohort for all endpoints expect cardiovascular death stroke or MI **(Figure 7.2)**. At 5 years in the development cohort 2683 (30.1%) had died, 1913 (22.4%) had cardiovascular death, stroke or MI events and 188 (2.5%) had major bleeding events. Whereas in the validation cohort at 5 years 1252 (30.6%) patients had died, 936 (23.8%) had cardiovascular death, stroke or MI and 188 and 98 (2.7%) had major bleeding events.

7.3.2 Development of prognostic models

In univariable modelling (Univariable log hazard ratio estimate shown in **Supplementary appendix 11.4.3**) I found increasing risk of all events with age and women were at greater risk than men of all-cause mortality, cardiovascular events and fatal or hospitalised bleeding. Patients whose index MI was STEMI or unspecified had lower risks of all endpoints compared with NSTEMI patients. All investigated cardiovascular and non-cardiovascular comorbidities were associated with increased risks of events or had non-significant associations. In particular patients with histories of heart failure, atrial fibrillation, stroke, peripheral arterial disease, type 2 diabetes, renal disease, cancer, chronic anaemia, bleeding disorders and hospitalised bleeding were at increased risk of all six endpoints. Underweight patients had greater risks of all-cause mortality and cardiovascular events compared with normal weight patients. Patients who had been prescribed oral anticoagulants were at increased risk of cardiovascular and bleeding events.

Checks of the proportional hazards assumptions (**Supplementary appendix 11.4.4)** showed no violations. The curves in log(-log) plots for categorical prognostic factors were parallel and time dependent coefficients for continuous prognostic factors were constant over time.

In multivariable **(Log hazard ratios displayed in Figure 7.3; hazard ratio estimates given in Supplementary appendix 11.4.5**) modelling**,** I identified 20 prognostic factors for inclusion in the cardiovascular death, stroke or MI model, 22 for inclusion in the cardiovascular death model, 24 for inclusion in the all-cause mortality model , 17 for CALIBER major bleeding, 17 for fatal or hospitalised bleeding and 16 for fatal or intracranial bleeding. There was no evidence of problematic multicollinearity between the prognostic factors in the multivariable models (variance inflation factors < 3). While the presence and direction of prognostic factor associations with cardiovascular events and bleeding outcomes were mostly concordant, the magnitude of prognostic factor associations (e.g. history of MI, stroke, diabetes) differed across endpoints. Systolic blood pressure was also included in all models and was modelled using restricted cubic splines due to a nonlinear relationship with the endpoints. An example curve demonstrating the association of systolic blood pressure with the risk of CV death, stroke or MI is shown in **Figure 7.4**. Patients with very low or very high systolic blood pressure had increased risk of events. The functions estimated for systolic blood pressure in each model are provided in **Supplementary appendix 11.4.6**.

7.3.3 Risk groups for all-cause mortality, atherothrombotic and bleeding events

The 16th, 50th and 84th percentiles of the linear predictors for each endpoint in the development cohort were used to split the patients into risk groups. These cut points were (−1,746, −2.972, and −4.077) for all-cause mortality, (−2.198, −3.298, and −4.252) for cardiovascular death, stroke or MI, (−2.439, −3.792, and −4.995) for cardiovascular death, (−4.159, −5.040 and −5.879) for fatal or intracranial bleeding, (−4.178, −5.118 and −5.989) for CALIBER major bleeding and (−3.437, −4.128 and −4.728) for hospitalised bleeding. The proportions of patients assigned to each risk group are shown in **Table 7.2.** By design 16%, 34%, 34% and 16% of patients in the development cohort were classified as highest, high, low and lowest risk respectively for each endpoint. When the cut points were applied to the validation cohort for all endpoints fewer than 16% patients were classified as lowest risk (from 14.5% for cardiovascular death, stroke or MI and cardiovascular death to 15.4% fatal or intracranial bleeding). More than 16% patients were assigned to the highest risk group for all endpoints (from 16.1% for fatal or intracranial bleeding to 18.5% for hospitalised bleeding). The expected numbers of patients were assigned to the high and low risk groups. This indicates worse outcomes and higher risks in the validation cohort compared with the development cohort.

In **Figure 7.5** I show the risk groupings for cardiovascular death, stroke or MI and CALIBER major bleeding across their respective linear predictor distributions in the validation cohort and the corresponding cumulative probability of events.

7.3.4 Internal validation

In internal validation, the models demonstrated high levels of discrimination and calibration (**Figure 7.6).** The c-indexes for all models were at least 0.70. The cardiovascular death model had the best discriminative ability (c-index=0.82). The observed event Kaplan-Meier curves were largely within the bounds of the model predicted values 95% confidence intervals for all risk groups in each of the endpoints.

7.3.5 Geographical validation - model discrimination

In the validation cohort the models discriminated risk well, c-index estimates were 0.81 [95% CI: 0.80, 0.82] for all-cause mortality, 0.75 [95% CI: 0.74, 0.77] for cardiovascular death, stroke or MI, 0.81 [95% CI: 0.80, 0.83] for cardiovascular death, 0.67 [95% CI: 0.64, 0.70] for fatal or hospitalised bleeding, 0.72 [95% CI: 0.67, 0.77] for CALIBER major bleeding, and 0.68 [95% CI: 0.61, 0.75] for fatal or intracranial bleeding. The hazard ratios for events across risk groups **(Table 7.3)** also demonstrated good discriminative ability of the models. The hazard ratios for the predicted highest: lowest risk group contrast were 41.0 [95% CI: 26.2, 64.0] for all-cause mortality, 11.5 [95% CI: 8.40, 15.5] for cardiovascular death, stroke or MI, 33.0 [95% CI: 18.9, 57.4] for cardiovascular death, 6.3 [95% CI: 3.7, 10.6] for hospitalised bleeding, 13.3 [95% CI: 4.8, 37.3] for CALIBER major bleeding and 3.1 [95% CI: 1.7, 5.9] for fatal or intracranial bleeding. For cardiovascular death, stroke or MI, CALIBER major bleeding, and fatal or intracranial bleeding the 95% confidence intervals for the hazard ratios comparing risks of events between low risk and lowest risk patients contain 1. This suggests that for these endpoints the models may struggle to distinguish between patients with lower risks.

7.3.6 Geographical validation - model calibration

The models were well calibrated **(Figure 7.7)**, in particular, models for all-cause mortality and cardiovascular death had very similar predicted and observed events in the validation cohort. The cardiovascular death, stroke or MI model underestimated events for patients in the lowest risk group. Patients in the low and lowest risk groups had very similar event rates. The fatal or hospitalised bleeding and fatal or intracranial bleeding models overestimated events for patients in the highest risk group and underestimated events for low risk patients. The CALIBER major bleeding model underestimated events for low risk patients.

7.3.7 Potential absolute benefits and harms in risk groups

I observed higher event rates in the validation cohort compared with the trial placebo group for all studied endpoints: cardiovascular death, stroke or MI (16.5% versus 9.04%), cardiovascular death (11.5% versus 3.39%), all-cause mortality (20.3% versus 5.16%), major bleeding (CALIBER 1.7% versus TIMI 1.26%) and fatal or intracranial bleeding (0.9% versus 0.6%). **Table 7.4** shows on an intention-to-treat basis in highest risk groups, for every 10,000 patients treated per year, 249 (95% CI: 228, 269) cardiovascular events may be prevented and 134 (95% CI: 87,181) major bleeding events may be caused, whereas in the lowest risk groups 28 (95% CI: 19, 37) cardiovascular events may be prevented and 9 (95% CI: 0, 20) major bleeding events may be caused. In the absence of risk stratification, overall an estimated 89 (95% CI: 83, 94) cardiovascular events prevented and 42 (95% CI: 32, 51) harmed patients per 10,000 treated per year. I also identified wide ranges in cardiovascular death and all-cause mortality events that may be prevented depending on risk. In their respective highest risk groups, the 3 year cumulative risk of fatal or intracranial bleeding was 2.2% and of fatal or hospitalised bleeding was 10.5%.

7.3.8 Potential net clinical benefits in individuals

A positive net clinical benefit was estimated in 93.5% of patients when cardiovascular death, stroke or MI events and CALIBER major bleeding were weighted equally, 63% if avoiding major bleeding was valued twice more than preventing cardiovascular events and 99.1% if preventing cardiovascular events was valued twice as high as avoiding major bleeding **(Figure 7.8, panel A).** The importance of using clinical characteristics updated at 1 year after acute MI compared with characteristics at acute MI discharge for 5 typical patients is illustrated in **Figure 7.8, panel B**. For example, using characteristics at discharge patient 5 would be considered suitable for prolonged DAPT under all weighting options (point '▽' on the figure), whereas using updated information 1 year post MI this patient would only be considered suitable for prolonged DAPT if they valued potential benefits over potential harms of treatment (point '▼' on the figure).

7.3.9 Web application

I developed a web-based tool which calculates personalised risk predictions following the input of baseline characteristics:

http://www.caliberresearch.org/prolonged_dapt_benefits_harms_risks

An example of the output from the web app is shown in **Figure 7.9** for an male patient aged 68, non-smoker, whose index MI was a STEMI and has type 2 diabetes, is overweight, systolic

blood pressure: 139mmHg, creatinine: 151mmol/L, haemoglobin: 14.3g/dL, white blood cell count: 9.3 10^9/L and HDL: total cholesterol ratio: 0.4.

The output displays that 90 out of 1000 patients with these characteristics would be expected to have cardiovascular death, stroke or MI at 3 years, which reduces to 76 out of 1000 when the PEGASUS-TIMI 54 estimates of the effect of prolonged dual antiplatelet therapy are applied. 9 out of 1000 patients with these characteristics would be expected to experience major bleeding at 3 years which increases to 21 out of 1000 when accounting for the estimated harms of prolonged dual antiplatelet therapy. This information is also displayed with Kaplan-Meier curves, with the option to overlay the predicted curves accounting for treatment effects and to display the expected event rates at different time points between 1 and 5 years.

7.3.10 Simplified approaches for risk stratification

Very little discrimination between predicted risks of events amongst patients with and without single high-risk factors for cardiovascular death, stroke or MI and CALIBER major bleeding endpoints was observed (**Figure 7.10).**

The distributions of HASBLED and CHA_2DS_2-VASc scores in the validation cohort are shown in **(Figure 7.11)**. The c-index for HASBLED was 0.64 (95% CI: 0.58, 0.70) and for CHA_2DS_2-VASc was 0.71 (95% CI: 0.69, 0.72). I observed generally increasing risks of cardiovascular death, stroke or MI and CALIBER major bleeding with increasing CHA_2DS_2-VASc and HASBLED scores respectively. However there were no bleeding events amongst patients with a HASBLED score of 0, 7 or 8 so I was unable to calculate Kaplan-Meier estimates for these groups.

7.4 Discussion

Using population-based linked electronic health records, I developed and validated prognostic models providing personalised estimates of risks of major cardiovascular and bleeding events in patients 12 months after an acute MI. With trial relative risks I estimated potential benefits and harms of prolonged DAPT across risk groups. In individuals, potential net clinical benefit was observed for the majority of patients, even when avoiding bleeding is considered twice as important as preventing cardiovascular events. However the magnitude of benefit must also be considered.

7.4.1 Potential benefits of prolonged dual antiplatelet therapy

The unselected study cohort experienced a much higher cardiovascular event rate than the PEGASUS-TIMI 54 trial placebo group despite the trial inclusion criteria enriching for high risk.[187] Thus the potential absolute benefit of prolonged DAPT may be greater than reported in

the trial. Importantly, potential benefits are comparable when the unselected real-world study population reported here is restricted to those meeting the trial inclusion and exclusion criteria (89 vs 101 events potentially prevented per 10,000 treated per year, respectively).[187] The models widely separated the risk of events: a clinician may treat nine times (28/249) as many patients in the lowest-risk, compared to the highest-risk group, to prevent one cardiovascular event. The models help clarify limitations of ignoring individual patient characteristics evaluated during the stable phase post-acute MI.

7.4.2 Potential bleeding harms of prolonged dual antiplatelet therapy

In balancing potential harms from prolonged DAPT, I show the importance of considering different bleeding endpoints. While all bleeding events studied may be considered serious (they all required hospitalisation), their event rates and potential harms vary significantly. I sought to approximate the PEGASUS-TIMI 54 trial primary safety endpoint, TIMI major bleeding, and successfully defined fatal bleeding, intracranial bleeding and transfusions, but not, in currently available EHR, acute haemoglobin change. Nonetheless, incidence of CALIBER major bleeding was comparable with TIMI major bleeding in the trial placebo arm.

7.4.3 Balancing potential benefits and harms in individuals

Patients and clinicians differ in how they value different benefits and harms, and I derived net clinical benefit estimates under different weighting scenarios. These can inform patient counselling and patient-doctor discussions tailored to patients risk covariates. Cost-effectiveness considerations are additionally important in determining the magnitude of net clinical benefit a given health system is willing to pay for.[213,214]

7.4.4 Need for multivariable risk prediction

The importance of tailoring risk to multiple patient characteristics, as opposed to single risk factors (e.g. the high-risk factors specified in the PEGASUS-TIMI 54 trial: age, diabetes, history of MI and renal disease)[215] is illustrated in **Figure 7.10.** In each case the risk distribution largely overlaps among people with and without simple binary prognostic factors. Furthermore, while simple, point-based scores for predicting bleeding risk prove valuable for other diseases,[6,216] they are unlikely to be useful for this population. For example, HASBLED[7] and CHA_2DS_2-VASc[6] did not perform as well as my models **(Figure 7.11)**. It is well-known categorising clinical information loses predictive value and systolic blood pressure has nonlinear U-shaped associations with endpoints (**Figure 7.4**).[217] Nonetheless in practice some simplification of the models will occur as absolute contraindications are established (e.g. although history of bleeding was adjusted for in the models, patients who bleed in the first year should be excluded from the prolonged DAPT treatment option).

7.4.5 Application in clinical practice

I demonstrate marked changes in the year following acute MI in the prevalence of major prognostic factors, including heart failure, renal disease and smoking, with consequent changes in net benefits. Good clinical practice dictates thorough evaluation of patients at the time of decision-making including up-to-date medical history and biomarkers. The models can be readily implemented in health systems with EHR.[90] I present a web-based tool to aid calculation of personalised atherothrombotic and bleeding risk predictions at: http://www.caliberresearch.org/prolonged_dapt_benefits_harms_risks. I envision that the risk prediction tool could be used by clinicians to assess their patients one year following an MI. This will aid clinicians to identify patients that are candidates for prolonged dual antiplatelet therapy and decisions could be made taking into account patients values towards treatment benefits and harms.

7.4.6 Methodological strengths

Few, if any, previous studies evaluated prognostic relevance of clinical data available at the decision point of surviving 1 year following an acute MI. Previous studies focused on factors measured in the acute hospitalised phase. Unlike voluntary disease registries or trial populations which may not reflect the entire population at risk, CALIBER is population-based. Therefore, estimates of risk obtained are likely to be representative of those observed in usual clinical practice. Routinely collected clinical information readily available in EHR enabled estimation of changing net clinical benefit of prolonged DAPT during patient journeys.

7.4.7 Limitations

The study has limitations. First, the information that is recorded as part of usual clinical practice is unlikely to have the same precision as that recorded as part of standardised research protocols. If the quality of information recording is low this will diminish the ability of the models to discriminate risk. Second, current large-scale population-based EHR data lack information on left ventricular function, number of diseased vessels, coronary stent type and diameter. However it is not known if such factors remain prognostically relevant when included in models with updated clinical information at one year post-MI. There is extensive evidence that stent type (bare metal, drug eluting, bioabsorbable) is an important predictor of prognosis from the time of acute myocardial infarction, and this study was unable to evaluate whether, in the stable phase 12 months after acute myocardial infarction, at a time when I demonstrate that clinical factors have changed, stent characteristics continue to provide incremental discrimination of risk. Previous nationwide studies of angiographic findings suggest only a modest predictor of subsequent events in stable patients.[218] Furthermore I did not validate the models in different health systems. However a study of 140,000 unselected

stable post-MI patients found similar rates of cardiovascular and bleeding events in England, France, Sweden and USA[199] and the effects of multiple prognostic factors reported in the present study were consistent across these 4 countries, suggesting potential geographic transportability of the models. I used multiple imputations where appropriate and the complete-case sensitivity analysis showed no important difference to the presented models. Multiply imputing MI type for those unspecified did not show any difference to the ST-elevation MI versus non ST-elevation MI contrast presented in the multivariable models.

7.4.8 Future research

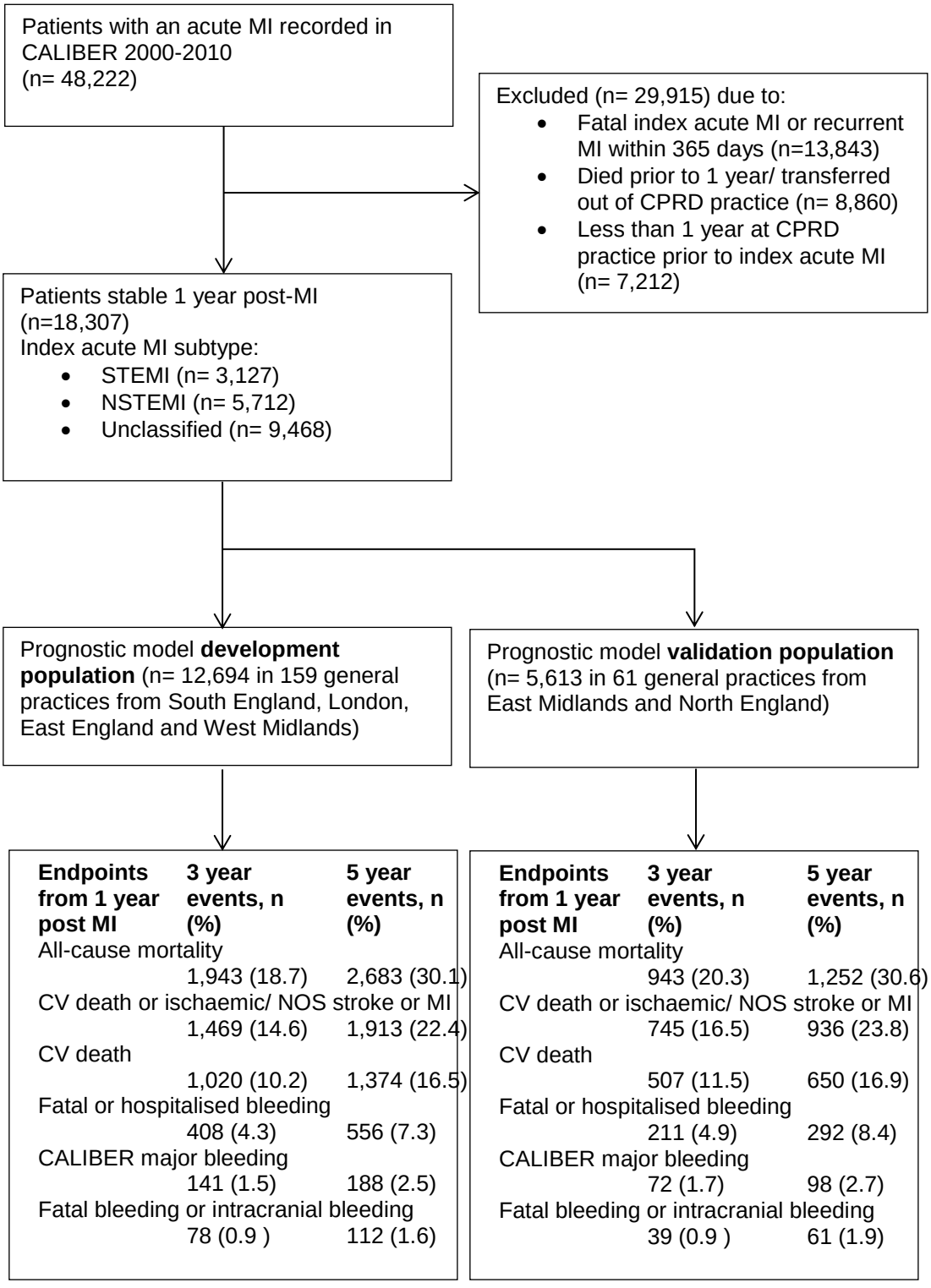
I propose three major avenues for future research. First, further validation studies are recommended in populations differing in background, treatment strategies and event rates to ensure generalisability.[218] I selected a study period to provide a relatively pure population, prior to the UK approval of ticagrelor and widespread occurrences of prolonged DAPT. Further research is required to test the models in more recent cohorts with a mix of clopidogrel, ticagrelor and prasugrel users and risk prediction models may be required for other harms (e.g. dyspnea in patients taking ticagrelor). Second, as with any prognostic model the question arises of what is their impact on clinical outcomes and costs[213,214] when implemented in clinical care.[193] This would involve developing clinical decision support systems using prognostic models, and evaluation through a cluster randomised controlled trial. Third, there is a need to identify biomarkers which distinguish bleeding from cardiovascular event risks and may further aid decision-making.

7.5 Conclusion

Using population-based EHRs I developed and validated prognostic models for benefits and different bleeding harms, relevant to the decision to prolong DAPT in patients stable 1 year after an acute MI. Personalised treatment decisions, based on individual patient risk profiles, can inform decision-making.

7.6 Tables and Figures

Figure 7.1: Study population flow diagram, endpoints and 3 & 5 year event rates



Endpoints from 1 year post MI	3 year events, n (%)	5 year events, n (%)
All-cause mortality	1,943 (18.7)	2,683 (30.1)
CV death or ischaemic/ NOS stroke or MI	1,469 (14.6)	1,913 (22.4)
CV death	1,020 (10.2)	1,374 (16.5)
Fatal or hospitalised bleeding	408 (4.3)	556 (7.3)
CALIBER major bleeding	141 (1.5)	188 (2.5)
Fatal bleeding or intracranial bleeding	78 (0.9)	112 (1.6)

Endpoints from 1 year post MI	3 year events, n (%)	5 year events, n (%)
All-cause mortality	943 (20.3)	1,252 (30.6)
CV death or ischaemic/ NOS stroke or MI	745 (16.5)	936 (23.8)
CV death	507 (11.5)	650 (16.9)
Fatal or hospitalised bleeding	211 (4.9)	292 (8.4)
CALIBER major bleeding	72 (1.7)	98 (2.7)
Fatal bleeding or intracranial bleeding	39 (0.9)	61 (1.9)

Note: MI= myocardial infarction, STEMI=ST-elevation myocardial infarction, NSTEMI= non-ST-elevation myocardial infarction, NOS=not otherwise specified, CV=cardiovascular

Table 7.1: Characteristics of population based samples at baseline defined as 1 year after their last acute MI

	Development cohort (n=12,694)	Validation cohort (n=5,613)
Demographics and behaviours		
Age (years)	70.1 (12.7)	69.1 (12.8)
Women	33.9%	35.6%
Ethnicity		
Asian	2.0%	1.1%
Black	0.5%	0.1%
Other	17.0%	14.1%
White	80.4%	84.6%
Index of multiple deprivation (highest quartile- most deprived)	15.4%	30.3%
Missing	0.1%	0.5%
Smoking status		
Ex-Smoker	45.5%	45.4%
Non-Smoker	33.9%	31.1%
Smoker	12.9%	14.3%
Missing	7.6%	9.1%
Excess alcohol	10.8%	15.4%
Cardiovascular diseases		
Index acute MI subtype		
NSTEMI	32.1%	29.1%
STEMI	17.4%	16.4%
Unspecified	50.5%	54.5%
MI (prior to index acute MI)	34.7%	38.7%
Revascularisation (any)	43.5%	33.0%
Primary PCI for index acute MI	24.8%	18.8%
PCI at any time prior to 1 year post MI	39.4%	31.0%
CABG at any time to 1 year post MI	4.1%	2.0%
Stroke	6.9%	8.1%
Atrial fibrillation	18.0%	17.9%
Heart failure	23.5%	28.0%
Peripheral arterial disease	9.8%	13.1%
Renal disease	13.6%	14.8%
Recent hospitalisation for acute renal disease	1.3%	1.0%
Non-cardiovascular diseases		
Diabetes		
Type 1	1.2%	0.9%
Type 2	16.7%	17%
Unspecified	1.5%	1.7%
COPD	9.1%	12.8%
Recent hospitalisation for acute COPD	1.1%	2.2%
Liver disease	0.4%	0.5%
Non-metastatic cancer	14.4%	13.2%
Metastatic cancer	1.0%	1.2%
Dementia	1.3%	2.0%
Chronic anaemia	14.3%	17.9%

	Development cohort (n=12,694)	Validation cohort (n=5,613)
Peptic ulcer	7.3%	10.2%
Bleeding diatheses and coagulation disorders	1.1%	1.1%
Hospitalised bleeding	6.5%	8.2%
Treatments prescribed		
Aspirin	87.0%	86.2%
Clopidogrel	50.5%	47.8%
Prolonged clopidogrel, post-baseline	2.7%	3.4%
Oral anticoagulant	9.9%	9.5%
Statin	88.5%	88.6%
Anti-hypertensive	96.4%	96.0%
Biomarkers*		
BMI (Continuous) (kg/m^2)	27.8 (5.1)	27.7 (5.1)
BMI (Categorical)		
Underweight	0.9%	1.1%
Normal	16.2%	17.3%
Overweight	23.3%	25.7%
Obese	16.1%	16.9%
Missing	43.5%	39.1%
SBP (mmHg)	133 (18.6)	132 (18.4)
Missing	5.4%	6.1%
DBP (mmHg)	75.3 (10.4)	74.6 (10.1)
Missing	5.4%	6.1%
Haemoglobin (g/dL)	13.4 (1.6)	13.3 (1.6)
Missing	43.3%	45.0%
White blood cell count ($10^9/L$)	7.60 (2.3)	7.68 (2.3)
Missing	45.4%	46.9%
Total cholesterol (mmol/L)	4.17 (1.0)	4.17 (1.0)
Missing	18.4%	18.0%
HDL cholesterol (mmol/L)	1.28 (0.4)	1.26 (0.4)
Missing	40.7%	40.1%
Creatinine (μmol/l) Median (IQR)	98 (84, 114)	99 (86, 117)
Missing	21.6%	21.8%
eGFR (ml/min)	65.5 (20.3)	64.7 (20.7)
eGFR < 60 ml/min	29.3%	31.4%
Missing	24.3%	23.7%

Note: *values are mean (SD) except where stated; MI=myocardial infarction, STEMI=ST-elevation myocardial infarction, NSTEMI=non-ST-elevation myocardial infarction, PCI= Percutaneous coronary intervention, CABG= Coronary artery bypass graft, COPD=chronic obstructive pulmonary disease, BMI= body mass index, SBP=systolic blood pressure, DBP= diastolic blood pressure, HDL= high-density lipoprotein, eGFR= estimated glomerular filtration rate

Figure 7.2: Comparison of all-cause mortality, cardiovascular and bleeding events in patents included in the development (n=12,694) and validation (n=5,613) cohorts

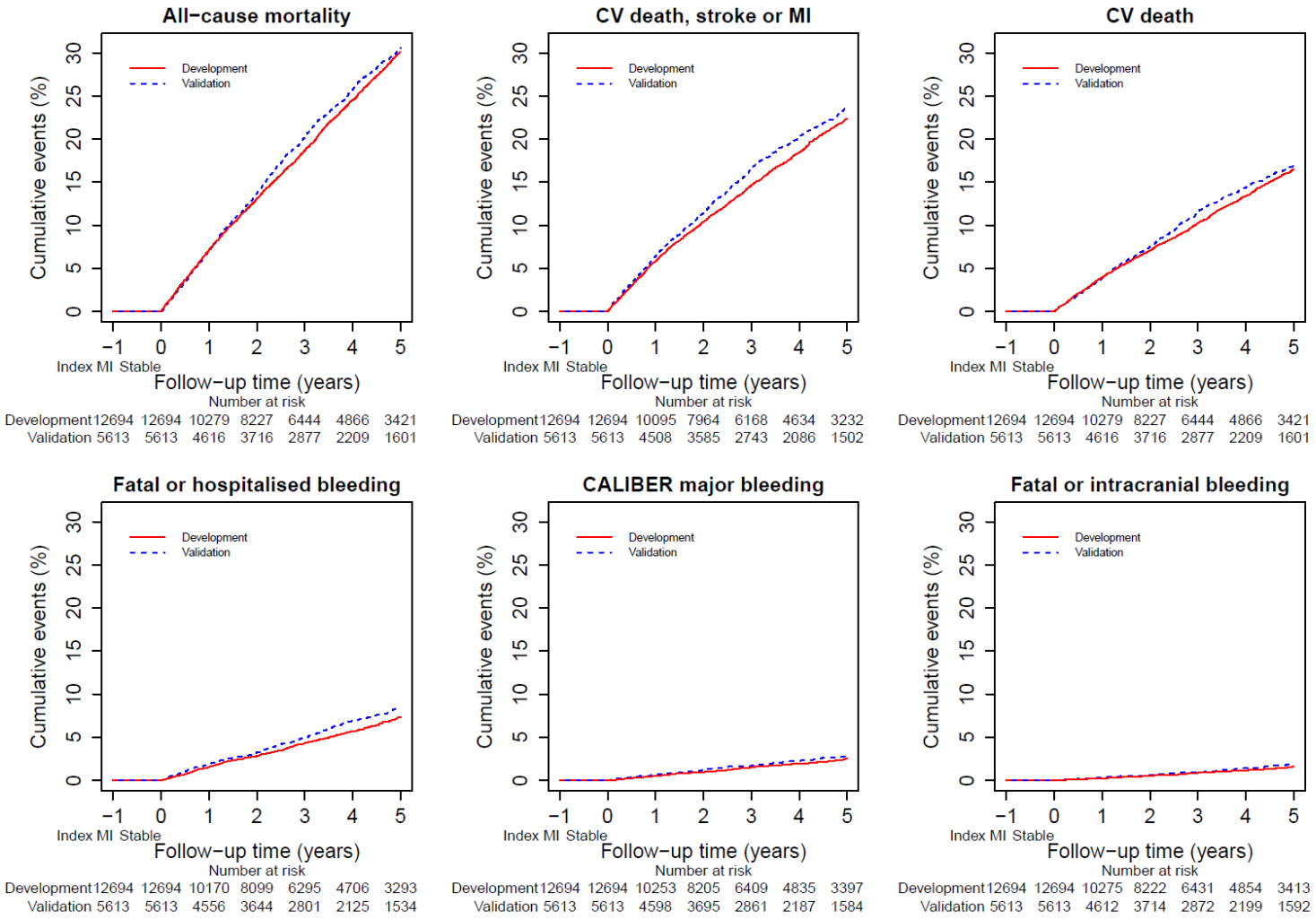
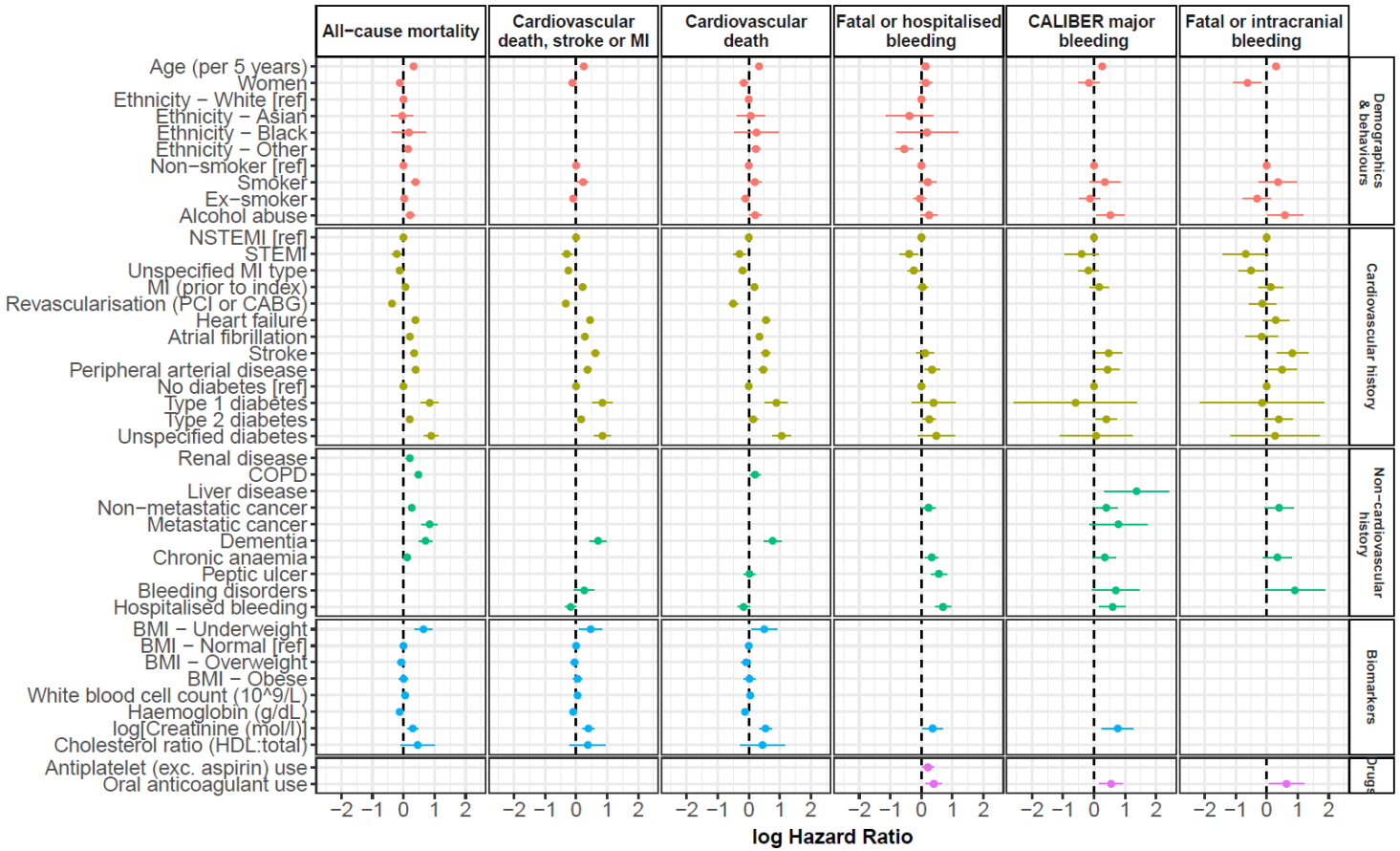


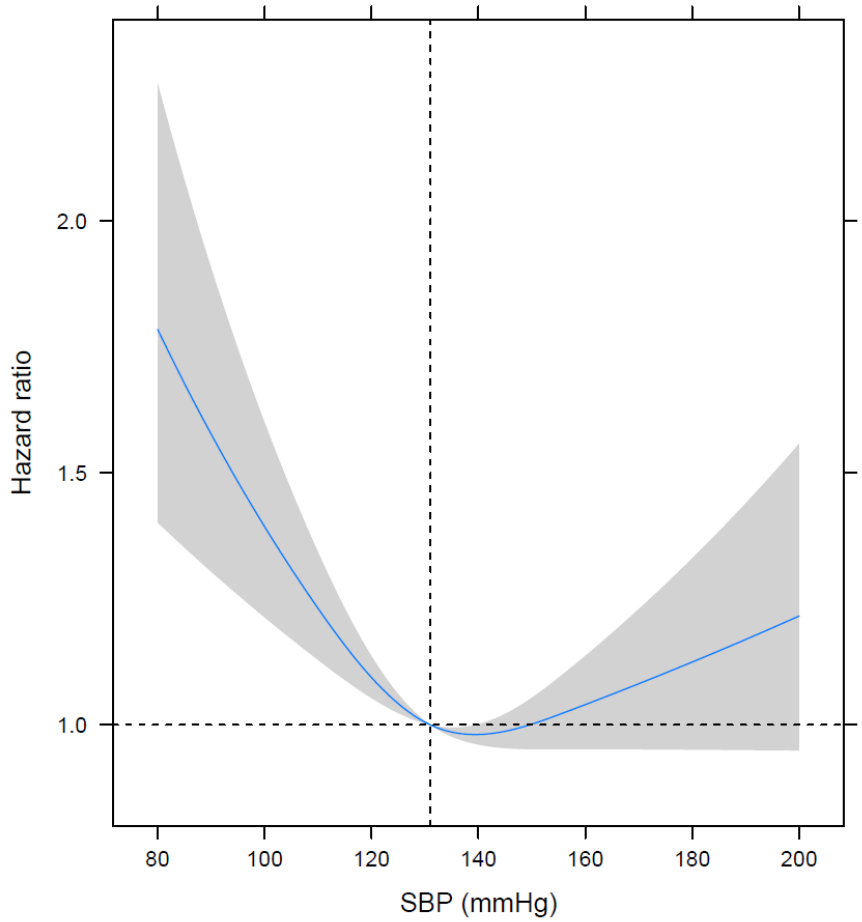
Figure 7.3: Prognostic factors (multivariable) for 5-year all-cause mortality, cardiovascular and bleeding endpoints



Note: CV= cardiovascular, MI= myocardial infarction, NSTEMI= non-ST-elevation myocardial infarction, STEMI= ST-elevation myocardial infarction, PCI= percutaneous coronary intervention, CABG= coronary artery bypass graft, BMI= body mass index, HDL= high-density lipoprotein; log hazard ratios compared with [ref] group for categorical factors or per

unit increase for continuous factors; For hazard ratios and 95% confidence intervals see **Supplementary appendix 11.4.5**. Systolic blood pressure was included in all models using restricted cubic splines [see **Supplementary Appendix 11.4.6**]

Figure 7.4: U-shaped association of systolic blood pressure (SBP) and 5 year cardiovascular death, stroke or myocardial infarction (MI) events [n=12,694, events=1,913]

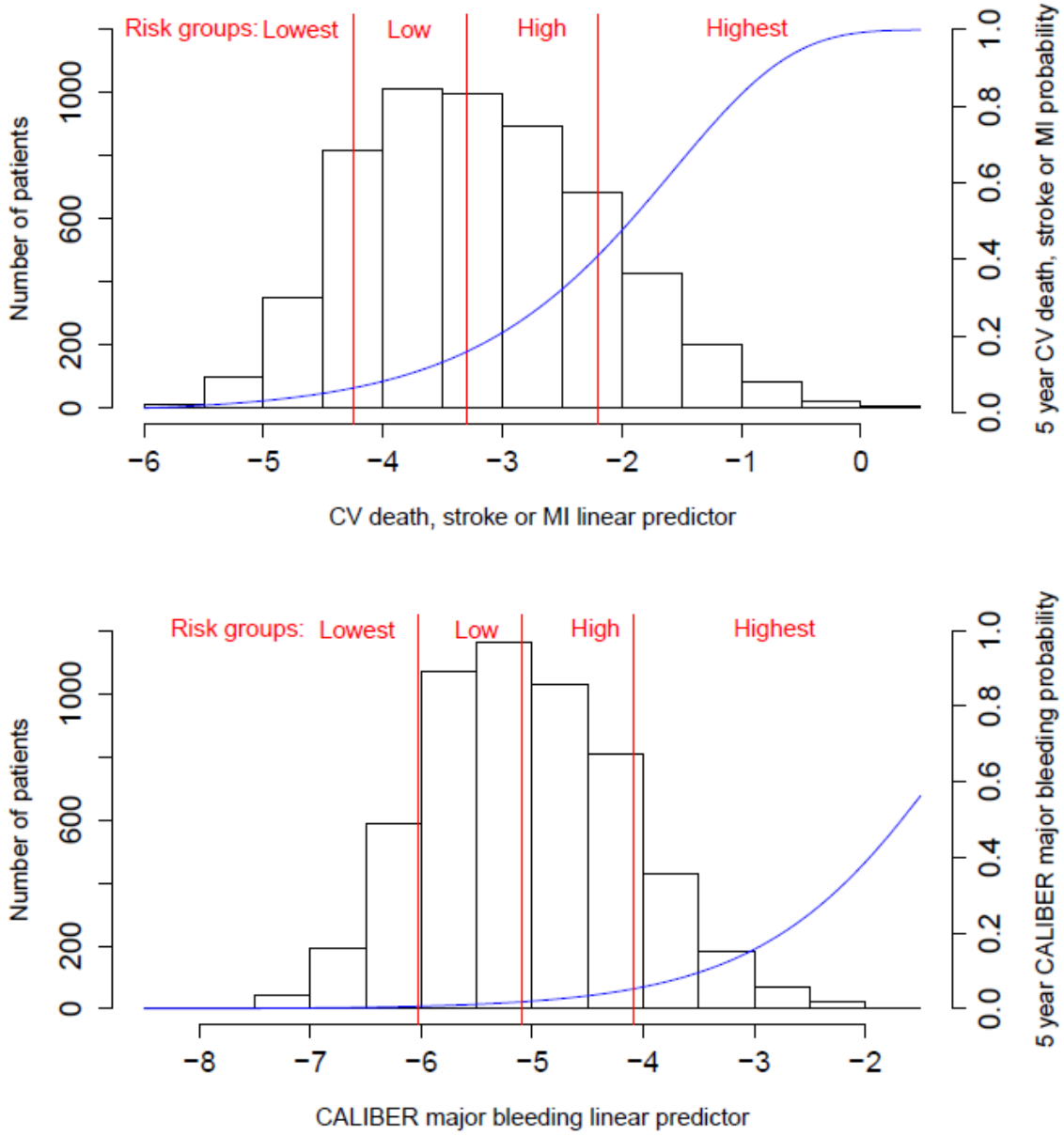


Note: Systolic blood pressure reference value = 136mmHg

Table 7.2: N(%) of patients allocated to each risk group in the development (n=12694) and validation (n=5613) cohorts for each endpoint

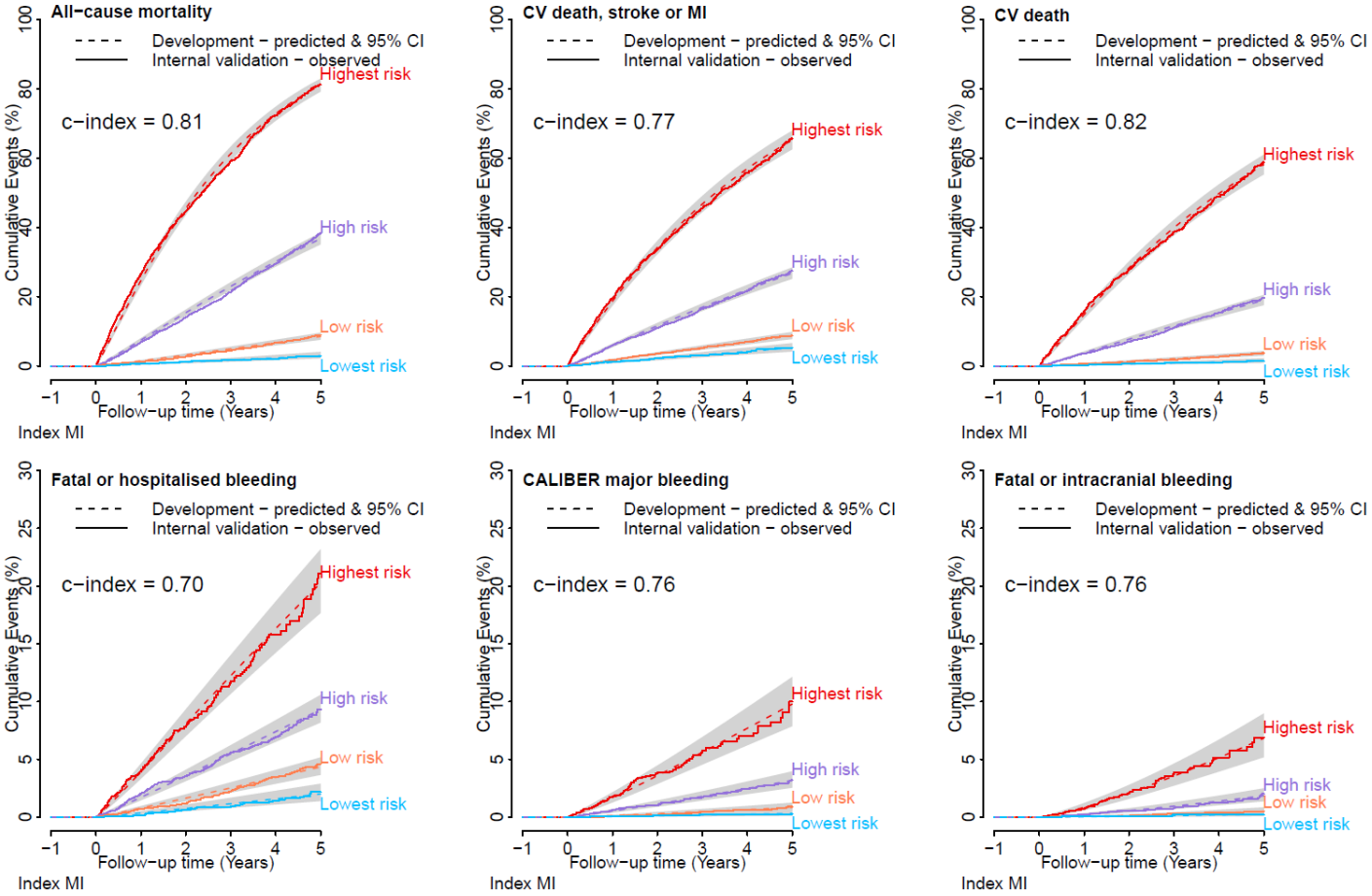
		Risk Group			
Endpoint	**Cohort**	**Highest**	**High**	**Low**	**Lowest**
All-cause mortality	Development	2031 (16.0)	4316 (34.0)	4316 (34.0)	2031 (16.0)
	Validation	974 (17.4)	1941 (34.6)	1874 (33.4)	824 (14.7)
Cardiovascular death, stroke or MI	Development	2031 (16.0)	4316 (34.0)	4316 (34.0)	2031 (16.0)
	Validation	993 (17.7)	1922 (34.2)	1886 (33.6)	812 (14.5)
Cardiovascular death	Development	2031 (16.0)	4316 (34.0)	4316 (34.0)	2031 (16.0)
	Validation	973 (17.3)	1933 (34.4)	1894 (33.7)	813 (14.5)
Hospitalised bleeding	Development	2031 (16.0)	4316 (34.0)	4316 (34.0)	2031 (16.0)
	Validation	1037 (18.5)	1880 (33.5)	1866 (33.2)	830 (14.8)
CALIBER major bleeding	Development	2031 (16.0)	4316 (34.0)	4316 (34.0)	2031 (16.0)
	Validation	939 (16.7)	1878 (33.5)	1940 (34.6)	856 (15.3)
Fatal or intracranial bleeding	Development	2031 (16.0)	4316 (34.0)	4316 (34.0)	2031 (16.0)
	Validation	901 (16.1)	1892 (33.7)	1958 (34.9)	862 (15.4)

Figure 7.5: Cumulative probability of cardiovascular death, stroke or MI (top) and CALIBER major bleeding (bottom) across their respective linear predictors in the validation cohort (n= 5613)



Note: CV= cardiovascular; MI= myocardial infarction

Figure 7.6: Internal validation: comparing observed events with model predictions in the development cohort



Note: CV=cardiovascular; MI= myocardial infarction; CI= confidence interval

Table 7.3: Hazard ratios comparing patient risk groups in development and validation cohorts for all-cause mortality, cardiovascular and bleeding end-points

	Hazard ratio (95% CI)		
	Highest vs. Lowest	**High vs. Lowest**	**Low vs. Lowest**
Potential Benefits			
All-cause mortality			
Development	54.9 (39.7, 75.8)	15.1 (10.9, 20.9)	2.9 (2.1, 4.1)
Validation	41.0 (26.2, 64.0)	11.8 (7.5, 18.4)	2.5 (1.6, 4.1)
Cardiovascular death, stroke or MI			
Development	19.6 (15.3, 25.1)	5.8 (4.5, 7.4)	1.7 (1.3, 2.2)
Validation	11.5 (8.4, 15.5)	3.3 (2.5, 4.5)	1.2 (0.9, 1.7)
Cardiovascular death			
Development	59.8 (37.4, 95.5)	14.5 (9.1, 23.2)	2.5 (1.5, 4.1)
Validation	33.0 (18.9, 57.4)	8.2 (4.7, 14.3)	1.7 (1.0, 3.2)
Potential harms			
Fatal or hospitalised bleeding			
Development	11.2 (7.5, 16.6)	4.8 (3.3, 7.1)	2.2 (1.5, 3.3)
Validation	6.3 (3.7, 10.6)	3.9 (2.4, 6.5)	2.1 (1.2, 3.5)
CALIBER major bleeding			
Development	32.1 (11.8, 87.6)	10.1 (3.7, 27.5)	2.6 (0.9, 7.5)
Validation	13.3 (4.8, 37.3)	3.7 (1.2, 10.6)	2.6 (0.9, 7.6)
Fatal bleeding or intracranial bleeding			
Development	28.9 (9.0, 92.7)	7.6 (2.4, 24.6)	2.0 (0.6, 7.1)
Validation	3.1 (1.7, 5.9)[a]	1.3 (0.7, 2.3)[b]	-

[a] Highest vs Low: No fatal bleeding or intracranial bleeding events in the validation cohort lowest risk group, thus we obtain inflated hazard ratio estimates when using lowest risk as the reference group. The low risk group is used as the reference group instead.

[b] High vs Low: No fatal bleeding or intracranial bleeding events in the validation cohort lowest risk group, thus we obtain inflated hazard ratio estimates when using lowest risk as the reference group. The low risk group is used as the reference group instead

Figure 7.7: Geographical validation: comparing observed events with model predictions in the validation cohort

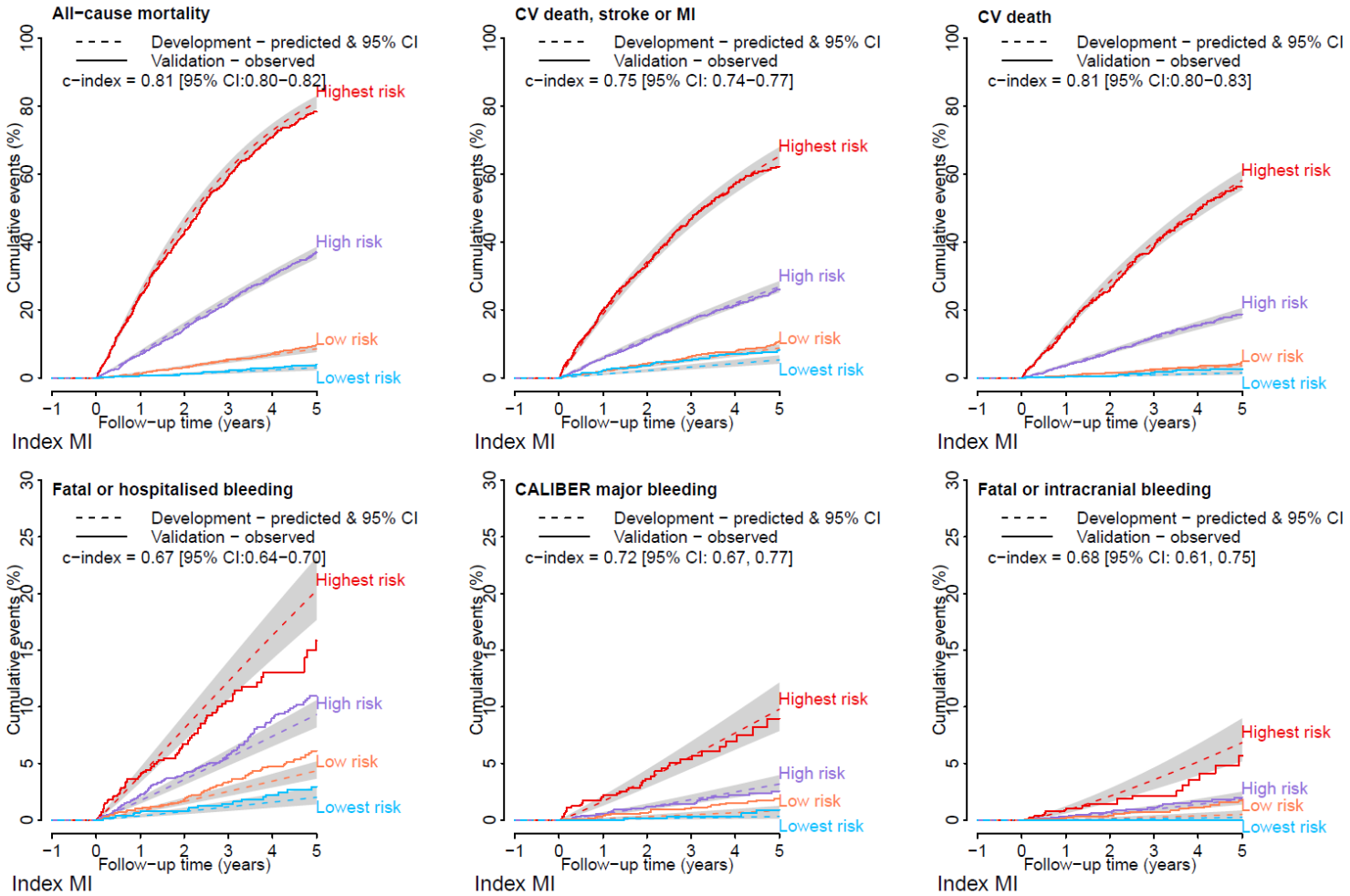
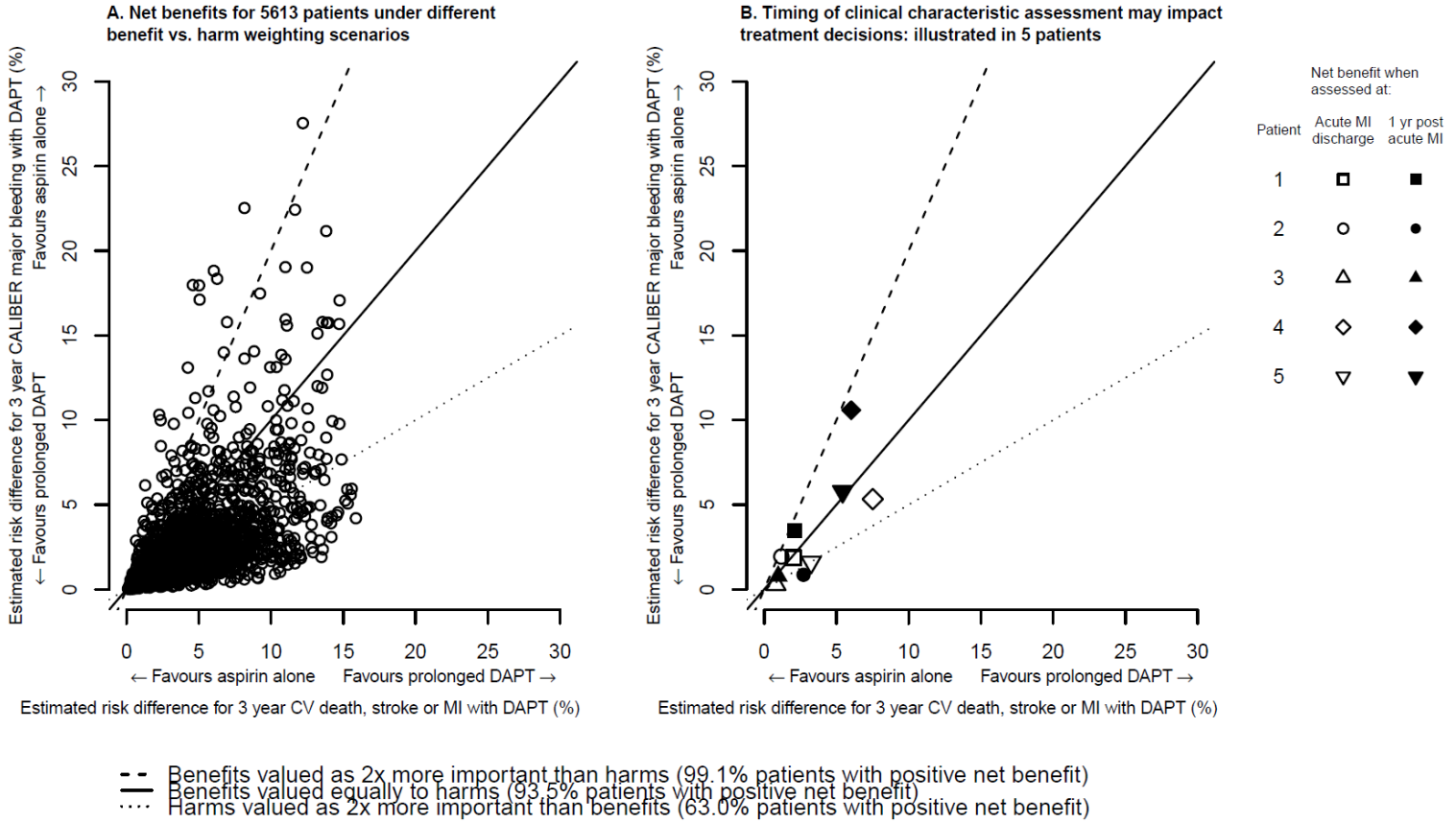


Table 7.4: Estimated events prevented and harms caused per 10,000 patients treated per year with prolonged dual antiplatelet therapy by predicted risk groups compared with all risk groups combined and the trial population. Calculated using PEGASUS-TIMI 54 trial relative risk estimates

	UNSELECTED POPULATION					TRIAL POPULATION
	Risk groups as defined by prognostic models in the validation cohort (n=5,613)					
	Lowest	Low	High	Highest	All	PEGASUS-TIMI 54 placebo arm
Potential benefits						
Cardiovascular death, stroke or MI						
3 year cumulative risk, % (95% CI)	5.2 (3.4, 6.9)	6.3 (5.0, 7.5)	17.1 (15.2, 19.0)	46.7 (42.7, 50.3)	16.5 (15.4, 17.6)	9.04
Events potentially prevented (95% CI) (ITT)	28 (19, 37)	34 (27, 41)	92 (81, 102)	249 (228, 269)	89 (83, 94)	42
Cardiovascular death						
3 year cumulative risk, % (95% CI)	1.5 (0.5, 2.6)	2.5 (1.7, 3.3)	12.2 (10.5, 13.9)	38.5 (34.6, 42.1)	11.5 (10.5, 12.4)	3.39
Events potentially prevented (95% CI) (ITT)	9 (3, 15)	15 (10, 19)	70 (60, 79)	219 (197, 239)	66 (60, 71)	18
All-cause mortality						
3 year cumulative risk, % (95% CI)	2.1 (0.9, 3.2)	5.4 (4.2, 6.5)	22.2 (20.1, 24.2)	59.2 (55.5, 62.5)	20.3 (19.1, 21.4)	5.16
Events potentially prevented (95% CI) (ITT)	8 (4, 12)	20 (16, 24)	82 (74, 89)	218 (204, 230)	75 (71, 79)	16
Potential harms						
CALIBER major bleeding						
3 year cumulative risk, % (95 % CI)	0.3 (0.0, 0.8)	1.0 (0.5, 1.5)	1.4 (0.8, 2.0)	5.4 (3.5, 7.2)	1.7 (1.3, 2.0)	1.26*(ITT); 1.06*(OT)
Harms potentially caused (95% CI) (ITT)	9 (0, 20)	26 (14, 39)	36 (21, 51)	134 (87, 181)	42 (32, 51)	31
Harms potentially caused (95% CI) (OT)	15 (0, 35)	46 (24, 68)	63 (36, 89)	236 (153, 318)	73 (56, 90)	47
Fatal bleeding or intracranial bleeding						
3 year cumulative risk, % (95 % CI)	0	0.7 (0.3, 1.1)	1.1 (0.5, 1.6)	2.2 (0.9, 3.4)	0.9 (0.6, 1.2)	0.60
Harms potentially caused (95% CI) (OT)	-	5 (2, 8)	8 (4, 11)	15 (7, 23)	7 (5, 9)	4
Fatal or hospitalised bleeding						
3 year cumulative risk, % (95 % CI)	1.4 (0.5, 2.3)	3.3 (2.4, 4.3)	5.8 (4.5, 7.0)	10.5 (8.0, 13.0)	4.9 (4.3, 5.6)	Not available†

Note: PEGASUS-TIMI 54 trial estimated relative risks [ticagrelor 60mg versus placebo; intention-to-treat (ITT) and on treatment (OT) estimates where available] for cardiovascular death, stroke or MI [ITT: 0.84 (95% CI: 0.74, 0.95), main report], cardiovascular death [ITT: 0.83 (95% CI: 0.68, 1.01), main report], all-cause mortality [ITT: 0.89 (0.76, 1.04), main report], TIMI major bleeding [ITT: 1.75 (95% CI:), appendix E; OT: 2.32 (95% CI: 1.68, 3.21), main report], fatal bleeding or intracranial bleeding [OT:1.20 (95% CI:0.73, 1.97), main report] were used to calculate CV events potentially prevented and bleeding harms potentially caused per 10,000 treated per year; *TIMI-Major bleeding; †No broad/ composite bleeding endpoint reported in PEGASUS-TIMI 54

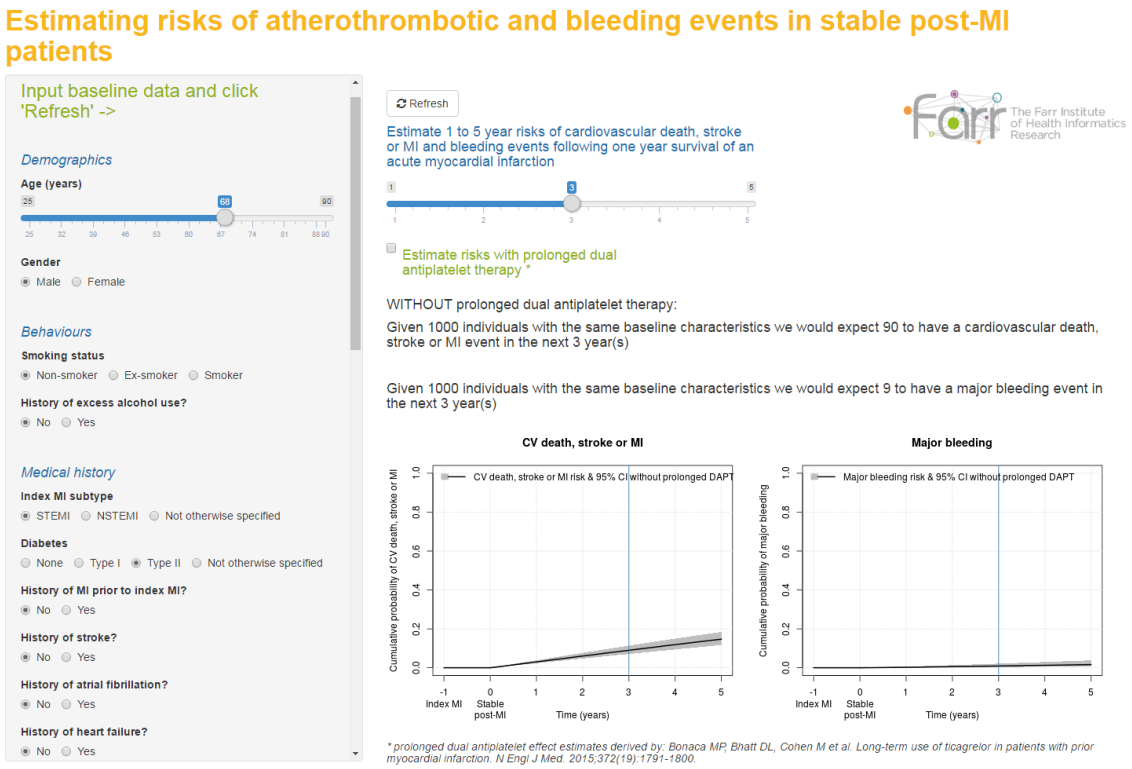
Figure 7.8: Net predicted risk for cardiovascular death, stroke or MI and CALIBER major bleeding with prolonged dual antiplatelet therapy



MI= Myocardial infarction; CV=cardiovascular; DAPT= dual antiplatelet therapy; The straight lines correspond to different scenarios of how patients or clinicians may value (or weight) benefits and harms. All patients to the right of a line applicable to their values would be estimated to have a positive net benefit with prolonged DAPT.

Panel A. Net benefit for 5613 patients under different benefit vs. harm weighting scenarios. Panel B. Timing of clinical characteristic assessment may impact treatment decisions shows the net benefits calculated using patient characteristics at discharge from acute MI and at 1 year post-acute MI for 5 typical patients in the validation cohort.

Figure 7.9: Screenshots of risk prediction application displaying 3 year predicted risks of CV death, stroke or MI and major bleeding for patient without prolonged dual antiplatelet therapy (top) and with prolonged dual antiplatelet therapy (bottom)



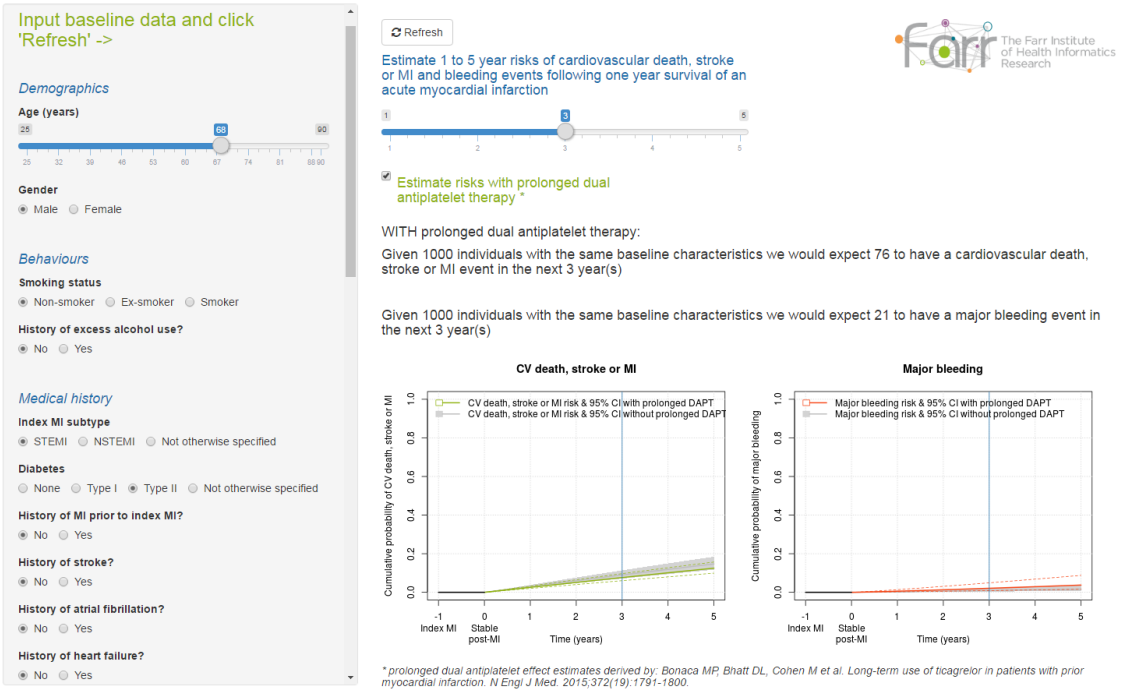
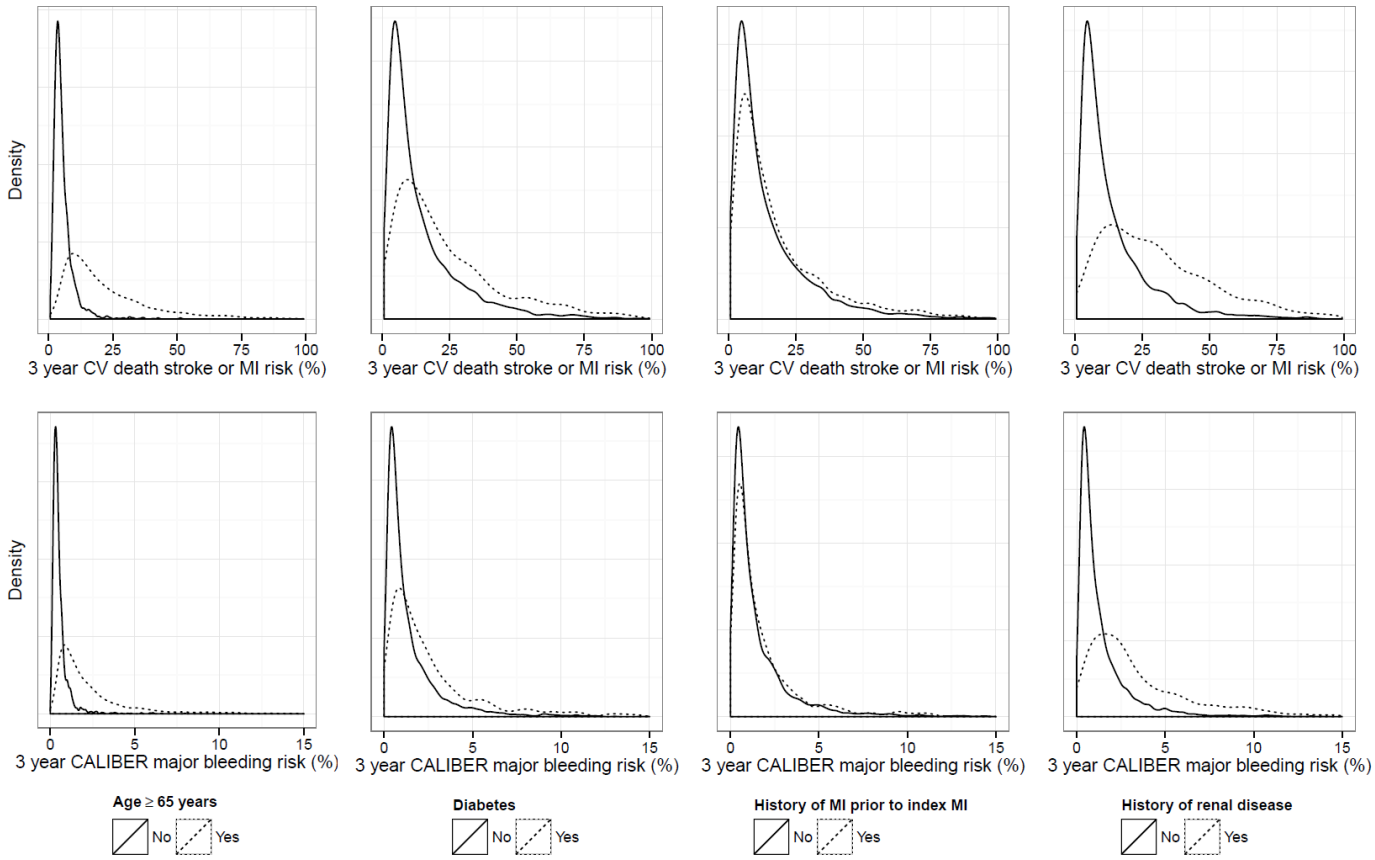
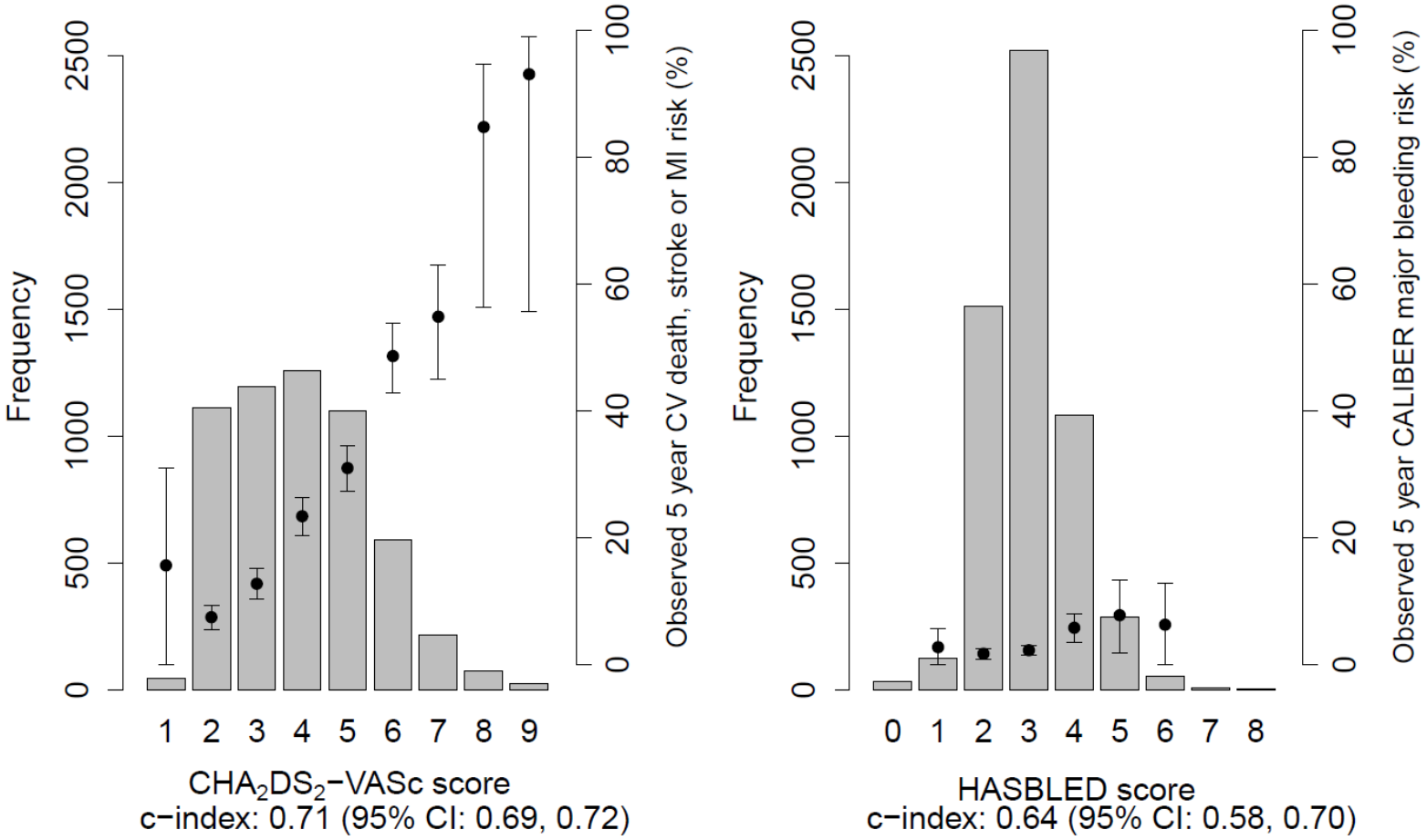


Figure 7.10: Overlap of 3 year predicted risks in the validation cohort (n=5613) based on multivariable models in those with and without categorical risk factors: age≥ 65, diabetes, history of MI and renal disease (used to define high risk in the PEGASUS-TIMI 54 trial)



Note: Each panel shows the distribution of predicted 3-year CV or bleeding risks in patients with and without 4 binary indicators of high CV risk (Age, diabetes, MI history, renal disease). The dotted curve is the distribution of risk for patients with the high risk factor and the solid curve is the distribution of risk without the high risk factor. I demonstrated that these 'high risk' factors alone are insufficient to separate patients who are truly at higher and lower risks. CV= cardiovascular; MI= myocardial infarction

Figure 7.11: The distribution of CHA_2DS_2-VASc and HASBLED scores and their observed risks of cardiovascular death, stroke or MI and CALIBER major bleeding respectively in the validation cohort (n=5613)



8 Predictors and outcomes of INR time in therapeutic range: a linked-electronic health record study

Chapter Summary

Background

Oral anticoagulants (OAC) are a widely used class of drugs primarily for stroke prevention in atrial fibrillation and treating venous thromboembolism. The therapeutic range for OAC treatment is particularly narrow. If overtreated, patients are at increased risk of major bleeding or if undertreated patients may be at increased risk of atherothrombotic events. The international normalised ratio (INR) is a widely used biomarker for patients being treated with OAC in order to measure treatment control. The most common measure of INR control is percent time in therapeutic range (TTR). The objectives of this analysis were to describe INR and TTR data within CALIBER within 4 disease populations indicated for oral anticoagulation, to assess patient baseline characteristics associated with time in therapeutic range, to determine whether a risk score developed for predicting treatment control are effective in the study population and to assess the association between TTR and risk of all-cause mortality, atherothrombotic and bleeding events.

Methods

Using CALIBER linked electronic health records, the study population comprised of 18823 patients prescribed OAC and at least 5 consecutive INR records sorted into atrial fibrillation, venous thromboembolism, heart valve replacement or other indication groups. The distribution of INR records and TTR in the study population was examined. The SAME-TT_2R_2 score was applied to the study population and their predicted INR control was compared observed TTR values. A wide range of clinical risk factors were assessed for association with TTR using linear regression models. Cox proportional hazard models were used to investigate association between TTR and the risk of outcomes.

Results

Of the 541770 INR records included in this analysis 309097 (57.1%) were in the range of 2-3. The mean (SD) TTR was 62.8% (21.4) and 9800 (52.1%) patients had a TTR greater than 65%. TTR varied across the indication groups from 66.2% (19.8) in atrial fibrillation patients to 51.2% (22.7) in heart valve replacement patients. The SAME-TT_2R_2 score did not perform well in the

study population (c-index≤0.55 in all models stratified by indication and with different TTR cut-offs). Older patients were found to have higher TTR, whilst higher deprivation, smoking and a number of comorbidities were associated with lower TTR. Patients with TTR>65% were found to be at lower risk of all-cause mortality, atherothrombotic events and bleeding compared to those with TTR≤65%.

Conclusion

In the absence of guidelines recommended targets for TTR amongst patients venous thromboembolism, heart valve replacement and other indications for oral anticoagulation, the study results supports maintaining treatment control as measured by TTR above 65%.

8.1 Introduction

Oral anticoagulants are a widely used class of drugs primarily for stroke prevention in atrial fibrillation, treating venous thromboembolism and prevent atherothrombotic events in patients who underwent heart valve replacement. Vitamin K antagonists (VKA) are currently the class of oral anticoagulants most commonly used for long-term treatment. The guidelines for VKA therapy control in atrial fibrillation, venous thromboembolism and heart valve replacement patients are summarised in the literature chapter **(Table 2.7).**

The therapeutic range for VKA treatment is particularly narrow. If over treated with VKAs, patients are at increased risk of major bleeding or if undertreated patients may be at increased risk of atherothrombotic events. VKAs are particularly sensitive drugs and their effectiveness may be altered by comorbidities, concomitant medications, diet and alcohol consumption[103]. Guidelines for treatment of patients with venous thromboembolism[33,34,106] suggests initial 3 months of VKA therapy with the potential to extend the duration for long term secondary prevention of atherothrombotic events in high risk patients. Patients with mechanical heart valve replacements are recommended to undergo lifelong VKA therapy.[107,108] Patients with bio-prosthetic heart valve replacements are recommended to undergo 3 months of VKA therapy.[107,108] The intensity of treatment defined by the therapeutic range is dependent on the presence of additional risk factors for atherothrombotic events.

The international normalised ratio (INR) is widely used biomarker for patients being treated with VKAs in order to measure treatment control and requires patients to undergo regular blood tests. The most common measure of INR control is percent time in therapeutic range (TTR). In calculating TTR a commonly used assumption is linear change between consecutive INR measures, known as Rosendaals method.[105] The SAME-TT_2R_2 score[117] is a clinical prediction

tool that was developed to estimate the likelihood of achieving good anticoagulation control defined by TTR in individual patients.

NICE[4] and ESC[5] guidelines for treating atrial fibrillation differ in recommending an ideal minimum TTR threshold, ranging from 65% to 70%. However there are no such recommendations for TTR for other diseases indicated to be treated with VKAs, such as venous thromboembolism and heart valve replacement.

The objectives of this chapter is to address the following objectives: First, to describe INR and TTR data within CALIBER within 4 distinct populations indicated for oral anticoagulation with VKA therapy (AF, VTE, heart valve replacement and other indications).,second to assess patient baseline characteristics associated with time in therapeutic range in the study population and whether my results agree with the results of previous studies, third to determine whether the SAME-TT_2R_2 score is effective for predicting TTR in the study population and fourth I will assess the association between TTR and outcomes (all-cause mortality, atherothrombotic events and bleeding).

8.2 Methods

8.2.1 Study population and index INR spells

I used CALIBER (1997-2010) linked electronic health records. Patients were included if they had at least one prescription for a VKA: warfarin, phenindione or acenocoumarol and a minimum of five consecutive INR records. INR records with value over 20 were not included in this analysis and INR values were considered to be consecutive if they were within 90 days. Sensitivity analyses were carried out in choosing this cut-off. Examples of INR spells with 50 day and 90 day cut-offs between consecutive INR records are shown in **Figure 8.1.** For patients 1 and 2 we see that a 50 day cut-off between consecutive INR records is not sufficient to capture their entire INR spells. For patient 3, although 50 days appears to be sufficient cut-off, the 90 day cut-off is not excessively long and the three distinct spells of INR monitoring are kept as such.

For patients with multiple valid spells of INR monitoring I used their first as their index INR spell.

8.2.2 Indication for oral anticoagulation and INR monitoring

Patients' indication for oral anticoagulation and INR monitoring was determined by records for atrial fibrillation, venous thromboembolism and heart valve replacement days prior to the start of the index INR spell using the following hierarchy:

1) **Venous thromboembolism** - a venous thromboembolism record in primary care or hospital admission records within 30 days prior to the start of the index INR spell
2) **Heart valve replacement** – a heart valve replacement record in primary care or hospital procedure records at any time prior to the start of the index INR spell
3) **Atrial fibrillation** – an atrial fibrillation record in primary care or hospital admissions records at any time prior to the start of the index INR spell
4) **Other indication** – none of the indications as described above

The other indication group represents patients treated with vitamin K therapy for less common indications. In the UK, small numbers of patients with chronic rheumatic heart disease, antiphospholipid syndromes and cardiomyopathy may be recommended to undergo anticoagulation therapy. The phenotypes for venous thromboembolism, heart valve replacement and atrial fibrillation are described in the Data chapter.

8.2.3 Overall distribution of INR data

I assessed the overall distribution of INR records in the study population. I also assessed the frequency of monitoring by estimating the mean time between INR tests at each individual INR value.

8.2.4 Time in therapeutic range

The time in therapeutic range (TTR) is the percent of a patient's INR spell spent within their given prescribed therapeutic range. Within electronic health records, data on the therapeutic range assigned to patients is not available. I therefore assumed a range of 2 to 3 for all patients since this is the most commonly used range for treating and preventing atherothrombotic events. High risk patients with heart valve replacements can be indicated to receive more intense VKA therapy, so in a sensitivity analysis I calculated the TTR using a range of 2.5 to 3.5 in this subgroup. Using Rosendaal's method[105] I applied linear interpolation between consecutive INR records. Assuming a linear trajectory we can estimate the number of days each patient spent within the therapeutic range and the TTR is expressed as the percentage of the INR spell spent within range. Examples of TTR calculated in three INR spells is shown in **Figure 8.2**.

I assessed the distribution of TTR for the index INR spells to provide an overview of the level of anticoagulation control within the study population as a whole and within each indication for oral anticoagulation.

8.2.5 Baseline characteristics

I described the characteristics of the study cohort at baseline as a whole and stratified by indication for oral anticoagulation. Broadly, the characteristics investigated were demographics and behaviours, characteristics of their index INR spell and vitamin K antagonist prescribing, cardiovascular and non-cardiovascular medical history, clinical biomarkers and prescribed medications prior to and during INR spell. Baseline characteristics were described using mean and standard deviation (SD) or median and interquartile range (IQR) for continuous variables and frequencies and percentages for categorical variables. Details of the phenotypes for the characteristics are described in the data chapter.

8.2.5.1 Application of the SAME-TT_2R_2 score

I calculated the SAME-TT_2R_2 score[117] (**Table 8.1**) for each patient at baseline, prior to the start of their INR spell. A single point is scored for each of the following patient characteristics: **s**ex (female), **a**ge<60years, **me**dical history (at least two of hypertension, diabetes, coronary disease, myocardial infarction, peripheral arterial disease, congestive heart failure, stroke, pulmonary disease, hepatic or renal disease), **t**reatment with interacting drugs (amiodarone, calcium channel blockers, beta-blockers or digoxin). Two points are scored for current **t**obacco use and non-Caucasian **r**ace. The SAME-TT_2R_2 score is the sum of the awarded points. The score therefore ranges from 0 to 8 and a high score represents poor predicted anticoagulation.

I described the distribution of the SAME-TT_2R_2 score in the study population produced boxplots of TTR stratified by SAME-TT_2R_2 score.

8.2.6 Endpoints

Patients were followed up for all-cause mortality, cardiovascular death stroke or MI, any bleeding and major bleeding. The composition of the phenotypes for these endpoints is described in the **Chapter 3**.

8.2.7 Statistical analysis

8.2.7.1 Predictors of TTR

Univariable linear regression models were used to assess the association between baseline characteristics and the time in therapeutic range for index INR spells. This included demographics, behaviours, medical history, clinical biomarkers and prescribed medications prior and during index INR spells. The effect estimates were calculated within indication subgroups and in the cohort as a whole.

8.2.7.2 Outcomes following first INR spell

I assessed the incidence of all-cause mortality, cardiovascular death, stroke or MI, any bleeding and major bleeding following INR monitoring stratified by TTR (≤65%, >65%) using Kaplan-Meier plots. The risk of 1 year outcomes was estimated using Cox proportional hazards models. TTR was fitted in models using restricted cubic splines to allow for non-linear effects. The number of spline knot points to be used in each model will be checked in sensitivity analyses. The models were also adjusted for age, gender and length of index INR spell.

Modelling assumptions

The presence of collinear variables in the multivariable models was checked using variance inflation factor estimates and any offending variables were removed.

In Cox proportional hazards models, the proportional hazards assumption was checked using log(-log) plots for categorical variables and plots of time dependent coefficients for continuous variables.

Missing Data

Missing covariate data was multiply imputed using the MICE (multiple imputation with chained equations).[211] 10 imputed data sets were generated.

8.3 Results

8.3.1 Study Population and characteristics by indication

The flow of patients into the study population is displayed in **Figure 8.3.** In CALIBER 46568 patients had at least one oral anticoagulant prescription. Of these patients, 18823 had at least one spell of at least 5 consecutive (within 90 days of each other) INR records in primary care data. Their first valid INR spell was used in the analysis consisting of 541770 INR records in total. The determined indication for oral anticoagulation and INR monitoring was atrial fibrillation for 10132 (53.8%) patients, venous thromboembolism for 3471 (18.4%) patients, heart valve replacement for 744 (4.0%) patients and other indication for 4476 patients (23.8%).

The characteristics of the study population are shown in **Table 8.2.** Overall the mean age of the study population was 69.8 year and 45.4% were women. Atrial fibrillation patients were older (mean age: 73.6 years) than venous thromboembolism, heart valve replacement and other indication patients. More than half of venous thromboembolism patients were women, compared with 42.6% of atrial fibrillation patients, 21.8% of heart valve replacement patients and 28% of other indication patients. Smokers were more prevalent in the venous

thromboembolism and other indication groups compared with the atrial fibrillation and heart valve replacement groups.

Medical histories differed amongst the indication groups. 61% of patients with atrial fibrillation had a history or hypertension compared with 55.2% of heart valve replacement patients, 42.6% of other indication patients and 39.7% of venous thromboembolism patients. 10% of atrial fibrillation and other indication patients had history of myocardial infarction compared 4.7% venous thromboembolism patients and 7.8% heart valve replacement patients. Heart failure was more common in heart valve replacement (33.1%) and atrial fibrillation (22.5%) patients than venous thromboembolism (6.5%) and other indication (13%) patients. Across all patients, history of any or major bleeding was 26.5% and 1.6% respectively, however the heart valve replacement patients had higher prevalence of bleeding history. Patients with venous thromboembolism had a lower prevalence of diabetes. More than 20% of venous thromboembolism and other indication patients had a history of cancer, compared with 16.3% atrial fibrillation patients and 9.1% heart valve replacement patients. Chronic obstructive pulmonary disease and renal disease was more prevalent in atrial fibrillation patients.

Clinical biomarker levels were generally similar amongst patients. The overall mean (SD) body mass index was 29kg/m^2 (6.23) and the mean (SD) systolic blood pressure was 136mmHg (18.8).

More than half of atrial fibrillation patients had been prescribed antiplatelets prior to their index INR spell compared with 27.4% of venous thromboembolism patients, 47.4% of heart valve replacement patients and 32.2% of other indication patients. Antiplatelet prescriptions were less common during index INR spells, 12.4% across all patients. Prescriptions for antiarrhythmic drugs digoxin and amiodarone were similar prior to and during index INR spells across all indications. Overall, 66.7% patients had a prescription for nonsteroidal anti-inflammatory drugs prior to their index INR spell compared to 8.5% during index INR spells. 30.5% of patients had a prescription for an antidepressant prior to their index INR spell and 17.4% had an antidepressant prescription during their index INR spell.

8.3.2 INR data in CALIBER

Of the 541770 INR records included in this analysis 309097 (57.1%) were in the range of 2-3. The distribution of INR records, stratified by indication is shown in **Figure 8.4**. The distribution of INR records was similar amongst the atrial fibrillation, venous thromboembolism and other indications, whereas for heart valve replacement patients the distribution was comparatively skewed to the right.

The mean number of days between INR records by INR value is displayed in **Figure 8.5.** Time between INR tests was highly variable but peaked between INR values of 2 and 3 (i.e. therapeutic range) and was generally lower at the extreme ends of the INR range. For INR records between 2 and 3, the mean (SD) time until next INR test was 26.8 (18.9) days. For INR records lower than 1 the mean (SD) time until next INR test was 9.2 (11.0) days and for INR records greater than 5 the mean (SD) time until the next INR test was 6.0 (7.2) days.

8.3.3 Index INR spell characteristics and time in therapeutic range

Overall the mean (SD) TTR was 62.8% (21.4) and 9800 (52.1%) patients had a TTR greater than 65%. The TTR varied across the indications, from 66.2% (19.8) in atrial fibrillation patients to 51.2% (22.7) in heart valve replacement patients. This is reflected in **Figure 8.6** where it is shown that the distribution of TTR for atrial fibrillation patients is more skewed to the right than for heart valve replacement patients.

Further characteristics of index INR spells are described in **Table 8.3**. In a sensitivity analysis, a therapeutic range of 2.5-3.5 was applied to heart valve replacement patients and the mean (SD) TTR was 49.5% (21.4).

The duration of index INR spells (time difference between first and last INR record in a spell) was variable amongst the indications. The median (IQR) duration was longest for atrial fibrillation and heart valve replacement patients, 477 (196, 959) and 582 (236, 1269) days respectively. For venous thromboembolism patients the median (IQR) duration was 162 (97, 239) days and for other indication the median (IQR) duration was 233 (118, 652) days. Overall the median (IQR) number of INR records in each INR spell was 17 (9, 35). Atrial fibrillation and heart valve replacement patients had higher numbers of INR records per spell compared with venous thromboembolism and other indication patients.

Heart valve replacement and atrial fibrillation patients had higher mean time prescribed to vitamin K antagonists prior to their index INR spell compared with venous thromboembolism and other indication patients. The duration of vitamin K antagonist prescribing was very similar to the duration of index INR spell duration. Overall patients were prescribed vitamin K antagonists for a median of 310 (167, 707) days during their index INR spell.

8.3.4 Application of the SAME-TT_2R_2 score

The majority of the study population had a SAME-TT_2R_2 score of 2 or 3 **(Table 8.4).** No patients were allocated the maximum score of 8 which indicates very high risk of poorly controlled anticoagulation, and only 4 patients (1 atrial fibrillation and 3 other indication) were scored 7.

No venous thromboembolism or heart valve replacement patients were allocated a score higher than 5.

For all indications there appeared to be a slight deceasing trend in mean TTR with increasing SAME-TT_2R_2 score **(Figure 8.7)**, with the exception that the mean (95% CI) TTR for other indication patients and a SAME-TT_2R_2 score of 7 was 71.7% (42.7, 100%), the highest of all of the subgroups. The mean (95% CI) TTR for patients with a SAME-TT_2R_2 score of zero was 68.5% (67.5, 69.6%) for atrial fibrillation patients, 62.9% (61.2, 64.6%) for venous thromboembolism patients, 51.4% (46.9, 55.8%) for heart valve replacement patients and 62.5% (60.9, 64.1%) for other indication patients. The mean (95% CI) TTR for patients with a SAME-TT_2R_2 score of five was 61.2% (57.8, 64.6%) for atrial fibrillation patients, 55.8% (46.9, 64.7%) for venous thromboembolism patients, 31.6% (23.8, 39.4%) for heart valve replacement patients and 52.4% (45.9, 58.9%) for other indication patients

C-indexes were estimated for logistic regression models using the SAME-TT_2R_2 score to predict patients achieving TTR>60, 65 and 70 across all patients and within the indication subgroups (**Table 8.5**). None of the models had a c-index higher than 0.55, suggesting that the SAME-TT_2R_2 score was not much better than chance at predicting TTR in the study population.

8.3.5 Predictors of time in therapeutic range

In using univariable linear regression models I assessed predictors of time in therapeutic range. The effect estimates for demographic and behaviour variables are shown in **Figure 8.8**. Patients in age categories >40 years had mean TTR approximately 10 points higher than patients aged 40 or below. This held true in the atrial fibrillation, venous thromboembolism and other indication populations. Women generally had lower TTR than men, across all patients the difference in mean TTR (95% CI) was -1.4 (-2.0, -0.8). Non-white patients had lower mean TTR than white patients, across all patients the difference was -3.5 (-6.4, -0.5).There did not appear to be any regional variation in mean TTR across England. Patients in the highest quintile of deprivation had lower mean TTR than patients with lower degrees of deprivation, -2.5 (-3.2, -1.7).

Smokers had lower mean TTR than non-smokers, -5.3 (-6.4, -4.3) and patients with a history of alcohol abuse had lower mean TTR than patients without a history of alcohol abuse, -2.1 (-3.0, -1.2).

The effect estimates for medical history variables are shown in **Figure 8.9**. Patients with hypertension had a higher mean TTR than patients without hypertension, 2.1 (1.5, 2.7). This difference was particularly apparent in the other indication group, 2.6 (1.3, 4.0). Atrial

fibrillation patients with a history of myocardial infarction had a lower mean TTR -2.3 (-3.6, -1.0), whereas other indication patients with a history of myocardial infarction had a higher mean TTR, 2.4 (0.2, 4.7). Similarly, atrial fibrillation and venous thromboembolism patients with a history of heart failure had a lower mean TTR, -2.4 (-3.3, -1.5) and -3.8 (-6.8, -0.9) respectively, whereas other indication patients with a history of heart failure had a higher mean TTR 2.2 (0.2, 4.2). Patients with a history of unstable angina had lower mean TTR. Type 1 diabetes patients had mean TTR -9.9 (-14.1, -5.6) points lower than patients without diabetes. Patients with a history of liver disease had a lower mean TTR, -5.4 (-8.7, -2.1), particularly in the heart valve replacement and other indications subgroups. Patients with a history of chronic obstructive pulmonary disease in the atrial fibrillation and venous thromboembolism subgroups had lower mean TTR, -6.2 (-7.5, -5.0) and -3.7 (-6.4, -0.9) respectively. Overall, patients with a history of cancer had a lower mean TTR, -2.5 (-3.2, -1.7), however in the heart valve replacement subgroup patients with cancer had a higher mean TTR, 2.9 (-2.7, 8.7). Patients with a history of chronic anaemia had lower mean TTR, -3.4 (-4.4, -2.5), however histories of either any bleeding or major bleeding did not appear to be associated with TTR.

The effect estimates for clinical biomarker variables are shown in **Figure 8.10.** BMI did not appear to be associated with TTR. There appeared to be a 'J-curve' association between both systolic and diastolic blood pressure and TTR. That is patients with low blood pressure (SBP≤120 mmHg, DBP≤60 mmHg) and high blood pressure (SBP>160 mmHg, DBP>100 mmHg) had lower TTR compared with patients with normal blood pressure (SBP 121-140mmHg, DBP 81-90 mmHg). There was an increasing association between haemoglobin and TTR. Compared to patients with haemoglobin 13-14g/dL, patients with haemoglobin ≤12g/dL had lower mean TTR, -4.8 (-5.9, -3.7), and patients with haemoglobin >14g/dL had higher mean TTR, 3.0 (2.2, 3.9). High white blood cell counts (>11 10^9/L vs. 5-11 10^9/L) and platelet counts (>450 10^9/L vs. 151-450 10^9/L) were associated with lower TTR, -3.4 (-4.8, -2.0) and -6.5 (-8.6, -4.3) respectively. Patients with low creatinine (≤80mmol/l vs. 81-110mmol/l) had lower mean TTR, -2.5 (-3.3, -1.6). Patients with high total cholesterol (>6.3 vs. ≤5.2) had lower mean TTR, -2.6 (-3.9, -1.3).

K-sparing diuretics, loop diuretics, insulin, antidiabetics and antidepressants prescribed prior to index INR spells were associated with lower TTR. On the contrary, antiplatelets, thiazides, ARBs, beta blockers, CCBs, statins and digoxin prescribed prior to the index INR spell were associated with higher TTR (**Figure 8.11).** Similarly, thiazides, ace inhibitors, ARBs, beta blockers, CCBs, statins and digoxin prescribed during index INR spells were associated with higher TTR, whereas prescriptions during index INR spells for insulin, amiodarone and antidepressants were associated with lower TTR (**Figure 8.12).**

8.3.6 The association between TTR and outcomes

All-cause mortality incidence was higher in patients whose TTR was ≤65% compared with patients whose TTR was>65% in the year following INR monitoring. The overall risk of 1 year all-cause mortality was 22.97% (95% CI: 21.97, 23.95%) for patients with TTR≤65% and 15.99 (95% CI: 15.08, 16.89) for patients with TTR>65%. The difference between the groups was similar across the indication subgroups (**Figure 8.13**).

There was a small difference in 1 year cardiovascular death, stroke or MI incidence between patients with TTR≤65% and TTR>65%, 10.69% (95% CI: 9.93, 11.44%) and 9.09% (95% CI: 8.37, 9.80%) respectively. The difference between TTR groups was more apparent for the subgroups of patients with atrial fibrillation and heart valve replacement than for patients with venous thromboembolism or other indications (**Figure 8.13**). The risk of 1 year cardiovascular death, stroke or MI in atrial fibrillation patients was 14.96% (95% CI: 13.62, 16.27%) and 11.42 (95% CI: 10.34, 12.49%) for TTR≤65% and TTR>65% respectively. For venous thromboembolism patients the risk of 1 year cardiovascular death, stroke or MI was 4.78% (95% CI: 3.70, 5.85%) and 3.85% (95% CI: 2.74, 4.94%) for TTR≤65% and TTR>65% respectively.

The incidence of bleeding events in the year following INR monitoring was similar in the TTR groups, 11.95% (95% CI: 11.12, 12.77%) and 11.48% (95% CI: 10.65, 12.31%) for TTR≤65%, and TTR>65% respectively. In the heart valve replacement subgroup, patients with TTR>65% had slightly higher incidence of bleeding at 1 year than patients with TTR≤65% **(Figure 8.14)**.

Similarly, for major bleeding events, one year incidence did not differ greatly between the TTR groups, 3.69% (3.23, 4.15%). Incidence of one year major bleeding was higher for patients with TTR>65% compared with patients with TTR≤65% in the heart valve replacement subgroup **(Figure 8.14)**.

In multivariable analysis I assessed the risk of 1 year events across the full range of TTR values. TTR was fitted in the Cox proportional hazards models using restricted cubic splines with three knots and models were adjusted for age, gender and index INR spell duration.

In all indication subgroups, the risk of one all-cause mortality was approximately halved for patients with 100% TTR compared to patients with 65% TTR **(Figure 8.15)**. The risk of one year all-cause mortality for atrial fibrillation, heart valve replacement and other indication patients peaked and plateaued at TTR of 50% or less. However, there was a more linear association between TTR and all-cause mortality for venous thromboembolism. The risks of cardiovascular death stroke or MI associated with TTR (**Figure 8.15)** were similar to those for all-cause mortality within the indication subgroups.

The association between TTR and bleeding events are shown in **Figure 8.16.** The risk of any bleeding or major bleeding in atrial fibrillation patients is similar for TTR between 0 and 65% and TTR>65% is associated with decreasing risk. The association between TTR and bleeding events appeared to have a reversed J-curve shape for patients with venous thromboembolism. That is, patients had increasing risk from 65% TTR to 0% TTR. For example patients with 20% TTR were estimated to have 2 times the risk of any bleeding and 3 times the risk of major bleeding compared to patients with 65% TTR. There also appeared to be increasing risks of bleeding events from 65% TTR to 100% TTR. Heart valve replacement patients had a similar shaped association between TTR and bleeding events, although the magnitude of effect estimates were much lower. TTR had a linear association with risks of bleeding events for patients with other indications.

8.4 Discussion

In this study I used linked electronic health record to investigate patterns and outcomes of TTR in a population-based cohort of patients in 4 disease populations indicated for oral anticoagulation. The average TTR was 62.8% across the study population. The highest mean TTR was observed in atrial fibrillation patients, who on average spent two thirds of their INR monitoring time within therapeutic range. Distribution of TTR for atrial fibrillation patients was skewed to the right, indicating more patients with higher TTR during their treatment monitoring period. The lowest TTR was observed in heart valve replacement patients, who on average spent only half the time within therapeutic range.

Similar to previous literature, demographic factors of age >40 years, men, white, and a lower deprivation were associated with higher TTR. Lifestyle factors of smoking and history of alcohol abuse were associated with lower TTR. In terms of clinical history, patients with hypertension had a higher TTR, as summarised in the literature (**Figure 2.2**). There appeared to be a 'J-curve' association between both systolic and diastolic blood pressure and TTR. Other disease histories (type 1 diabetes, liver disease, COPD, anaemia) were associated with lower TTR. It is notable that the association between a CVD history (myocardial infarction, heart failure) and TTR was different among patients with atrial fibrillation, compared with patients with other indications. It may be related to the observation that the prescriptions of K-sparing diuretics, loop diuretics, prescriptions prior to index INR spells were associated with lower TTR. This is however to be examined in future. The prescriptions of anti-hypertensive treatment, such as thiazides, ARBs, beta blockers, CCBs, statins, digoxin were associated with higher TTR. Anti-diabetic drug treatment, for example insulin, was associated with lower TTR.

The SAME-TT_2R_2 score did not perform well to predict patients TTR in the data, as increasing in SAME-TT_2R_2 score was observed with little to null change in TTR (c-indexes ranging from 0.52-0.55 across the studied disease populations and TTR cut-off values). However, I found that the predictors used to define the SAME-TT_2R_2 score were associated with lower TTR. The simple point-based scheme and weighting of predictors in SAME-TT_2R_2 may not be sufficient to detect small changes in TTR.

Patients with a TTR lower than the guideline recommended threshold of 65% had a higher risk of all-cause mortality and cardiovascular death, stroke or MI and this was evident across each of the studied disease populations. There was a small bleeding risk increase in patients with TTR lower than 65%. TTR above guideline recommended threshold appeared to be effective in indicating beneficial oral anticoagulation, suggested by a higher atherothrombotic risk reduction and lower bleeding risk.

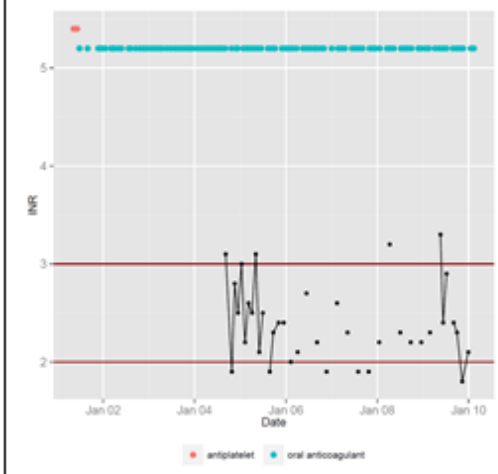
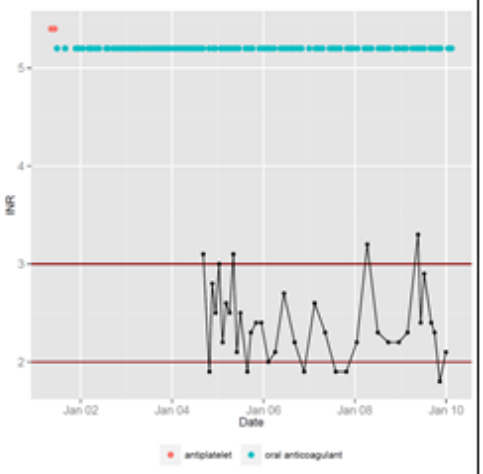
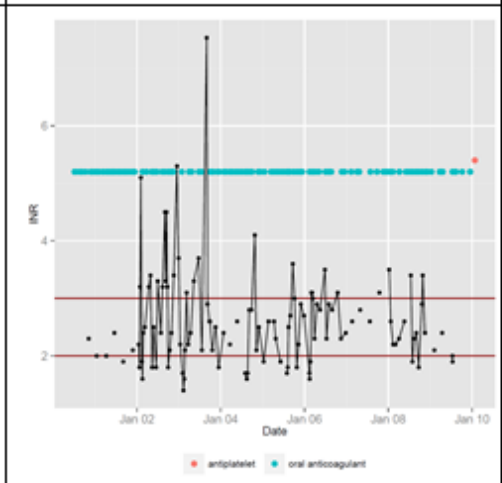
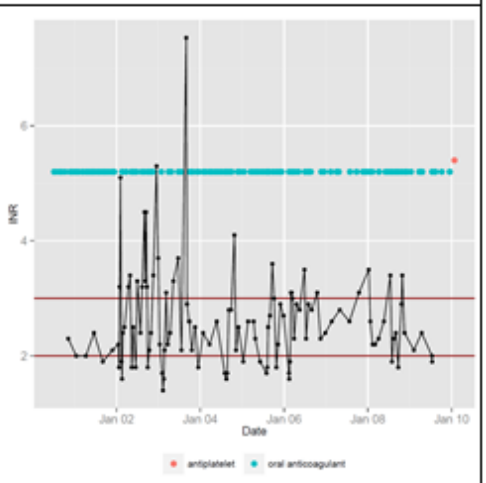
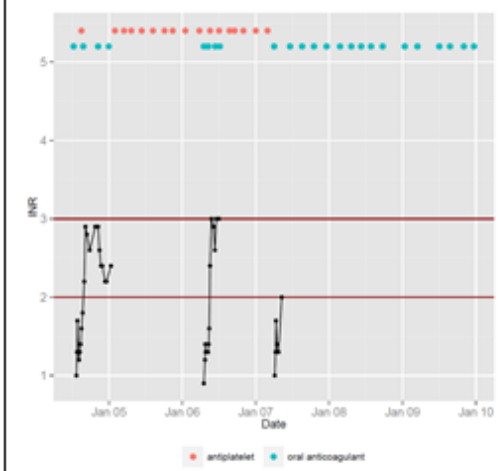
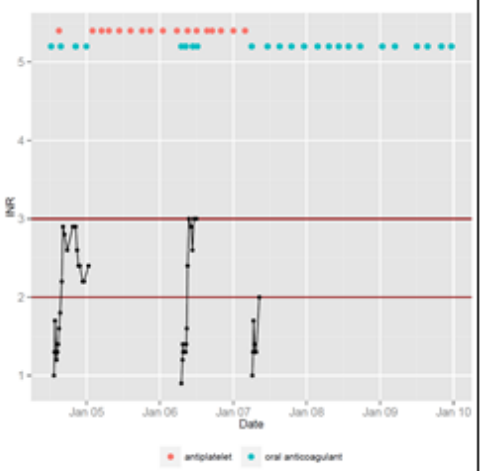
The real-world data of CALIBER has the strength to compare TTR amongst 4 disease populations indicated for oral anticoagulants. A limitation is not all CPRD general practices electronically receive INR records from anticoagulation clinics, therefore the depth and quality of INR data in CALIBER may be lower compared with studies set in anticoagulation clinics. While I restricted the analysis to patients with at least 5 consecutive INR records (the median number of records per patient spell was 17 (IQR: 9, 35)), the TTR estimates for patients with fewer INR records in their oral anticoagulation spell may still not be as accurate as for patients with a larger number of INR records. This is in particular because initially in oral anticoagulation, INR may not be stable until a patients' therapeutic dose is determined. However this data reflects real-world practice and we do not have sufficient evidence to recommend that patients should be monitored more frequently. The precise INR therapeutic range for the patients in the study population was assumed to be 2-3 for all patients. This may have resulted in misspecification of TTR for some heart valve replacement patients whose recommended therapeutic range can vary between 2-3 or 2.5-3.5 depending on their type of valve replacement and their risk profile. However, in my sensitivity analysis, a shifted therapeutic range of 2.5-3.5 was applied to the heart valve replacement population and the change in TTR was small.

My study observed that for patients who maintained a TTR of 65% or greater had improved long term prognosis and reduced bleeding risk. This was true not only for atrial fibrillation patients, but also for patients with venous thromboembolism, heart valve replacements and other indications for oral anticoagulation. In the absence of guidelines recommended targets for TTR amongst these patient groups, the study results supports maintaining oral

anticoagulation treatment control as measured by TTR above 65%. Future randomised clinical trials should investigate thresholds of TTR for effective oral anticoagulation and treatment strategies to improve and maintain a high TTR.

8.5 Tables and Figures

Figure 8.1: INR spells with 50 vs. 90 day cut-off for consecutive INR records

Patient	50 days between consecutive INR records	90 days between consecutive INR records
1		
2		
3		

Note: Green dots represent records for oral anticoagulant prescriptions, orange dots represent records for antiplatelet prescriptions;

Figure 8.2: Examples of time in therapeutic range calculated for three patients

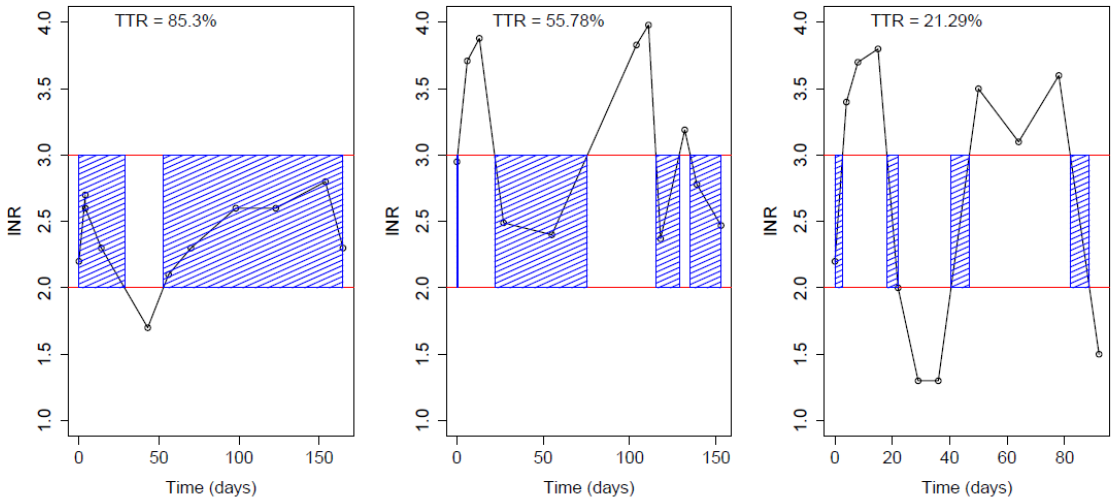
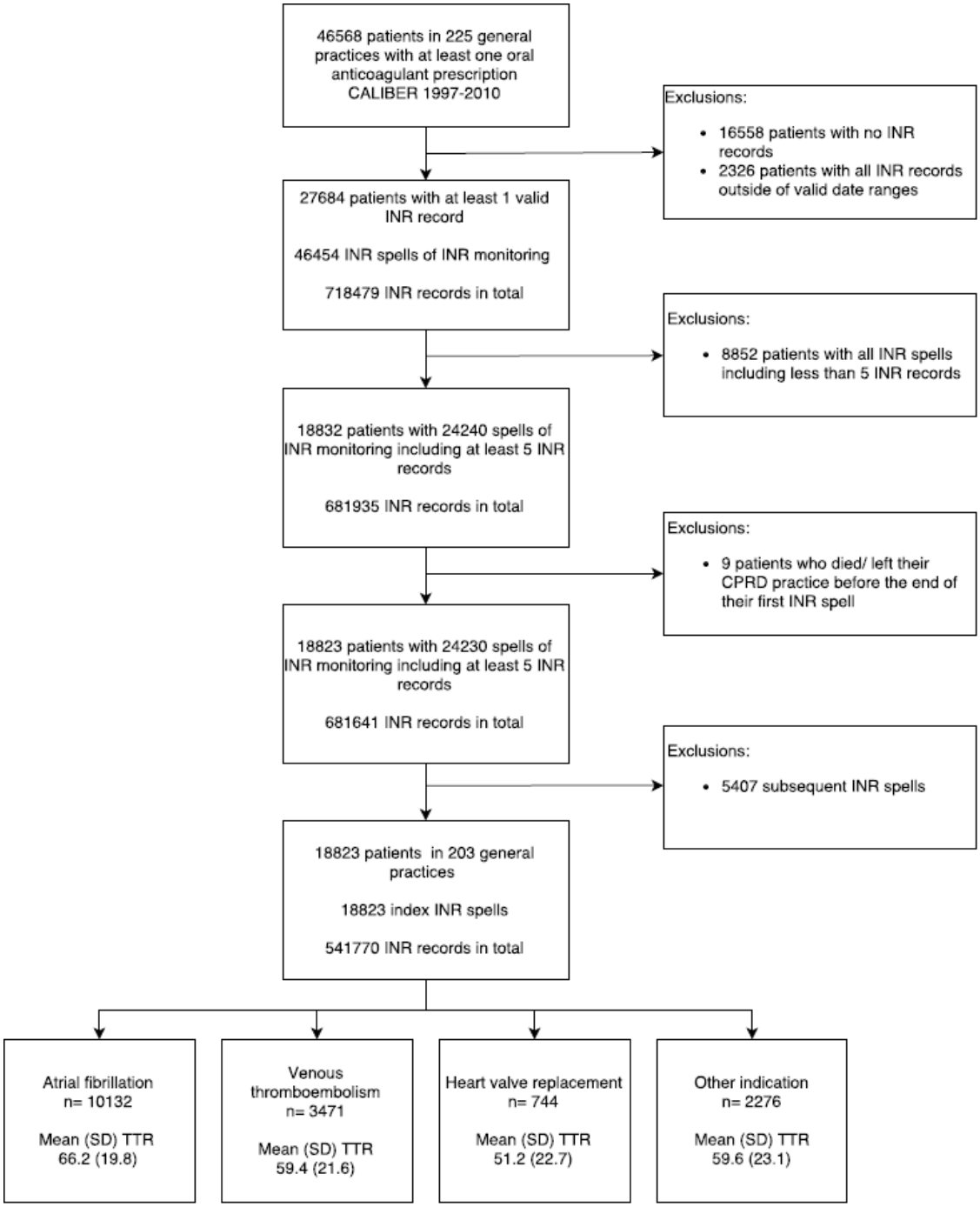


Table 8.1: Same-TT_2R_2 score

Risk factor	Points
Sex (female)	1
Age (<60 years)	1
Medical history (at least two of: hypertension, diabetes, myocardial infarction, peripheral arterial disease, congestive heart failure, history of stroke, pulmonary disease, liver disease, renal disease)	1
Treatment (interacting medications e.g. amiodarone	1
Tobacco use (within two years)	2
Race (non-Caucasian)	2

Figure 8.3: Flow of patients into the study population



Note: INR= international normalised ratio; SD= standard deviation; TTR= time in therapeutic range

Table 8.2: Baseline characteristics of the study population (n=18823) stratified by indication

		Atrial fibrillation (n= 10132)	Venous thrombo-embolism (n= 3471)	Heart valve replacement (n= 744)	Other indication (n= 4476)	All patients (n=18823)
Demographics and behaviours						
Age (Years)		73.6 (9.95)	65.0 (15.61)	64.9 (12.01)	65.8 (15.35)	69.8 (13.3)
Women		4316 (42.6)	1832 (52.8)	244 (32.8)	2147 (48)	8539 (45.4)
Ethnicity: Non-white		89 (0.9)	44 (1.3)	14 (1.9)	50 (1.1)	197 (1)
Ethnicity: White		7729 (76.3)	2664 (76.8)	627 (84.3)	3356 (75)	14376 (76.4)
	Missing %	22.8	22	13.8	23.9	22.6
Practice region	*North West*	1038 (10.2)	245 (7.1)	84 (11.3)	403 (9)	1770 (9.4)
	North East	212 (2.1)	65 (1.9)	11 (1.5)	55 (1.2)	343 (1.8)
	Yorkshire & The Humber	575 (5.7)	136 (3.9)	45 (6)	224 (5)	980 (5.2)
	West Midlands	1091 (10.8)	327 (9.4)	91 (12.2)	420 (9.4)	1929 (10.2)
	East Midlands	538 (5.3)	142 (4.1)	30 (4)	278 (6.2)	988 (5.2)
	East Of England	1415 (14)	496 (14.3)	97 (13)	581 (13)	2589 (13.8)
	London	457 (4.5)	84 (2.4)	33 (4.4)	174 (3.9)	748 (4)
	South West	2426 (23.9)	1178 (33.9)	180 (24.2)	1256 (28.1)	5040 (26.8)
	South Central	1399 (13.8)	500 (14.4)	113 (15.2)	648 (14.5)	2660 (14.1)
	South East Coast	981 (9.7)	298 (8.6)	60 (8.1)	437 (9.8)	1776 (9.4)
Highest quintile of deprivation (IMD)		1941 (19.2)	655 (18.9)	144 (19.4)	1002 (22.4)	3742 (19.9)
	Missing %	0.4	0.2	0.4	0.5	0.4
Alcohol abuse		1408 (13.9)	445 (12.8)	104 (14)	559 (12.5)	2516 (13.4)
	Missing %	0.3	0.2	0.1	0.3	0.3
Smoking status	*Non-Smoker*	4421 (43.6)	1570 (45.2)	309 (41.5)	1847 (41.3)	8147 (43.3)
	Ex-Smoker	4104 (40.5)	1170 (33.7)	298 (40.1)	1514 (33.8)	7086 (37.6)
	Smoker	718 (7.1)	405 (11.7)	64 (8.6)	493 (11)	1680 (8.9)
	Missing %	8.8	9.4	9.8	13.9	10.1
Medical history						
Hypertension		6194 (61.1)	1377 (39.7)	411 (55.2)	1906 (42.6)	9888 (52.5)
Myocardial infarction		975 (9.6)	164 (4.7)	58 (7.8)	449 (10)	1646 (8.7)
Heart failure		2279 (22.5)	224 (6.5)	246 (33.1)	581 (13)	3330 (17.7)
Unstable angina		534 (5.3)	97 (2.8)	60 (8.1)	166 (3.7)	857 (4.6)
Ischaemic or unspecified stroke		685 (6.8)	170 (4.9)	52 (7)	386 (8.6)	1293 (6.9)
Peripheral arterial disease		1117 (11)	116 (3.3)	34 (4.6)	385 (8.6)	1652 (8.8)
Any bleeding		2726 (26.9)	875 (25.2)	236 (31.7)	1156 (25.8)	4993 (26.5)
Major bleeding		137 (1.4)	65 (1.9)	24 (3.2)	81 (1.8)	307 (1.6)
Chronic anaemia		1074 (10.6)	443 (12.8)	140 (18.8)	551 (12.3)	2208 (11.7)
Diabetes	*Unspecified*	74 (0.7)	15 (0.4)	5 (0.7)	34 (0.8)	128 (0.7)
	Type 1	41 (0.4)	11 (0.3)	4 (0.5)	40 (0.9)	96 (0.5)
	Type 2	1275 (12.6)	292 (8.4)	86 (11.6)	453 (10.1)	2106 (11.2)
Cancer		1652 (16.3)	746 (21.5)	68 (9.1)	941 (21)	3407 (18.1)
Chronic obstructive pulmonary disease		979 (9.7)	259 (7.5)	60 (8.1)	352 (7.9)	1650 (8.8)
Renal disease		1441 (14.2)	382 (11)	87 (11.7)	459 (10.3)	2369 (12.6)
Liver disease		59 (0.6)	45 (1.3)	10 (1.3)	52 (1.2)	166 (0.9)

		Atrial fibrillation (n= 10132)	Venous thrombo-embolism (n= 3471)	Heart valve replacement (n= 744)	Other indication (n= 4476)	All patients (n=18823)
Clinical biomarkers						
Body mass index		28.8 (6.08)	29.8 (6.65)	28.7 (5.66)	29.1 (6.40)	29 (6.23)
	Missing %	58.4	66.5	61.7	66.5	61.9
Systolic blood pressure (mmHg)		137 (18.7)	136 (18.1)	134 (18.1)	136 (19.8)	136 (18.8)
	Missing %	12	29.8	13.4	28	19.1
Diastolic blood pressure (mmHg)		79.2 (10.9)	79.1 (10.7)	75.9 (10.8)	78.8 (10.6)	79 (10.8)
	Missing %	12	29.8	13.4	28	19.1
Heart rate (bpm)		80.8 (19.7)	80.1 (17.7)	78.6 (18.7)	81.2 (18.7)	80.7 (19.3)
	Missing %	78.9	89.3	85.8	88.8	83.4
Haemoglobin (g/dL)		13.9 (1.61)	13.2 (1.85)	13.2 (1.93)	13.4 (1.87)	13.7 (1.75)
	Missing %	41.8	47.7	46.4	48.9	44.8
Total white blood cell count (10^9/L)		7.49 (2.47)	7.89 (3.13)	7.62 (2.49)	7.63 (2.71)	7.59 (2.66)
	Missing %	43.8	49.9	48.4	50.8	46.7
Platelets (10^9/L)		248 (79.9)	281 (107.1)	267 (112.3)	271 (99.6)	259 (92)
	Missing %	43.7	49.8	48.5	51.4	46.9
Creatinine (mmol)		101.9 (32.5)	96.9 (35.0)	101.8 (44.7)	99.2 (36.8)	101 (34.4)
	Median (IQR)	97 (84, 112)	90 (78, 107)	97 (85, 111)	94 (80, 110)	95 (82, 111)
	Missing %	28.2	45.3	36.3	43.7	35.3
Hba1c (mmol/mol)		55.1 (15.4)	57.0 (18.0)	55.5 (13.6)	58.5 (17.7)	56.1 (16.3)
	Missing %	87.4	92.3	89.8	90.1	89
High-density lipoproteins (mmol)		1.39 (0.419)	1.43 (0.474)	1.36 (0.405)	1.39 (0.457)	1.39 (0.433)
	Missing %	60.4	77.7	62.4	74.7	67.1
Total cholesterol (mmol)		4.72 (1.11)	5.14 (1.17)	4.71 (1.13)	4.96 (1.28)	4.82 (1.16)
	Missing %	44.8	68.9	43.1	63.3	53.6
Prescribed medication before and during INR spell						
Antiplatelets	*pre*	5914 (58.4)	950 (27.4)	353 (47.4)	1443 (32.2)	8660 (46)
	during	1437 (14.2)	252 (7.3)	132 (17.7)	521 (11.6)	2342 (12.4)
Digoxin	*pre*	4546 (44.9)	118 (3.4)	159 (21.4)	406 (9.1)	5229 (27.8)
	during	4863 (48)	155 (4.5)	144 (19.4)	478 (10.7)	5640 (30)
Amiodarone	*pre*	1353 (13.4)	42 (1.2)	153 (20.6)	117 (2.6)	1665 (8.8)
	during	1354 (13.4)	40 (1.2)	125 (16.8)	181 (4)	1700 (9)
Thiazides	*pre*	4355 (43)	1005 (29)	228 (30.6)	1336 (29.8)	6924 (36.8)
	during	2187 (21.6)	445 (12.8)	109 (14.7)	658 (14.7)	3399 (18.1)
Loop diuretics	*pre*	4034 (39.8)	833 (24)	400 (53.8)	1370 (30.6)	6637 (35.3)
	during	4287 (42.3)	714 (20.6)	341 (45.8)	1307 (29.2)	6649 (35.3)
K-sparing diuretics and aldosterone antagonists	*pre*	1468 (14.5)	290 (8.4)	207 (27.8)	519 (11.6)	2484 (13.2)
	during	1324 (13.1)	165 (4.8)	161 (21.6)	458 (10.2)	2108 (11.2)
K-sparing diuretics with other diuretics	*pre*	1217 (12)	297 (8.6)	146 (19.6)	403 (9)	2063 (11)
	during	458 (4.5)	87 (2.5)	70 (9.4)	146 (3.3)	761 (4)
Beta blockers	*pre*	6074 (59.9)	1022 (29.4)	410 (55.1)	1588 (35.5)	9094 (48.3)
	during	5190 (51.2)	536 (15.4)	337 (45.3)	1180 (26.4)	7243 (38.5)
Ace inhibitors	*pre*	6078 (60)	1447 (41.7)	466 (62.6)	2145 (47.9)	10136 (53.8)

		Atrial fibrillation (n= 10132)	Venous thrombo-embolism (n= 3471)	Heart valve replacement (n= 744)	Other indication (n= 4476)	All patients (n=18823)
	during	4940 (48.8)	722 (20.8)	395 (53.1)	1514 (33.8)	7571 (40.2)
Angiotensin receptor blockers	*pre*	1544 (15.2)	292 (8.4)	107 (14.4)	419 (9.4)	2362 (12.5)
	during	1692 (16.7)	242 (7)	114 (15.3)	416 (9.3)	2464 (13.1)
Calcium channel blockers	*pre*	4281 (42.3)	788 (22.7)	252 (33.9)	1191 (26.6)	6512 (34.6)
	during	3244 (32)	465 (13.4)	144 (19.4)	817 (18.3)	4670 (24.8)
Statins	*pre*	4350 (42.9)	826 (23.8)	399 (53.6)	1392 (31.1)	6967 (37)
	during	4916 (48.5)	748 (21.5)	458 (61.6)	1527 (34.1)	7649 (40.6)
Insulin	*pre*	626 (6.2)	160 (4.6)	44 (5.9)	293 (6.5)	1123 (6)
	during	621 (6.1)	131 (3.8)	53 (7.1)	275 (6.1)	1080 (5.7)
Anti-diabetics	*pre*	945 (9.3)	209 (6)	56 (7.5)	378 (8.4)	1588 (8.4)
	during	1063 (10.5)	192 (5.5)	60 (8.1)	372 (8.3)	1687 (9)
Nonsteroidal anti-inflammatory drugs	*pre*	6699 (66.1)	2543 (73.3)	429 (57.7)	2889 (64.5)	12560 (66.7)
	during	858 (8.5)	293 (8.4)	67 (9)	385 (8.6)	1603 (8.5)
Antidepressants	*pre*	2695 (26.6)	1270 (36.6)	215 (28.9)	1570 (35.1)	5750 (30.5)
	during	1564 (15.4)	646 (18.6)	126 (16.9)	931 (20.8)	3267 (17.4)

Figure 8.4: Distribution of 541770 INR records for 18823 patients stratified by indication

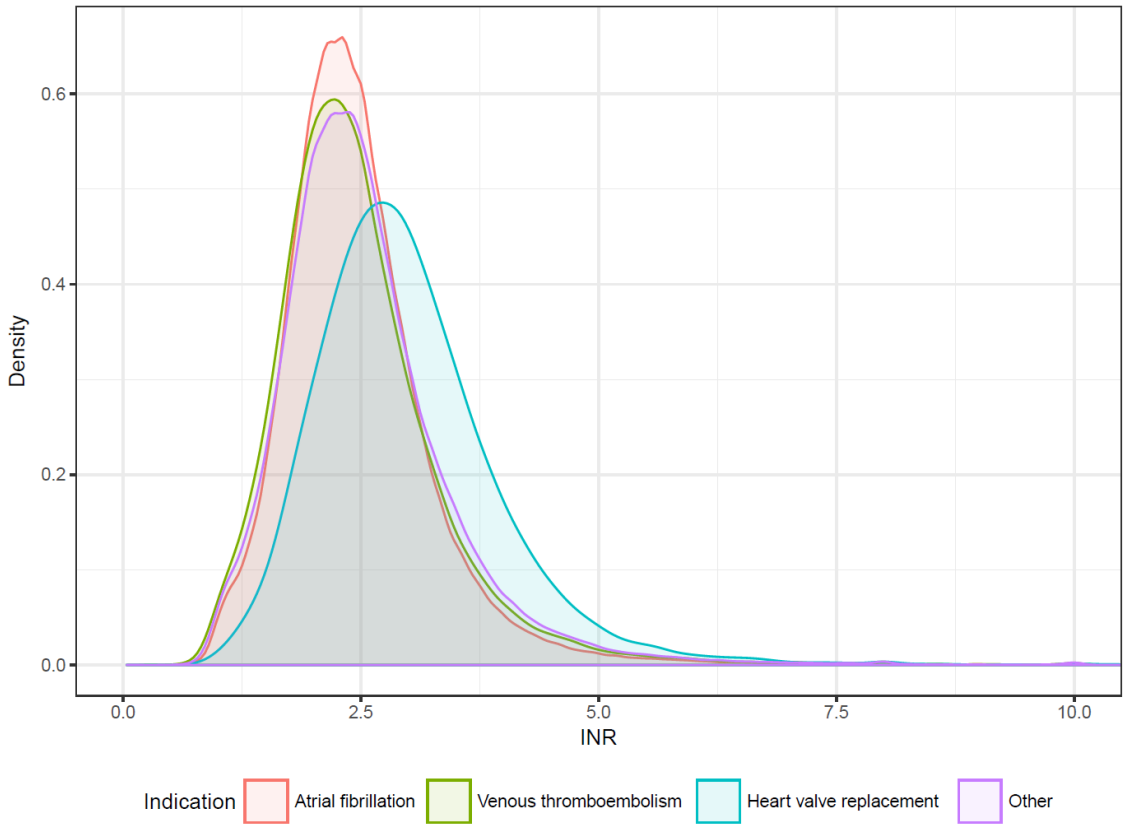
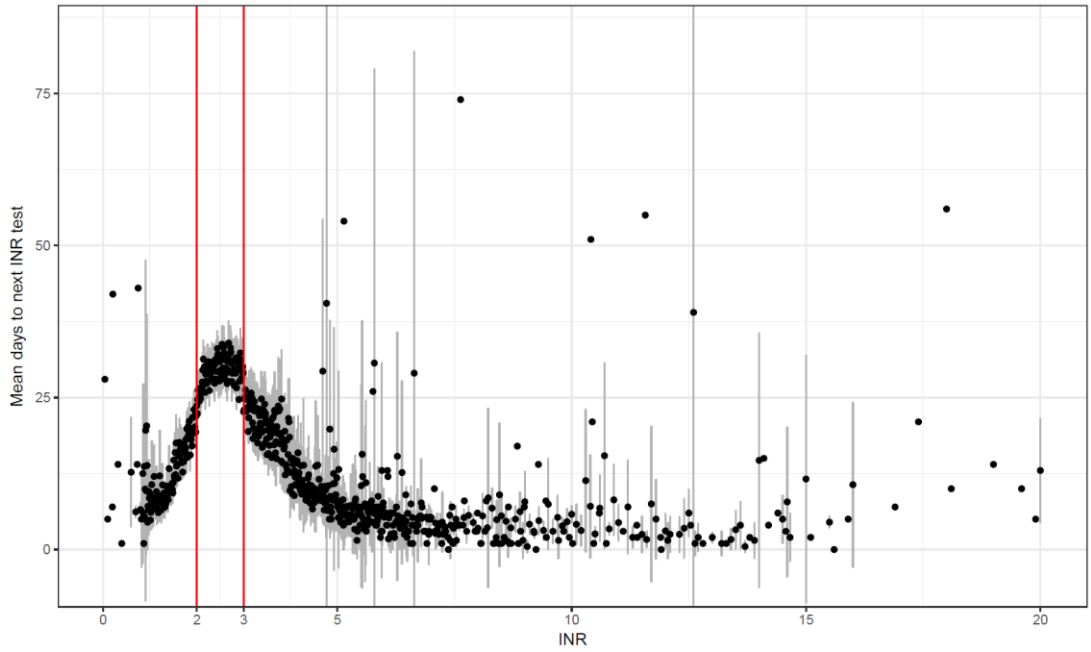


Figure 8.5: Mean days to next INR test by INR value from 541770 INR records for 18823 patients.



Note: INR= International normalised ratio; grey lines represent 95% confidence interval; INR values only captured once in the data do not have a confidence interval

Figure 8.6: The distribution of percent time in therapeutic range for 18823 patients stratified by indication

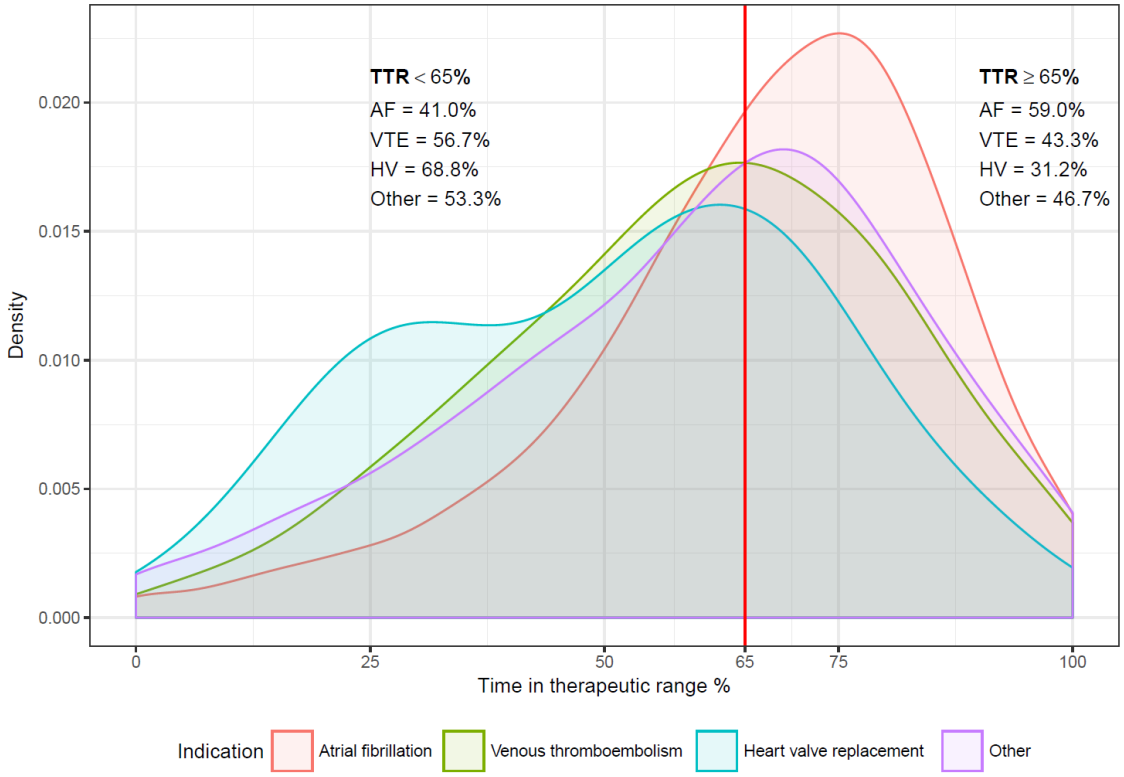


Table 8.3: Characteristics of index INR spells and vitamin K antagonist prescribing prior to and during index INR spells

	Atrial fibrillation (n= 10132)	Venous thrombo-embolism (n= 3471)	Heart valve replacement (n= 744)	Other indication (n= 4476)	All patients (n=18823)
Index INR spell characteristics					
Time in therapeutic range **(2-3)**	66.2 (19.8)	59.4 (21.6)	51.2 (22.7)	59.6 (23.1)	62.8 (21.4)
Median (IQR)	69.3 (55.5, 80.2)	61.3 (44.9, 75.5)	53.7 (32.0, 68.9)	63.0 (44.8, 76.3)	66.1 (50.1, 78.3)
Time in therapeutic range **(2.5-3.5)**	36.3 (19.8)	32.6 (21.7)	49.5 (21.4	34.8 (21.7)	35.8 (20.9)
Median (IQR)	35.8 (23.0, 48.3)	30.6 (15.7, 46.2)	50.7 (35.9, 63.4)	30.6 (15.7, 46.2)	35.1 (21.0, 48.9)
Index INR spell duration (days)	698 (687)	294 (449)	908 (894)	532 (684)	592 (679)
Median (IQR)	477 (196, 959)	162 (97, 239)	582 (236, 1269)	233 (118, 652)	312 (142, 798)
Number of INR records	31.5 (30.7)	19.4 (22.9)	47.2 (51.0)	26.8 (33.6)	28.8 (31.8)
Median (IQR)	21.0 (11, 41)	13.0 9, 19)	27.5 (13, 61)	14.0 (8, 31)	17 (9, 35)
Mean time between INR records (days)	24.3 (12.61)	15.3 (7.41)	21.9 (10.86)	21.1 (11.60)	21.8 (12)
Median (IQR)	22.0 (15.4, 30.5)	13.9 (10.2, 18.7)	19.7 (14.4, 27.2)	18.8 (12.9, 26.8)	19.4 (13.2, 27.5)
Vitamin K antagonist prescribing					
Prior to INR spell (days)	366.1 (601.4)	53.6 (88.3)	516.5 (787.4)	294.2 (511.6)	297 (546)
Median (IQR)	90 (90, 410)	0 (0, 90)	90 (0, 698)	90 (90, 266)	90 (0, 246)
During INR spell (days)	643 (606)	304 (394)	785 (756)	478 (563)	547 (586)
Median (IQR)	449 (211, 868)	196 (137, 267)	544 (229, 1072)	244 (144, 575)	310 (167, 707)

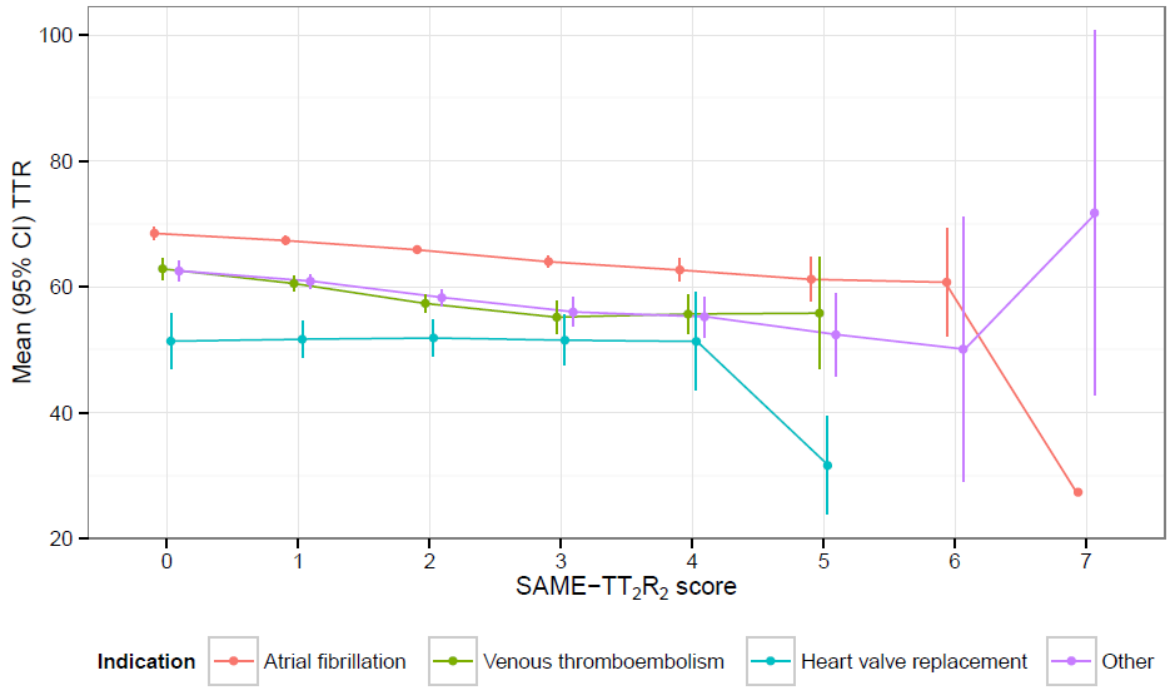
INR= international normalised ratio; Mean (SD) for continuous variables unless stated otherwise

Table 8.4: The distribution of the SAME-TT_2R_2 score within the study population

SAME-TT_2R_2 score	Atrial fibrillation (n= 10132)	Venous thrombo-embolism (n= 3471)	Heart valve replacement (n= 744)	Other indication (n= 4476)	All patients (n=18823)
0	1337 (13.2)	605 (17.4)	119 (16)	736 (16.4)	2797 (14.9)
1	3099 (30.6)	1451 (41.8)	247 (33.2)	1737 (38.8)	6534 (34.7)
2	3365 (33.2)	944 (27.2)	218 (29.3)	1334 (29.8)	5861 (31.1)
3	1751 (17.3)	295 (8.5)	114 (15.3)	404 (9)	2564 (13.6)
4	414 (4.1)	163 (4.7)	30 (4)	213 (4.8)	820 (4.4)
5	147 (1.5)	13 (0.4)	16 (2.2)	44 (1)	220 (1.2)
6	18 (0.2)	0 (0)	0 (0)	5 (0.1)	23 (0.1)
7	1 (0)	0 (0)	0 (0)	3 (0.1)	4 (0)

Reported numbers are n (%)

Figure 8.7: Mean (95% CI) time in therapeutic range across the range of SAME-TT_2R_2 scores stratified by indication



Note: The SAME-TT_2R_2 score is specified in **Table 8.1.** No confidence interval for atrial fibrillation and SAME-TT_2R_2 score 7 due to there being only one patient in this subgroup. CI= confidence interval; TTR= time in therapeutic range (%)

Table 8.5: C-indexes for models using the SAME-TT_2R_2 score to predict TTR>60, 65 and 70 in the study population

Endpoint	All patients	Atrial fibrillation	Venous thromboembolism	Heart valve replacement	Other indication
TTR>60	0.53	0.54	0.54	0.53	0.55
TTR>65	0.53	0.55	0.55	0.52	0.55
TTR>70	0.54	0.55	0.55	0.54	0.54

Figure 8.8: Univariable linear regression estimates for the difference (95% CI) in TTR in demographics and behaviour subgroups

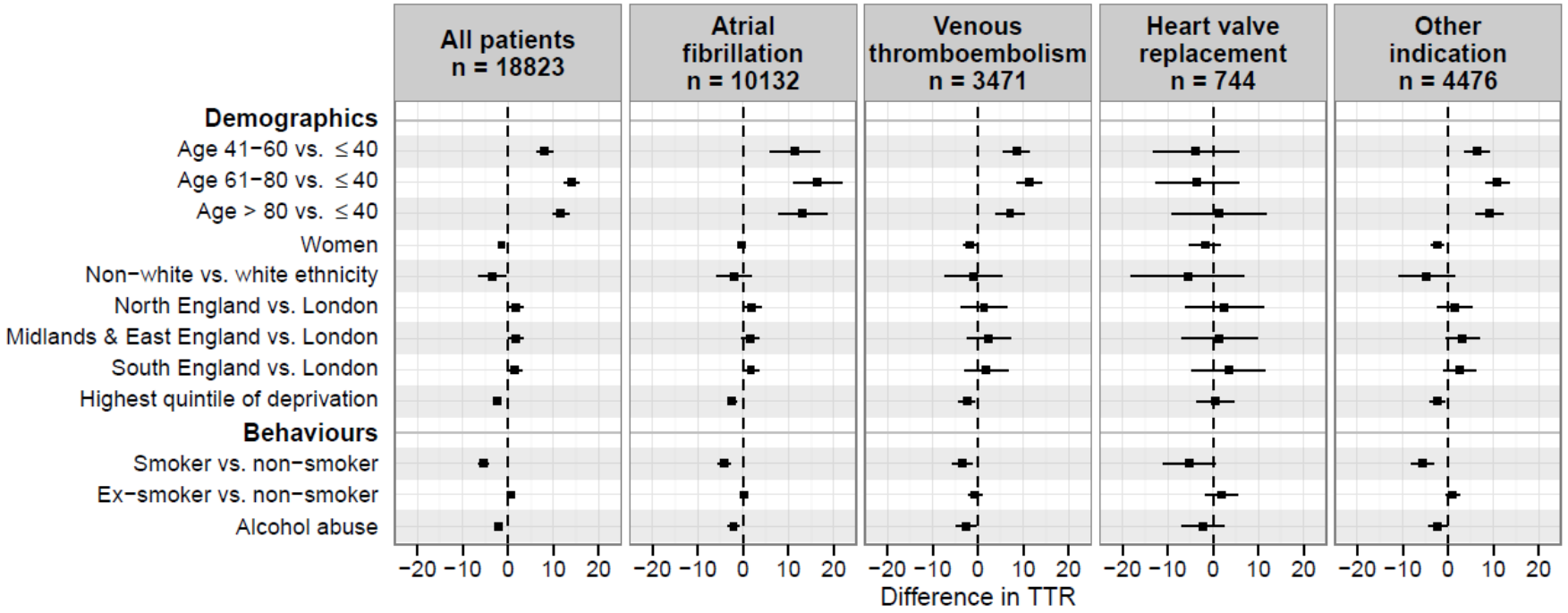


Figure 8.9: Univariable linear regression estimates for the difference (95% CI) in TTR in medical history subgroups

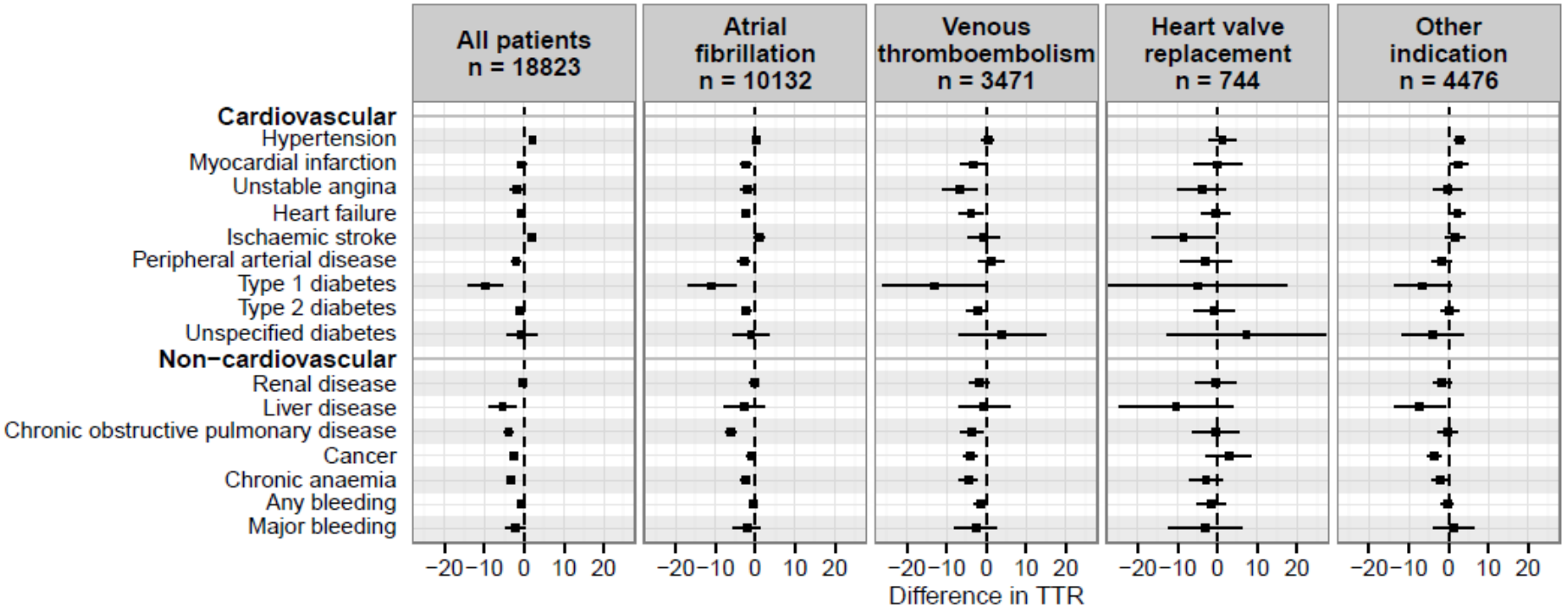


Figure 8.10: Univariable linear regression estimates for the difference (95% CI) in TTR in clinical biomarker subgroups

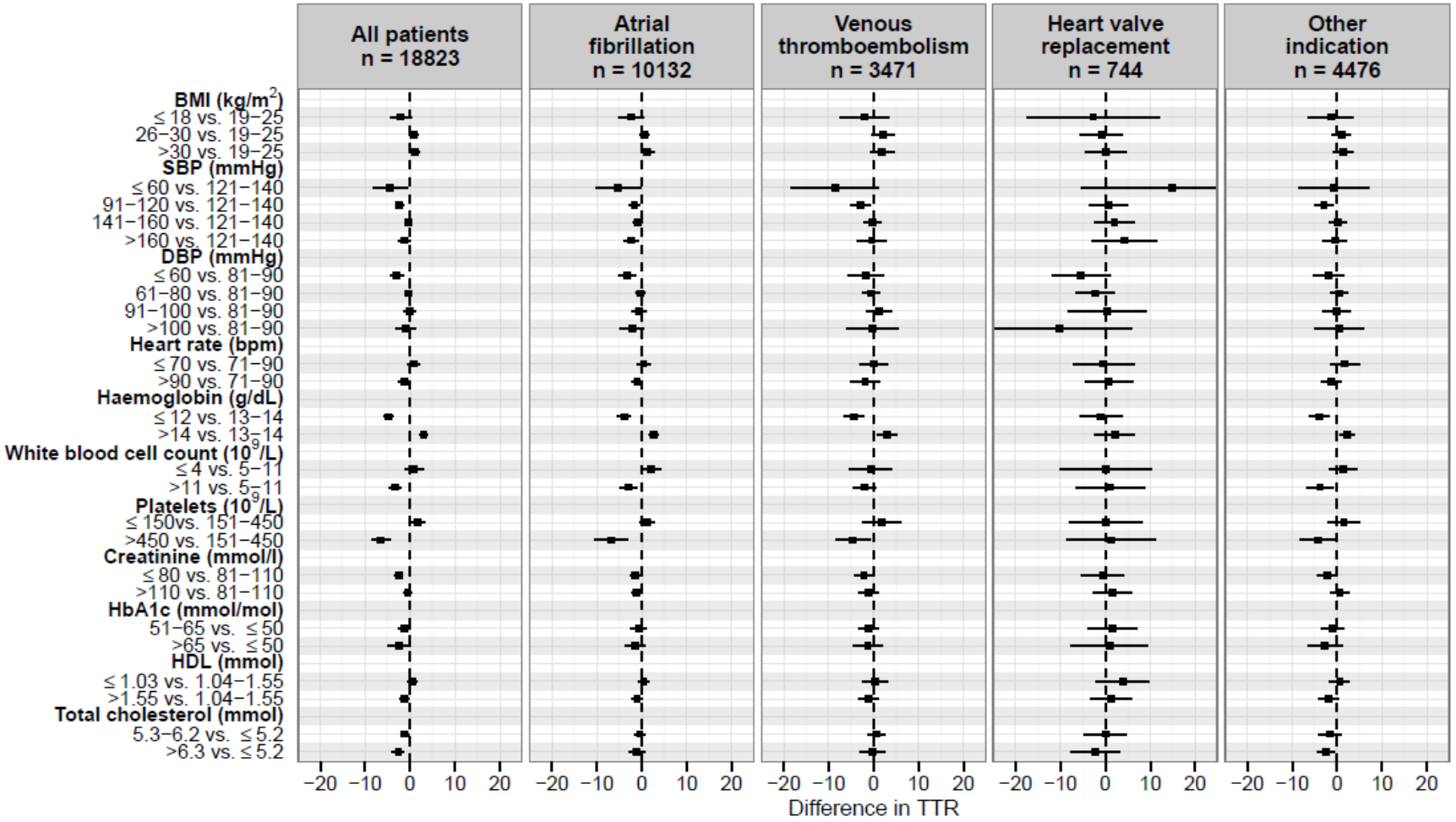


Figure 8.11: Univariable linear regression estimates for the difference (95% CI) in TTR in subgroups defined by prescribed medication prior to index INR spell

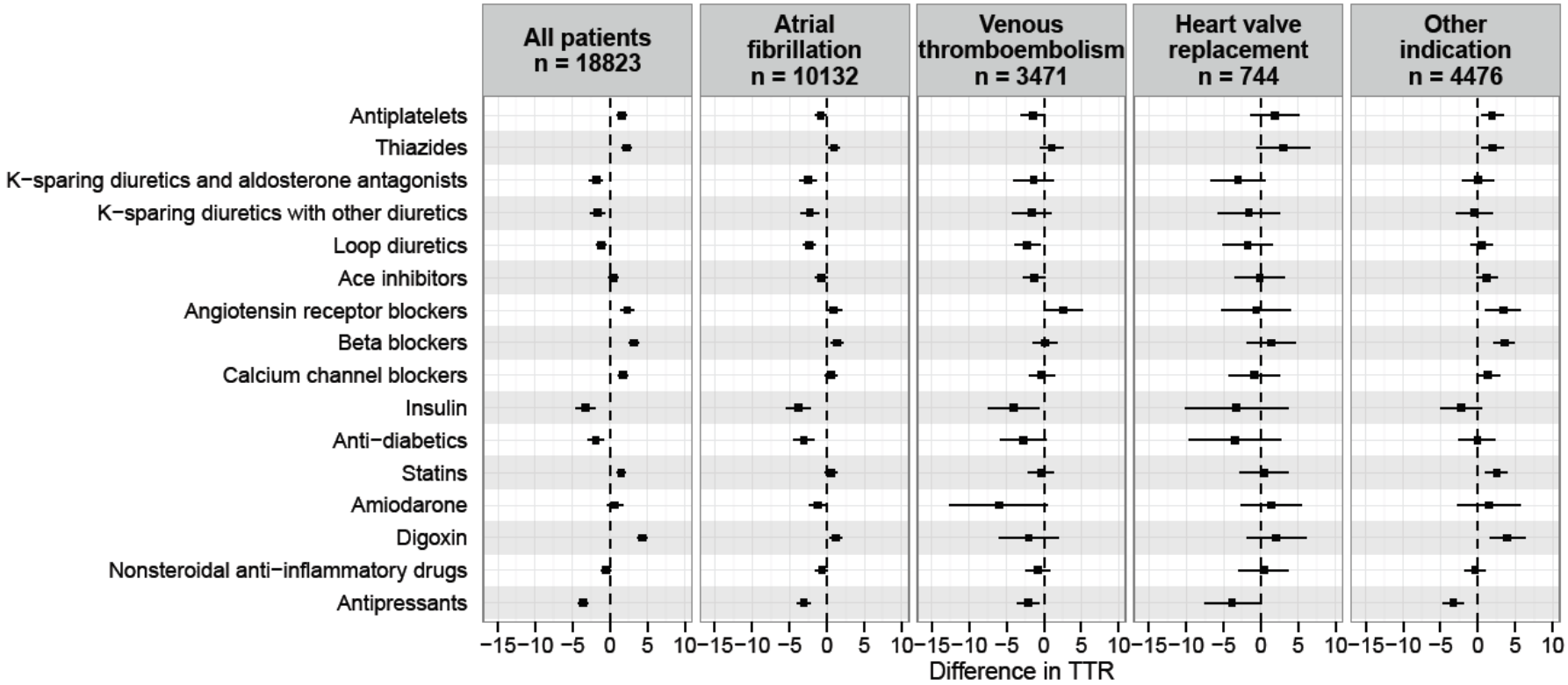


Figure 8.12: Univariable linear regression estimates for the difference (95% CI) in TTR in subgroups defined by prescribed medication during index INR spells

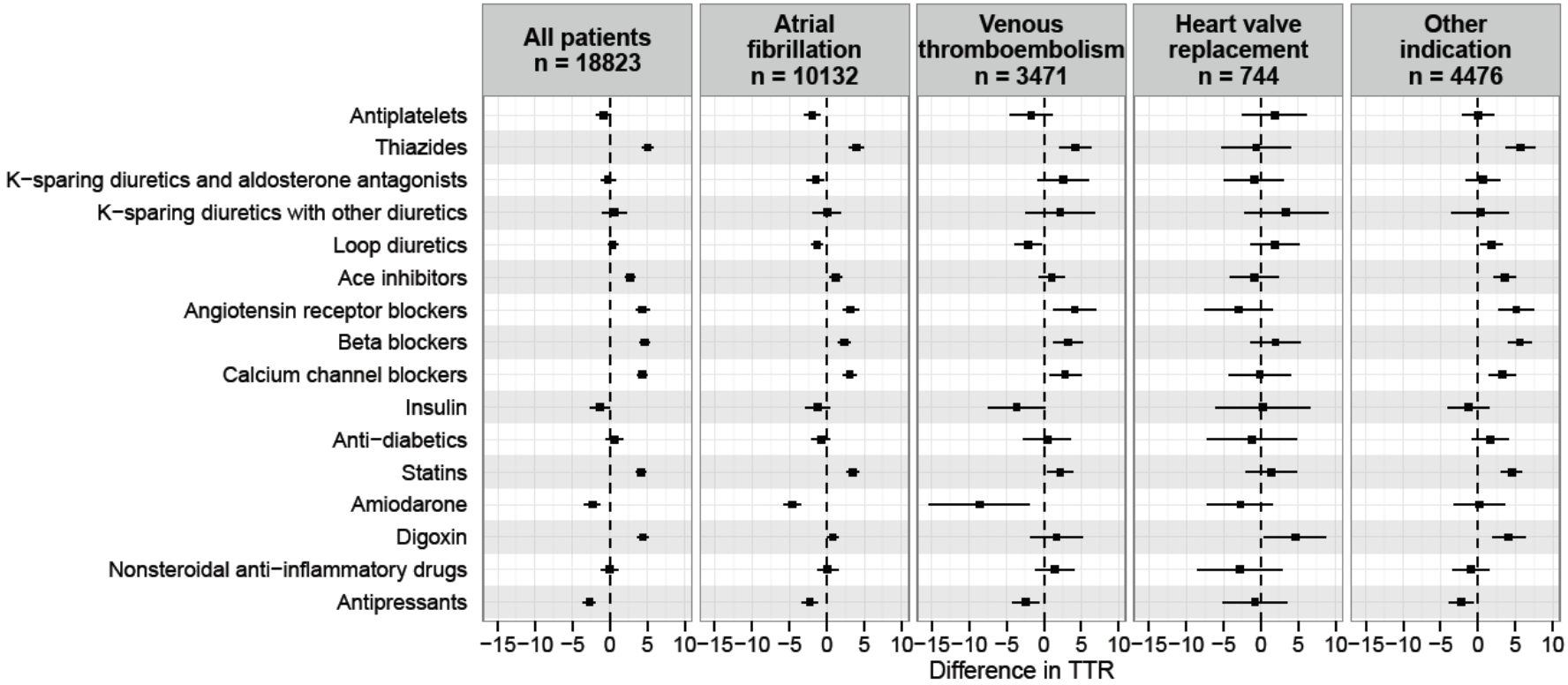


Figure 8.13: All-cause mortality and cardiovascular death, stroke or MI events stratified by TTR above and below 65%

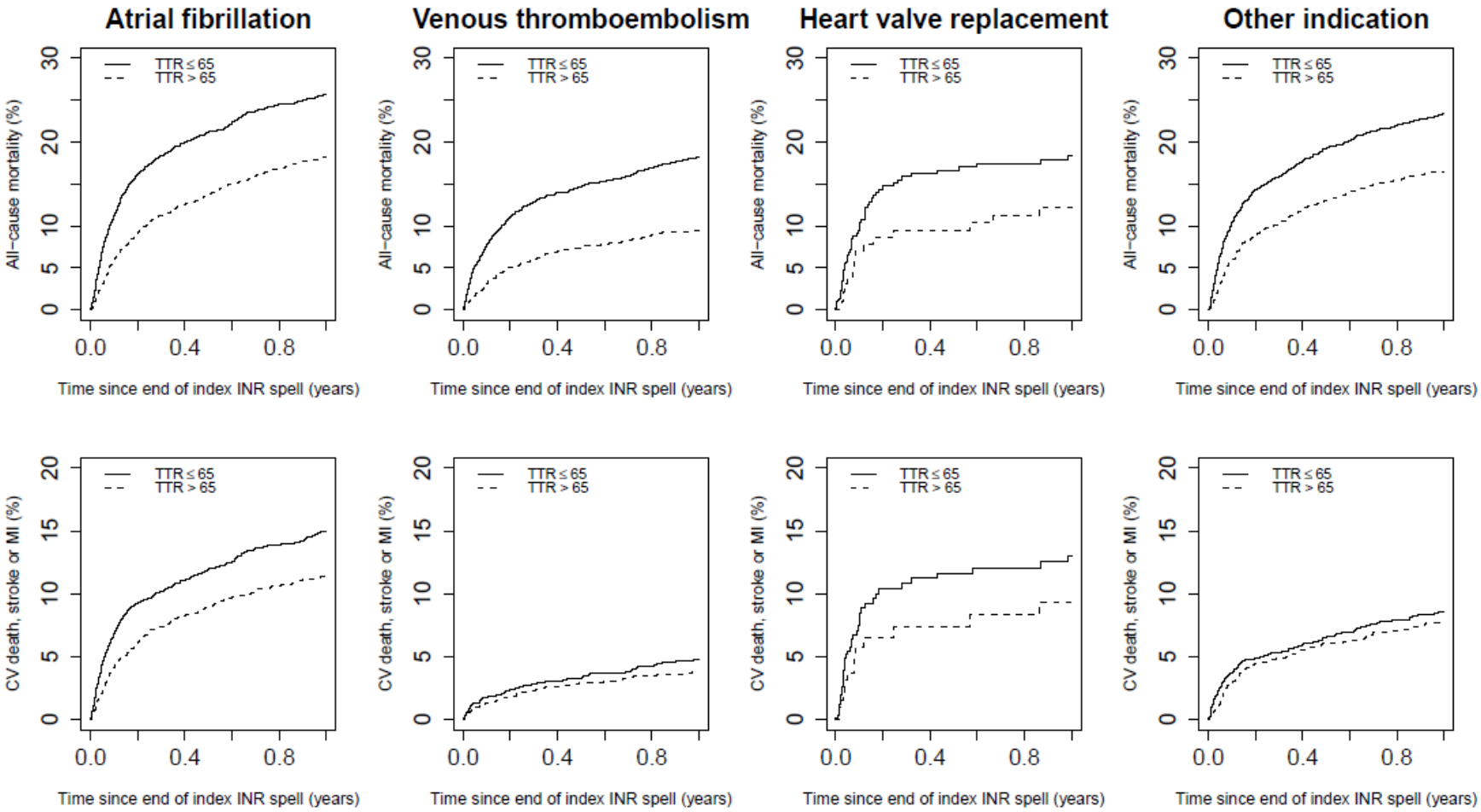


Figure 8.14: Any bleeding and major bleeding events stratified by TTR above and below 65%

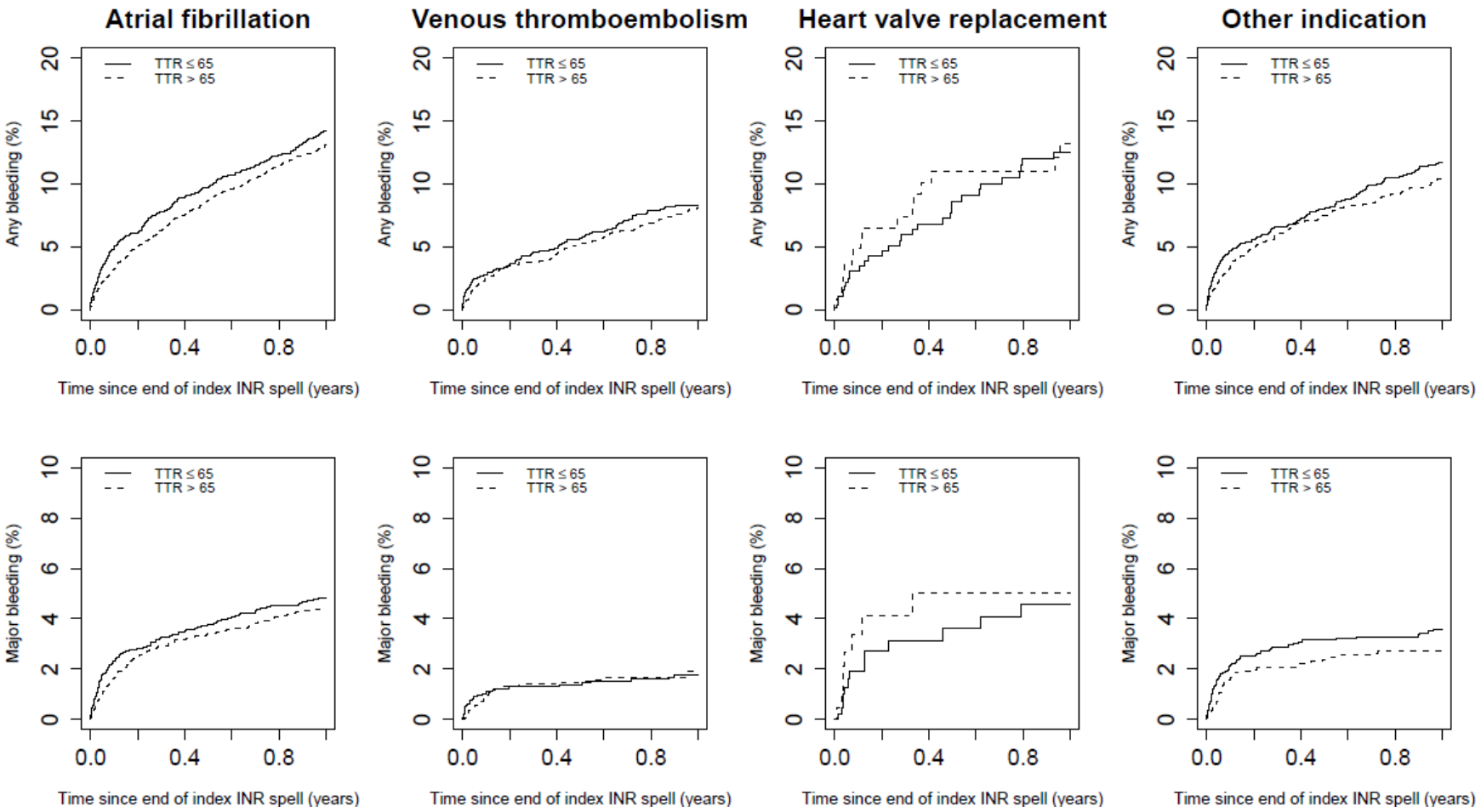
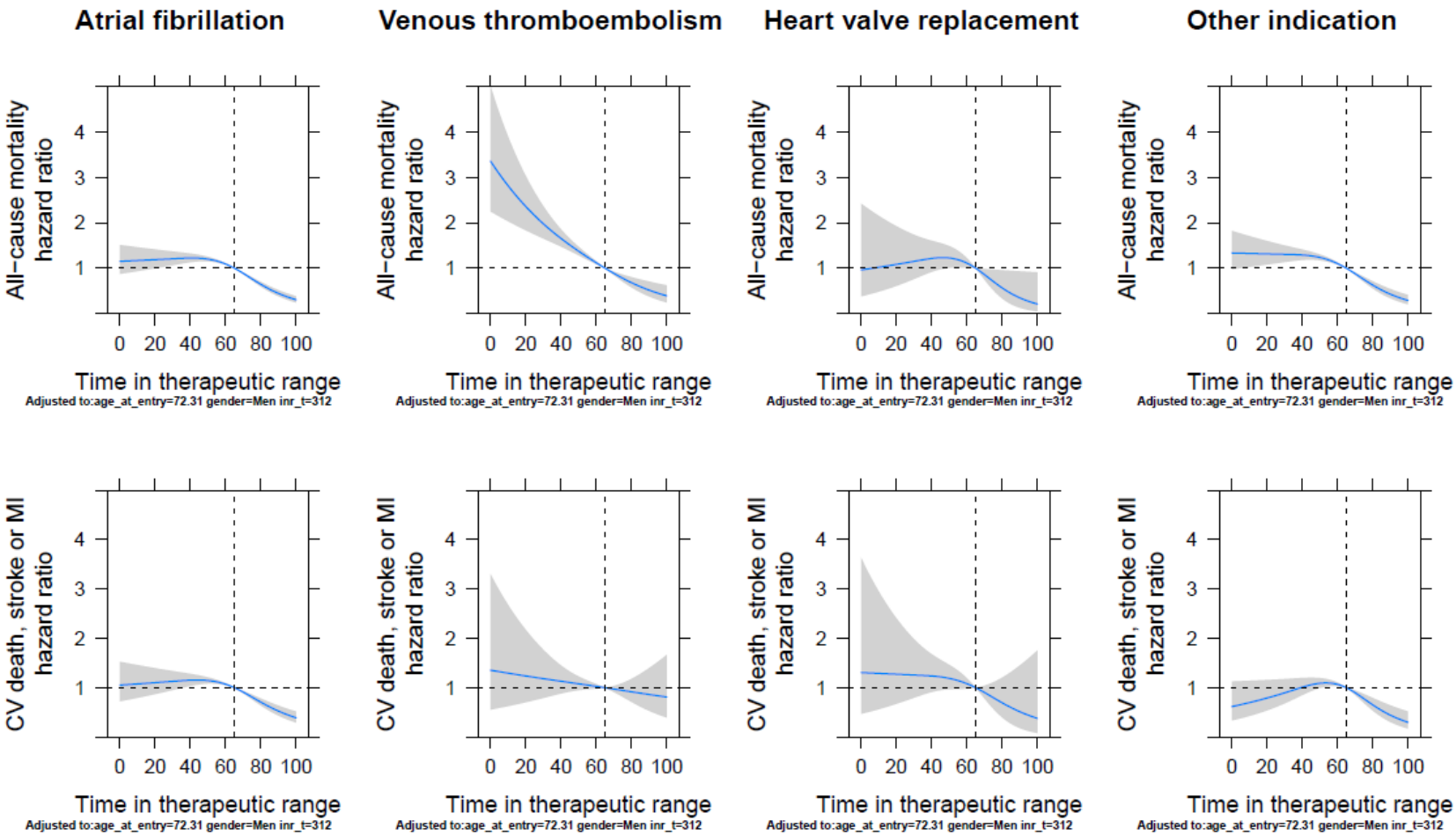
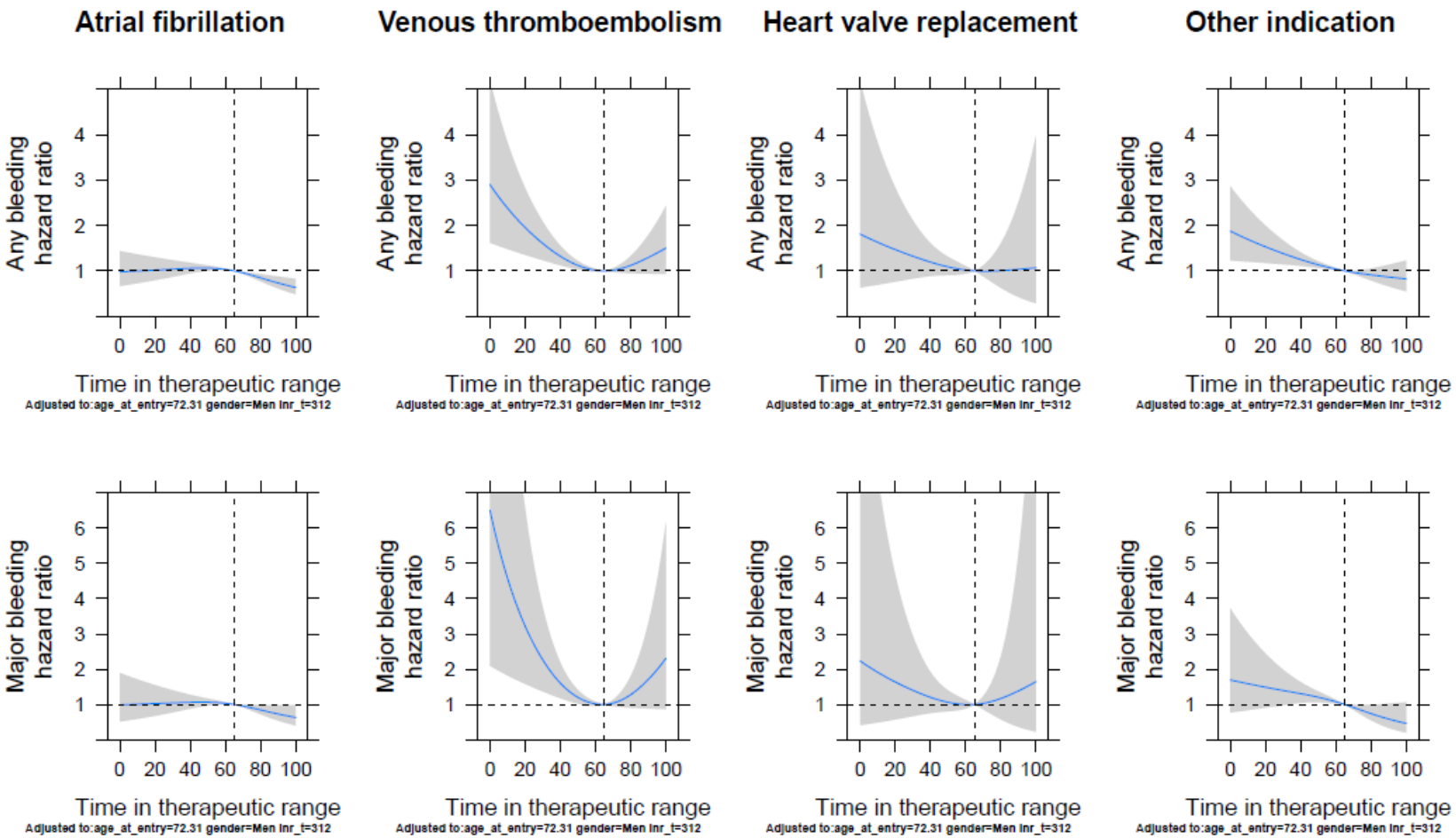


Figure 8.15: Risk of all-cause mortality and cardiovascular death, stroke or MI by index INR spell time in therapeutic range



Hazard ratios estimated for 1 year endpoints using Cox proportional hazards models adjusted for age, gender and length of index INR spell. Time in therapeutic range was fitted using restricted cubic splines with 3 knot points in each model. The reference value for time in therapeutic range is 65%.

Figure 8.16: Risk of any bleeding and major bleeding death, stroke or MI by index INR spell time in therapeutic range



Hazard ratios estimated for 1 year endpoints using Cox proportional hazards models adjusted for age, gender and length of index INR spell. Time in therapeutic range was fitted using restricted cubic splines with 3 knot points in each model. The reference value for time in therapeutic range is 65%.

9 The predictive value of measures of INR control for atherothrombotic and bleeding outcomes: a population-based linked electronic health record study

Chapter Summary

Background

In the previous chapter I examined the association between TTR and all-cause mortality, atherothrombotic events and bleeding and found patients with higher TTR have lower risks of events. Despite this association prognostic models seldom use measures of INR control as a covariate for assessing atherothrombotic and bleeding risk in the relevant patient populations. Measures of INR control beyond TTR, such as INR variability, have also been shown to be associated with outcomes. The objectives of this chapter were to illustrate INR control within the study population, to estimate predictors of INR variability, and to estimate the predictive value of range of measures of INR control in prognostic models for atherothrombotic and bleeding events.

Methods

Using CALIBER linked electronic health records, the study population comprised of 18823 patients prescribed OAC and at least 5 consecutive INR records. I examined the distribution of a range of measures of INR control, including TTR, INR variability, INR trajectory,. A linear regression model was used to estimate associations between baseline risk factors and INR variability. Prognostic models were constructed using logistic regression for one year outcomes, adjusting for standard cardiovascular risk factors and each measure of INR control in turn. Net reclassification improvement and integrated discrimination improvement statistics were used to estimate the predictive value of each measure of INR control.

Results

Smoking, excess alcohol consumption, prior heart valve replacement, cancer were among the risk factors found to be associated with increased INR variability during oral anticoagulation. Over-anticoagulation, high INR variability, high maximum INR and an increasing INR trajectory were found to be associated with higher risks of all-cause mortality, atherothrombotic and bleeding events. In the presence of standard cardiovascular risk factors INR variability, % time

above therapeutic range, improved the predictive performance of prognostic models for the studied endpoints.

Conclusions

I demonstrated that measures of INR control, such as INR variability, maximum INR and TTR can improve the predictive performance of prognostic models for all-cause mortality, atherothrombotic and bleeding events in patients previously treated with oral anticoagulants. Such measures would be can be implemented within electronic health records with simple algorithms to enable continuous monitoring and updating of patients risks.

9.1 Introduction

Oral anticoagulants are a widely used class of drugs primarily for stroke prevention in atrial fibrillation and other cardiovascular diseases. The international normalised ratio (INR) is a regular blood test for patients being treated with vitamin K antagonists (VKA) a class of oral anticoagulants in order to observe treatment control. INR is a standardised calculation of prothrombin time: the time taken for blood to clot. In healthy people not treated with oral anticoagulants a normal INR is between 0.8 and 1.2. For patients being treated with oral anticoagulants the therapeutic range for most diseases is between 2 and 3. If over treated patients are at risk of major bleeding, and if under treated patients are at risk of blood clotting which could result in stroke or other serious atherothrombotic events. Given the severity of outcomes associated with unstable INR, patients are monitored regularly for the duration of their treatment with oral anticoagulants. The frequency of INR tests is dependent on the stage of anticoagulation, presence of bleeding risk factors, concomitant prescriptions with known interactions with oral anticoagulants, and poor INR stability.[104] Depending on the indication for oral anticoagulation patients may be treated from 3 months (e.g. venous thromboembolism) to lifelong (e.g. for atrial fibrillation with high stroke risk).

There are many ways repeated INR tests may be summarised for assessment, the most common being percent time in therapeutic range (TTR). Cohort studies have shown that sub-therapeutic INR and INR stability is associated with increased risks of all-cause mortality, atherothrombotic events, and bleeding events.[112,171] Prognostic models and risk scores have widely been adopted for assessing patient's risks associated with oral anticoagulation. However within such risk assessments, measures of INR are seldom included as a prognostic factor. HAS-BLED,[7] a simple point-based bleeding risk score includes labile INR (defined as TTR<65%) as a risk factor. It has been shown that modification of the risk scores, ATRIA,

HEMORR$_2$HAGES and ORBIT to include labile INR resulted in improvement in prediction of major bleeding events in atrial fibrillation patients.[146] Previous studies of INR variability[140-143] and wider ranges of measures of INR control[139,144,145] have demonstrated scope for improving upon TTR; in particular INR variability has shown stronger associations with bleeding and atherothrombotic events. However the value these measures may add to prognostic models with standard cardiovascular and bleeding risk factors is unknown.

Using an unselected cohort from primary care-hospital-death registry linked electronic health records I sought to investigate which measures of INR control are optimal to predict the risks of both the benefits and harms of oral anticoagulation therapy. The objectives of the study were **1)** To illustrate overall distribution of INR control within the study population, **2)** To estimate predictors of INR variability and **3)** To estimate the predictive value of a range of measures of INR control for atherothrombotic and bleeding events; Is TTR the best measure and do measures of INR control work well for predicting both bleeding and atherothrombotic endpoints and add value to standard risk factors?

9.2 Methods

9.2.1 Study population and index INR spell

I used CALIBER (1997-2010) linked electronic health records. Patients were included if they had at least one prescription for an oral anticoagulant (vitamin K antagonists: warfarin, phenindione or acenocoumarol) and a minimum of five consecutive INR records. INR values were considered to be consecutive if they were within 90 days. If a patient had multiple valid spells of INR monitoring I used their first as their index INR spell.

9.2.2 Measures of INR control

The measures of INR control investigated in the study are described below including how to calculate them and the rationale for their inclusion.

9.2.2.1 Time in therapeutic range, above range and below range

The time in therapeutic range (TTR) is the percent of the INR spell spent within the range of 2 to 3. The trajectory between adjacent INR measures is assumed to be linear following Rosendaal's method.[105] The time above therapeutic range (TAR) is the percent of the INR spell spent >3. This is calculated in a similar manner to TTR except any time below therapeutic range (<2) is not considered. The time below therapeutic range (TBR) is the percent of the INR spell spent <3. This is calculated in a similar manner to TTR except any time above therapeutic range (<3) is not considered.

Examples of TTR, TAR and TBR are shown in **Tables and** Figures

Figure 9.1**.**

9.2.2.2 TTR control group

I grouped patients by TTR whilst utilising information on whether patients out of range INR measures were predominantly above range or below range. Patients with a TTR ≥ 80% were grouped as *stable*. Patients with TTR<80% and all of their out of range INR measures >3 were grouped as *over-anticoagulated*. Patients with TTR<80% and all of their out of range INR measures <2 were grouped as *under-anticoagulated*. The remaining patients with TTR<80% and a mixture of out of range INR values >3 and <2 were grouped as *erratic*.

The advantage of this grouping method compared with TTR alone is that it accounts for the direction of patients out of range measures.

Examples of index INR spells for patients in each TTR control group are shown in **Figure 9.2**.

9.2.2.3 Mean INR and INR variability

I used the mean INR as a measure of average location of INR values within spells. The variability of INR spells was calculated using the standard deviation of the measures within a spell.

Examples of patients with different means and standard deviation for their INR spells are shown in **Figure 9.3**.

9.2.2.4 Single INR values

I also used single values within INR spells to determine whether they alone held any predictive value for future events, rather than 'time consuming' summaries of longitudinal data. I analysed the first, last, minimum and maximum values individually from each INR spell. An example of these single INR values for a patient's INR spell is shown in **Figure 9.4**.

9.2.2.5 INR trajectory

I estimated the trajectory of INR spells by fitting linear models to the INR measures by time. I initially fitted quadratic models. If the quadratic coefficient for time was significant ($p<0.05$) I used the sign of coefficients to determine if the trajectory was *upwards quadratic* or *downwards quadratic*. For INR spells with a non-significant quadratic coefficient for time I fit a simple linear model for INR measures vs. time. If the coefficient for time was significant ($p<0.05$) I used the sign of the coefficient to determine if the trajectory was *increasing* or *decreasing*. The trajectories of INR spells with non-significant coefficients for time were classified as *stable*.

However this method does not account for the location of INR measures. For example for a patient with an increasing trajectory, this method does not capture if the INR is increasing having started low, or was initially high and continued to increase, the latter of which might considered to be a significantly worse scenario.

Examples of patients INR spells with the different trajectories described are shown in **Figure 9.5**.

9.2.3 Baseline characteristics

I described the characteristics of the study cohort at baseline stratified by TTR control group (stable, erratic, over anticoagulated, under anticoagulated). Broadly, the characteristics investigated were demographics and behaviours, characteristics of their index INR spell and oral anticoagulant exposure, cardiovascular and non-cardiovascular medical history, clinical biomarkers and prescribed medications prior to and during INR spell. Baseline characteristics were described using mean and standard deviation (SD) or median and interquartile range (IQR) for continuous variables and frequencies and percentages for categorical variables.

The phenotypes for the characteristics are described in **chapter 3**.

9.2.4 Endpoints

Patients were followed up for 1 year for all-cause mortality, cardiovascular death stroke or MI, any bleeding and major bleeding. The composition of the phenotypes for these endpoints is described in **chapter 3.**

9.2.5 Statistical analysis

9.2.5.1 Baseline predictors of measures of INR variability

Multivariable linear regression models were used to estimate the effects of baseline characteristics on INR variability. The model was adjusted for demographics and behaviours, characteristics of their index INR spell and oral anticoagulant exposure, cardiovascular and non-cardiovascular medical history, clinical biomarkers and prescribed medications prior to and during INR spell.

9.2.5.2 Outcomes following first INR spell

To estimate the association of the measures of INR control with 1 year mortality, cardiovascular and bleeding endpoints following an INR spell I used multivariable logistic regression models to calculate odds ratios and 95% confidence intervals (CI), adjusted for standard cardiovascular risk factors and confounders measured prior to and during INR spells: Demographics and behaviours (Age, sex, index of multiple deprivation, smoking status and

history of excess alcohol consumption), INR spell characteristics (length of index INR spell and exposure to oral anticoagulants prior to index INR spell), cardiovascular disease history (atrial fibrillation, myocardial infarction, heart failure, heart valve replacement, unstable angina, peripheral arterial disease, venous thromboembolism, ischaemic or unspecified stroke, bleeding and major bleeding), non-cardiovascular disease history (diabetes, cancer, chronic obstructive pulmonary disease and renal disease) and clinical biomarkers (BMI, systolic blood pressure, high density lipoproteins, serum creatinine, haemoglobin and white blood cell count).

9.2.5.3 Estimating the predictive value of measures of INR control

In order to evaluate the change in performance of prediction models with the inclusion of each of the measures of INR control I first fit a multivariable logistic 'base model' adjusting for the standard risk factors and confounders only (described in **chapter 9.2.5.2**). Next models were fitted that comprised of the base model with the inclusion each of the measures of INR control. For ease of interpretation and presentation the continuous measures of INR control were categorised into quintiles, with the middle quintile being used as the reference group. For the measures of INR control that were already categorical, TTR control group and INR trajectory, their respective 'Stable' groups were used as reference categories in multivariable models. I compared each of the new models with the base model using the net reclassification improvement and the integrated discrimination improvement measures.

The **net reclassification improvement**[219-221] (NRI) is the sum of the net change in proportion of patients correctly reclassified upwards (i.e. patients that did have an event and are predicted to have no event with the base model and predicted to have an event with the new model) and the net change in proportion of patients correctly reclassified downwards (i.e. patients that did not have an event and are predicted to have am event with the base model and predicted to have no event with the new model).

Let *M1* and *M2* denote the individual predicted probabilities of events estimated from model 1 (base model) and model 2 (base model plus measure of INR control) respectively. The NRI is calculated as follows:

$$NRI = P(M2 > M1|event) - P(M2 < M1|event) + P(M2 < M1|no\ event) - P(M2 > M1|no\ event)$$

Note that this is the continuous version of NRI as opposed to the category-based NRI which requires the specification of cut-points and risk groups.

The **Integrated discrimination improvement**[219] (IDI) is the difference in discrimination slope between a model with and without an added risk factor. The discrimination slope of a model is the difference in mean predicted probabilities between those that do and don't have the event of interest. The IDI therefore can be expressed as:

$$IDI = (\bar{p}_{event} - \bar{p}_{no\ event})_{M2} - (\bar{p}_{event} - \bar{p}_{no\ event})_{M1}$$

Model prediction comparison

To visualise the changes in individual patient's predictions with the additions of measures of INR control, I plotted predictions from the base model versus predictions for measure of INR control models.

Model calibration

The calibration of all models was assessed using Hosmer-Lemeshow tests and plots of observed events versus model predictions in decile groups.

Modelling assumptions

In univariable analyses evidence of non-linear associations between dependent and independent variables was checked and where appropriated fitted in the multivariable models using restricted cubic splines. Presence of collinear variables in the multivariable models was checked and any offending variables were removed.

Missing Data

Missing covariate data was multiply imputed using the MICE (multiple imputation with chained equations).[211] 10 imputed data sets were generated.

9.3 Results

9.3.1 Demographics and baseline characteristics of study population

I identified 18823 patients in CALIBER who had been prescribed an oral anticoagulant and had at least 5 consecutive (within 90 days) INR records **(Figure 9.6)**. Of these patients, 11677 had an index INR spell classed as erratic, 4123 were stable, 2168 were under anticoagulated and 855 were over anticoagulated (**Table 9.1**). Patients in the erratic and stable groups were on average older, 70.1 years and 71.2 years respectively, than those who were over-anticoagulated or under-anticoagulated, 68.7 and 66.0 years respectively. The majority of all four groups were men; however women were slightly more prevalent in the erratic and under-anticoagulated groups. Patients in the erratic group had higher proportions of excess alcohol

consumption history (14.3%) and current smokers (10.6%) compared with those in the stable group (11.7% and 7.5%).

A history of atrial fibrillation was most prevalent in the stable group (64.5%), whilst a history of myocardial infarction was most prevalent in the over anticoagulated group (10.3%). Heart failure history was most prevalent in the erratic (19.3%) and over-anticoagulated groups (19.2%). More than 25% of patients had a bleeding event of any severity prior to their index INR spell whilst 2% or less had experienced major bleeding in each group. Chronic anaemia was most common in the over-anticoagulated group (14.6%). Type 2 diabetes, cancer and chronic obstructive pulmonary disease were more prevalent in the erratic group than the others.

At baseline the levels of clinical biomarkers, including systolic blood pressure, body mass index, haemoglobin, and creatinine were similar amongst the groups.

The stable group had lower percentages of patients who had a history of prescriptions for antidepressants, anti-diabetics and insulin prior to index INR spells compared with the other groups. A history of antiplatelet prescribing was highest in the stable and erratic groups.

Across most classes of drugs, prescribing was markedly lower during index INR spells compared with prior to INR spells in all groups. Exceptions to this were increased prescribing of angiotensin receptor blockers, statins and anti-diabetics in both the stable and erratic groups and insulin in the erratic group. 10% of the erratic group were prescribed nonsteroidal anti-inflammatory drugs (NSAIDs) during their index INR spell compared with <7% in the three remaining groups. The erratic group also had higher prescribing of antiplatelets, K-sparing diuretics, loop diuretics, ace inhibitors, calcium channel blockers and antidepressants during index INR spells compared with the other groups. Patients in the stable group had a higher percentage of prescriptions for thiazides and beta blockers during index INR spells compared with the other groups.

9.3.2 Characteristics of index INR spells and oral anticoagulation

The median duration of INR monitoring for patients in the stable and erratic TTR control groups were the longest of the four groups with both exceeding a year at 425 (IQR: 174, 896) and 399 (IQR: 168, 936) days respectively, whilst the median duration was less than 6 months for the over-anticoagulated and under-anticoagulated group, at 154 (IQR: 82.5, 344) and 120 (IQR: 70, 184) days respectively. As a result the exposure to oral anticoagulants during index INR spells was also highest in the erratic and stable TTR control groups. Median duration of

oral anticoagulant exposure prior to the start of index INR spells was highest in the over-anticoagulated group.

9.3.3 Measures of INR control

The distributions and relationships between the continuous measures of INR control are shown in a matrix plot, **Supplementary appendix 11.5.1**. The distribution of INR trajectories in the study population were 13462 (71.5%) stable, 1731 (9.2%) increasing, 1092 (5.8%) upwards quadratic, 862 (4.6%) decreasing and 1676 (8.9%) downwards quadratic.

For the regression analysis the continuous measures of INR control were grouped into quintiles (**Table 9.2**).

9.3.4 The association between baseline characteristics and INR variability

The associations between patient baseline characteristics and log INR spell standard deviation were estimated using a multivariable linear regression model and the exponentiated coefficients are shown in **Figure 9.7**.

Smokers compared with non-smokers had a 7.5% (95% CI: 4.7%, 10.3%) increase in INR SD. Patients with a history of excess alcohol consumption had a 6.7% increase in INR SD (95% CI: 4.6%, 8.9%). Having a history of heart valve replacement was associated with a 11.3% (95% CI: 7.4%, 15.3%) increase in INR SD. Other cardiovascular an non-cardiovascular disease histories associated with a significant increase in INR SD included peripheral arterial disease 3.0% (95% CI: 0.3%, 5.8%), major bleeding 6.1% (95% CI: 0.6%, 11.9%), chronic anaemia 2.6% (95% CI: 0.3%, 4.9%), cancer 6.5% (95% CI: 4.6%, 8.4%) and chronic obstructive pulmonary disease 9.2% (95% CI: 6.5%, 11.9%). One unit increase in haemoglobin was associated with a small decrease in INR SD 0.989 (95% CI: 0.982, 0.997).

Patients had increased INR SD if prior to their index INR spell they add a history of prescriptions for loop diuretics 3.2% (95% CI: 1.1%, 5.2%), antiplatelets 5.4% (95% CI: 3.8%, 7.1%) and antidepressants 1.9% (95% CI: 0.2%, 3.6%). Patients had lower INR SD if they were prescribed at any time prior to INR index spell statins 0.959 (95% CI: 0.939, 0.979), digoxin 0.963 (95% CI: 0.942, 0.985) or amiodarone 0.963 (95% CI: 0.937, 0.989).

The drugs prescribed during INR spells that were associated with increased variability were loop diuretics 3.1% (95% CI: 1.1%, 5.1%), k-sparing diuretics and aldosterone antagonists 5.1% (95% CI: 2.2%, 8.2%), ace inhibitors 5.6% (95% CI: 2.5%, 8.8%), nonsteroidal anti-inflammatory drugs 7.1% (95% CI: 4.6%, 9.7%), antidepressants 7.0% (95% CI: 5.0%, 9.1%), digoxin 3.8% (95% CI: 1.6%, 6.1%) and amiodarone 15.2% (95% CI: 12.2%, 18.3%). Only hypertension and

heart failure drugs were associated with lower INR SD (0.947 (95% CI: 0.918, 0.977)) when prescribed during the INR spells

9.3.5 The base prognostic models

At one year following the end of their index INR spell 2686 (14.3%) patients had died, 1304 (6.9%) had cardiovascular death, stroke or MI, 1446 (7.7%) had any bleeding and 478 (2.5%) had major bleeding. Multivariable logistic 'base' prognostic models were developed for each endpoint which included covariates for demographics and cardiovascular risk factors and no measures of INR control. Systolic blood pressure and body mass index were modelled using restricted cubic splines with 3 knots. The base models estimates are displayed in **Supplementary appendix 11.5.2.**

9.3.6 The association between measures of INR control and one year endpoints

Each of the measures of INR control was included in each base model in turn. The adjusted odds ratios for the association between measures of INR control and all-cause mortality and cardiovascular death, stroke or MI are shown in **Figure 9.8** and for any bleeding and major bleeding are shown in **Figure 9.9**.

9.3.6.1 Time in therapeutic range stability group

Patients had higher odds of all-cause mortality and cardiovascular death, stroke or MI one year after their index INR spell if their INR spell was classed as erratic or over-anticoagulated compared to those whose INR spells were stable. Odds of any bleeding did not differ significantly between TTR stability groups, however patients had nearly double the odds of major bleeding if their INR spell was classed as over-anticoagulated compared with stable, OR: 1.90 (95% CI: 1.24, 2.91).

9.3.6.2 Time in therapeutic range

There was a decreasing trend in the odds of all four investigated endpoints as TTR increased and the trend appeared to be most pronounced for all-cause mortality. Patients in the lowest quintile of TTR [0.0%, 45.4%), compared to those in the 3rd quintile [60.5%, 70.9%) had significantly higher odds of all-cause mortality, OR: 1.47 (95% CI: 1.28, 1.69), any bleeding, OR: 1.32 (95% CI: 1.11, 1.58) and major bleeding, OR: 1.38 (95% CI: 1.02, 1.87). Patients in the highest quintile of TTR [81.0%, 100%], compared to those in the 3rd quintile had significantly lower odds of all-cause mortality, OR: 0.50 (95% CI: 0.43, 0.58) and cardiovascular death stroke or MI OR: 0.60 (95% CI: 0.49, 0.73).

9.3.6.3 Time above therapeutic range

There was an increasing trend association between % time above therapeutic range and the odds of the four investigated endpoints. These trends appeared to be modest for the bleeding endpoints compared with the clearer trends for all-cause mortality and cardiovascular death, stroke or MI endpoints. Patients in the 1st quintile [0.0%, 1.5%) of time above therapeutic range compared with the 3rd quintile [8.5%, 14.9%) had significantly lower odds of all-cause mortality and major bleeding. Patients in the 5th quintile [25.4, 100.0] compared with the 3rd quintile had significantly higher odds of all-cause mortality and cardiovascular death stroke or MI.

9.3.6.4 Time below therapeutic range

There were no apparent trends between % time below therapeutic range and the one year endpoints. The odds of cardiovascular death, stroke or MI, any bleeding and major bleeding did not differ significantly between the quintiles of % time below therapeutic range. However the 1st quintile [0.0%, 5.0%) and 5th quintile [33.8%, 100.0%] had lower odds of one year all-cause mortality when compared to the 3rd quintile, OR's: 0.84 (95% CI: 0.73, 0.97) and 0.86 (95% CI: 0.74, 0.99) respectively.

9.3.6.5 Mean INR

Patients with mean INR in the 4th quintile [2.5, 2.7) and 5th quintile [2.7, 8.2) had higher odds of one year all-cause mortality OR: 1.34 (95% CI: 1.17, 1.54) and 2.27 (95% CI: 1.98, 2.60) respectively compared with patients in the 3rd quintile [2.4, 2.5). The odds of one year cardiovascular death, stroke or MI was also significantly higher in patients whose mean INR was in the 5th quintile. The odds of any bleeding and major bleeding did not differ significantly amongst the mean INR quintiles.

9.3.6.6 INR standard deviation

Increasing INR standard deviation was associated with higher odds of one year events. Patients in whose INR standard deviation was in the 1st quintile [0.000, 0.458) had lower odds of all-cause mortality OR: 0.66 (95% CI: 0.56, 0.78) and cardiovascular death, stroke or MI OR: 0.74 (95% CI: 0.60, 0.92) than patients in the 3rd quintile [0.584, 0.710). Patients with greatest variability during their INR spell, in the 5th quintile [0.922, 5.917], had increased odds of all-cause mortality OR: 2.67 (95% CI: 2.34, 3.05), cardiovascular death, stroke or MI OR: 1.80 (95% CI: 1.51, 2.14), any bleeding OR: 1.32 (95% CI: 1.11, 1.57) and major bleeding OR: 1.59 (95% CI: 1.19, 2.12).

9.3.6.7 Minimum INR

There were no distinct trends between minimum INR values and one year outcomes. Patients whose minimum INR fell within the range of the 2nd quintile [1.11, 1.31) had increased odds of one year all-cause mortality, OR: 1.17 (95% CI: 1.02, 1.34) compared with those whose minimum INR was in the 3rd quintile [1.31, 1.51). Patients with minimum INR values in the 4th quintile [1.51, 1.71) and 5th quintile [1.71, 4.00] had reduced odds of one year all-cause mortality. There was no significant association between minimum INR values and cardiovascular death, stroke or MI or the bleeding outcomes.

9.3.6.8 Maximum INR

Increasing maximum INR values was associated with higher odds of one year outcomes. Patients whose INR exceeded 5.21 (i.e. 5th quintile) during their index INR spell had increased odds of one year all-cause mortality OR: 3.01 (95% CI: 2.63, 3.44), cardiovascular death, stroke or MI 2.15 (95% CI: 1.81, 2.57), any bleeding OR: 1.18 (95% CI: 1.00, 1.39) and major bleeding OR: 1.54 (95% CI: 1.17, 2.02) compared with patients whose maximum INR value was in the 3rd quintile [3.61, 4.22). Patients whose maximum INR value did not exceed 3.1 (1st quintile) had lower odds of one year all-cause mortality OR: 0.67 (95% CI: 0.57, 0.79), any bleeding OR: 0.84 (95% CI: 0.70, 1.00) and major bleeding OR: 0.67 (95% CI: 0.48, 0.95) compared with patients whose maximum INR value was in the 3rd quintile.

9.3.6.9 First INR

The first INR measure in patients index INR spells showed no significant association with one year cardiovascular death, stroke or MI and the bleeding outcomes. However patients whose first INR was in the 5th quintile [2.91, 18.10] had increased odds of one year all-cause mortality OR: 1.24 (95% CI: 1.08, 1.43) compared with patients whose first INR was in the 3rd quintile [1.91, 2.31).

9.3.6.10 Last INR

The last INR measure in index INR spells appeared to have a U-shaped relationship with one year outcomes. Patients whose last INR value was in the 1st quintile [0.6, 2.0) (i.e. below range) had increased odds of one year all-cause mortality OR: 1.38 (95% CI: 1.18, 1.61) and cardiovascular death, stroke or MI, OR 1.22 (95% CI: 1.00, 1.49) compared with patients whose last INR was in the 3rd quintile [2.31, 2.51). Patients whose last INR value was in the 5th quintile [2.91, 19.20] (i.e. above range) had increased odds of all-cause mortality, OR: 2.83 (95% CI: 2.43, 3.29), cardiovascular death, stroke or MI, OR: 1.89 (95% CI: 1.55, 2.31), any bleeding OR: 1.58 (95% CI: 1.31, 1.91) and major bleeding OR: 2.34 (95% CI: 1.68, 3.26).

9.3.6.11 INR trajectory

Patients with decreasing or downwards quadratic INR trajectory had lower odds of one year all-cause mortality OR: 0.79 (95% CI: 0.62, 0.99) and 0.67 (95% CI: 0.56, 0.81) respectively compared with patients who had a stable INR trajectory. Patients who had an increasing or upwards quadratic trajectory had higher odds of one year all-cause mortality OR: 1.33 (95% CI: 1.15, 1.54) and 1.64 (95% CI: 1.39, 1.93) respectively. Upwards quadratic INR trajectories were also associated with increased odds of cardiovascular death, stroke or MI OR: 1.34 (95% CI: 1.07, 1.69) and downwards quadratic INR trajectories were associated with lower odds of any bleeding, OR: 0.80 (95% CI: 0.65, 0.99) compared with stable trajectories. The odds of major bleeding did not differ significantly between the INR trajectory groups.

9.3.7 The predictive value of measures of INR control

The IDI **(Figure 9.10)** and NRI **(Figure 9.11)** estimates for measures of INR control demonstrated the additional predictive value compared with the base models. The breakdown of estimated IDI and NRI values by events and non-events are displayed in **Supplementary appendix 11.5.3** for all-cause mortality, **Supplementary appendix 11.5.4** for cardiovascular death, stroke or MI, **Supplementary appendix 11.5.5** for any bleeding and **Supplementary appendix 11.5.6** for major bleeding.

9.3.7.1 All-cause mortality

The inclusion of each of the measures of INR control resulted in a statistically significant improvement to the one year all-cause mortality model, according to both IDI and NRI metrics. The inclusion of maximum INR value and INR standard deviation into the all-cause mortality model gave the largest IDI estimates, 0.0375 (95% CI: 0.0337, 0.0413) and 0.0353 (95% CI: 0.0317, 0.0388) respectively. For both of these measures their inclusion in the model increased the mean predicted probability for patients that died at one year by ≈0.03 and decreased the mean predicted probabilities for patients that did not die at one year by 0.005. Similarly, NRI estimates for maximum INR (0.48 (95% CI: 0.44, 0.52)) and INR standard deviation (0.46 (95% CI: 0.42, 0.50)) were the largest of the measures of INR control. With maximum INR included in the model there were net increases of 0.12 (95% CI: 0.08, 0.15) of the patients that died at one year who had a higher predicted probability and 0.36 (95% CI: 0.35, 0.38) of patients that were alive at one year had a lower predicted probability, compared with predicted probabilities from the base model. With INR standard deviation included in the model there were net increases of 0.15 (95% CI: 0.11, 0.18) of the patients that died at one year had a higher predicted probability and 0.31 (95% CI: 0.30, 0.33) of patients that were alive at one year had a lower predicted probability. Time in therapeutic range had modest predictive value, IDI: 0.0159 (95% CI: 0.0137, 0.0181) and NRI: 0.30 (95% CI: 0.27, 0.34). Percent time below

therapeutic range, minimum INR, first INR and INR trajectory yielded the smallest IDI and NRI estimates.

9.3.7.2 Cardiovascular death, stroke or MI

All measures of INR control provided significant improvement to the cardiovascular death, stroke or MI model, except first percent time below therapeutic range according to the IDI estimate (0.0004 (95% CI: 0.0000, 0.0008)), first INR and INR trajectory according to the NRI estimates, 0.03 (95% CI: -0.02, 0.09) and 0.04 (95% CI: 0.00, 0.08), respectively. Maximum INR and INR standard deviation had the largest IDI, 0.0102 (95% CI: 0.0078, 0.0125) and 0.0093 (95% CI: 0.0071, 0.0116) respectively. Both variables were associated with an increase in mean predicted probability of around 0.009 for patients who had cardiovascular death, stroke or MI at one year and a decrease in mean predicted probability of 0.0007 in patients who did not have an event. However, according to the NRI estimates both maximum INR and INR standard deviation increased misspecification of patients who had events but had large increases in correctly reclassifying patients who did not have an event. For example, with INR standard deviation included in the model, of patients that had an event there was a net decrease, -0.18 (95% CI:-0.24, -0.13), of patients with higher predicted values compared with the base model. This was countered by a net increase of 0.41 (95% CI: 0.40, 0.42) of patients who did not have an event who had a decreased predicted value in the new model resulting in an overall NRI value of 0.23 (95% CI: 0.17, 0.28) for INR standard deviation. TTR group had a large net increase for events 0.52 (95% CI: 0.47, 0.56) countered by a large net decrease for non-events -0.30 (95% CI: -0.31, -0.29).

9.3.7.3 Any bleeding

Of the measures of INR control, last INR had the best predictive value in the any bleeding model according to the IDI estimates. The mean predicted probability for those who had any bleeding at one year increased by 0.0020 and for those who did not experience any bleeding decreased by 0.0002, overall IDI: 0.0021 (95% CI: 0.0013, 0.0030). The next best measures of INR control by the IDI estimates were time in therapeutic range, 0.0009 (95% CI: 0.0003, 0.0014), and INR standard deviation, 0.0008 (95% CI: 0.0003, 0.0014). The NRI estimates suggest INR standard deviation had the biggest impact on predictive value for any bleeding. While there was a net decrease of proportion of patients who had a bleeding event and a larger prediction in the INR standard deviation model compared with the base model, -0.07 (95% CI: -0.12, -0.02), there was a net increase, 0.20 (95% CI: 0.19, 0.22), of proportion of patients who didn't have a bleeding event and had lower predictions. Therefore the overall NRI estimate was 0.13 (95% CI: 0.08, 0.18). According to both IDI and NRI estimates, percent

time below therapeutic range, mean INR and minimum INR added no statistically significant predictive value to the base model.

9.3.7.4 Major bleeding

Both IDI and NRI estimates suggested that of the measures of INR control, last INR was the best addition to the major bleeding base model. With the inclusion of last INR in the base model patients that had major bleeding at one year had a 0.0031 increase in mean predicted probability and patients that did not have major bleeding at one year had a 0.0001 decrease in mean predicted probability, overall IDI: 0.0032 (95% CI: 0.0019, 0.0044). The NRI estimate for last INR suggested no significant net increase in the proportion of patients who had a major bleeding event and higher predictions, -0.03 (95% CI: -0.12, 0.06) but a net increase in the proportion of patients who did not have a major bleeding event and had lower predictions, 0.28 (95% CI: 0.27, 0.29), therefore the overall NRI was 0.25 (95%CI: 0.16, 0.34). The predictive values of TTR group, time above therapeutic range and INR standard deviation for major bleeding were also reasonable. The IDI estimates suggested time in therapeutic range, time below therapeutic range, mean INR and minimum INR did not add any statistically significant predictive value to the base model, while NRI estimates suggested only time in therapeutic range and minimum INR added no predictive value.

9.3.8 Comparison of predictions between models with different measures of INR control

Figure 9.12 displays the individual model predictions compared between base models for all-cause mortality and cardiovascular death, stroke or MI and their respective models adjusted for selected measures of INR control. The line through y=x represents no change in predicted probability between the base model and the models including measures of INR control. Points above the line represent increased predicted probabilities and points below the line represent decreased predicted probabilities from the base model to the new model.

The predictions from models including measures of INR control with poorer predictive value - TBR for all-cause mortality and first INR for cardiovascular death, stroke or MI – did not differ greatly than those from the base models. For the all-cause mortality adjusted for INR standard deviation compared with the base model there were general increases in predicted probability for patients that had events and decreases in probability for patients that did not have events.

The individual model predictions compared between base models for any bleeding and major bleeding and their respective models adjusted for selected measures of INR control are displayed in **Supplementary appendix 11.5.7**.

9.3.9 Model calibration

There was evidence of miscalibration in the all-cause mortality base model; however calibration improved with the inclusion of measures of INR control, in particular INR standard deviation. The cardiovascular death, stroke or MI base model was well calibrated and calibration was further improved with the inclusion of maximum INR in the model. However calibration was slightly worsened when adjusting for TTR.

The any bleeding models were well calibrated. However calibration appeared slightly worse in models adjusted with measures of INR control compared with the base model. The major bleeding models had good calibration and improved with the inclusion of measures on INR control.

The calibration of the base models and a selection of models adjusted for measures of INR control are displayed in **Supplementary appendix 11.5.8** (all-cause mortality and cardiovascular death, stroke or MI) and **Supplementary appendix 11.5.9** (any bleeding and major bleeding)**.**

9.4 Discussion

In the analysis of 18823 patients undergoing INR monitoring for oral anticoagulation, using linked electronic health records, I sought to estimate the associations between patient characteristics and INR stability and to assess the predictive value of various measures of INR control, some of which have not previously been investigated, for all-cause mortality, atherothrombotic and bleeding outcomes.

9.4.1 Overall distribution of TTR and INR control and their predictors

The majority of index INR spells in the study population were classed as erratic and over 20% were classed as stable. Under or over-anticoagulation for the entire duration of an INR spell was much less frequent. Prior heart valve replacement, although uncommon in the study population, was associated with higher INR variability. Smoking, as a modifiable factor was an important predictor of INR control. Patients who were recorded as smokers at the time of their index INR spell had higher variability compared with non-smokers, whilst ex-smokers did not differ significantly from non-smokers in INR variability. Major bleeding and chronic anaemia was associated with worse INR variability. A number of drugs were associated with lower INR variability when prescribed during the index INR spells however it is unlikely a causal relationship. Due to warfarin's known interaction with other drugs, patients on concomitant medication may have been subject to more intense monitoring during anticoagulation.

Knowledge of the patient characteristics which may influence the maintenance of INR control and high INR variability can help target patients for closer monitoring during anticoagulation.

9.4.2 Predictive value of measures of INR control for and bleeding and major bleeding

As previous studies have shown TTR and INR variability are associated with increased odds bleeding events[140-143,145] and shown their predictive value within a small anticoagulation clinic based cohort.[144] I have demonstrated this in CALIBER data for two bleeding endpoints, any bleeding and major bleeding at one year following oral anticoagulation. I also identified increased odds of any bleeding and major bleeding events using simpler point estimates of INR spells: a maximum INR record greater than 5 and a final INR record greater than 2.9 i.e. outside of therapeutic range.

Further to demonstrating the independent associations of measures of INR control with bleeding events I examined and quantified how model predictions may be improved with their inclusion. According to the NRI estimates, inclusion of measures of INR control such as TTR, INR standard deviation, maximum INR and last INR improved the predictions for patients that didn't have events, whilst there were negligible changes to predictions for patients who had bleeding events. Identifying patients who are at low risk of bleeding is important and may prevent patients being removed from oral anticoagulation prematurely. Categorised TTR stability group was the only measure of INR control which improved predictions for patients that had any bleeding or major bleeding events.

9.4.3 Predictive value of measures of INR control for atherothrombotic events and all-cause mortality

I found evidence of associations between all-cause mortality and cardiovascular death, stroke or MI and each of the investigated measures of INR control. Previous studies have shown associations between INR variability and TTR and risks of death and thrombotic events.[140-143,145] However no studies have investigated whether INR control provides predictive value to in addition to standard risk factors for atherothrombotic events. This was demonstrated in the analyses and furthermore I found a high initial INR value (>2.9 compared with 2.3-2.5) is associated with increased odds of all-cause mortality. The most novel summary measure, INR trajectory also had an association with all-cause mortality and atherothrombotic events.

Further to estimating relative risks, I assessed and quantified how the measures of INR control improved predictions of all-cause mortality and cardiovascular death, stroke or MI. For all-cause mortality, addition of INR standard deviation, maximum INR or time above range to the base model improved predicted probabilities for both patients that did and didn't have events. For cardiovascular death, stroke or MI the addition of INR standard deviation and maximum

INR improved predictions predominantly for those that did not have events and performed better than the standard measure of INR control, TTR.

9.4.4 Electronic health record implications

The use of electronic health records for research has become increasingly widespread. However such resources are often designed for administrative purposes and have to be adapted for use in research. It is important to generate useful and valid algorithms for defining diseases within electronic health records.[38] In this study I have demonstrated a number of ways to harness INR data for analysis in the linked electronic health records. This can aid future studies of INR in electronic health records - in particular the development of risk prediction models for anticoagulated patients which utilise INR data tailored to the endpoints of interest.

Within electronic health records calculating time in therapeutic range may be inaccurate due to patients prescribed therapeutic ranges not be specified in the data. Researchers can reasonably assume a therapeutic range depending on the indication for oral anticoagulation however in EHR it may be more appropriate to use measures of INR control that are not dependent on the unspecified time in therapeutic range, such as variability.

9.4.5 Clinical implications

Current guidelines suggest for atrial fibrillation patients with vitamin K antagonists advise maintaining a TTR above 65%.[4,5] In this study I have found that alternative measures of INR control, such as INR variability, % time above range and maximum INR may be improved predictors of adverse events associated with oral anticoagulation.

While it may seem time intensive for clinicians to assess a range of non-standard measures of INR control, algorithms could be implemented within primary care EHR systems to provide up-to-date measures of INR control within electronic health records. This would provide clinicians with a more thorough overview of patients anticoagulation resulting in more effective monitoring during and post-anticoagulation. Examples of this may be flagging patients who have an INR record exceeding 5, have high INR variability, a final INR record outside of the therapeutic range, and overall TTR indicating whether out of range measures are primarily above or below range.

9.4.6 Limitations

This study had important limitations. Not all CPRD general practices electronically receive INR records from anticoagulation clinics, therefore the depth and quality of INR data in CALIBER may be lower compared with studies set in anticoagulation clinics. The precise indication for anticoagulation for the patients in the study population was unknown. As a result the

therapeutic range for patients was unknown and assumed to be 2-3 for all patients. This may have resulted in misspecification of some patients INR stability groups, TTR, TAR and TBR. Last INR was one of the stronger performing measures of INR control across the investigated endpoints. However there is potential that its effect may be biased due to reverse causality i.e. patients had an event and were subsequently discontinued from anticoagulation, a problem worsened by lag in updating electronic health records. I did not include a washout period between the end of anticoagulation and the start of follow-up. However <1% of patients had recorded events within 7 days of their last INR record. The use of NRI and IDI measures have been criticised for their potential to provide false positive results within incorrectly specified prediction models[222,223] However in this study I aimed to demonstrate and compare the relative contribution to prediction models amongst a multiple measures of INR control as opposed to showing the absolute predictive value of a single measure.

9.4.7 Further work

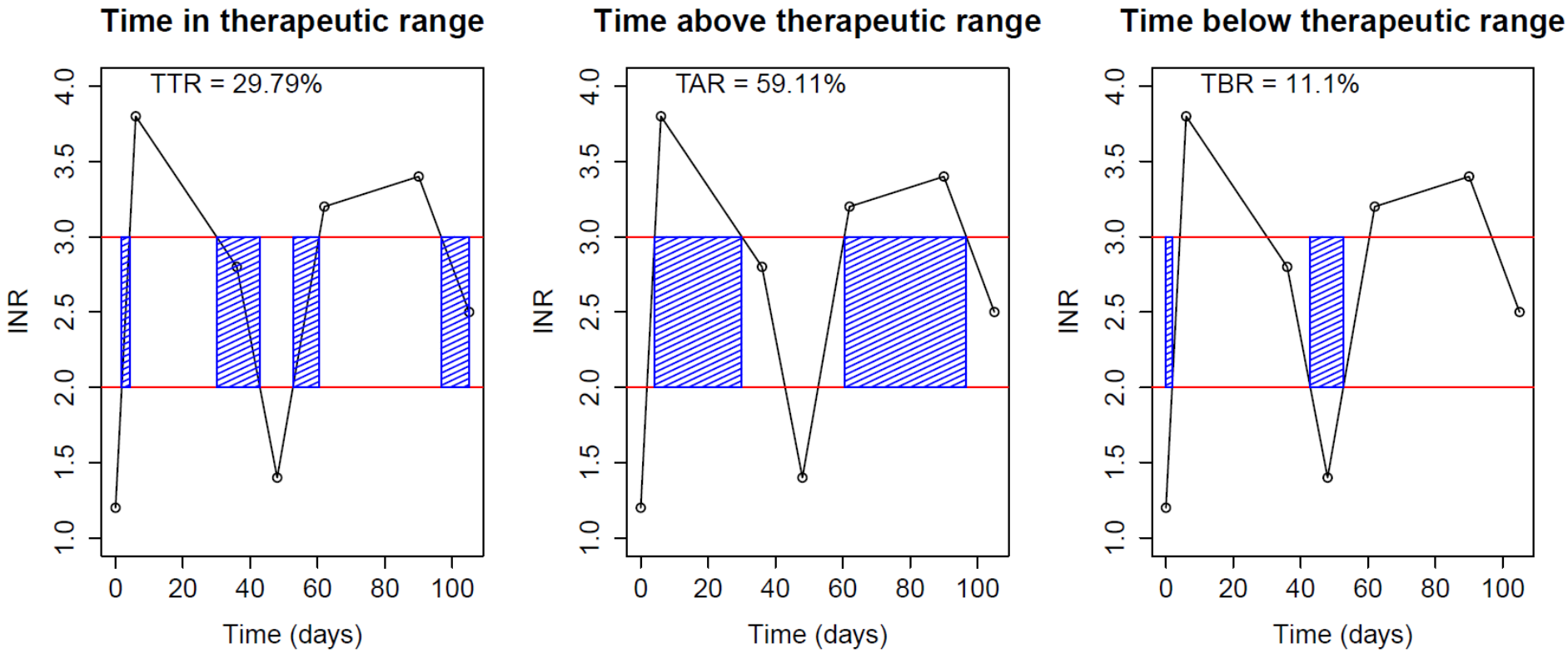
Further avenues of research in this area that I propose includes assessing the predictive value of measures of INR control within relevant subgroups, such as diseases (e.g. atrial fibrillation and venous thromboembolism) and age groups. A previous study has shown the use of electronic health records to dynamically update patients predicted survival as blood pressure records were updated using joint modelling longitudinal and survival analysis. I demonstrated the trajectory of INR enhanced prediction of our endpoints. Therefore it may be appropriate to carry out such analysis with INR data, and develop models which dynamically assess patients risks of adverse events during anticoagulation. Genotypes (CYP2CP*3) have also been shown to be associated with increased bleeding risks in patients treated with warfarin[224] and therefore an assessment of their predictive value in prognostic models would be of interest.

9.5 Conclusion

Where INR data is available it should be used for more accurate prognostic assessment of patients for bleeding, atherothrombotic and all-cause mortality risks. In particular INR variability and maximum INR improved the specificity of models, ensuring better discrimination of lower risk patients. Such measures would be can be implemented within electronic health records with simple algorithms to enable continuous monitoring and updating of patients risks.

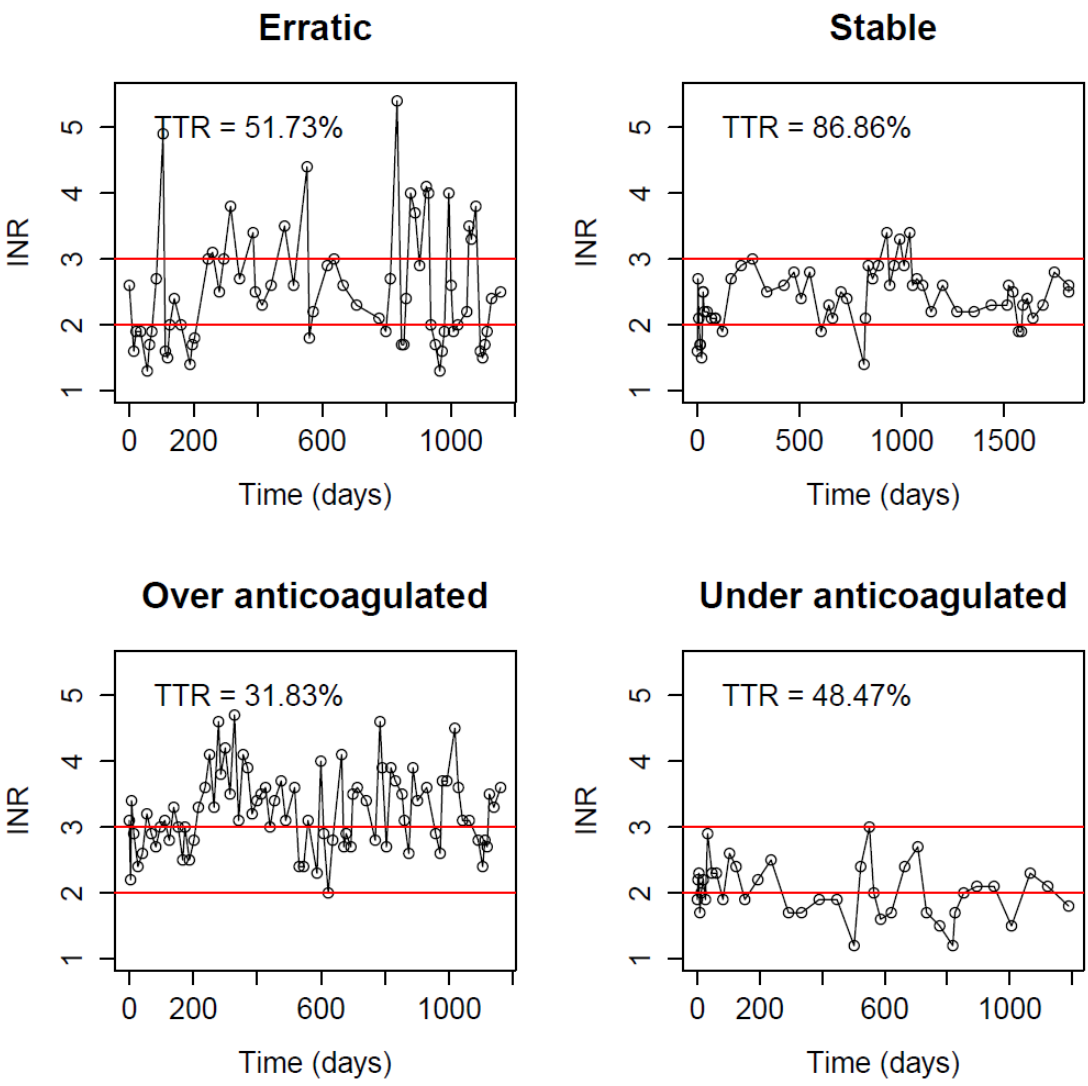
9.6 Tables and Figures

Figure 9.1: Examples of calculating time in therapeutic range, time above therapeutic range and time below therapeutic range



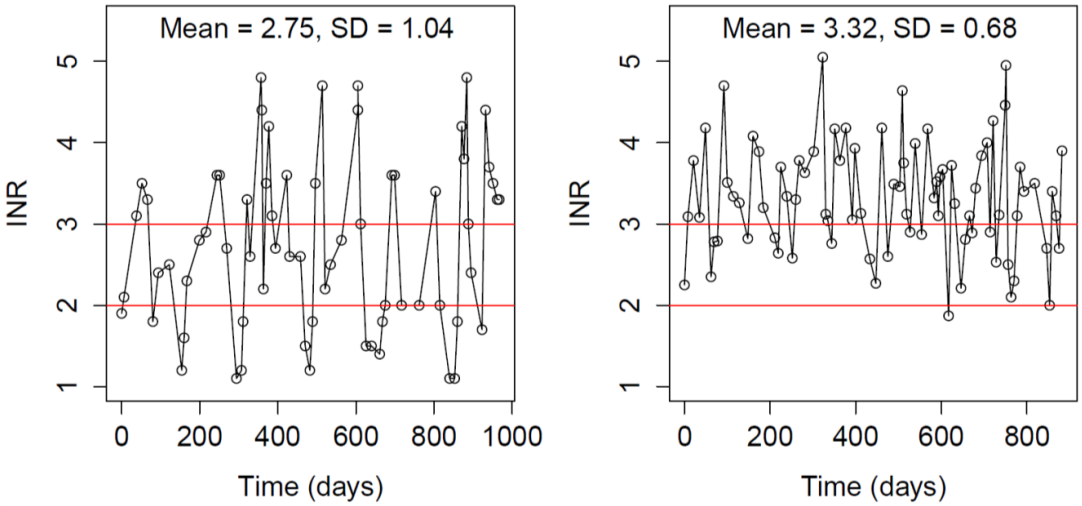
INR= international normalised ratio; TTR= Time in therapeutic range; TAR= time above therapeutic range; TBR= time below therapeutic range

Figure 9.2: Examples of time in therapeutic range control groups



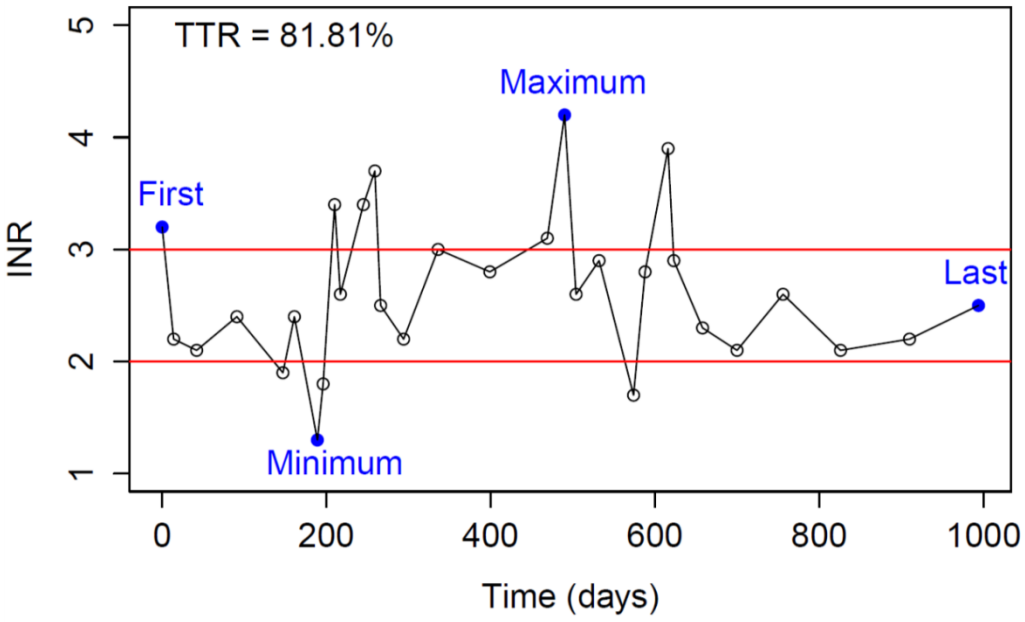
INR= international normalised ratio; TTR= Time in therapeutic range

Figure 9.3: Examples of mean INR and INR variability (standard deviation)



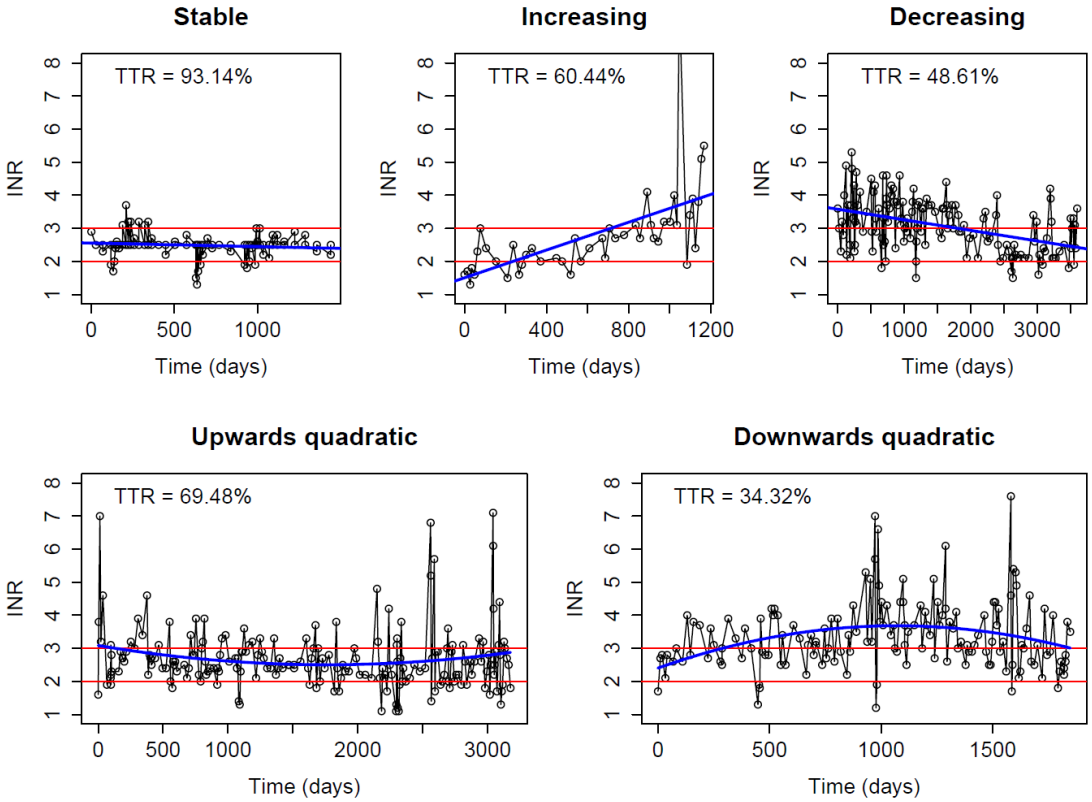
INR= international normalised ratio; SD= standard deviation

Figure 9.4: Examples of single INR values



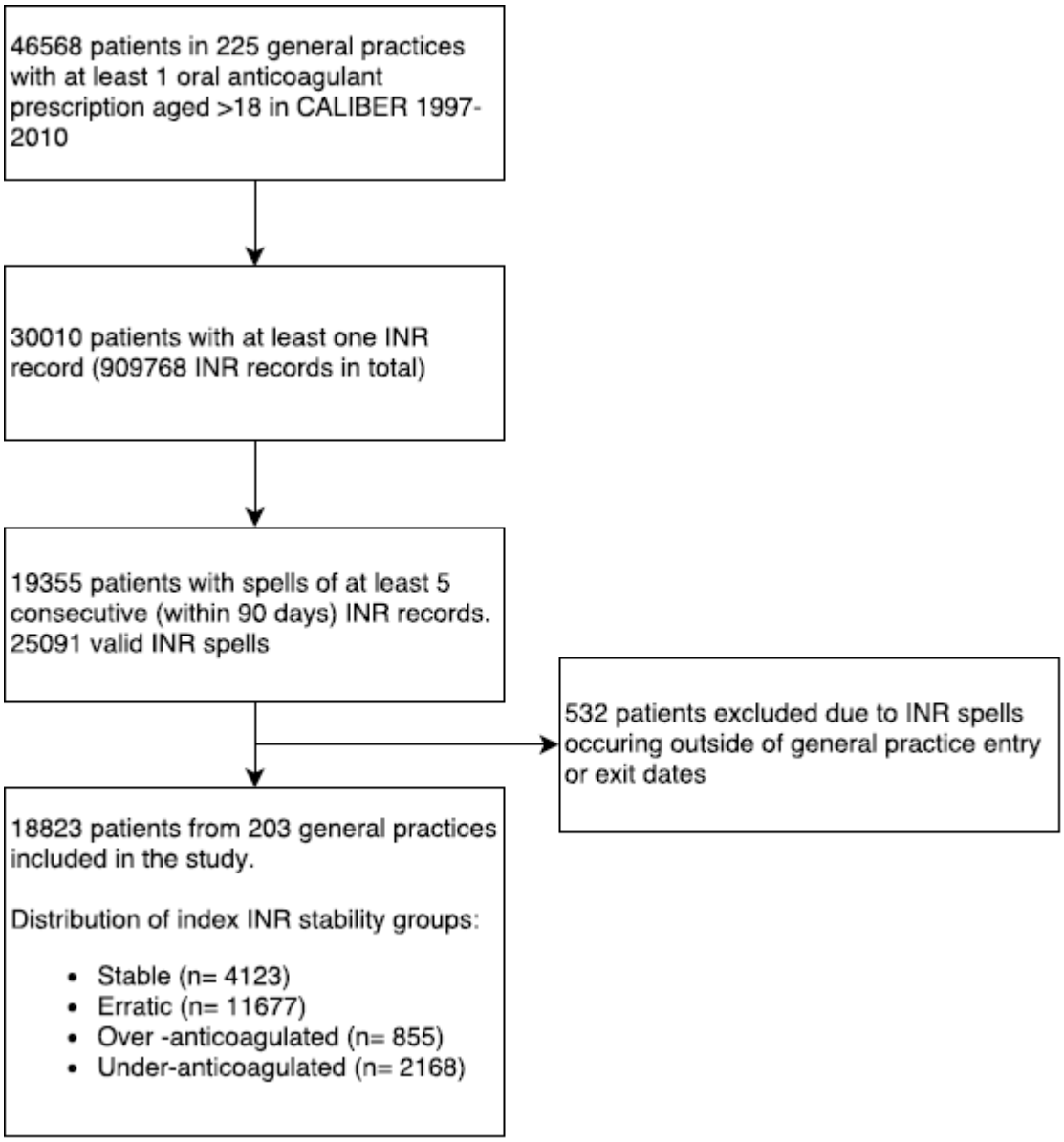
INR= international normalised ratio; TTR= Time in therapeutic range

Figure 9.5: Examples of INR trajectory groups



INR= international normalised ratio; TTR= Time in therapeutic range

Figure 9.6: Study population flow diagram



INR= international normalised ratio

Table 9.1: Study population baseline characteristics by TTR control group

		Erratic (n=11677)	Over-anticoagulated (n=855)	Stable (n=4123)	Under-anticoagulated (n=2168)
Demographics and behaviours					
Age (Years)		70.1 (13.2)	68.7 (13.4)	71.2 (11.3)	66.0 (16.0)
Women		5432 (46.5)	359 (42.0)	1743 (42.3)	1005 (46.4)
Highest quintile of deprivation		2358 (20.2)	177 (20.7)	738 (17.9)	469 (21.6)
Excess alcohol		1661 (14.2)	111 (13.0)	482 (11.7)	262 (12.1)
	Missing %	0.3	0.2	0.2	0.2
Smoking status	*Non-Smoker*	4835 (41.4)	376 (44.0)	1951 (47.3)	985 (45.4)
	Ex-Smoker	4469 (38.3)	321 (37.5)	1550 (37.6)	746 (34.4)
	Smoker	1105 (9.5)	61 (7.1)	282 (6.8)	232 (10.7)
	Missing %	10.9	11.3	8.2	9.5
Index INR spell characteristics and oral anticoagulation					
Number of INR readings in index spell		35.2 (35.65)	13.5 (16.02)	23.5 (23.66)	10.1 (5.68)
	Median (IQR)	22 (12, 44)	8 (6, 14)	15 (8, 29)	9 (6, 12)
Duration of index INR spell (days)		673 (722)	298 (394)	651 (670)	160 (154)
	Median (IQR)	399 (168.0, 936)	154 (82.5, 344)	425 (174.0, 896)	120 (70.0, 184)
VKA exposure prior to INR spell (days)		232 (506)	477 (773)	338 (612)	122 (349)
	Median (IQR)	28 (0, 152)	56 (28, 626)	28 (28, 364)	28 (0, 28)
VKA exposure during INR spell (days)		550 (618)	239 (344)	543 (592)	131 (139)
	Median (IQR)	310 (127, 755)	118 (55, 288)	338 (139, 736)	92 (49, 163)
Time in therapeutic range (%)		57.3 (16.87)	49.6 (22.88)	88.3 (5.96)	49.0 (22.98)
	Median (IQR)	60.9 (47.3, 70.6)	55.4 (31.2, 69.5)	87.1 (83.2, 92.4)	54.5 (33.2, 67.9)
Medical history					
Atrial Fibrillation		6755 (57.8)	387 (45.3)	2658 (64.5)	962 (44.4)
Myocardial infarction		1066 (9.1)	88 (10.3)	339 (8.2)	153 (7.1)
Heart failure		2248 (19.3)	164 (19.2)	656 (15.9)	262 (12.1)
Heart valve replacement		511 (4.4)	128 (15.0)	77 (1.9)	36 (1.7)
Unstable angina		584 (5)	40 (4.7)	157 (3.8)	76 (3.5)
Ischaemic or unspecified stroke		1010 (8.6)	64 (7.5)	405 (9.8)	173 (8)
Peripheral arterial disease		868 (7.4)	73 (8.5)	227 (5.5)	125 (5.8)
Venous thromboembolism		3741 (32)	321 (37.5)	1169 (28.4)	944 (43.5)
Any bleeding		3145 (26.9)	240 (28.1)	1059 (25.7)	549 (25.3)
Major bleeding		207 (1.8)	17 (2)	45 (1.1)	38 (1.8)
Chronic anaemia		1441 (12.3)	125 (14.6)	420 (10.2)	222 (10.2)
Diabetes	*Unspecified*	90 (0.8)	5 (0.6)	23 (0.6)	10 (0.5)
	Type 1	70 (0.6)	5 (0.6)	11 (0.3)	10 (0.5)
	Type 2	1372 (11.7)	87 (10.2)	422 (10.2)	225 (10.4)
Cancer		2246 (19.2)	137 (16)	650 (15.8)	374 (17.3)
Chronic obstructive pulmonary disease		1175 (10.1)	66 (7.7)	251 (6.1)	158 (7.3)
Renal disease		1521 (13)	113 (13.2)	505 (12.2)	230 (10.6)
Clinical biomarkers					
Body mass index		28.9 (6.26)	29.4 (6.53)	29.2 (5.88)	29.0 (6.68)
	Missing %	61.9	64.9	60.6	63.6
Systolic blood pressure (mmHg)		136 (19.2)	135 (18.4)	136 (17.9)	135 (18.5)

		Erratic (n=11677)	Over-anticoagulated (n=855)	Stable (n=4123)	Under-anticoagulated (n=2168)
Missing %		18.7	23.2	17.2	23.9
Diastolic blood pressure (mmHg)		78.9 (11.0)	78.3 (11.1)	79.5 (10.4)	79.0 (10.5)
Missing %		18.7	23.2	17.2	23.9
Heart rate (bpm)		81.1 (19.5)	78.1 (18.4)	80.5 (19.2)	80.3 (18.0)
Missing %		83.4	84	81.8	86.3
Haemoglobin (g/dL)		13.6 (1.78)	13.4 (1.88)	14.0 (1.57)	13.6 (1.80)
Missing %		43.7	49.7	45.5	47.3
Total white blood cell count (10^9/L)		7.65 (2.77)	7.75 (2.66)	7.38 (2.27)	7.65 (2.66)
Missing %		45.8	51	47.3	49
Platelets (10^9/L)		261 (93.3)	267 (97.4)	248 (83.3)	268 (96.4)
Missing %		45.9	51.2	47.5	49.2
Creatinine (mmol)		101.1 (36.4)	103.5 (40.5)	99.9 (26.2)	97.7 (35.1)
Median (IQR)		95 (82, 111)	98 (83, 114)	96 (84, 111)	93 (79, 107)
Missing %		34.4	41.8	33.4	41.5
Hba1c (mmol/mol)		56.8 (16.9)	52.8 (13.7)	54.2 (13.7)	57.2 (17.5)
Missing %		88.5	89.2	89.6	90.4
High-density lipoproteins (mmol)		1.40 (0.442)	1.35 (0.390)	1.38 (0.421)	1.39 (0.427)
Missing %		67.7	67.7	63.9	69.6
Total cholesterol (mmol)		4.82 (1.16)	4.83 (1.16)	4.78 (1.12)	4.87 (1.25)
Missing %		53.6	56.5	50	59.2
Prescribed medication pre- and during INR spell					
Thiazides	*pre*	4375 (37.5)	290 (33.9)	1578 (38.3)	681 (31.4)
	during	2139 (18.3)	127 (14.9)	869 (21.1)	264 (12.2)
K-sparing diuretics and aldosterone antagonists	*pre*	1644 (14.1)	132 (15.4)	489 (11.9)	219 (10.1)
	during	1536 (13.2)	81 (9.5)	375 (9.1)	116 (5.4)
K-sparing diuretics with other diuretics	*pre*	1361 (11.7)	106 (12.4)	405 (9.8)	191 (8.8)
	during	536 (4.6)	32 (3.7)	154 (3.7)	39 (1.8)
Loop diuretics	*pre*	4447 (38.1)	324 (37.9)	1303 (31.6)	563 (26.0)
	during	4675 (40.0)	261 (30.5)	1254 (30.4)	459 (21.2)
Beta blockers	*pre*	5657 (48.4)	408 (47.7)	2156 (52.3)	873 (40.3)
	during	4658 (39.9)	263 (30.8)	1730 (42)	592 (27.3)
Hypertension and heart failure drugs	*pre*	7147 (61.2)	517 (60.5)	2492 (60.4)	1211 (55.9)
	during	6420 (55)	398 (46.5)	2290 (55.5)	802 (37)
Ace inhibitors	*pre*	6428 (55)	457 (53.5)	2172 (52.7)	1079 (49.8)
	during	5012 (42.9)	302 (35.3)	1669 (40.5)	588 (27.1)
Angiotensin receptor blockers	*pre*	1446 (12.4)	103 (12)	591 (14.3)	222 (10.2)
	during	1596 (13.7)	86 (10.1)	618 (15)	164 (7.6)
Calcium channel blockers	*pre*	4121 (35.3)	282 (33)	1458 (35.4)	651 (30)
	during	2992 (25.6)	160 (18.7)	1141 (27.7)	377 (17.4)
Antiplatelets	*pre*	5552 (47.5)	333 (38.9)	1944 (47.2)	831 (38.3)
	during	1560 (13.4)	68 (8)	451 (10.9)	263 (12.1)
Statins	*pre*	4325 (37)	336 (39.3)	1603 (38.9)	703 (32.4)
	during	4940 (42.3)	333 (38.9)	1787 (43.3)	589 (27.2)
Insulin	*pre*	730 (6.3)	58 (6.8)	202 (4.9)	133 (6.1)

		Erratic (n=11677)	Over-anticoagulated (n=855)	Stable (n=4123)	Under-anticoagulated (n=2168)
	during	768 (6.6)	39 (4.6)	186 (4.5)	87 (4)
Anti-diabetics	*pre*	1053 (9)	69 (8.1)	289 (7)	177 (8.2)
	during	1144 (9.8)	58 (6.8)	333 (8.1)	152 (7)
Nonsteroidal anti-inflammatory drugs	*pre*	7831 (67.1)	548 (64.1)	2727 (66.1)	1454 (67.1)
	during	1174 (10.1)	42 (4.9)	283 (6.9)	104 (4.8)
Antidepressants	*pre*	3752 (32.1)	244 (28.5)	1047 (25.4)	707 (32.6)
	during	2355 (20.2)	123 (14.4)	507 (12.3)	282 (13)
Digoxin	*pre*	3307 (28.3)	194 (22.7)	1341 (32.5)	387 (17.9)
	during	3735 (32)	179 (20.9)	1361 (33)	365 (16.8)
Amiodarone	*pre*	1089 (9.3)	94 (11)	345 (8.4)	137 (6.3)
	during	277 (10.9)	74 (8.7)	239 (5.8)	110 (5.1)

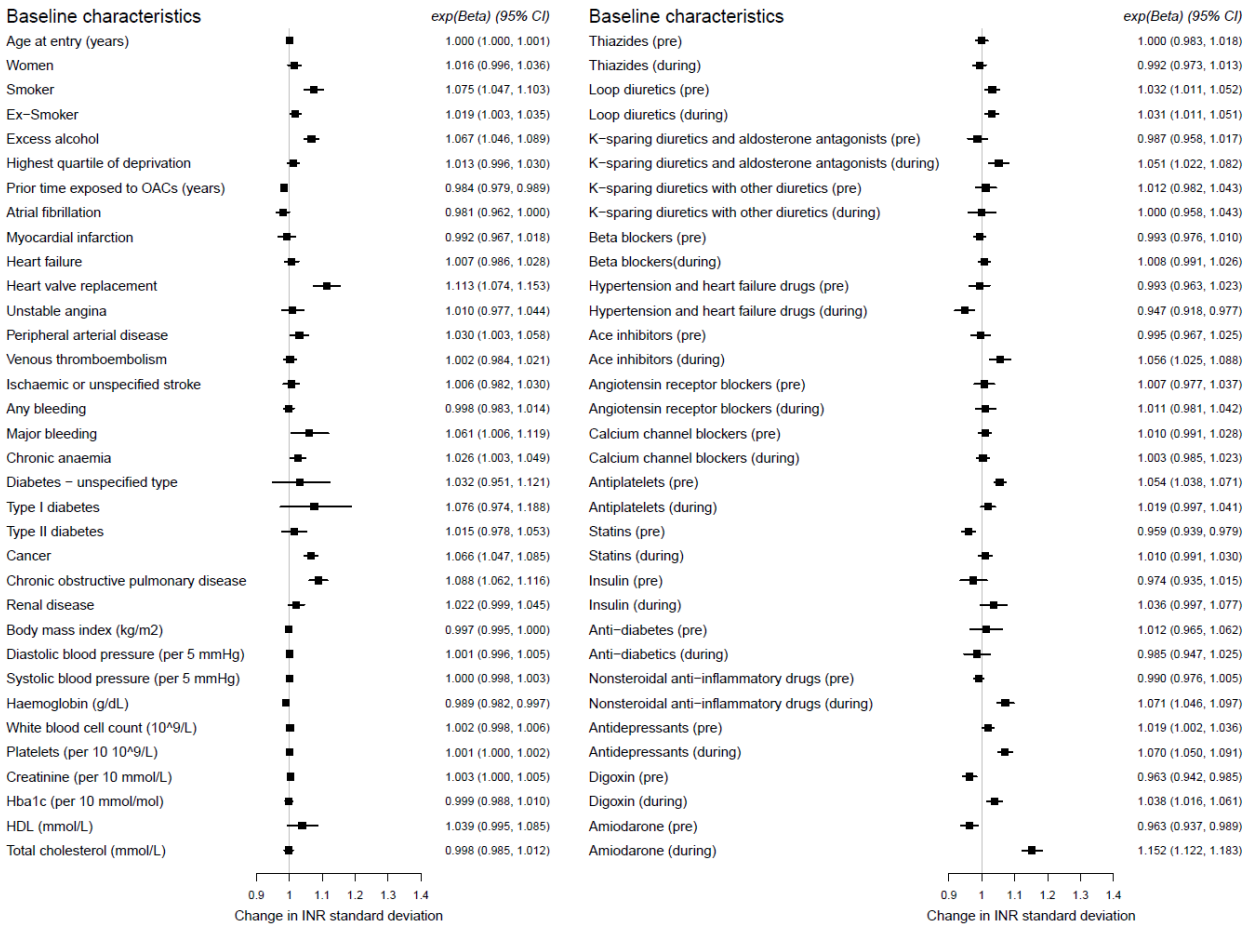
Mean (standard deviation) for continuous variables and n (%) for categorical variables unless otherwise stated; INR= international normalised ratio

Table 9.2: Quintiles of measures of INR control

Measure of INR control	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Time in therapeutic range (%)	[0.0, 45.4)	[45.4, 60.5)	[60.5, 70.9)	[70.9, 81.0)	[81.0, 100.0]
Time above therapeutic range (%)	[0.0, 1.5)	[1.5, 8.5)	[8.5, 14.9)	[14.9, 25.4)	[25.4, 100.0]
Time below therapeutic range (%)	[0.0, 5.0)	[5.0, 12.5)	[12.5, 20.6)	[20.6, 33.8)	[33.8, 100.0]
Mean INR	[0.9, 2.2)	[2.2, 2.4)	[2.4, 2.5)	[2.5, 2.7)	[2.7, 8.2]
INR standard deviation	[0.00, 0.46)	[0.46, 0.58)	[0.58, 0.71)	[0.71, 0.92)	[0.92, 5.92]
Minimum INR	[0.0, 1.1)	[1.1, 1.3)	[1.3, 1.5)	[1.5, 1.7)	[1.7, 4.0)
Maximum INR	[0.9, 3.1)	[3.1, 3.6)	[3.6, 4.2)	[4.2, 5.2)	[5.2, 20.0)
First INR	[0.6, 1.4)	[1.4, 1.9)	[1.9, 2.3)	[2.3, 2.9)	[2.9, 18.1)
Last INR	[0.9, 2.0)	[2.0, 2.3)	[2.3, 2.5)	[2.5, 2.9)	[2.9, 19.2)

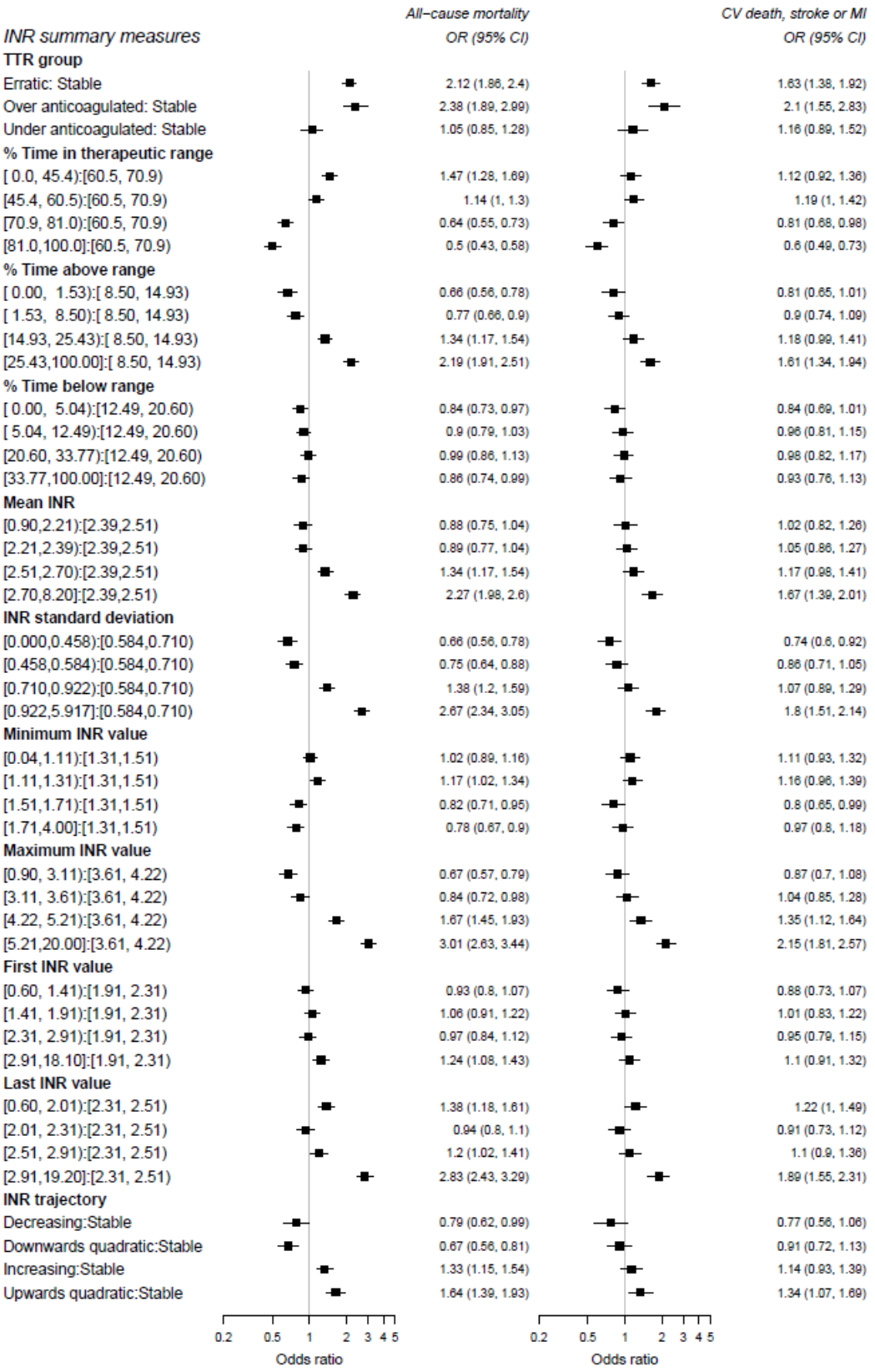
INR= International normalised ratio

Figure 9.7: Association between baseline characteristics and INR standard deviation



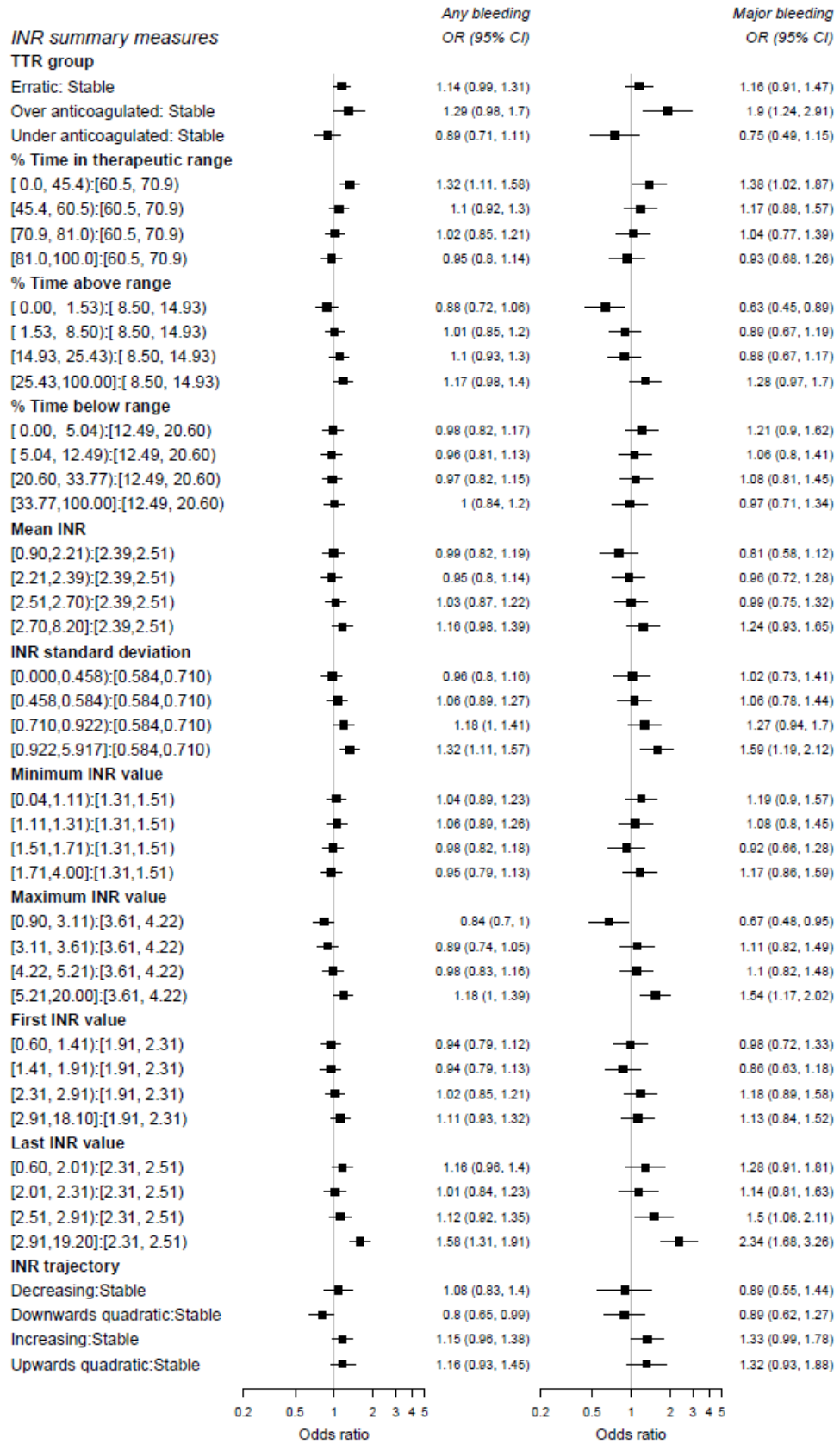
OAC= oral anticoagulant; HDL= high-density lipoproteins; pre= drugs at least one prescription at any time prior to index INR spell; during= drugs prescribed during: index INR spell

Figure 9.8: Association between measures of INR control and one year all-cause mortality (2686 events) and cardiovascular death, stroke or myocardial infarction (1304 events)



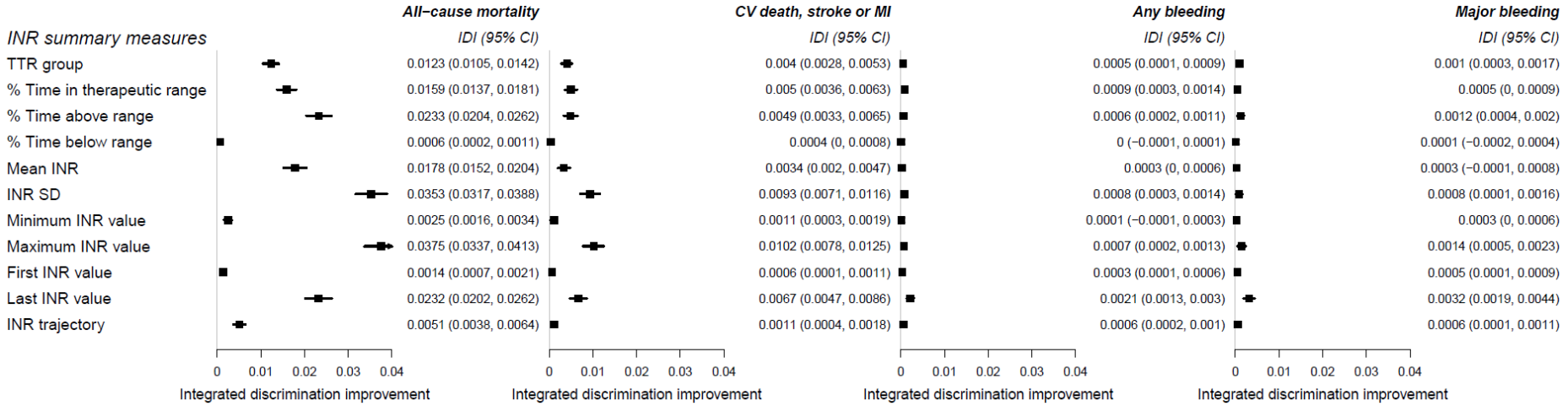
INR= international normalised ratio; TTR= time in therapeutic range; MI= myocardial infarction; OR= odds ratio; CI= confidence interval

Figure 9.9: Association between measures of INR control with any bleeding (1446 events) and major bleeding (478 events)



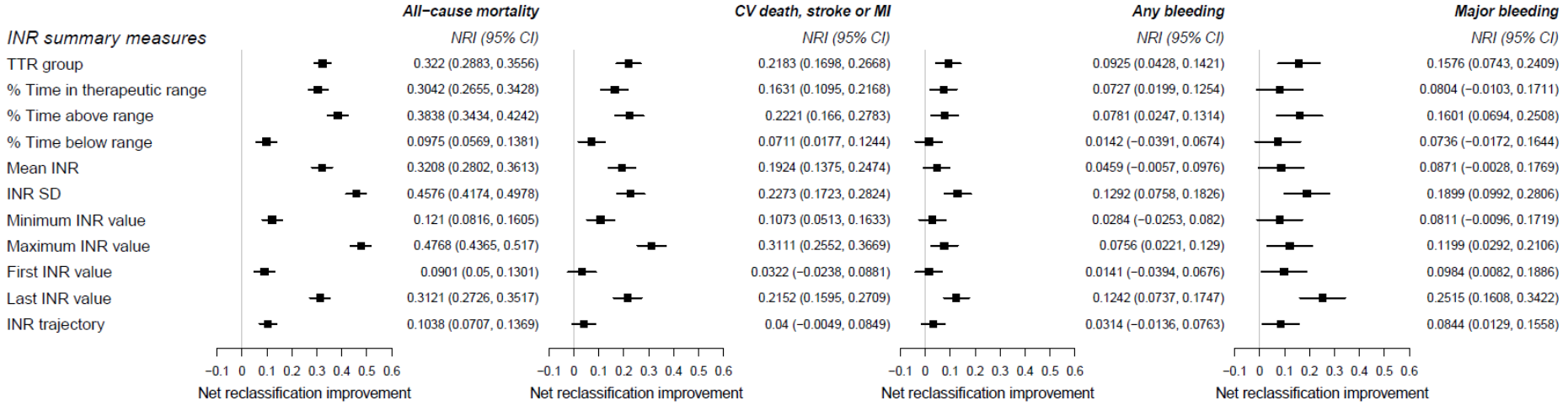
INR= international normalised ratio; TTR= time in therapeutic range; OR= odds ratio; CI= confidence interval

Figure 9.10: Integrated discrimination improvement with the inclusion of measures of INR control in models for 1 year all-cause mortality, cardiovascular death, stroke or MI, any bleeding and major bleeding events



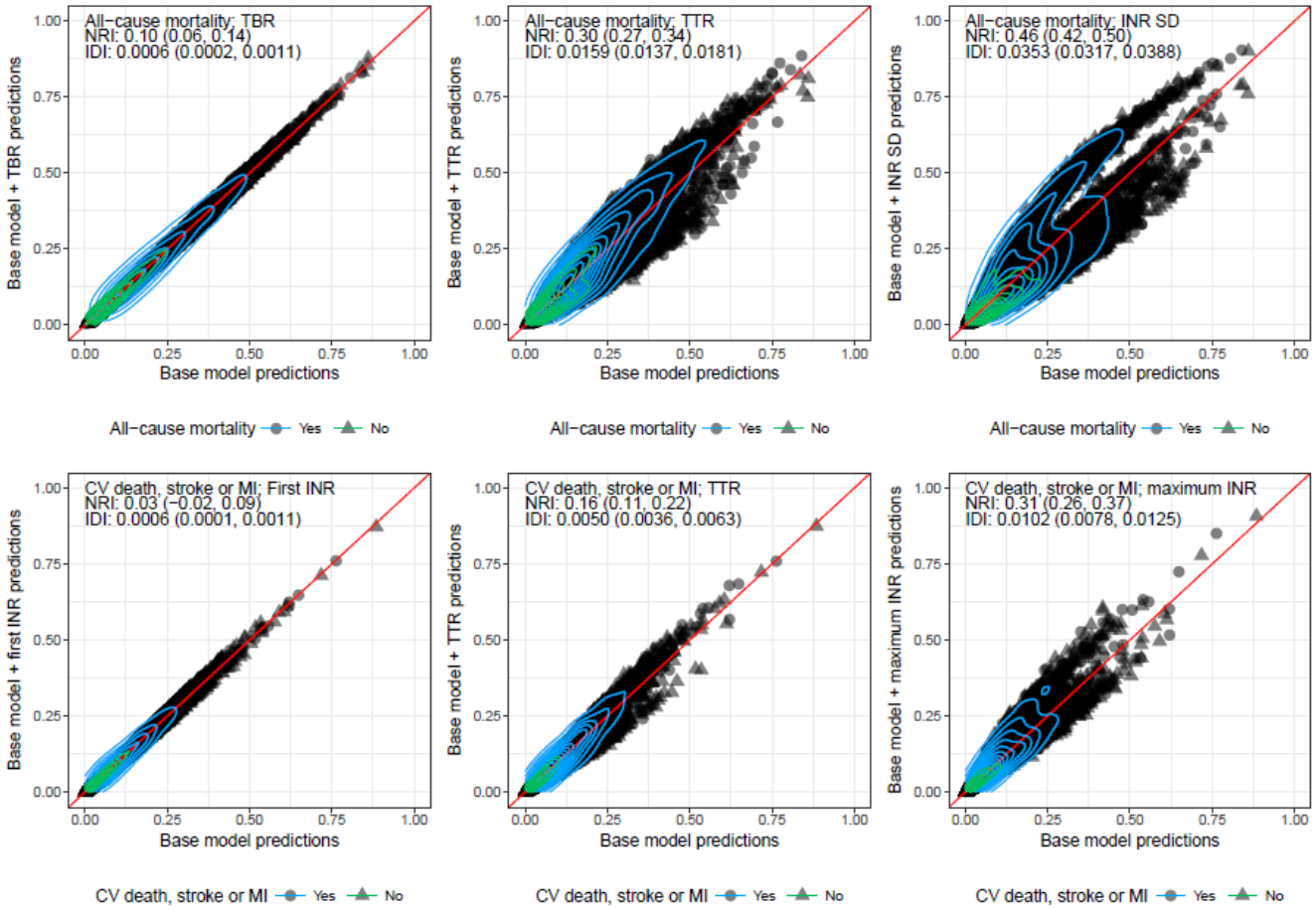
INR= international normalised ratio; TTR= time in therapeutic range; SD= standard deviation; IDI= integration discrimination improvement; CI= confidence interval; MI= myocardial infarction

Figure 9.11: Net reclassification improvement with the inclusion of measures of INR control in models for 1 year all-cause mortality, cardiovascular death, stroke or MI, any bleeding and major bleeding events



INR= international normalised ratio; TTR= time in therapeutic range; SD= standard deviation; NRI= net reclassification improvement; CI= confidence interval; MI= myocardial infarction

Figure 9.12: Predicted all-cause mortality and cardiovascular death, stroke or MI; comparing base models with models including measures of INR control



TBR= time below therapeutic range; NRI= net reclassification improvement; IDI= integrated discrimination improvement; TTR= time in therapeutic range; SD= standard deviation; CV= cardiovascular; MI= myocardial infarction;

10 Conclusion

10.1 Thesis overview

In my thesis, using routinely collected linked electronic health records (EHR) in England, I performed analyses that address the gaps in our current understanding of the benefits and harms associated with antithrombotic therapies with a focus on approaches to personalise treatment strategies.

I undertook a review on the current literature to identify relevant areas of research relating to antithrombotic therapies, summarising the various study designs and populations, therapeutic intervention, exposures, primary and secondary outcomes, analytical methods, conclusion and implications.

I provided an overview of the CALIBER linked electronic health records platform and each of the data sources linked data sources: primary care data from CPRD, hospital admissions from HES, myocardial infarction disease registry data from MINAP and cause-specific mortality data from ONS. I described accessing the data through applying and gaining study approval from the Independent Science Advisory Committee (ISAC) and the methods employed to curate the routinely collected patient information into research ready datasets.

10.2 Summary of key findings and impact

10.2.1 Chapter 4: Antithrombotic drug use and inferring drug indications

In Chapter 4 my aims were (1) to investigate the prevalence of antithrombotic therapy prescribing in CALIBER; (2) to determine how many of patient's first antithrombotic prescription had an indication recorded within a reasonable timeframe.

Through cross-referencing 20 cardiovascular and non-cardiovascular indications for prescribing antithrombotic therapies, against the first antithrombotic therapy prescription for 277,598 patients I was able to infer an indication for 45.7%, using clinical diagnosis codes recorded within the time frame of 180 days prior to, to 30 days following the prescription. To my knowledge, this was the first study of its kind for antithrombotic drugs with a more comprehensive search of indications for prescriptions through the use of linked EHR compared with using a single data source. My findings can inform monitoring of antithrombotic drug use, the patients issued prescriptions and for what reason. This is important due to the potentially serious bleeding side effects associated with antithrombotic drugs.

10.2.2 Chapter 5: Developing bleeding phenotypes in CALIBER

My objectives in this chapter were (1) to determine the extent that population-based EHRs can replicate the definitions of bleeding used in trials; (2) to determine if bleeding severity can be defined using bleeding diagnosis codes and other supporting information available within EHR.

I presented and applied methodology to develop phenotyping algorithms in EHR which define bleeding events and determine which available data were useful and usable markers of bleeding severity. The identified markers of bleeding severity were bleeding anatomical site (intracranial or ruptured abdominal aortic aneurysm), hospitalisation, duration of hospitalisation, transfusion and number of bleeding records on a single date. Amongst patients with no bleeding diagnosis in primary care or hospital records, I inferred a small number of bleeding cases using transfusion, iron deficiency anaemia, haemoglobin and endoscopy records. Of the bleeding events identified within the study population only 13.2% were identified in more than one of primary care, hospital records and death registry records, highlighting the importance of using linked electronic health records to improve case ascertainment.

The bleeding phenotyping algorithms offer a scalable approach to the analysis of harms of antithrombotic therapies using EHR. The algorithms could potentially be integrated within EHR systems used by general practitioners and clinicians to aid monitoring of bleeding in the general population.

10.2.3 Chapter 6: Incidence and prognosis of bleeding events in four common cardiovascular diseases

The objectives of this chapter were (1) to estimate the incidence of major bleeding and any bleeding across patients with atrial fibrillation, acute myocardial infarction, unstable and stable angina who are on different antiplatelet and anticoagulation regimens; (2) to assess time trends in bleeding incidence along with the changes in antithrombotic management (3) to evaluate the association between bleeding severity and long term prognosis in terms of all-cause mortality, atherothrombotic events and recurrent bleeding.

Applying the bleeding phenotypes developed in chapter 5 I found that amongst 128815 patients, 27259 (21.2%) had at least one bleeding event, with 5 year risks of bleeding of 29.1%, 21.9%, 25.3% and 23.4% following diagnoses of atrial fibrillation, acute myocardial infarction, unstable angina and stable angina respectively. Incidence of bleeding doubled between 1998 and 2010. Patients with hospitalised bleeding and primary care bleeding, with or without markers of severity, were at increased risk of all-cause mortality and atherothrombotic events

compared to those with no bleeding. Patients with bleeding diagnosed in primary care presenting with markers of bleeding severity had doubled risk of all-cause mortality compared with patients with no bleeding record. The prognosis was worse for patients with bleeding diagnosed in hospital care. This study demonstrated that bleeding is a commonly occurring public health problem among atrial fibrillation and coronary disease patients. It is associated with high mortality. Closer monitoring of individual patient bleeding risks using prognostic models may contribute to the prevention of avoidable harms associated with antithrombotic therapies.

10.2.4 Chapter 7: Development and validation of prognostic models for atherothrombotic events and bleeding in stable myocardial infarction survivors

The objectives of the chapter were to (1) develop and validate prognostic models for atherothrombotic and bleeding risks using a cohort of patients who were atherothrombotic event-free one year following a MI and (2) demonstrate how these models may be used to aid treatment decisions.

Using population-based electronic health records (EHRs) for patients evaluated 1 year after acute myocardial infarction (MI), I developed (n = 12694 patients) and validated (n = 5613) prognostic models for cardiovascular (cardiovascular death, MI or stroke) events and three different bleeding endpoints. I applied trial effect estimates to determine potential benefits and harms of dual antiplatelet therapy (DAPT) and the net clinical benefit of individuals. Prognostic models for cardiovascular events (c-index: 0.75 (95% CI: 0.74, 0.77)) and bleeding (c index 0.72 (95% CI: 0.67, 0.77)) were well calibrated. 3-year risk of cardiovascular events was 5.2% in the lowest and 46.7% in the highest-risk individuals, while for major bleeding, it was 0.3% in the lowest and 5.4% in the highest-risk patients. For every 10000 patients treated per year, we estimated 249 (95% CI: 228, 269) cardiovascular events prevented and 134 (95% CI: 87, 181) major bleeding events caused in the highest-risk patients, and 28 (95% CI: 19, 37) cardiovascular events prevented and 9 (95% CI: 0, 20) major bleeding events caused in the lowest-risk patients. There was a net clinical benefit of prolonged DAPT in 63–99% patients depending on how benefits and harms were weighted.

Given patients baseline characteristics change over the one year following an acute MI, the developed and validated prognostic models may help to personalise decisions for prolonged DAPT 1-year following acute MI using up to date information. Web calculators and web tools integrated within EHR systems used by clinicians can easily compute the estimated risks based on individual patient profiles.

10.2.5 Chapter 8: Predictors and outcomes of INR time in therapeutic range

The objectives of this chapter were (1) to describe the international normalised ratio (INR) and time in therapeutic range (TTR) data available within CALIBER within 4 distinct populations indicated for oral anticoagulation with vitamin K antagonist (VKA) therapy (atrial fibrillation, venous thromboembolism, heart valve replacement or other indication); (2) to assess patient baseline characteristics associated with time in therapeutic range in our study population and whether they agree with the conclusions made in previous literature; (3) to assess the strength of the SAME-TT_2R_2 score for predicting TTR in our population; (4) to assess all-cause mortality, atherothrombotic and bleeding outcomes following oral anticoagulation with TTR below the recommended levels.

I found that the average TTR was 62.8% across the study population. The highest mean TTR was observed in atrial fibrillation patients, who on average spent two thirds of their INR monitoring time within therapeutic range. Similar to previous literature, demographic factors of age >40 years, men, white, and a lower deprivation were associated with higher TTR. I found that SAME-TT_2R did not perform well to predict patients TTR in my data, as the simple point-based scheme and weighting of predictors in SAME-TT_2R_2 may not be sufficient to detect small changes in TTR.

My study observed that for patients who maintained a TTR of 65% or greater had improved long term prognosis and reduced bleeding risk. This was true not only for atrial fibrillation patients, but also for patients with venous thromboembolism, heart valve replacements and other indications for oral anticoagulation. In the absence of guidelines recommended targets for TTR amongst the non-atrial fibrillation patient groups, the study results supports maintaining oral anticoagulation treatment control as measured by TTR above 65%.

10.2.6 Chapter 9: The predictive value of measures of INR control for atherothrombotic and bleeding outcomes

The objectives in this chapter were (1) To illustrate overall distribution of INR control within our cohort; (2) To estimate predictors of INR variability and (3) To estimate the predictive value of a range of measures of INR control for atherothrombotic and bleeding events; Is TTR the best measure and do measures of INR control work well for predicting both bleeding and atherothrombotic endpoints and add value to standard risk factors?

In my analysis of 18823 patients undergoing INR monitoring during oral anticoagulation, I investigated the INR control measures of time in therapeutic range (%), time above therapeutic range (%), time below therapeutic range (%), mean INR, INR variability, minimum INR, maximum INR, first INR, last INR.

The majority of index INR spells in our study population were classed as erratic and over 20% were classed as having stable INR control. Lower TTR and higher INR variability were associated with increased odds of bleeding and major bleeding defined by the algorithm developed in **chapter 5**. We found evidence of associations between all-cause mortality and cardiovascular death, stroke or MI and each of the investigated measures of INR control.

Where INR data is available it should be used for more accurate prognostic assessment of patients for bleeding, atherothrombotic and all-cause mortality risks. In particular INR variability and maximum INR improved the specificity of models, ensuring better discrimination of lower risk patients. Such measures would be can be implemented within electronic health records with simple algorithms to enable continuous monitoring and updating of patients risks.

10.3 Limitations

Further to the limitations identified within each chapter there were broad limitations encountered within this thesis. The version of CALIBER used in this thesis covers records from 1997 to 2010, which may be considered out of date, especially for research in medical areas with ever-changing guidelines and recommended treatments.

The studies are observational in design, and are subject to bias, such as misclassification of exposures, information bias or confounding that failed to be accounted in the analysis. Missing data is an issue for observational cohorts; in particular for electronic health records it is difficult to know what the missingness mechanism of data may be therefore not all methods for handling missing data may be appropriate.

Aspirin prescribing and therefore antithrombotic prescribing will have been underestimated in our data, due to the fact that aspirin may be bought over the counter.

10.4 Overall impact

The thesis addressed the many considerations regarding the prescribing and use of antithrombotic therapies in clinical practice, using real-world population based linked electronic health records. The bleeding phenotypes are scalable and may be used by researchers in future electronic health record studies. The prognostic models developed may be used in clinical practice to aid refinement of long-term treatment decisions for myocardial infarction survivors. I have highlighted the prognostic importance of measures of INR control and recommend their use in models for cardiovascular events and bleeding where the data is available.

Beyond the theme of antithrombotic therapy, this thesis demonstrates a framework on how population-based linked electronic health records may be harnessed to study a widely used class of medication, procedures or interventions, develop definitions for related clinical events, and investigate ways to personalise and optimise treatment strategies through increasing benefits and reducing harms in the relevant populations.

11 Supplementary appendix

This chapter contains the supplementary material for chapters 3, 5 and 7.

11.1 Supplementary Appendix (Chapter 3)

11.1.1 CALIBER data dictionary

CPRD patient and general practice data fields:

patid: Encrypted CPRD patient identifier

pracid: unique CPRD general practice identifier

prac_region: the UK region in which the general practice is situated

prac_lcd: date of last data collection from general practice

prac_uts: date at which general practice data considered to be of up-to-standard research quality

gender: patient gender

dob: patient date of birth

frd: patients date of first registration with the general practice

crd: patients date of current registration with the general practice

tod: date of transfer out of CPRD general practice

toreason: patient transfer out reason

deathdate: patient date of death

date_entry: the date which is the latest of crd, prac_uts

date_exit: the date which is the earliest of tod, deathdate, prac_lcd

ethnicity: patient ethnicity

CPRD diagnosis data fields:

patid: Encrypted CPRD patient identifier

eventdate: date of record

medcode: CPRD unique diagnosis code

CPRD prescription data fields:

patid: Encrypted CPRD patient identifier

eventdate: date of record

consid: The identifier for the consultation, used in combination with pracid

prodcode: CPRD unique product code

staffid: The identifier for the staff at the general practice who input the data

textid: The identifier for freetext information regarding the prescription, if available

bnfcode: code referring to the BNF chapter and section

qty: The quantity of the product prescribed

ndd: The numeric daily dose prescribed

numdays: Length of prescription/ therapy (days)

numpacks: The number of product packs prescribed

packtype: The size or type of pack

issueseq: The sequence number for repeat prescriptions

CPRD clinical biomarkers data fields:

patid: Encrypted CPRD patient identifier

eventdate: date of record

data1: depending on data type – usually biomarker measurement

data2: depending on data type – usually biomarker measurement units

HES admission data fields:

patid: Encrypted CPRD patient identifier

date_admission: date of hospital admission

date_discharged: date of hospital discharge

icd: ICD10 code for the hospital admission

epi_primary: Indicates if the ICD-10 code is the primary diagnosis in the episode

hosp_primary: Indicates if the ICD-10 code is the primary diagnosis in the hospitalisation

HES procedure data fields:

patid: Encrypted CPRD patient identifier

date_admission: date of hospital admission

date_procedure: date of procedure

date_discharged: date of hospital discharge

opcs: OPCS-4 code for the procedure

ONS mortality data fields:

patid: Encrypted CPRD patient identifier

dod: date of death

cod: ICD-10 code recorded as the underlying cause of death

cod1-cod15: ICD-10 codes for up to 15 secondary causes of death

11.1.2 My ISAC application for the prognostic modelling study

ISAC APPLICATION FORM

PROTOCOLS FOR RESEARCH USING THE CLINICAL PRACTICE RESEARCH DATALINK (CPRD)

ISAC use only: Protocol Number Date submitted		**IMPORTANT** **If you have any queries, please contact ISAC Secretariat**

1. Study Title
Prognostic models for risk of serious atherothrombotic and bleeding events and death in patients who survived after myocardial infarction

2. Principal Investigator (full name, job title, organisation & e-mail address for correspondence regarding this protocol)

3. Affiliation (full address)

4. Protocol's Author (if different from the principal investigator)

5. List of all investigators/collaborators (*please list the names, affiliations and e-mail addresses* of all collaborators*, other than the principal investigator)

Please note that your ISAC application form and protocol **must be copied to all e-mail addresses listed above at the time of submission of your application to the ISAC mailbox. Failure to do so will result in delays in the processing of your application.*

6. Type of Institution (please tick one box below)

Academia	☒	Research Service Provider	☐	Pharmaceutical Industry	☐
NHS	☐	Government Departments	☐	Others	☐

7. Financial Sponsor of study

Pharmaceutical Industry *(please specify)*	☒ **AstraZeneca**	Academia*(please specify)*	☐
Government / NHS *(please specify)*	☐	None	☐
Other *(please specify)*	☐		

8. Data source *(please tick one box below)*

Sponsor has on-line access	☐	Purchase of ad hoc dataset	☒
Commissioned study	☐		
Other	☐ *(please specify)*		

9. Has this protocol been peer reviewed by another Committee?
Yes* ☐ No ☒
** Please state in your protocol the name of the reviewing Committee(s) and provide an outline of the review process and outcome.*

10. Type of Study *(please tick all the relevant boxes which apply)*

Adverse Drug Reaction/Drug Safety	☐	Drug Use	☐	Disease Epidemiology	☐
Drug Effectiveness	☐	Pharmacoeconomic	☐	Other	☒

Prognostic modelling

11. This study is intended for:

Publication in peer reviewed journals	☒	Presentation at scientific conference	☒
Presentation at company/institutional meetings	☒	Other	

12. Does this protocol also seek access to data held under the CPRD Data Linkage Scheme?

Yes ☒ No ☐

13. If you are seeking access to data held under the CPRD Data Linkage Scheme*, please select the source(s) of linked data being requested.

☒ Hospital Episode Statistics ☐ Cancer Registry Data**
☐ MINAP ☒ ONS Mortality Data
☒ Index of Multiple Deprivation/ Townsend Score
☐ Mother Baby Link ☐ Other: (please specify)

** As part of the ISAC review of linkages, the protocol may be shared - in confidence - with a representative of the requested linked data set(s) and summary details may be shared - in confidence - with the Confidentiality Advisory Group of the Health Research Authority.*

***Please note that applicants seeking access to cancer registry data must provide consent for publication of their study title and study institution on the UK Cancer Registry website. Please contact the CPRD Research Team to discuss this requirement further.*

14. If you are seeking access to data held under the CPRD Data Linkage Scheme, have you already discussed your request with a member of the Research team?

Yes ☒ No* ☐

**Please contact the CPRD Research Team to discuss your requirements before submitting your application.*

Please list below the name of the person/s at the CPRD with whom you have discussed your request.

15. If you are seeking access to data held under the CPRD Data Linkage Scheme, please provide the following information:

The number of linked datasets requested: **4**

A synopsis of the purpose(s) for which the linkages are required:
The study requires description of patient characteristics, medication use, anthropometric measurements and laboratory tests included in CPRD.
Further, the study requires details of hospital admissions, length of stay and reason for admission provided by HES.
Finally, it requires cause-specific mortality from ONS to assess fatalities and causes of death.

Is linkage to a local dataset with <1 million patients being requested?

Yes* ☐ No ☒

** If yes, please provide further details:*

16. If you have requested linked data sets, please indicate whether the Principal Investigator or any of the collaborators listed in response to question 5 above, have access to any of the linked datasets in a patient identifiable form, or associated with a patient index.

Yes* ☐ No ☒

** If yes, please provide further details:*

17. Does this protocol involve requesting any additional information from GPs?

Yes* ☐ No ☒

* *Please indicate what will be required:*

Completion of questionnaires by the GP[ψ] Yes ☐ No ☐
Provision of anonymised records (e.g. hospital discharge summaries) Yes ☐ No ☐
Other (please describe)

[ψ] *Any questionnaire for completion by GPs or other health care professional must be approved by ISAC before circulation for completion.*

18. Does this protocol describe a purely observational study using CPRD data (this may include the review of anonymised free text)?

Yes* ☒ No** ☐

** Yes: If you will be using data obtained from the CPRD Group, this study does not require separate ethics approval from an NHS Research Ethics Committee.*
*** No: You may need to seek separate ethics approval from an NHS Research Ethics Committee for this study. The ISAC will provide advice on whether this may be needed.*

19. Does this study involve linking to patient *identifiable* data from other sources?

Yes ☐ No ☒

20. Does this study require contact with patients in order for them to complete a questionnaire?

Yes ☐ No ☒

N.B. Any questionnaire for completion by patients must be approved by ISAC before circulation for completion.

21. Does this study require contact with patients in order to collect a sample?

Yes* ☐ No ☒

** Please state what will be collected*

22. Experience/expertise available

Please complete the following questions to indicate the experience/expertise available within the team of researchers actively involved in the proposed research, including analysis of data and interpretation of results

	Previous GPRD/CPRD Studies	Publications using GPRD/CPRD data
None	☐	☐
1-3	☐	☐
> 3	☒	☒

Is statistical expertise available within the research team? Yes ☒ No ☐

Is statistical expertise available within the research team? Yes ☒ No ☐

If yes, please outline level of experience

Is experience of handling large data sets (>1 million records) available within the research team? Yes ☒ No ☐

If yes, please outline level of experience

The team at UCL has used the CALIBER dataset previously demonstrating their experience of handling large datasets.

Is UK primary care experience available within the research team? Yes ☒ No ☐

If yes, please outline level of experience

The researchers in the team have conducted many research projects in UK primary care and using linked electronic health records through CALIBER (dataset linking CPRD, HES, MINAP and ONS data).

23. References relating to your study

Please list up to 3 references (most relevant) relating to your proposed study.

1. Mehran R, Pocock SJ, Nikolsky E, Clayton T, Dangas GD, Kirtane AJ, et al. A Risk Score to Predict Bleeding in Patients With Acute Coronary Syndromes. Journal of the American College of Cardiology. 2010;55(23):2556-66.
2. Ducrocq G, Wallace JS, Baron G, Ravaud P, Alberts MJ, Wilson PWF, et al. Risk score to predict serious bleeding in stable outpatients with or at risk of atherothrombosis. European heart journal. 2010;31(10):1257-65.
3. Rapsomaniki E, Shah A, Perel P, Denaxas S, George J, Nicholas O, et al. Prognostic models for stable coronary artery disease based on electronic health record cohort of 102 023 patients. European heart journal. 2013.

PROTOCOL CONTENT CHECKLIST

In order to help ensure that protocols submitted for review contain adequate information for protocol evaluation, ISAC have produced instructions on the content of protocols for research using CPRD data. These instructions are available on the CPRD website (www.cprd.com/ISAC). All protocols using CPRD data which are submitted for review by ISAC must contain information on the areas detailed in the instructions. IF you do not feel that a specific area required by ISAC is relevant for your protocol, you will need to justify this decision to ISAC.

Applicants must complete the checklist below to confirm that the protocol being submitted includes all the areas required by ISAC, or to provide justification where a required area is not considered to be relevant for a specific protocol. Protocols will not be circulated to ISAC for review until the checklist has been completed by the applicant.

Please note, your protocol will be returned to you if you do not complete this checklist, or if you answer 'no' and fail to include justification for the omission of any required area.

	Included in protocol?		
Required area	Yes	No	If no, reason for omission
Lay Summary (max.200 words)	☒	☐	
Background	☒	☐	
Objective, specific aims and rationale	☒	☐	
Study Type *Descriptive* *Hypothesis Generating* *Hypothesis Testing*	☒	 ☐ ☐ ☐	
Study Design	☒	☐	
Sample size/power calculation *(Please provide justification of sample size in the protocol)*	☒	☐	
Study population *(including estimate of expected number of relevant patients in the CPRD)*	☒	☐	
Selection of comparison group(s) or controls	☐	☒	N/A
Exposures, outcomes and covariates *Exposures are clearly described* *Outcomes are clearly described*	☒ ☒ ☒	☐ ☐ ☐	
Use of linked data *(if applicable)*	☒	☐	
Data/ Statistical Analysis Plan *There is plan for addressing confounding* *There is a plan for addressing missing data*	☒ ☒ ☒	☐ ☐ ☐	
Patient/ user group involvement †	☒	☐	
Limitations of the study design, data sources and analytic methods	☒	☐	
Plans for disseminating and communicating study results	☒	☐	

† ***It is expected that many studies will benefit from the involvement of patient or user groups in their planning and refinement, and/or in the interpretation of the results and plans for further work. This is particularly, but not exclusively true of studies with interests in the impact on quality of life. Please indicate whether or not you intend to engage patients in any of the ways mentioned above.***

Voluntary registration of ISAC approved studies:
Epidemiological studies are increasingly being included in registries of research around the world, including those primarily set up for clinical trials. To increase awareness amongst researchers of ongoing research, ISAC encourages voluntary registration of epidemiological research conducted using MHRA databases. This will not replace information on ISAC approved protocols that may be published in its summary minutes or annual report. It is for the applicant to determine the most appropriate registry for their study. Please inform the ISAC secretariat that you have registered a protocol and provide the location.

Prognostic models for risk of serious atherothrombotic and bleeding events and death in patients who survived after myocardial infarction

Lay Summary
Patients who survive a heart attack (acute myocardial infarction) have an increased risk of death in the following years. This risk is largely due to another heart attack or other causes related to their heart and circulatory disease. National guidelines recommend that specific medicines are prescribed to patients following a heart attack to reduce the risk of further attacks, other circulatory disease and death. One of these recommendations is that all patients receive medicines that decrease the formation of blood clots (antiplatelet) either aspirin and/or another antiplatelet. However, one of the side effects of these drugs is bleeding.

This proposed observational study will use linked data covering hospital admissions and primary care data to develop algorithms (prognostic models) that can be used to predict the risk of cardiovascular mortality and morbidity and bleeding among patients who survived a heart attack. The purpose of the study proposed is to inform clinicians on the safety to continue therapy when monitoring this type of patients.

Background

Post-myocardial infarction (MI) patients are at high risk of further cardiovascular disease (CVD) events and death. Antiplatelet agents (e.g. aspirin, clopidogrel) are amongst the front-line treatments recommended for the management of atherothrombotic risk in after MI. However, their use is associated with increased risk of bleeding. Thus clinicians with their patients have to continuously weigh the benefits of antiplatelet therapies against the risk of bleeding, particularly as their improved survival post-MI as a result of medical advances exposes them to long periods of antiplatelet treatment.

Existing prognostic models for bleeding in post-MI survivors are restricted to short-term / in-hospital bleeds (e.g. ACUITY/HORIZONS-AMI (1)), have been developed for different populations (e.g. HAS-BLED (2) for patients with atrial fibrillation), or lacked data for important prognostic factors (e.g. REACH(3) lacked data on antiplatelet use and International Normalised Ratio [INR] levels). Previous models for MI/coronary death and all-cause mortality have been developed for stable angina patients and patients who survived after acute coronary syndrome ([ACS], MI or unstable angina) in CALIBER (4), but it is not known whether better performance might be achieved if models were restricted to post-MI survivors .

Using contemporary, linked electronic health record data (study period 2000-2014) we aim to develop algorithms to predict atherothrombotic and bleeding risks in post-MI survivors over a much longer period than previously possible. When making treatment decisions clinicians and patients need to be able to weight up the risk of bleeding against the risk of further CV events. We will thus enable evaluation of these risks separately and by combining them under a competing risks framework.

The new models for atherothrombotic and bleeding outcomes will be extended to dynamic risk models. The majority of prognostic models developed to date are limited to estimating baseline risks. However, dynamic models will allow doctors to monitor the changing risks of bleeding with time to make up-to-date decisions with respect to the safety of continuing therapy or receiving coronary artery bypass graft (CABG).

We will evaluate the performance of the new models in the extended coronary artery disease (CAD) survivor population that covers stable angina and patients who survived ACS, different time periods (depending on availability of data at the time of submission), and in patients from electronic health records of one or more other countries.

Overall aim
To develop and validate prognostic risk models of atherothrombotic events and bleeding for post-MI survivors.

Specific objectives

1) Simple prognostic model development in the post-MI survivor population:
a) To develop a simple prognostic model for atherothrombotic events using clinical covariates that are readily available in UK general practice.
b) To develop a simple prognostic model for bleeding events using clinical covariates that are readily available in UK general practice.

c) To demonstrate how these risks can be used to aid treatment and procedure decisions, that take into account both the potential benefits and harms of interventions.
d) To assess the clinical utility of the models using hypothetical cost effectiveness scenarios.

2) Validation of simple prognostic models developed

a) To validate the simple prognostic models internally
b) To validate the simple prognostic models internally on a wider stable cohort following various cardiovascular events and/or procedures.

3) Dynamic prognostic model development in the post-MI survivor population:

a) To extend the simple prognostic models that use baseline information into dynamic prognostic models and utilise routinely collected, continuous biomarker trajectories to estimate changing risks over time.
b) To compare dynamic prognostic models with simple prognostic models – in particular, the differences in utility.

4) External validation of prognostic models developed

a) To validate the prognostic models using international electronic health record cohorts compromising of patients with stable angina and/or stable post-MI (i.e. extended stable CAD population)
b) To validate the prognostic models in trial data (Thrombolysis in Myocardial Infarction [TIMI] trial)

Study Type

Prognostic model development and validation.

Study Design

A retrospective observational cohort study using primary care data from Clinical Practice Research Datalink (CPRD) linked to secondary care data from Hospital Episode Statistics (HES) and Office for National Statistics (ONS) mortality.

The overall cohort will be patients registered at CPRD practices to permit data linkage who were discharged alive from hospital following admission for MI between January 2000 and July 2014. The post-MI patients will be those diagnosed with an MI in HES between January 2000 and July 2014, registered in a linked CPRD general practice at least 12 months previously and with no subsequent myocardial infarction within the first 12 months. This cohort will be used to assess outcomes and costs.

Sample Size

Between 2000 and 2010 there were 35,858 patients with MI in CALIBER (linked electronic heath records of CPRD, HES, MINAP and ONS) (5), of which approximately 16,000 had continuous registration for >12 prior to and after their index MI, and survived with no recurrent MI for the 12 month period following their index MI. These 16,000 post-MI 1-year survivors experienced approximately 2,400 CVD events and 1,300 non-CVD deaths over a mean follow-up of 3 (SD=2) years. The sample size and number of events will be considerably higher when data up to July 2014 is used.

Study Population

The model development study population consists of patients aged ≥18 years who had an MI and did not have a recurrent MI or were deregistered from the GP practice in the 12 months following their index MI. Patients must also have had continuous registration in a participating primary practice for the 12 months prior to their index MI.

In addition to internal validation on the model development population, the model will also be validated on wider cohort including those with stable angina, stable following a stroke or transient ischaemic attack, peripheral arterial disease, abdominal aortic aneurysm, chronic heart disease and heart failure.

Outcomes

1) Atherothrombotic events within 1 year after index MI

- Composite of CVD death, non-fatal MI or non-fatal ischemic/unspecified stroke (ICD 10 codes are listed in Appendix 1)
- Cardiovascular death

These events have established phenotypes in the CALIBER dataset.

2) All-cause mortality within 1 year after index MI

3) Long-term severe/moderate bleeding events

Long term (>1 year after index myocardial infarction) bleeding that required transfusion or hospitalisation (ICD 10 codes are listed in Appendix 2).

Risk factors/ Covariates

We will evaluate which covariates are good predictors of cardiovascular outcomes and mortality. Covariates will be defined according to methods agreed among the CALIBER research group and documented in the CALIBER data manual.

Demographics	Age, sex, ethnicity, smoking status, year of entry into study population
Clinical diagnoses/ procedures	Previous major bleed, previous coronary intervention, diabetes, chronic obstructive pulmonary disease (COPD), history of heart failure, renal disease, cancer, peripheral arterial disease, peptic ulcer, atrial fibrillation, dementia
Pharmacological interventions	Aspirin use, antiplatelet use, oral anticoagulant use, duration of antithrombotic therapy (ATT) use, antihypertensive drug use, statin therapy, ACE inhibitors, beta-blockers
Biomarkers	Systolic blood pressure, diastolic blood pressure, body mass index (BMI), heart rate, haemoglobin, white blood cell count, total and high-density lipoprotein (HDL) cholesterol, creatinine, estimated glomerular filtration rate (eGFR), INR

Use of linked data

CPRD Gold: all tables from all practices.
Linked with HES inpatient admissions including primary discharge diagnosis per hospitalization and procedure information.
Linked with ONS mortality including underlying and ranked causes of death.
Linked with ONS socioeconomic status (IMD score)

Data analysis

A. Primary analyses

PROGRESS guidance (6) for prognostic model development and validation will be followed in this study.

- **Prognostic model development**

Preliminary univariate analyses using Cox models will be carried out to determine the effect of each variable on atherothrombotic and bleeding risk and to inform initial inclusion of variables in the multivariate model.

For the multivariate regression model stepwise selection methods will be employed to ensure parsimony, with age, sex and year of entry being kept in the model regardless of their significance.

For model adjustment, we will examine covariate missing data patterns. We will impute baseline covariates with non-informative missing patterns based on multiple imputation. As a sensitivity analysis, to take full advantage of the longitudinal nature of the data we will use multi-level imputation, as implemented in the R package 'mice' for continuous covariates, based on a multivariate normal assumption.

Clinical utility will be assessed by estimating the event-free years per number of life years saved in general practices that use the prognostic models.

- **Internal validation of prognostic models**

The performance of the models developed will be examined internally (with cross-validation to correct for 'optimism') in terms of their calibration and discrimination (Harrell's C-index).

To determine the generalizability of our developed models, we will also internally validate them using a wider cohort including those with stable angina, stable following a stroke or transient ischaemic attack, peripheral arterial disease, abdominal aortic aneurysm, chronic heart disease and heart failure.

- **Development of dynamic prognostic models**

Joint modeling of longitudinal and survival data allows us make predictions using a longitudinal biomarker (modeled using linear mixed effects models) on a survival outcome – in this case, time until bleeding and atherothrombotic events.
Graphical methods will be used to determine which longitudinal risk factors may indicate changing risk over time. Due to computationally intensive methods we will use high performance computing resources available at the Farr Institute where necessary.

- **External validation of prognostic models**

The developed models will be externally validated using data from international (Sweden, TIMI trial) electronic health records in order to demonstrate the generalizability of the models. The external validation population will compromise of the extended stable CAD population – that is, patients with stable angina and/or stable post-MI.

B. Secondary analyses

To account for changing guidelines between 2000 and 2014 analyses restricted to the 2005-2014 period will be performed. This will entail developing and validating the models described above on the subcohort identified for that period.

Limitations

- The nature of the electronic health record data has limitations, in particular, not all MIs will have been coded and therefore some patients eligible for inclusion may be missed out and not all the event will have been coded.
- Not all CPRD practices have given permission for linkage outside primary care (i.e. to HES) and this will reduce potential sample size.
- CPRD only include records of prescription issued and not whether the patient collected the medicine from pharmacy.
- Costs are not captured in CPRD so unit prices will be multiplied with resource use and based on national listings and published literature. Unit prices from one calendar year will be used.

Plans for disseminating and communicating study results

We will publish the results of these studies in peer reviewed medical journals. Additional strategies will also be applied for communicating study results to the general public, including:

- Study results will be included in the Farr 'Citizen Science' initiative - in which general public are invited to take part in analysing and visualising health record data.
- Study results will contribute to current projects of the Farr Institute in the development of a European Roadmap on eHealth Innovations enabling patient and citizen in health.
- Study results will be communicated to our partners from charities for patient wellbeing, such as the National Voices and British Heart Foundation, to be used for information, services and training provided for heart disease patients and their family.
- Media and web-based social media campaign strategies will be applied for disseminating research results to the general public.

Results of the proposed studies will be presented through national and international conferences in informatics (including those hosted by Farr Institute), and in national and international clinical meetings (for example, European Society of Cardiology, America Heart Association).

11.2 Supplementary Appendix (Chapter 5)

11.2.1 Bleeding ICD-10 codes

code	term	Anatomical site
D683	Haemorrhagic disorder due to circulating anticoagulants	blood
D69	Purpura and other haemorrhagic conditions	blood
D698	Other specified haemorrhagic conditions	blood
D699	Haemorrhagic condition, unspecified	blood
I230	Haemopericardium as current complication following acute myocardial infarction	circulatory
I312	Haemopericardium, not elsewhere classified	circulatory
I711	Thoracic aortic aneurysm, ruptured	circulatory
I713	Abdominal aortic aneurysm, ruptured	circulatory
I715	Thoracoabdominal aortic aneurysm, ruptured	circulatory
I718	Aortic aneurysm of unspecified site, ruptured	circulatory
H922	Otorrhagia	ear
N02	Recurrent and persistent haematuria	genitourinary
N026	Recurrent and persistent haematuria, dense deposit disease	genitourinary
N028	Recurrent and persistent haematuria, other	genitourinary
N029	Recurrent and persistent haematuria, unspecified	genitourinary
N421	Congestion and haemorrhage of prostate	genitourinary
N501	Vascular disorders of male genital organs	genitourinary
N836	Haematosalpinx	genitourinary
N837	Haematoma of broad ligament	genitourinary
N857	Haematometra	genitourinary
N897	Haematocolpos	genitourinary
N898	Other specified noninflammatory disorders of vagina	genitourinary
N908	Other specified noninflammatory disorders of vulva and perineum	genitourinary
N921	Excessive and frequent menstruation with irregular cycle	genitourinary
N925	Other specified irregular menstruation	genitourinary
N926	Irregular menstruation, unspecified	genitourinary
N93	Other abnormal uterine and vaginal bleeding	genitourinary
N938	Other specified abnormal uterine and vaginal bleeding	genitourinary
N939	Abnormal uterine and vaginal bleeding, unspecified	genitourinary
N950	Postmenopausal bleeding	genitourinary
R31	Unspecified haematuria	genitourinary
R31X	Unspecified haematuria	genitourinary
K625	Haemorrhage of anus and rectum	GI (lower)
K921	Melaena	GI (lower)
K922	Gastrointestinal haemorrhage, unspecified	GI (unspecified)
I850	Oesophageal varices with bleeding	GI (upper)
I983	Oesophageal varices with bleeding in diseases classified elsewhere	GI (upper)
K250	Gastric ulcer ; Acute with haemorrhage	GI (upper)
K252	Gastric ulcer ; Acute with both haemorrhage and perforation	GI (upper)
K254	Gastric ulcer ; Chronic or unspecified with haemorrhage	GI (upper)
K256	Gastric ulcer ; Chronic or unspecified with both haemorrhage and perforation	GI (upper)
K260	Duodenal ulcer ; Acute with haemorrhage	GI (upper)
K262	Duodenal ulcer ; Acute with both haemorrhage and perforation	GI (upper)

code	term	Anatomical site
K264	Duodenal ulcer ; Chronic or unspecified with haemorrhage	GI (upper)
K266	Duodenal ulcer ; Chronic or unspecified with both haemorrhage and perforation	GI (upper)
K270	Peptic ulcer, site unspecified ; Acute with haemorrhage	GI (upper)
K272	Peptic ulcer, site unspecified ; Acute with both haemorrhage and perforation	GI (upper)
K274	Peptic ulcer, chronic or unspecified with haemorrhage	GI (upper)
K276	Chronic or unspecified with both haemorrhage and perforation	GI (upper)
K280	Gastrojejunal ulcer ; Acute with haemorrhage	GI (upper)
K282	Gastrojejunal ulcer ; Acute with both haemorrhage and perforation	GI (upper)
K284	Gastrojejunal ulcer ; Chronic or unspecified with haemorrhage	GI (upper)
K286	Gastrojejunal ulcer ; Chronic or unspecified with both haemorrhage and perforation	GI (upper)
K290	Acute haemorrhagic gastritis	GI (upper)
K661	Haemoperitoneum	GI (upper)
K920	Haematemesis	GI (upper)
I61	Intracerebral haemorrhage	IC (intracerebral)
I610	Intracerebral haemorrhage in hemisphere, subcortical	IC (intracerebral)
I611	Intracerebral haemorrhage in hemisphere, cortical	IC (intracerebral)
I612	Intracerebral haemorrhage in hemisphere, unspecified	IC (intracerebral)
I613	Intracerebral haemorrhage in brain stem	IC (intracerebral)
I614	Intracerebral haemorrhage in cerebellum	IC (intracerebral)
I615	Intracerebral haemorrhage, intraventricular	IC (intracerebral)
I616	Intracerebral haemorrhage, multiple localized	IC (intracerebral)
I618	Other intracerebral haemorrhage	IC (intracerebral)
I619	Intracerebral haemorrhage, unspecified	IC (intracerebral)
I60	Subarachnoid haemorrhage	IC (subarachnoid)
I600	Subarachnoid haemorrhage from carotid siphon and bifurcation	IC (subarachnoid)
I601	Subarachnoid haemorrhage from middle cerebral artery	IC (subarachnoid)
I602	Subarachnoid haemorrhage from anterior communicating artery	IC (subarachnoid)
I603	Subarachnoid haemorrhage from posterior communicating artery	IC (subarachnoid)
I604	Subarachnoid haemorrhage from basilar artery	IC (subarachnoid)
I605	Subarachnoid haemorrhage from vertebral artery	IC (subarachnoid)
I606	Subarachnoid haemorrhage from other intracranial arteries	IC (subarachnoid)
I607	Subarachnoid haemorrhage from intracranial artery, unspecified	IC (subarachnoid)
I608	Other subarachnoid haemorrhage	IC (subarachnoid)
I609	Subarachnoid haemorrhage, unspecified	IC (subarachnoid)
I62	Other nontraumatic intracranial haemorrhage	IC (subarachnoid)
I620	Subdural haemorrhage (acute)(nontraumatic)	IC (subdural)
I621	Nontraumatic extradural haemorrhage	IC (extradural)
I629	Intracranial haemorrhage (nontraumatic), unspecified	IC (unspecified)
S064	Epidural haemorrhage	IC (extradural)
R233	Spontaneous ecchymoses	NOS
R58	Haemorrhage, not elsewhere classified	NOS
R58X	Haemorrhage, not elsewhere classified	NOS
T81	Haemorrhage and haematoma complicating a procedure NEC	NOS
H313	Choroidal haemorrhage and rupture	ocular
H356	Retinal haemorrhage	ocular

code	term	Anatomical site
H431	Vitreous haemorrhage	ocular
H450	Vitreous haemorrhage in diseases classified elsewhere	ocular
J942	Haemothorax	respiratory (lower)
R040	Epistaxis	respiratory (upper)
R041	Haemorrhage from throat	respiratory (upper)
R042	Haemoptysis	respiratory (lower)
R048	Haemorrhage from other sites in respiratory passages	respiratory (unspecified)
R049	Haemorrhage from respiratory passages, unspecified	respiratory (unspecified)

GI= gastrointestinal; IC=intracranial; NOS= not otherwise specified

11.2.2 Bleeding Read codes

code	term	Anatomical site
D31..00	Purpura and other haemorrhagic conditions	blood
D31X.00	Haemorrhagic condition, unspecified	blood
D31y.00	Other specified haemorrhagic conditions	blood
D31yz00	Other specified haemorrhagic condition NOS	blood
D31z.00	Haemorrhagic condition NOS	blood
Dyu3300	[X]Other specified haemorrhagic conditions	blood
G360.00	Haemopericardium/current comp folow acut myocard infarct	circulatory
G530.00	Haemopericardium	circulatory
G711.00	Thoracic aortic aneurysm which has ruptured	circulatory
G711.11	Ruptured thoracic aortic aneurysm	circulatory
G713.00	Abdominal aortic aneurysm which has ruptured	circulatory
G713.11	Ruptured abdominal aortic aneurysm	circulatory
G713000	Ruptured suprarenal aortic aneurysm	circulatory
G715.00	Ruptured aortic aneurysm NOS	circulatory
G715000	Thoracoabdominal aortic aneurysm, ruptured	circulatory
G723500	Ruptured popliteal artery aneurysm	circulatory
7303000	Drainage of haematoma of external ear	ear
7303200	Drainage haematoma ext ear control cavity c bolster suture	ear
F501G00	Haemorrhagic otitis externa	ear
F503100	Haematoma of pinna	ear
C063000	Thyroid haemorrhage	endocrine
C12y100	Haemorrhage of parathyroid	endocrine
1584	Heavy episode of vaginal bleeding	genitourinary
1A45.00	Blood in urine - haematuria	genitourinary
K167.00	Haemorrhage into bladder wall	genitourinary
K16y200	Bladder haemorrhage	genitourinary
K197.00	Haematuria	genitourinary
K197000	Painless haematuria	genitourinary
K197100	Painful haematuria	genitourinary
K197300	Frank haematuria	genitourinary
K197400	Clot haematuria	genitourinary
K19y400	Bleeding from urethra	genitourinary
K19y411	Urethral bleeding	genitourinary
K221100	Prostatic haemorrhage	genitourinary
K275100	Corpus cavernosum haematoma	genitourinary
K275200	Corpus cavernosum haemorrhage	genitourinary
K286100	Scrotal haemorrhage	genitourinary
K286400	Testicular haemorrhage	genitourinary
K286v00	Male genital haematoma NOS	genitourinary
K286w00	Male genital haemorrhage NOS	genitourinary
K537.00	Haematoma of the broad ligament	genitourinary
K55y300	Haemorrhage of cervix	genitourinary
K566.00	Vaginal haematoma	genitourinary
K56y100	Haemorrhage of vagina	genitourinary
K575.00	Haematoma of vulva	genitourinary
K59yx00	Dysfunctional uterine haemorrhage NOS	genitourinary

code	term	Anatomical site
K59yy00	Functional uterine haemorrhage NOS	genitourinary
K5E..00	Other abnormal uterine and vaginal bleeding	genitourinary
K5E0.00	Abnormal uterine bleeding unrelated to menstrual cycle	genitourinary
K5E1.00	Abnormal uterine bleeding, unspecified	genitourinary
K5E2.00	Abnormal vaginal bleeding, unspecified	genitourinary
K5Ez.00	Abnormal uterine and vaginal bleeding, unspecified	genitourinary
Kyu9D00	[X]Other specified abnormal uterine and vaginal bleeding	genitourinary
196B.00	Painful rectal bleeding	GI (lower)
196C.00	Painless rectal bleeding	GI (lower)
19E4.12	C/O - melaena	GI (lower)
19E6.00	Blood in faeces	GI (lower)
19E6.11	Blood in faeces symptom	GI (lower)
4737.11	Melaena - O/E of faeces	GI (lower)
4762	Faeces: fresh blood present	GI (lower)
4762.11	Blood in faeces	GI (lower)
479..11	Faeces occult blood test	GI (lower)
J510900	Bleeding diverticulosis	GI (lower)
J573.00	Haemorrhage of rectum and anus	GI (lower)
J573000	Rectal haemorrhage	GI (lower)
J573011	Rectal bleeding	GI (lower)
J573012	PRB - Rectal bleeding	GI (lower)
J573100	Anal haemorrhage	GI (lower)
J573z00	Haemorrhage of rectum and anus NOS	GI (lower)
J681.00	Melaena	GI (lower)
J681.11	Blood in stool	GI (lower)
J681.12	Altered blood in stools	GI (lower)
J681.13	Blood in stools altered	GI (lower)
J68z100	Intestinal haemorrhage NOS	GI (lower)
S740100	Liver haematoma and contusion without open wound into cavity	GI (lower)
SE22300	Haematoma of rectus sheath	GI (lower)
SE23111	Perianal haematoma	GI (lower)
J68..00	Gastrointestinal haemorrhage	GI (NOS)
J68z.00	Gastrointestinal haemorrhage unspecified	GI (NOS)
J68z.11	GIB - Gastrointestinal bleeding	GI (NOS)
J68zz00	Gastrointestinal tract haemorrhage NOS	GI (NOS)
1994	Vomiting blood - fresh	GI (upper)
1994.11	Blood in vomit - symptom	GI (upper)
1995	Vomiting blood - coffee ground	GI (upper)
4A23.00	Vomit: frank blood present	GI (upper)
4A23.11	Blood in vomit O/E	GI (upper)
4A5..00	Vomit occult blood	GI (upper)
4A5..11	Occult blood in vomit	GI (upper)
4A51.00	Vomit occult blood positive	GI (upper)
4A5Z.00	Vomit occult blood NOS	GI (upper)
7619100	Gastrotomy and ligation of bleeding point of stomach	GI (upper)
G850.00	Oesophageal varices with bleeding	GI (upper)
G852000	Oesophageal varices with bleeding in diseases EC	GI (upper)

code	term	Anatomical site
J10y000	Haemorrhage of oesophagus	GI (upper)
J110100	Acute gastric ulcer with haemorrhage	GI (upper)
J110111	Bleeding acute gastric ulcer	GI (upper)
J110300	Acute gastric ulcer with haemorrhage and perforation	GI (upper)
J111100	Chronic gastric ulcer with haemorrhage	GI (upper)
J111111	Bleeding chronic gastric ulcer	GI (upper)
J111300	Chronic gastric ulcer with haemorrhage and perforation	GI (upper)
J11y100	Unspecified gastric ulcer with haemorrhage	GI (upper)
J11yy00	Unspec gastric ulcer; unspec haemorrhage and/or perforation	GI (upper)
J120100	Acute duodenal ulcer with haemorrhage	GI (upper)
J120300	Acute duodenal ulcer with haemorrhage and perforation	GI (upper)
J121100	Chronic duodenal ulcer with haemorrhage	GI (upper)
J121111	Bleeding chronic duodenal ulcer	GI (upper)
J121300	Chronic duodenal ulcer with haemorrhage and perforation	GI (upper)
J12y100	Unspecified duodenal ulcer with haemorrhage	GI (upper)
J12y300	Unspecified duodenal ulcer with haemorrhage and perforation	GI (upper)
J12yy00	Unspec duodenal ulcer; unspec haemorrhage and/or perforation	GI (upper)
J130100	Acute peptic ulcer with haemorrhage	GI (upper)
J130300	Acute peptic ulcer with haemorrhage and perforation	GI (upper)
J131100	Chronic peptic ulcer with haemorrhage	GI (upper)
J13y100	Unspecified peptic ulcer with haemorrhage	GI (upper)
J13y300	Unspecified peptic ulcer with haemorrhage and perforation	GI (upper)
J140100	Acute gastrojejunal ulcer with haemorrhage	GI (upper)
J14y100	Unspecified gastrojejunal ulcer with haemorrhage	GI (upper)
J150000	Acute haemorrhagic gastritis	GI (upper)
J680.00	Haematemesis	GI (upper)
J680.11	Vomiting of blood	GI (upper)
J68z000	Gastric haemorrhage NOS	GI (upper)
J68z200	Upper gastrointestinal haemorrhage	GI (upper)
G620.00	Extradural haemorrhage - nontraumatic	IC (extradural)
S626.00	Epidural haemorrhage	IC (extradural)
G600.00	Ruptured berry aneurysm	IC (intracerebral)
G61..00	Intracerebral haemorrhage	IC (intracerebral)
G610.00	Cortical haemorrhage	IC (intracerebral)
G611.00	Internal capsule haemorrhage	IC (intracerebral)
G612.00	Basal nucleus haemorrhage	IC (intracerebral)
G613.00	Cerebellar haemorrhage	IC (intracerebral)
G614.00	Pontine haemorrhage	IC (intracerebral)
G615.00	Bulbar haemorrhage	IC (intracerebral)
G616.00	External capsule haemorrhage	IC (intracerebral)
G617.00	Intracerebral haemorrhage, intraventricular	IC (intracerebral)
G618.00	Intracerebral haemorrhage, multiple localized	IC (intracerebral)
G61X.00	Intracerebral haemorrhage in hemisphere, unspecified	IC (intracerebral)
G61X000	Left sided intracerebral haemorrhage, unspecified	IC (intracerebral)
G61X100	Right sided intracerebral haemorrhage, unspecified	IC (intracerebral)
G61z.00	Intracerebral haemorrhage NOS	IC (intracerebral)
Gyu6200	[X]Other intracerebral haemorrhage	IC (intracerebral)

code	term	Anatomical site
Gyu6F00	[X]Intracerebral haemorrhage in hemisphere, unspecified	IC (intracerebral)
G60..00	Subarachnoid haemorrhage	IC (subarachnoid)
G601.00	Subarachnoid haemorrhage from carotid siphon and bifurcation	IC (subarachnoid)
G602.00	Subarachnoid haemorrhage from middle cerebral artery	IC (subarachnoid)
G603.00	Subarachnoid haemorrhage from anterior communicating artery	IC (subarachnoid)
G604.00	Subarachnoid haemorrhage from posterior communicating artery	IC (subarachnoid)
G605.00	Subarachnoid haemorrhage from basilar artery	IC (subarachnoid)
G606.00	Subarachnoid haemorrhage from vertebral artery	IC (subarachnoid)
G60z.00	Subarachnoid haemorrhage NOS	IC (subarachnoid)
Gyu6100	[X]Other subarachnoid haemorrhage	IC (subarachnoid)
G621.00	Subdural haemorrhage - nontraumatic	IC (subdural)
G623.00	Subdural haemorrhage NOS	IC (subdural)
G62..00	Other and unspecified intracranial haemorrhage	IC (unspecified)
G62z.00	Intracranial haemorrhage NOS	IC (unspecified)
2F65.00	O/E - haemorrhagic bullae	NOS
Ryu7300	[X]Haemorrhage, not elsewhere classified	NOS
S750100	Spleen haematoma without mention of open wound into cavity	NOS
S751100	Spleen haematoma with open wound into cavity	NOS
SE...11	Haematoma with intact skin	NOS
SE33011	Subungal haematoma	NOS
SE45.11	Haematoma of leg	NOS
SE4z.11	Haematoma NOS	NOS
SE4z.12	Intramuscular haematoma NOS	NOS
SK02.00	Secondary and recurrent haemorrhage	NOS
SK02.11	Secondary and recurrent haemorrhage	NOS
SK02.12	Secondary and recurrent haemorrhage	NOS
2BB5.00	O/E - retinal haemorrhages	ocular
2BB8.00	O/E - vitreous haemorrhages	ocular
F404500	Intra-ocular haemorrhage	ocular
F424300	Retinal pigment epithelium haemorrhagic detachment	ocular
F42y.11	Haemorrhage - retinal	ocular
F42y100	Superficial retinal haemorrhage	ocular
F42y300	Deep retinal haemorrhage	ocular
F42y400	Subretinal haemorrhage	ocular
F42y500	Retinal haemorrhage NOS	ocular
F436.00	Choroidal haemorrhage and rupture	ocular
F436000	Unspecified choroidal haemorrhage	ocular
F436100	Expulsive choroidal haemorrhage	ocular
F436z00	Choroidal haemorrhage or rupture NOS	ocular
F437200	Haemorrhagic choroidal detachment	ocular
F4Ey000	Haemorrhage of eyelid	ocular
F4H4100	Optic nerve sheath haemorrhage	ocular
F4K2800	Vitreous haemorrhage	ocular
F4K7.00	Retrobulbar haemorrhage	ocular
FyuH400	[X]Vitreous haemorrhage in diseases classified elsewhere	ocular
7004100	Evacuation of haematoma from temporal lobe of brain	procedure
7004200	Evacuation of haematoma from cerebellum	procedure

code	term	Anatomical site
7004300	Evacuation of intracerebral haematoma NEC	procedure
7008200	Aspiration of haematoma of brain tissue	procedure
7017000	Evacuation of subdural haematoma	procedure
7032000	Evacuation of extradural haematoma	procedure
7736000	Evacuation of perianal haematoma	procedure
7D05200	Evacuation of haematoma of vulva	procedure
7D1C000	Evacuation of haematoma from vagina	procedure
7G2H400	Liposuction removal of haematoma	procedure
7M0G000	Aspiration of haematoma of organ NOC	procedure
7M0G400	Evacuation of haematoma NEC	procedure
C154200	Adrenal haemorrhage	renal
K138100	Renal artery haemorrhage	renal
K138300	Intrarenal haematoma	renal
K13y800	Perirenal haematoma	renal
S760100	Kidney haematoma without mention of open wound into cavity	renal
S760111	Renal haematoma without mention of open wound into cavity	renal
S761100	Kidney haematoma with open wound into cavity	renal
172..00	Blood in sputum - haemoptysis	respiratory (lower)
R063.00	[D]Haemoptysis	respiratory (lower)
R063000	[D]Cough with haemorrhage	respiratory (lower)
R063100	[D]Pulmonary haemorrhage NOS	respiratory (lower)
R063z00	[D]Haemoptysis NOS	respiratory (lower)
1C62.00	Has nose bleeds - epistaxis	respiratory (upper)
2D25.00	O/E - epistaxis	respiratory (upper)
2DE7.00	O/E - throat haemorrhage	respiratory (upper)
R047.00	[D]Epistaxis	respiratory (upper)
R047.11	[D]Nosebleed	respiratory (upper)
R048.00	[D]Throat haemorrhage	respiratory (upper)

GI= gastrointestinal; IC=intracranial; NOS= not otherwise specified

11.2.3 Bleeding complications coded in MINAP

Category	Definition	Notes
0	None	
1	Intracranial bleed	Of any severity. Should be confirmed by scanning
2	Retroperitoneal haemorrhage	Of any severity. Should be confirmed by scanning
3	Any bleed with haemoglobin (Hb) fall >5g	From any site except options 1 and 2.
4	Any bleed with Hb fall >3g and <5g	
5	Any bleed with Hb fall <3g	
9	Unknown	

11.2.4 Transfusion OPCS codes

code	term
X33	Other blood transfusion
X331	Intra-arterial blood transfusion
X332	Intravenous blood transfusion of packed cells
X333	Intravenous blood transfusion of platelets
X338	Other specified other blood transfusion
X339	Unspecified other blood transfusion

11.2.5 Transfusion Read codes

code	term
7L14.00	Other blood transfusion
7L14000	Intraarterial blood transfusion
7L14100	Intravenous blood transfusion of packed cells
7L14300	Intravenous blood transfusion NEC
7L14311	Blood transfusion
7L14y00	Other specified other blood transfusion
7L14z00	Other blood transfusion NOS
TAy0.00	Mismatched blood transfused
TB1y000	Blood transfusion with complication, without blame
ZV58200	[V]Blood transfusion, without reported diagnosis

11.2.6 Haematoma removal OPCS codes

code	term
A052	Evacuation of haematoma from temporal lobe of brain
A053	Evacuation of haematoma from cerebellum
A054	Evacuation of intracerebral haematoma NEC
A103	Aspiration of haematoma of tissue of brain
A401	Evacuation of extradural haematoma
A411	Evacuation of subdural haematoma
D041	Drainage of haematoma of external ear
H531	Evacuation of perianal haematoma
P093	Evacuation of haematoma from vulva
P271	Evacuation of haematoma from vagina
Y221	Aspiration of haematoma of organ NOC

11.2.7 Bleeding cessation procedure OPCS codes

code	term
E05	Surgical arrest of bleeding from internal nose
E051	Cauterisation of internal nose
E052	Ligation of artery of internal nose
E053	Embolisation of artery of internal nose
E054	Laser therapy of internal nose
E058	Other specified surgical arrest of bleeding from internal nose
E059	Unspecified surgical arrest of bleeding from internal nose
E203	Surgical arrest of postoperative bleeding of adenoid
F162	Surgical arrest of postoperative bleeding from tooth socket
F365	Surgical arrest of postoperative bleeding from tonsillar bed
T032	Reopening of chest and re-exploration of intrathoracic operation site and surgical arrest of postoperative bleeding
T301	Reopening of abdomen and re-exploration of intra-abdominal operation site and surgical arrest of postoperative bleeding
V032	Reopening of cranium and re-exploration of intracranial operation site and surgical arrest of postoperative bleeding
Y321	Re-exploration of organ and surgical arrest of postoperative bleeding NOC

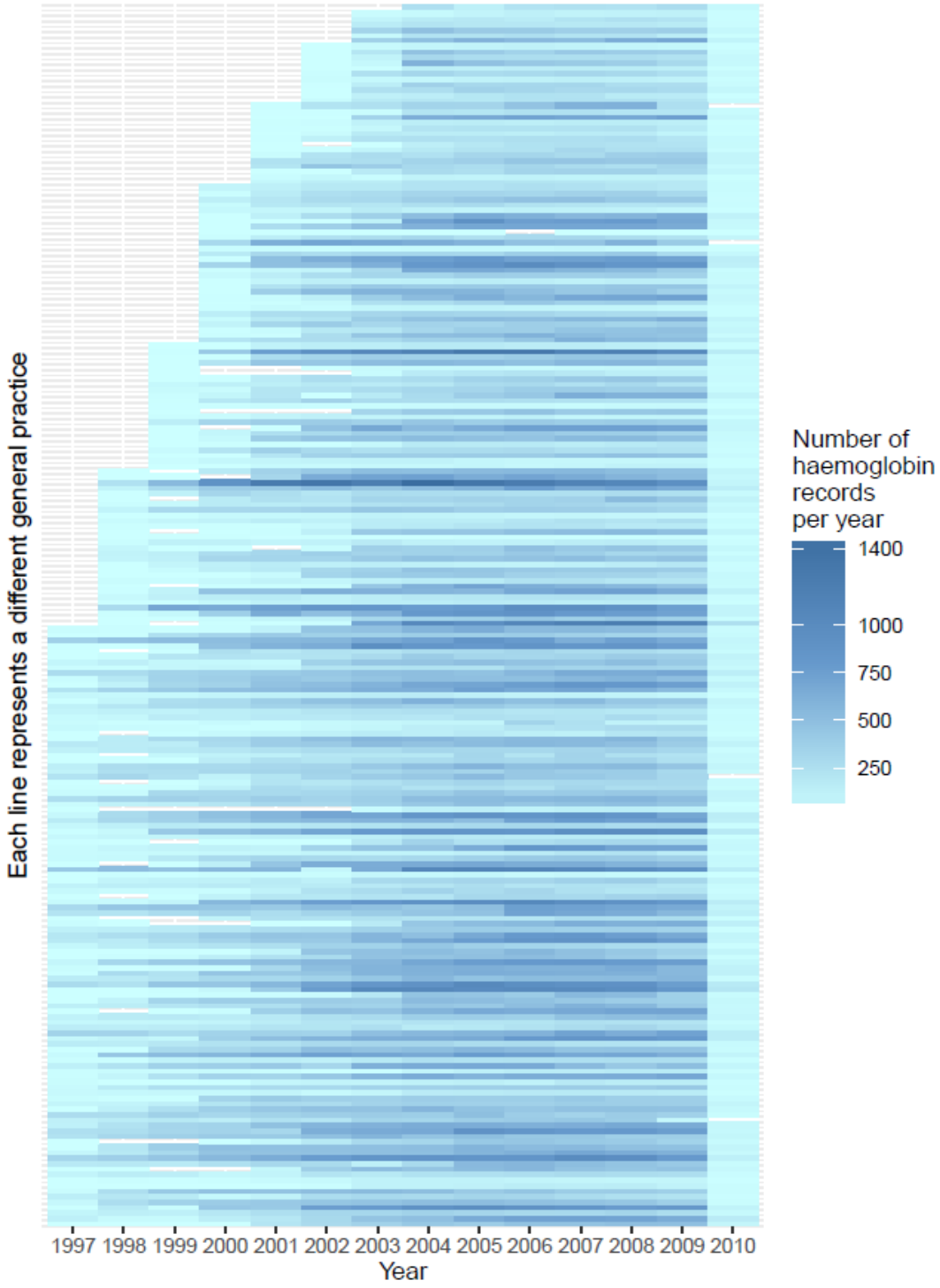
11.2.8 Endoscopy OPCS-4 codes

code	term
A18	Diagnostic endoscopic examination of ventricle of brain
A188	Other specified diagnostic endoscopic examination of ventricle of brain
A189	Unspecified diagnostic endoscopic examination of ventricle of brain
E25	Diagnostic endoscopic examination of pharynx
E253	Diagnostic endoscopic examination of nasopharynx NEC
E258	Other specified diagnostic endoscopic examination of pharynx
E259	Unspecified diagnostic endoscopic examination of pharynx
E36	Diagnostic endoscopic examination of larynx
E368	Other specified diagnostic endoscopic examination of larynx
E369	Unspecified diagnostic endoscopic examination of larynx
E37	Diagnostic microendoscopic examination of larynx
E378	Other specified diagnostic microendoscopic examination of larynx
E379	Unspecified diagnostic microendoscopic examination of larynx
E49	Diagnostic fibreoptic endoscopic examination of lower respiratory tract
E498	Other specified diagnostic fibreoptic endoscopic examination of lower respiratory tract
E499	Unspecified diagnostic fibreoptic endoscopic examination of lower respiratory tract
E51	Diagnostic endoscopic examination of lower respiratory tract using rigid bronchoscope
E518	Other specified diagnostic endoscopic examination of lower respiratory tract using rigid bronchoscope
E519	Unspecified diagnostic endoscopic examination of lower respiratory tract using rigid bronchoscope
E63	Diagnostic endoscopic examination of mediastinum
E638	Other specified diagnostic endoscopic examination of mediastinum
E639	Unspecified diagnostic endoscopic examination of mediastinum
G16	Diagnostic fibreoptic endoscopic examination of oesophagus
G162	Diagnostic fibreoptic endoscopic ultrasound examination of oesophagus
G168	Other specified diagnostic fibreoptic endoscopic examination of oesophagus
G169	Unspecified diagnostic fibreoptic endoscopic examination of oesophagus
G19	Diagnostic endoscopic examination of oesophagus using rigid oesophagoscope
G198	Other specified diagnostic endoscopic examination of oesophagus using rigid oesophagoscope
G199	Unspecified diagnostic endoscopic examination of oesophagus using rigid oesophagoscope
G45	Diagnostic fibreoptic endoscopic examination of upper gastrointestinal tract
G452	Fibreoptic endoscopic ultrasound examination of upper gastrointestinal tract
G458	Other specified diagnostic fibreoptic endoscopic examination of upper gastrointestinal tract
G459	Unspecified diagnostic fibreoptic endoscopic examination of upper gastrointestinal tract
G55	Diagnostic endoscopic examination of duodenum
G558	Other specified diagnostic endoscopic examination of duodenum
G559	Unspecified diagnostic endoscopic examination of duodenum
G65	Diagnostic endoscopic examination of jejunum
G658	Other specified diagnostic endoscopic examination of jejunum
G659	Unspecified diagnostic endoscopic examination of jejunum
G80	Diagnostic endoscopic examination of ileum
G803	Diagnostic endoscopic balloon examination of ileum

code	term
G808	Other specified diagnostic endoscopic examination of ileum
G809	Unspecified diagnostic endoscopic examination of ileum
H22	Diagnostic endoscopic examination of colon
H228	Other specified diagnostic endoscopic examination of colon
H229	Unspecified diagnostic endoscopic examination of colon
H25	Diagnostic endoscopic examination of lower bowel using fibreoptic sigmoidoscope
H258	Other specified diagnostic endoscopic examination of lower bowel using fibreoptic sigmoidoscope
H259	Unspecified diagnostic endoscopic examination of lower bowel using fibreoptic sigmoidoscope
H28	Diagnostic endoscopic examination of sigmoid colon using rigid sigmoidoscope
H288	Other specified diagnostic endoscopic examination of sigmoid colon using rigid sigmoidoscope
H289	Unspecified diagnostic endoscopic examination of sigmoid colon using rigid sigmoidoscope
H68	Diagnostic endoscopic examination of enteric pouch using colonoscope
H682	Diagnostic endoscopic examination of colonic pouch using colonoscope NEC
H684	Diagnostic endoscopic examination of ileoanal pouch using colonoscope NEC
H688	Other specified diagnostic endoscopic examination of enteric pouch using colonoscope
H689	Unspecified diagnostic endoscopic examination of enteric pouch using colonoscope
H69	Diagnostic endoscopic examination of enteric pouch using fibreoptic sigmoidoscope
H692	Diagnostic endoscopic examination of colonic pouch using fibreoptic sigmoidoscope NEC
H694	Diagnostic endoscopic examination of ileoanal pouch using fibreoptic sigmoidoscope NEC
H698	Other specified diagnostic endoscopic examination of enteric pouch using fibreoptic sigmoidoscope
H699	Unspecified diagnostic endoscopic examination of enteric pouch using fibreoptic sigmoidoscope
H70	Diagnostic endoscopic examination of enteric pouch using rigid sigmoidoscope
H702	Diagnostic endoscopic examination of colonic pouch using rigid sigmoidoscope NEC
H704	Diagnostic endoscopic examination of ileoanal pouch using rigid sigmoidoscope NEC
H708	Other specified diagnostic endoscopic examination of enteric pouch using rigid sigmoidoscope
H709	Unspecified diagnostic endoscopic examination of enteric pouch using rigid sigmoidoscope
J09	Diagnostic endoscopic examination of liver using laparoscope
J098	Other specified diagnostic endoscopic examination of liver using laparoscope
J099	Unspecified diagnostic endoscopic examination of liver using laparoscope
J17	Endoscopic ultrasound examination of liver
J178	Other specified endoscopic ultrasound examination of liver
J179	Unspecified endoscopic ultrasound examination of liver
J43	Diagnostic endoscopic retrograde examination of bile duct and pancreatic duct
J438	Other specified diagnostic endoscopic retrograde examination of bile duct and pancreatic duct
J439	Unspecified diagnostic endoscopic retrograde examination of bile duct and pancreatic duct
J44	Diagnostic endoscopic retrograde examination of bile duct
J448	Other specified diagnostic endoscopic retrograde examination of bile duct
J449	Unspecified diagnostic endoscopic retrograde examination of bile duct

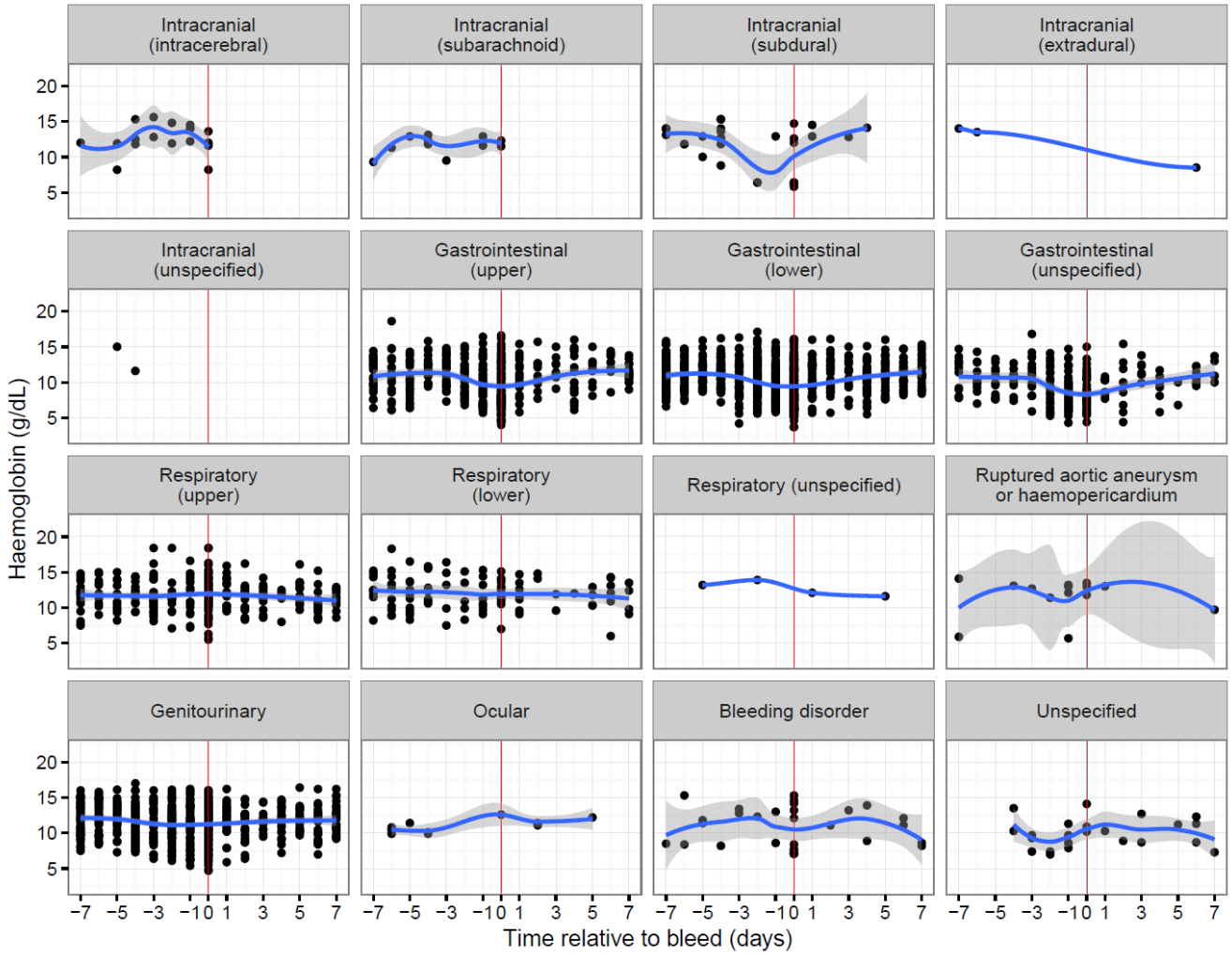
code	term
J45	Diagnostic endoscopic retrograde examination of pancreatic duct
J458	Other specified diagnostic endoscopic retrograde examination of pancreatic duct
J459	Unspecified diagnostic endoscopic retrograde examination of pancreatic duct
J53	Endoscopic ultrasound examination of bile duct
J538	Other specified endoscopic ultrasound examination of bile duct
J539	Unspecified endoscopic ultrasound examination of bile duct
J74	Endoscopic ultrasound examination of pancreas
J748	Other specified endoscopic ultrasound examination of pancreas
J749	Unspecified endoscopic ultrasound examination of pancreas
M11	Diagnostic endoscopic examination of kidney
M113	Diagnostic endoscopic retrograde examination of kidney NEC
M118	Other specified diagnostic endoscopic examination of kidney
M119	Unspecified diagnostic endoscopic examination of kidney
M30	Diagnostic endoscopic examination of ureter
M308	Other specified diagnostic endoscopic examination of ureter
M309	Unspecified diagnostic endoscopic examination of ureter
M45	Diagnostic endoscopic examination of bladder
M455	Diagnostic endoscopic examination of bladder using rigid cystoscope
M458	Other specified diagnostic endoscopic examination of bladder
M459	Unspecified diagnostic endoscopic examination of bladder
M77	Diagnostic endoscopic examination of urethra
M778	Other specified diagnostic endoscopic examination of urethra
M779	Unspecified diagnostic endoscopic examination of urethra
M85	Diagnostic endoscopic examination of urinary diversion
M851	Endoscopic examination of intestinal conduit
M858	Other specified diagnostic endoscopic examination of urinary diversion
M859	Unspecified diagnostic endoscopic examination of urinary diversion
Q18	Diagnostic endoscopic examination of uterus
Q188	Other specified diagnostic endoscopic examination of uterus
Q189	Unspecified diagnostic endoscopic examination of uterus
Q39	Diagnostic endoscopic examination of fallopian tube
Q398	Other specified diagnostic endoscopic examination of fallopian tube
Q399	Unspecified diagnostic endoscopic examination of fallopian tube
Q50	Diagnostic endoscopic examination of ovary
Q508	Other specified diagnostic endoscopic examination of ovary
Q509	Unspecified diagnostic endoscopic examination of ovary
T11	Diagnostic endoscopic examination of pleura
T118	Other specified diagnostic endoscopic examination of pleura
T119	Unspecified diagnostic endoscopic examination of pleura
T43	Diagnostic endoscopic examination of peritoneum
T433	Diagnostic endoscopic ultrasound examination of peritoneum
T438	Other specified diagnostic endoscopic examination of peritoneum
T439	Unspecified diagnostic endoscopic examination of peritoneum

11.2.9 The number of haemoglobin records per year in 224 CPRD general practices (1997-2010)



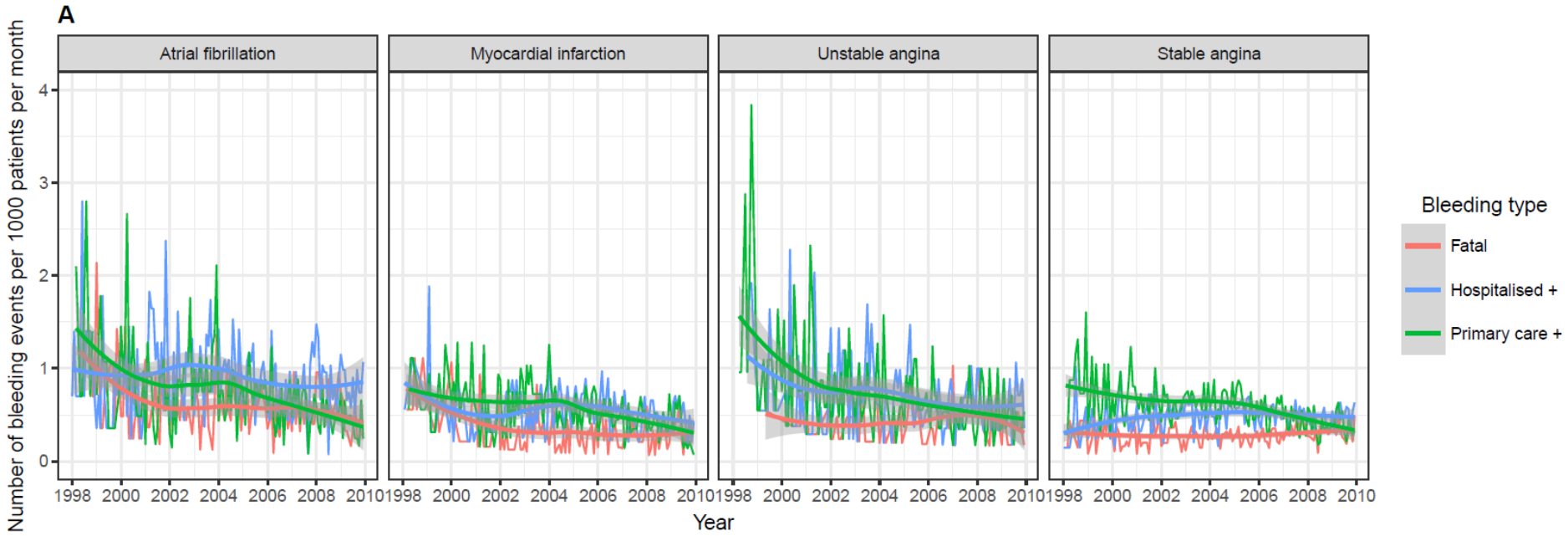
Note: 2010 data includes up to March only

11.2.10 Haemoglobin records within +/- 7 days of HES bleeding events with fitted LOESS curves by anatomical site

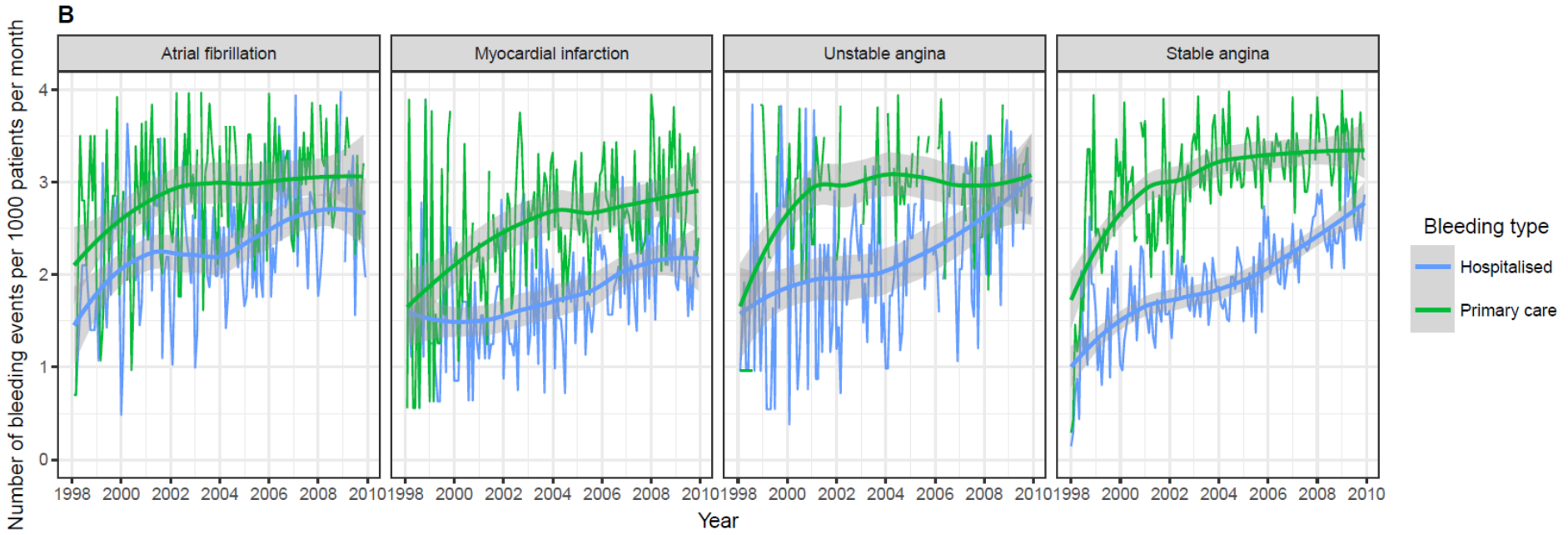


11.3 Supplementary appendix (chapter 6)

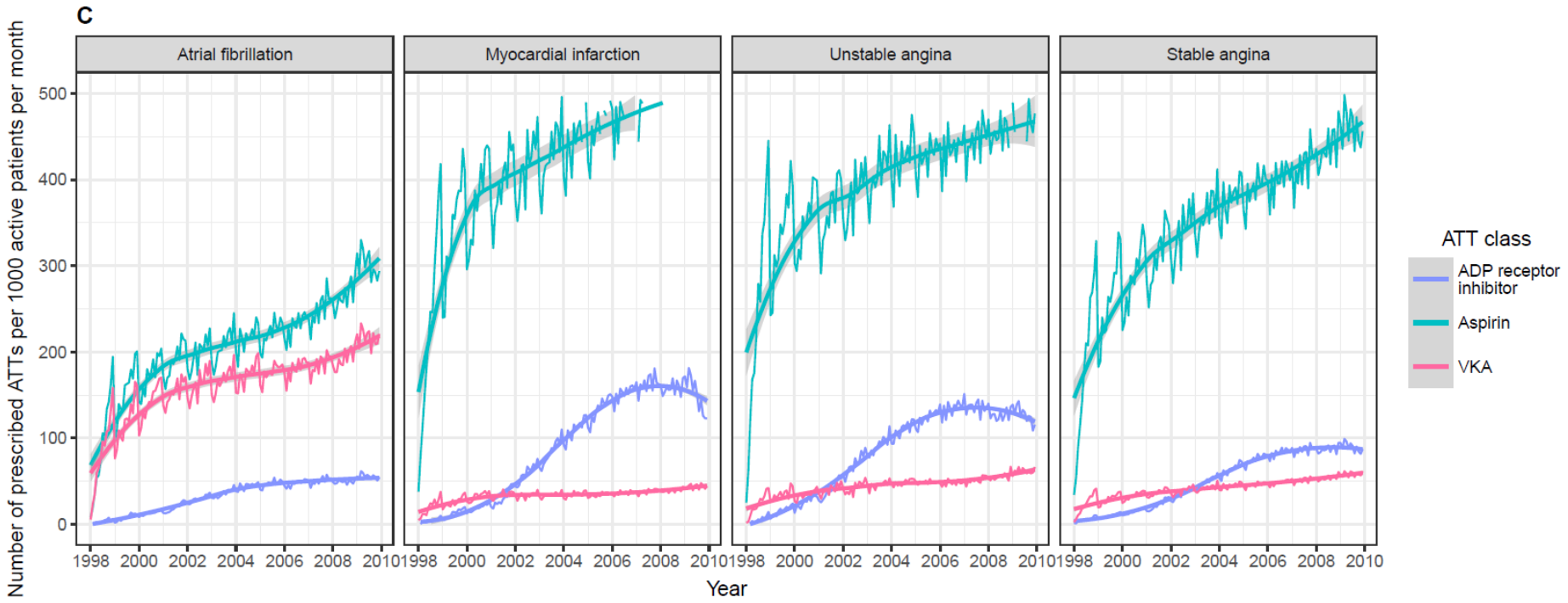
11.3.1 Fatal, hospitalised+ and primary care+ bleeding event rates 1998-2010 stratified by initial cardiovascular disease



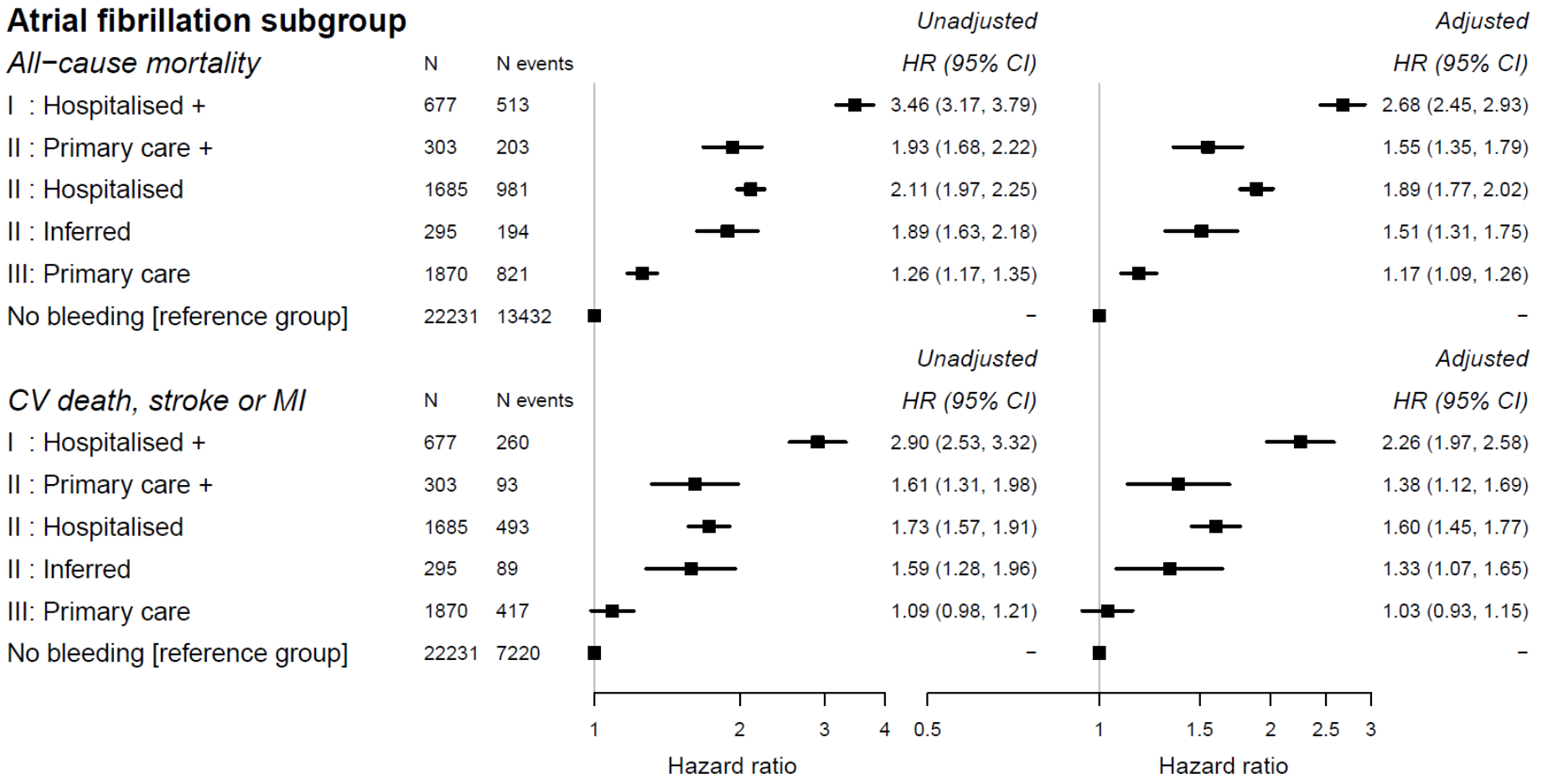
11.3.2 Hospitalised and primary care bleeding event rate 1998-2010 stratified by initial cardiovascular disease



11.3.3 Antithrombotic therapy prescribing rates 1998-2010 stratified by initial cardiovascular disease



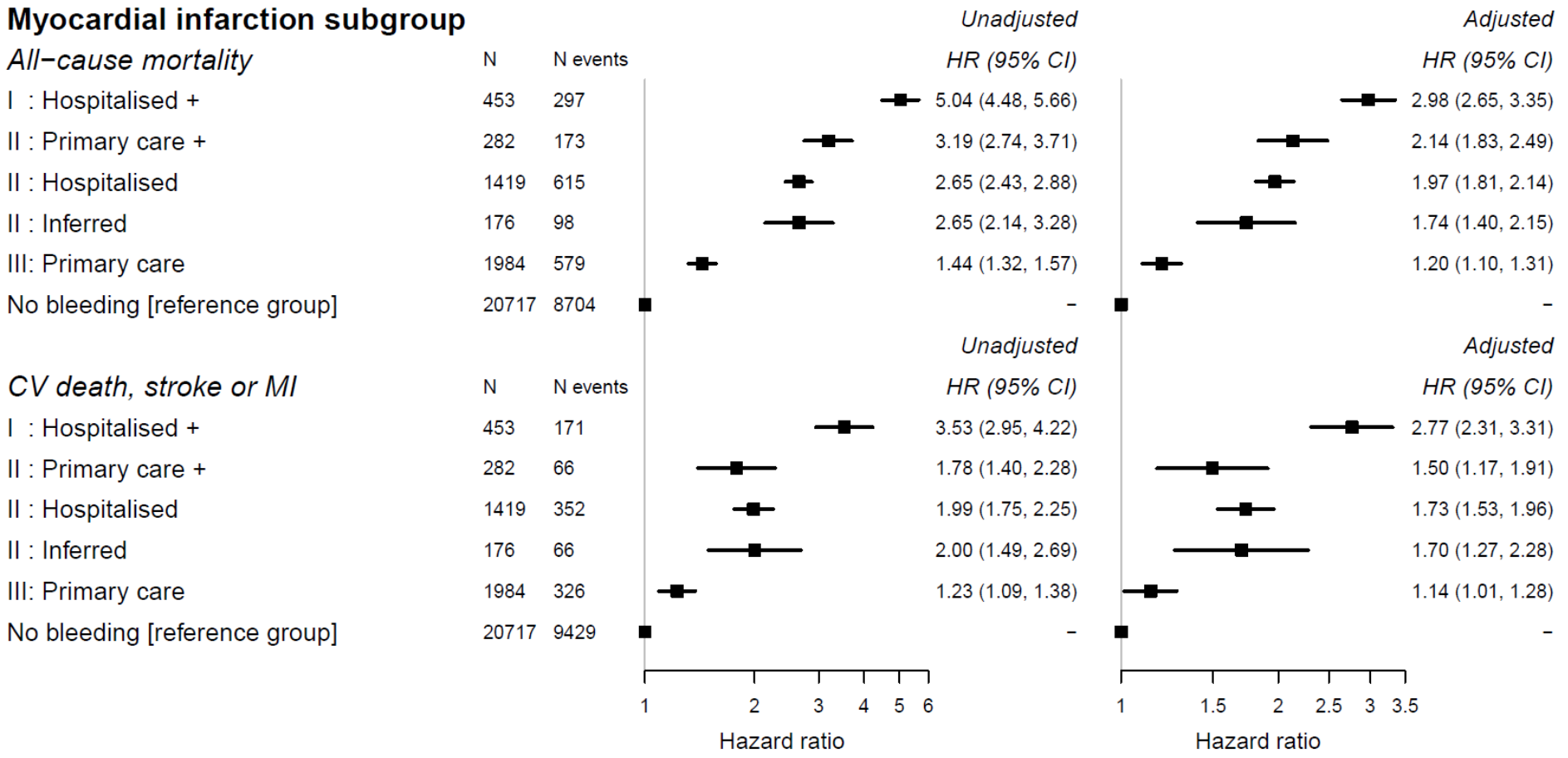
11.3.4 The association between non-fatal bleeding severity classes and time to all-cause mortality and cardiovascular death, stroke or myocardial infarction in the atrial fibrillation subgroup



Note:

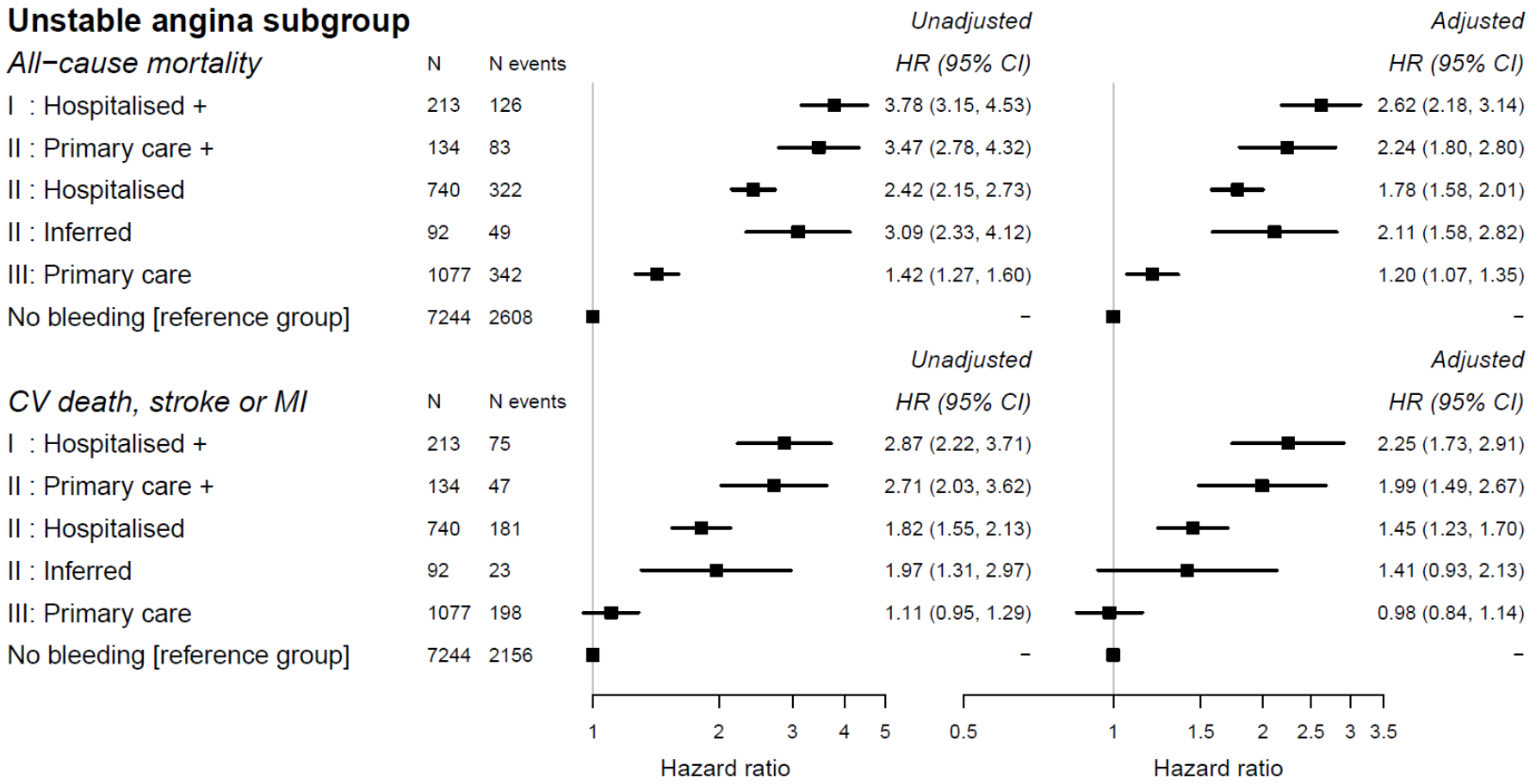
Adjusted estimates are adjusted for age, sex and comorbidities; CV= cardiovascular; MI= myocardial infarction

11.3.5 The association between non-fatal bleeding severity classes and time to all-cause mortality and cardiovascular death, stroke or myocardial infarction in the myocardial infarction subgroup



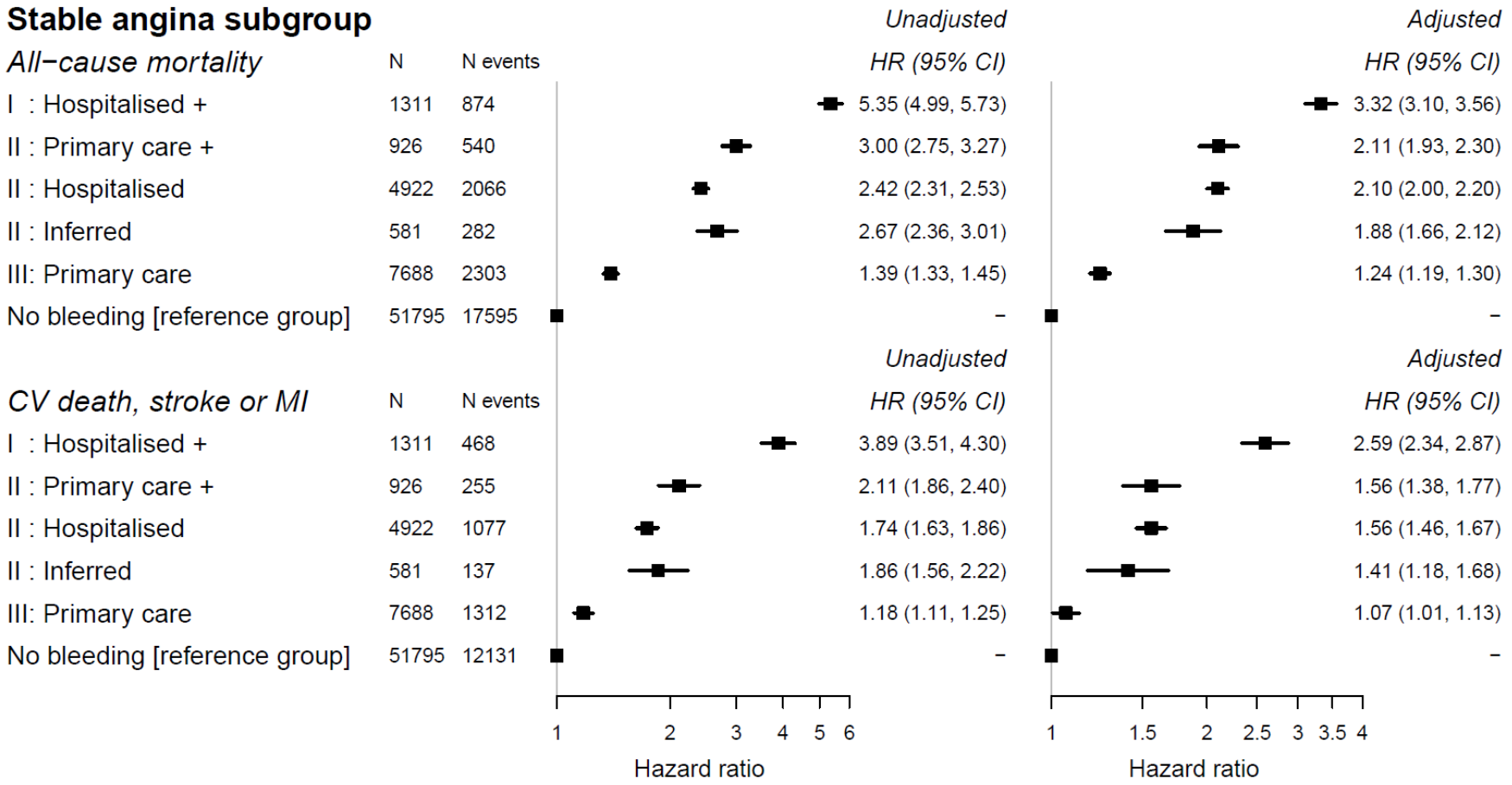
Note: Adjusted estimates are adjusted for age, sex and comorbidities; CV= cardiovascular; MI= myocardial infarction

11.3.6 : The association between non-fatal bleeding severity classes and time to all-cause mortality and cardiovascular death, stroke or myocardial infarction in the unstable angina subgroup



Note: Adjusted estimates are adjusted for age, sex and comorbidities; CV= cardiovascular; MI= myocardial infarction

11.3.7 The association between non-fatal bleeding severity classes and time to all-cause mortality and cardiovascular death, stroke or myocardial infarction in the stable angina subgroup



Note: Adjusted estimates are adjusted for age, sex and comorbidities; CV= cardiovascular; MI= myocardial infarction

11.4 Supplementary appendix (Chapter 7)

11.4.1 Completed TRIPOD checklist

Section	Item		Checklist Item	Page
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	210
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	210
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	211
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	211
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	212
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	212
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	212
	5b	D;V	Describe eligibility criteria for participants.	212
	5c	D;V	Give details of treatments received, if relevant.	n/a
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	213; Table 3.4; Table 3.5
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	n/a
Predictors	7a	D;V	Clearly define all predictors used in developing the multivariable prediction model, including how and when they were measured.	213; Table 3.3
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	n/a
Sample size	8	D;V	Explain how the study size was arrived at.	212
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	214
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	213
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	213
	10c	V	For validation, describe how the predictions were calculated.	214
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	214
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	n/a
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	214
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	212;
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	Figure 7.1
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	Table 7.1

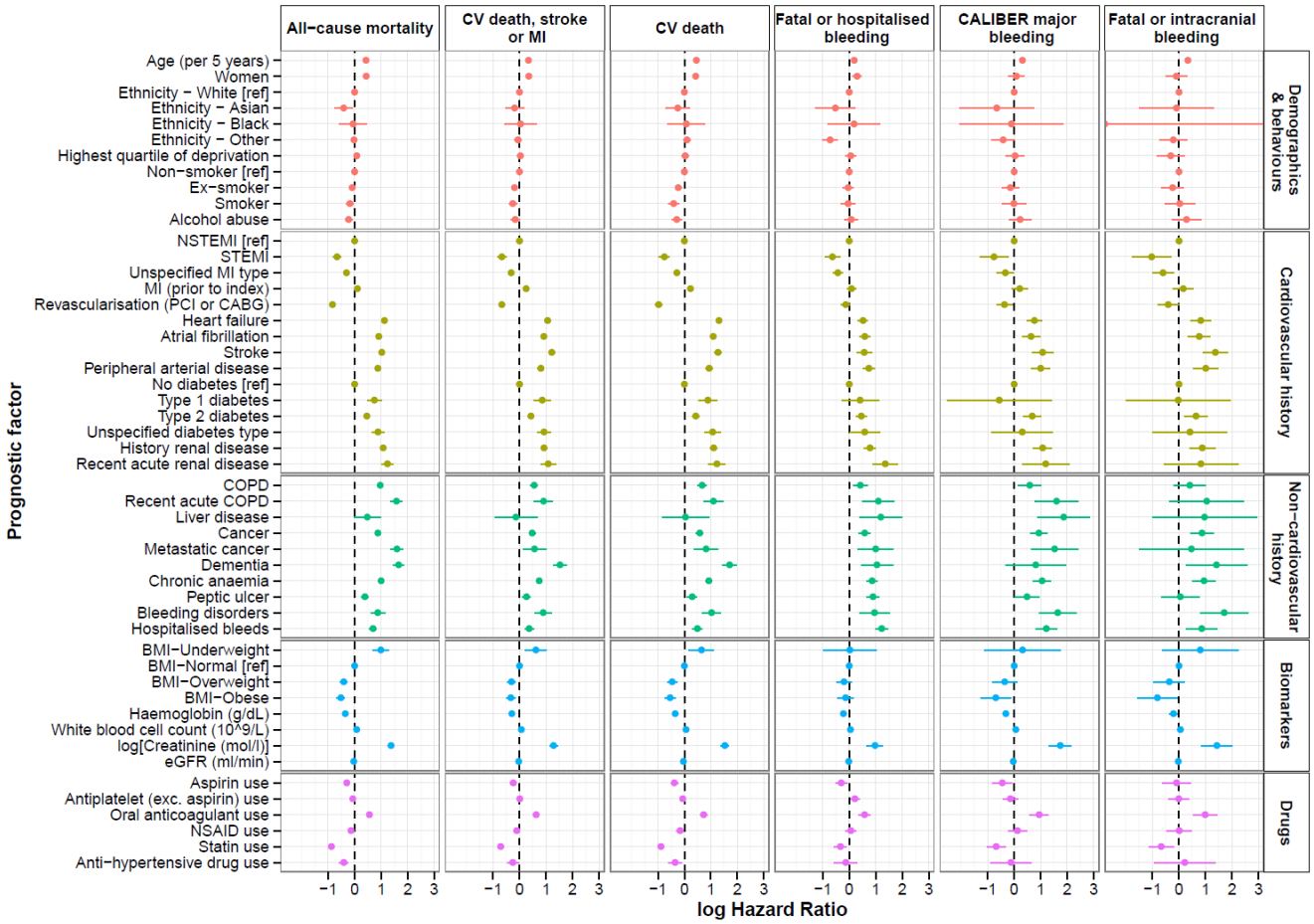
Section	Item		Checklist Item	Page
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	Table 7.1; Figure 7.2
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	Figure 7.1
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	Suppl. 11.4.3
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	Suppl. 11.4.5 Suppl. 11.4.6
	15b	D	Explain how to use the prediction model.	220
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	219
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	n/a
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	223
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	221
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	222
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	223-224
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	220; Figure 7.9
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	

11.4.2 Patient characteristics at index acute MI discharge and 1 year post-index acute MI in the development (n=12,694) and validation (n=5,613) cohorts

	Development (n=12,694)		Validation (n=5,613)	
	At index acute MI discharge	1 year post-MI	At index acute MI	1 year post-MI
Age	69.1 (12.7)	70.1 (12.7)	68.1 (12.8)	69.1 (12.8)
Smoking status				
Ex-smoker	40.1	49.3	40.3	50.0
Non-smoker	42	36.7	38.7	34.2
Smoker	17.9	14.0	21.1	15.8
Alcohol abuse	9.8	10.8	14.3	15.4
Cardiovascular diseases				
Revascularisation (PCI)	24.8	43.5	18.8	33.0
Heart failure	18.2	23.5	21.4	28.0
Atrial fibrillation	14.9	18.0	14.8	17.9
Stroke	6.0	6.9	7.0	8.1
Peripheral arterial disease	8.2	9.8	11	13.1
Diabetes				
Type 1	1.3	1.2	0.9	0.9
Type 2	15.2	16.7	15.5	17
Unspecified	1.2	1.5	1.4	1.7
Renal disease	8.7	13.6	9.5	14.8
Non-cardiovascular diseases				
COPD	7.8	9.1	10.9	12.8
Liver disease	0.3	0.4	0.5	0.5
Non-metastatic cancer	12.9	14.4	11.3	13.2
Metastatic cancer	0.8	1.0	0.7	1.2
Dementia	0.8	1.3	1.4	2.0
Chronic anaemia	11	14.3	14	17.9
Peptic ulcer	6.6	7.3	9.3	10.2
Bleeding diatheses and coagulation disorders	0.8	1.1	0.9	1.1
Hospitalised bleeding	4.3	6.5	5.7	8.2
Biomarkers				
BMI (Continuous)	28.0 (5.0)	27.8 (5.1)	27.9 (5.2)	27.7 (5.1)
BMI (Categorical)				
Underweight	1.1	1.5	1.6	1.8
Normal	27.1	28.7	27.9	28.4
Overweight	41.8	41.2	42.1	42.2
Obese	30.1	28.6	28.4	27.7
SBP (mmHg)	145 (16.3)	133 (18.6)	144 (17.0)	132 (18.4)
Haemoglobin (g/dL)	13.9 (1.62)	13.4 (1.6)	13.9 (1.61)	13.3 (1.6)
White blood cell count (10^9/L)	7.69 (2.19)	7.60 (2.3)	7.72 (2.19)	7.68 (2.3)
Total cholesterol (mmol/L)	5.32 (1.16)	4.17 (1.0)	5.38 (1.14)	4.17 (1.0)
HDL cholesterol (mmol/L)	1.34 (0.40)	1.28 (0.4)	1.33 (0.37)	1.26 (0.4)
Creatinine (mol/l) Median (IQR)	95 (83, 110)	98 (84, 114)	97 (85, 113)	99 (86, 117)

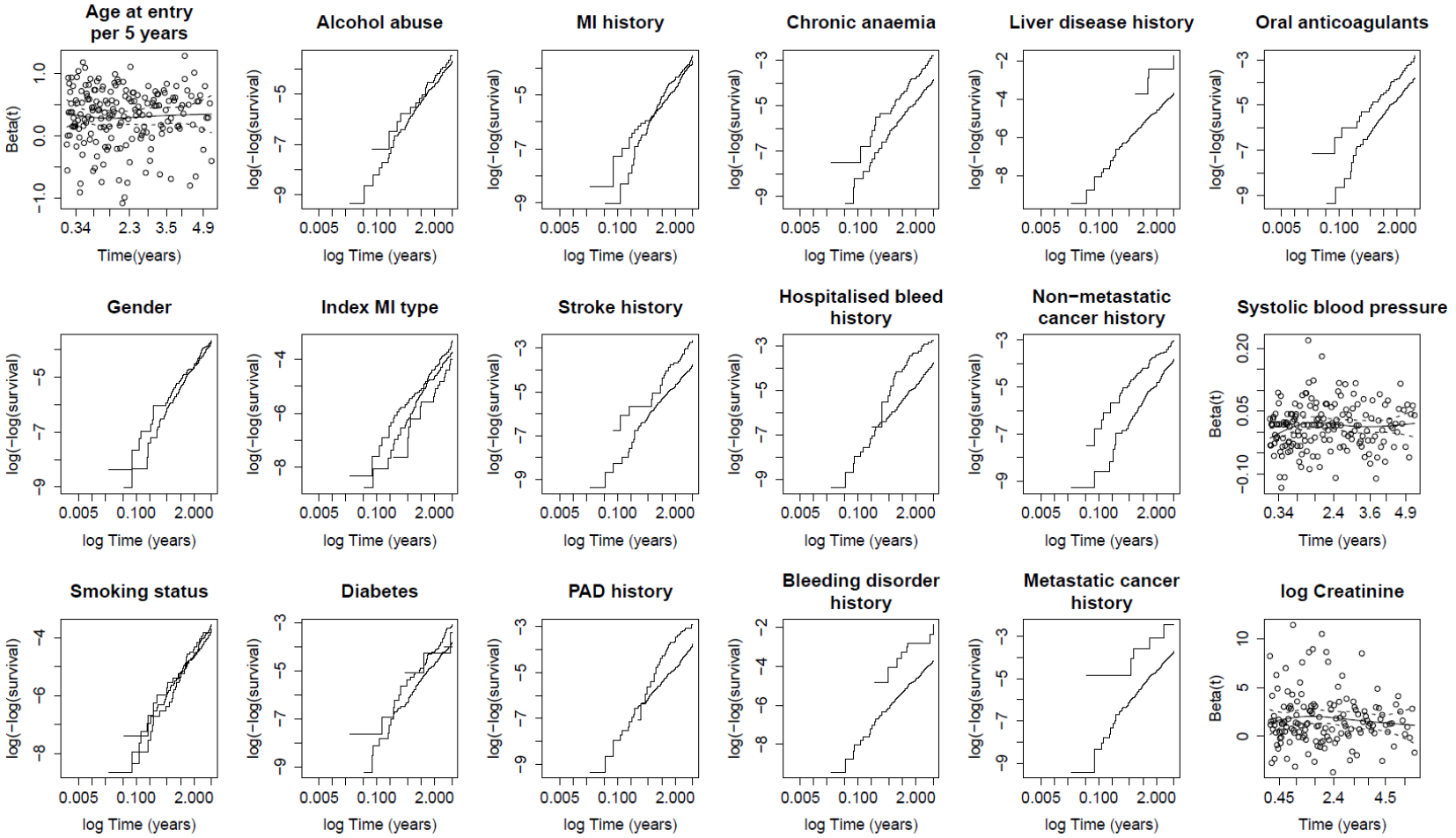
Note: Categorical prognostic factors are presented as %, continuous prognostic factors are presented as mean(SD) unless stated otherwise

11.4.3 Univariable effects of prognostic factors on 5 year all-cause mortality, cardiovascular and bleeding endpoints



Note: CV= cardiovascular, MI= myocardial infarction, NSTEMI= non-ST-elevation myocardial infarction, STEMI= ST-elevation myocardial infarction, COPD= chronic obstructive pulmonary disease, BMI= body mass index, eGFR= estimated glomerular filtration rate, NSAID= non-steroidal anti-inflammatory drugs; (log hazards compared with reference group or per unit increase for continuous prognostic factors and 95% confidence intervals)

11.4.4 Univariable proportional hazards assumption checks for the CALIBER major bleeding outcome



Note: For continuous variables, Beta(t) is a time dependent coefficient and should remain constant over time if the proportional hazards assumption have not been violated**.** For categorical variables, proportional hazards have not been violated if the log(-log(survival)) curves remain parallel over time.; MI= myocardial infarction; PAD= peripheral arterial disease

11.4.5 Multivariable model prognostic hazard ratios and 95% confidence intervals for all-cause mortality, cardiovascular and bleeding endpoints

Prognostic factor (measured at 1 year post-MI)	All-cause mortality	Cardiovascular death, stroke or MI	Cardiovascular death	CALIBER major bleeding	Fatal or hospitalised bleeding	Fatal bleeding or intracranial bleeding
Age (per 5 years)	1.38 (1.35, 1.42)	1.28 (1.24, 1.32)	1.39 (1.34, 1.45)	1.3 (1.19, 1.42)	1.14 (1.09, 1.19)	1.35 (1.20, 1.52)
Women	0.89 (0.81, 0.98)	0.89 (0.79, 0.99)	0.85 (0.75, 0.97)	0.85 (0.61, 1.2)	1.15 (0.95, 1.4)	0.54 (0.34, 0.85)
Ethnicity						
White [ref]	1	NA	1	NA	1	NA
Asian	0.96 (0.68, 1.36)	NA	1.07 (0.68, 1.68)	NA	0.68 (0.32, 1.44)	NA
Black	1.19 (0.7, 2.04)	NA	1.28 (0.63, 2.6)	NA	1.20 (0.45, 3.22)	NA
Other	1.16 (1.05, 1.28)	NA	1.26 (1.1, 1.44)	NA	0.58 (0.43, 0.77)	NA
Smoking status						
Non-smoker [ref]	1	1	1	1	1	1
Ex-smoker	1.02 (0.93, 1.12)	0.91 (0.82, 1.01)	0.9 (0.79, 1.02)	0.88 (0.63, 1.23)	0.96 (0.79, 1.17)	0.73 (0.47, 1.14)
Smoker	1.47 (1.28, 1.7)	1.25 (1.06, 1.48)	1.22 (0.99, 1.5)	1.42 (0.87, 2.32)	1.23 (0.92, 1.64)	1.44 (0.77, 2.68)
Alcohol abuse	1.24 (1.07, 1.43)	1.25 (1.07, 1.47)	1.23 (1.01, 1.5)	1.69 (1.08, 2.65)	1.28 (0.97, 1.68)	1.80 (1.02, 3.19)
Index acute MI type						
NSTEMI [ref]	1	1	1	1	1	1
STEMI	0.81 (0.7, 0.93)	0.74 (0.63, 0.87)	0.74 (0.61, 0.91)	0.67 (0.39, 1.16)	0.67 (0.5, 0.91)	0.51 (0.24, 1.06)
Unspecified	0.88 (0.81, 0.97)	0.78 (0.7, 0.86)	0.82 (0.73, 0.93)	0.84 (0.61, 1.15)	0.78 (0.65, 0.95)	0.61 (0.40, 0.91)
MI (prior to index acute MI)	1.07 (0.98, 1.16)	1.23 (1.12, 1.35)	1.2 (1.08, 1.34)	1.18 (0.88, 1.59)	1.04 (0.87, 1.23)	1.14 (0.77, 1.68)
Previous revascularisation	0.69 (0.63, 0.76)	0.72 (0.64, 0.8)	0.61 (0.53, 0.7)	NA	NA	0.87 (0.56, 1.34)
Heart failure	1.47 (1.35, 1.6)	1.57 (1.42, 1.74)	1.74 (1.55, 1.97)	NA	NA	1.34 (0.88, 2.04)
Atrial fibrillation	1.23 (1.13, 1.34)	1.33 (1.2, 1.47)	1.41 (1.26, 1.59)	NA	NA	0.85 (0.51, 1.42)
Stroke	1.41 (1.26, 1.58)	1.86 (1.64, 2.11)	1.73 (1.49, 2)	1.6 (1.06, 2.43)	1.13 (0.84, 1.51)	2.28 (1.38, 3.78)
Peripheral arterial disease	1.48 (1.33, 1.64)	1.45 (1.28, 1.64)	1.6 (1.38, 1.84)	1.55 (1.06, 2.27)	1.41 (1.11, 1.79)	1.64 (1.01, 2.69)
Diabetes						
No diabetes	1	1	1	1	1	1

Prognostic factor (measured at 1 year post-MI)	All-cause mortality	Cardiovascular death, stroke or MI	Cardiovascular death	CALIBER major bleeding	Fatal or hospitalised bleeding	Fatal bleeding or intracranial bleeding
Type 1 diabetes	2.33 (1.75, 3.09)	2.33 (1.7, 3.2)	2.44 (1.68, 3.52)	0.55 (0.08, 4)	1.48 (0.73, 3)	0.87 (0.12, 6.40)
Type 2 diabetes	1.22 (1.11, 1.35)	1.18 (1.05, 1.32)	1.15 (1, 1.33)	1.49 (1.06, 2.09)	1.29 (1.05, 1.59)	1.48 (0.95, 2.30)
Unspecified diabetes	2.44 (1.93, 3.1)	2.34 (1.79, 3.07)	2.89 (2.13, 3.92)	1.07 (0.33, 3.41)	1.62 (0.9, 2.89)	1.31 (0.31, 5.45)
Renal disease	1.23 (1.1, 1.37)	NA	NA	NA	NA	NA
COPD	1.62 (1.45, 1.8)	NA	1.23 (1.04, 1.44)	NA	NA	NA
Liver disease	NA	NA	NA	3.94 (1.4, 11.11)	NA	NA
Non-metastatic cancer	1.31 (1.19, 1.44)	NA	NA	1.48 (1.04, 2.12)	1.26 (1.01, 1.56)	1.49 (0.95, 2.34)
Metastatic cancer	2.32 (1.81, 2.98)	NA	NA	2.19 (0.86, 5.58)	NA	NA
Dementia	2.03 (1.64, 2.53)	2.03 (1.56, 2.64)	2.16 (1.62, 2.87)	NA	NA	NA
Chronic anaemia	1.13 (1.02, 1.25)	NA	NA	1.42 (0.99, 2.03)	1.40 (1.13, 1.73)	1.42 (0.89, 2.25)
Peptic ulcer	NA	NA	1.02 (0.85, 1.24)	NA	1.75 (1.37, 2.24)	NA
Bleeding diatheses or coagulation disorders	NA	1.3 (0.95, 1.79)	NA	2.02 (0.94, 4.32)	NA	2.49 (0.95, 6.48)
Hospitalised bleeding	NA	0.84 (0.71, 1)	0.85 (0.69, 1.03)	1.82 (1.2, 2.78)	2.01 (1.56, 2.58)	NA
BMI						
Underweight	1.91 (1.43, 2.53)	1.6 (1.11, 2.3)	1.65 (1.09, 2.5)	NA	NA	NA
Normal [ref]	1	1	1	NA	NA	NA
Overweight	0.93 (0.82, 1.05)	0.95 (0.84, 1.08)	0.92 (0.78, 1.08)	NA	NA	NA
Obese	1 (0.88, 1.15)	1.05 (0.91, 1.22)	1.02 (0.85, 1.24)	NA	NA	NA
White blood cell count (10^9/L)	1.06 (1.03, 1.08)	1.05 (1.02, 1.07)	1.05 (1.02, 1.08)	NA	NA	NA
Haemoglobin (g/dL)	0.88 (0.85, 0.92)	0.91 (0.88, 0.95)	0.89 (0.85, 0.93)	NA	NA	NA
log Creatinine (µmol/l)	1.35 (1.15, 1.59)	1.49 (1.25, 1.77)	1.72 (1.41, 2.1)	2.15 (1.29, 3.57)	1.44 (1.05, 1.97)	NA
Cholesterol ratio (HDL:total)	1.58 (0.92, 2.71)	1.46 (0.82, 2.59)	1.56 (0.77, 3.15)	NA	NA	NA
Antiplatelet	NA	NA	NA	NA	1.23 (1.03, 1.47)	NA
Oral anticoagulant	NA	NA	NA	1.74 (1.19, 2.52)	1.49 (1.17, 1.9)	1.89 (1.08, 3.33)

Note: All models are adjusted for systolic blood pressure using restricted cubic splines. Functions are shown in **Supplementary appendix 11.4.6**; MI=myocardial infarction, NSTEMI= non-ST-elevation myocardial infarction, STEMI= ST-elevation myocardial infarction, COPD= chronic obstructive pulmonary disease, HDL= high-density lipoprotein

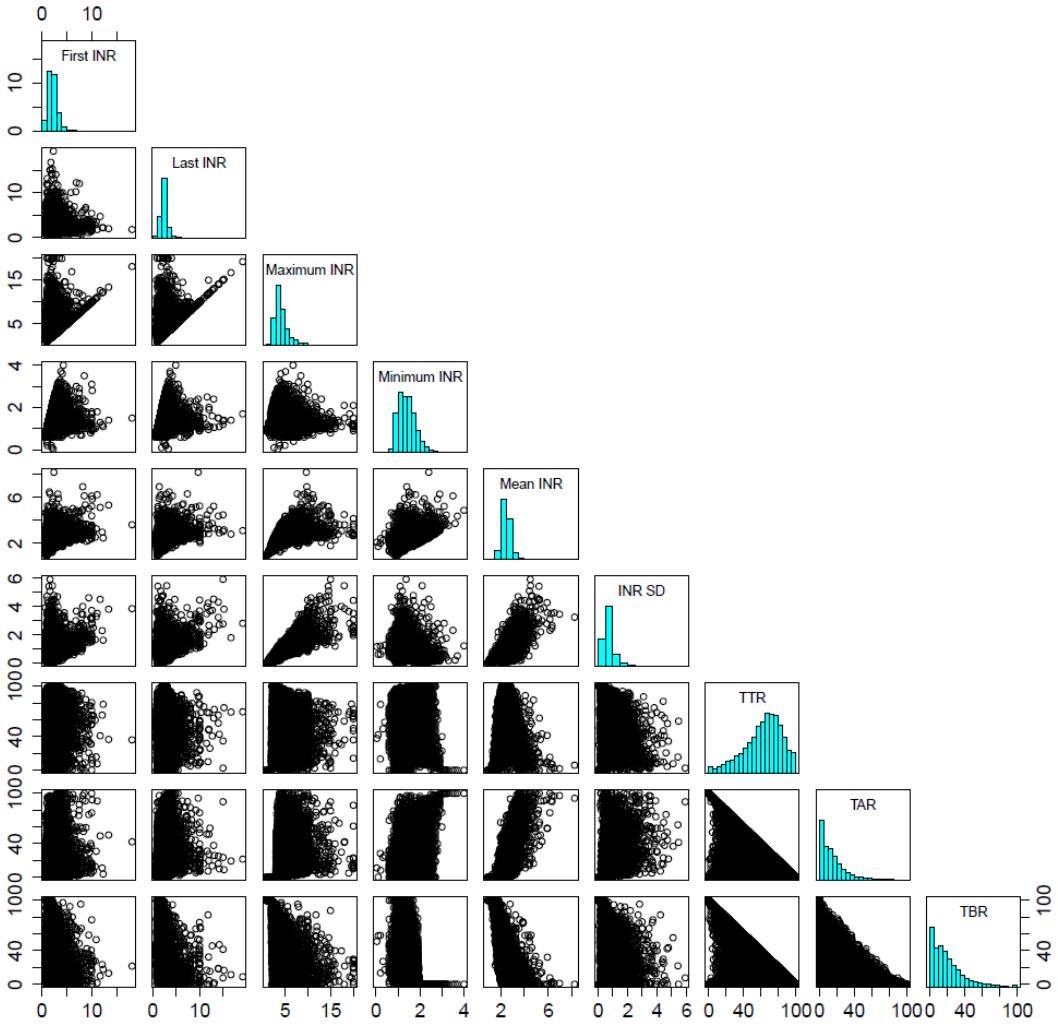
11.4.6 Linear predictor functions for systolic blood pressure in the multivariable models

Endpoint	Function
All-cause mortality	$0.011 \times SBP - 3.566 \times 10^{-6} \times \max(SBP - 110,0)^3 + 6.562 \times 10^{-6} \times \max(SBP - 131,0)^3 - 2.996 \times 10^{-6} \times \max(SBP - 156,0)^3$
Cardiovascular death, stroke or MI	$0.012 \times SBP - 5.619 \times 10^{-6} \times \max(SBP - 110,0)^3 + 1.034 \times 10^{-5} \times \max(SBP - 131,0)^3 - 4.720 \times 10^{-6} \times \max(SBP - 156,0)^3$
Cardiovascular death	$0.015 \times SBP - 6.383 \times 10^{-6} \times \max(SBP - 110,0)^3 + 1.174 \times 10^{-5} \times \max(SBP - 131,0)^3 - 5.362 \times 10^{-6} \times \max(SBP - 156,0)^3$
CALIBER major bleeding	$0.006 \times SBP - 5.722 \times 10^{-6} \times \max(SBP - 110,0)^3 + 1.0529 \times 10^{-5} \times \max(SBP - 131,0)^3 - 4.807 \times 10^{-6} \times \max(SBP - 156,0)^3$
Fatal or hospitalised bleeding	$0.010 \times SBP - 6.231 \times 10^{-6} \times \max(SBP - 110,0)^3 + 1.146 \times 10^{-5} \times \max(SBP - 131,0)^3 - 5.234 \times 10^{-6} \times \max(SBP - 156,0)^3$
Fatal bleeding or intracranial bleeding	$0.002 \times SBP - 5.176 \times 10^{-6} \times \max(SBP - 110,0)^3 + 9.524 \times 10^{-6} \times \max(SBP - 131,0)^3 - 4.348 \times 10^{-6} \times \max(SBP - 156,0)^3$

Note: Systolic blood pressure was modelled using restricted cubic splines with 3 knots in the multivariable models. The functions in this table described the estimated relationship between systolic blood pressure and the studied endpoints

11.5 Supplementary appendix (Chapter 9)

11.5.1 Distributions and matrix plot of measures of INR control



INR= International normalised ratio; SD= standard deviation; TTR=time in therapeutic range; TAR= time above therapeutic range; TBR= time below therapeutic range

11.5.2 Base prognostic models for one year all-cause mortality, cardiovascular death, stroke or myocardial infarction, any bleeding and major bleeding

	Odds ratios (95% confidence intervals) for one year endpoints			
Variable	**All-cause mortality**	**Cardiovascular death, stroke or MI**	**Any bleeding**	**Major bleeding**
Intercept	0.059 (0.020, 0.174)	0.001 (0.000, 0.004)	0.018 (0.005, 0.061)	0.001 (0.000, 0.008)
Age (years)	1.046 (1.040, 1.051)	1.062 (1.054, 1.070)	1.021 (1.015, 1.027)	1.041 (1.029, 1.052)
Women	0.898 (0.802, 1.005)	1.018 (0.857, 1.209)	0.787 (0.679, 0.913)	0.790 (0.615, 1.015)
Highest quintile of deprivation	1.187 (1.066, 1.322)	1.224 (1.062, 1.409)	1.224 (1.071, 1.398)	1.230 (0.984, 1.538)
Log index INR spell duration (days)	1.056 (1.014, 1.100)	1.128 (1.067, 1.193)	0.990 (0.941, 1.041)	1.123 (1.030, 1.224)
Prior exposure to oral anticoagulants (years)	0.989 (0.960, 1.019)	1.002 (0.966, 1.040)	0.975 (0.938, 1.014)	1.002 (0.943, 1.063)
Smoker vs. non-smoker	1.342 (1.117, 1.613)	1.452 (1.143, 1.844)	1.040 (0.826, 1.309)	1.285 (0.878, 1.879)
Ex-smoker vs. non-smoker	1.109 (1.000, 1.229)	0.985 (0.861, 1.126)	0.984 (0.863, 1.121)	0.941 (0.754, 1.176)
History of excess alcohol consumption	0.842 (0.732, 0.969)	0.872 (0.720, 1.057)	0.998 (0.847, 1.174)	0.703 (0.510, 0.970)
Atrial fibrillation	0.747 (0.668, 0.837)	0.953 (0.819, 1.107)	0.886 (0.768, 1.021)	0.930 (0.733, 1.180)
Myocardial infarction	1.427 (1.240, 1.643)	1.900 (1.608, 2.244)	1.046 (0.862, 1.268)	1.109 (0.817, 1.505)
Heart failure	1.879 (1.690, 2.089)	2.112 (1.851, 2.409)	1.163 (1.009, 1.341)	1.306 (1.044, 1.634)
Heart valve replacement	0.766 (0.593, 0.990)	1.218 (0.904, 1.642)	0.878 (0.649, 1.187)	1.127 (0.707, 1.795)
Unstable angina	0.900 (0.736, 1.101)	0.927 (0.725, 1.186)	1.138 (0.886, 1.461)	1.138 (0.759, 1.707)
Peripheral arterial disease	1.411 (1.215, 1.637)	1.493 (1.240, 1.797)	0.920 (0.743, 1.139)	0.969 (0.692, 1.356)
Venous thromboembolism	0.952 (0.846, 1.071)	0.716 (0.605, 0.848)	0.875 (0.753, 1.017)	0.787 (0.606, 1.020)
Ischaemic or unspecified stroke	1.377 (1.198, 1.584)	1.853 (1.574, 2.180)	1.050 (0.872, 1.265)	1.121 (0.836, 1.504)
Any bleeding	0.995 (0.900, 1.100)	0.975 (0.851, 1.117)	1.764 (1.571, 1.982)	1.443 (1.183, 1.761)
Major bleeding	1.246 (0.917, 1.693)	1.152 (0.762, 1.741)	1.022 (0.706, 1.479)	1.366 (0.784, 2.378)
Unspecified diabetes	2.501 (1.657, 3.775)	1.914 (1.149, 3.189)	1.381 (0.785, 2.429)	1.111 (0.405, 3.050)
Type 1 diabetes	1.461 (0.818, 2.609)	1.017 (0.425, 2.437)	1.426 (0.710, 2.867)	1.011 (0.245, 4.172)
Type 2 diabetes	1.157 (1.010, 1.326)	1.205 (1.014, 1.433)	0.967 (0.808, 1.156)	1.100 (0.830, 1.457)
Cancer	2.314 (2.091, 2.560)	0.863 (0.738, 1.008)	1.211 (1.058, 1.387)	1.325 (1.063, 1.652)
Chronic obstructive pulmonary disease	1.776 (1.551, 2.032)	1.086 (0.894, 1.318)	1.101 (0.913, 1.328)	1.067 (0.785, 1.449)
Renal disease	0.809 (0.703, 0.930)	0.850 (0.709, 1.017)	0.738 (0.614, 0.887)	0.778 (0.581, 1.043)
HDL (mmol/L)	0.867 (0.729, 1.032)	0.780 (0.645, 0.943)	1.111 (0.854, 1.444)	1.189 (0.850, 1.664)
Creatinine (mmol/L)	1.003 (1.002, 1.005)	1.004 (1.002, 1.006)	1.002 (1.000, 1.004)	1.003 (1.000, 1.005)
Haemoglobin (g/dL)	0.885 (0.855, 0.915)	0.979 (0.932, 1.028)	0.982 (0.935, 1.031)	0.924 (0.848, 1.007)
Total white blood cell count (10^9/L)	1.032 (1.011, 1.053)	1.028 (1.005, 1.052)	1.003 (0.973, 1.034)	1.018 (0.970, 1.068)

INR= international normalised ratio; HDL= high density lipoproteins; the models were also adjusted for systolic blood pressure and body mass index using restricted cubic splines

11.5.3 All-cause mortality; integrated discrimination improvement and net reclassification improvement estimates

Measure of INR control	IDI – events [a]	IDI – non-events [b]	IDI (95% CI)	NRI - events (95% CI) [c]	NRI – non-events (95% CI) [d]	NRI (95% CI)
TTR group	0.0107	0.0016	0.0123 (0.0105, 0.0142)	0.60 (0.57, 0.63)	-0.28 (-0.30, -0.27)	0.32 (0.29, 0.36)
Time in therapeutic range	0.0139	0.002	0.0159 (0.0137, 0.0181)	0.35 (0.31, 0.39)	-0.05 (-0.06, -0.03)	0.30 (0.27, 0.34)
Time above therapeutic range	0.0201	0.0032	0.0233 (0.0204, 0.0262)	0.12 (0.08, 0.15)	0.27 (0.25, 0.28)	0.38 (0.34, 0.42)
Time below therapeutic range	0.0006	0.0001	0.0006 (0.0002, 0.0011)	-0.08 (-0.12, -0.04)	0.18 (0.16, 0.19)	0.10 (0.06, 0.14)
Mean INR	0.0153	0.0025	0.0178 (0.0152, 0.0204)	0.05 (0.01, 0.08)	0.27 (0.26, 0.29)	0.32 (0.28, 0.36)
INR standard deviation	0.0305	0.0047	0.0353 (0.0317, 0.0388)	0.15 (0.11, 0.18)	0.31 (0.30, 0.33)	0.46 (0.42, 0.50)
Minimum INR value	0.0022	0.0003	0.0025 (0.0016, 0.0034)	0.27 (0.24, 0.31)	-0.15 (-0.17, -0.14)	0.12 (0.08, 0.16)
Maximum INR value	0.0324	0.0051	0.0375 (0.0337, 0.0413)	0.12 (0.08, 0.15)	0.36 (0.35, 0.38)	0.48 (0.44, 0.52)
First INR value	0.0012	0.0002	0.0014 (0.0007, 0.0021)	-0.18 (-0.22, -0.14)	0.27 (0.26, 0.29)	0.09 (0.05, 0.13)
Last INR value	0.02	0.0032	0.0232 (0.0202, 0.0262)	-0.18 (-0.22, -0.15)	0.50 (0.48, 0.51)	0.31 (0.27, 0.35)
INR trajectory	0.0044	0.0007	0.0051 (0.0038, 0.0064)	-0.57 (-0.60, -0.54)	0.67 (0.66, 0.68)	0.10 (0.07, 0.14)

INR= international normalised ratio; IDI = Integrated discrimination improvement; NRI= net reclassification improvement; CI= confidence interval; [a] Mean probability increase for events patients; [b] Mean probability decrease for non-events patients; [c] proportion of events patients with increased prediction probability- proportion of events patients with decreased prediction probability; [d] proportion of non-events patients with decrease prediction probability- proportion of non-events patients with increased prediction probability

11.5.4 Cardiovascular death, stroke or MI; integrated discrimination improvement and net reclassification improvement estimates

Measure of INR control	IDI – events [a]	IDI – non-events [b]	IDI (95% CI)	NRI - events (95% CI) [c]	NRI – non-events (95% CI) [d]	NRI (95% CI)
TTR group	0.0038	0.0003	0.0040 (0.0028, 0.0053)	0.52 (0.47, 0.56)	-0.30 (-0.31, -0.29)	0.22 (0.17, 0.27)
Time in therapeutic range	0.0046	0.0003	0.0050 (0.0036, 0.0063)	0.31 (0.26, 0.36)	-0.15 (-0.16, -0.13)	0.16 (0.11, 0.22)
Time above therapeutic range	0.0046	0.0003	0.0049 (0.0033, 0.0065)	-0.01 (-0.06, 0.05)	0.23 (0.22, 0.25)	0.22 (0.17, 0.28)
Time below therapeutic range	0.0004	0	0.0004 (0.0000, 0.0008)	0.32 (0.27, 0.37)	-0.25 (-0.26, -0.24)	0.07 (0.02, 0.12)
Mean INR	0.0031	0.0002	0.0034 (0.0020, 0.0047)	-0.19 (-0.25, -0.14)	0.39 (0.37, 0.40)	0.19 (0.14, 0.25)
INR standard deviation	0.0087	0.0006	0.0093 (0.0071, 0.0116)	-0.18 (-0.24, -0.13)	0.41 (0.40, 0.42)	0.23 (0.17, 0.28)
Minimum INR value	0.001	0.0001	0.0011 (0.0003, 0.0019)	0.10 (0.05, 0.16)	0.00 (-0.01, 0.02)	0.11 (0.05, 0.16)
Maximum INR value	0.0095	0.0007	0.0102 (0.0078, 0.0125)	-0.07 (-0.12, -0.01)	0.38 (0.37, 0.39)	0.31 (0.26, 0.37)
First INR value	0.0006	0	0.0006 (0.0001, 0.0011)	0.11 (0.06, 0.16)	-0.08 (-0.09, -0.06)	0.03 (-0.02, 0.09)
Last INR value	0.0062	0.0005	0.0067 (0.0047, 0.0086)	-0.12 (-0.17, -0.06)	0.33 (0.32, 0.35)	0.22 (0.16, 0.27)
INR trajectory	0.001	0.0001	0.0011 (0.0004, 0.0018)	-0.60 (-0.64, -0.56)	0.64 (0.63, 0.65)	0.04 (0.00, 0.08)

INR= international normalised ratio; IDI = Integrated discrimination improvement; NRI= net reclassification improvement; CI= confidence interval; [a] Mean probability increase for events patients; [b] Mean probability decrease for non-events patients; [c] proportion of events patients with increased prediction probability- proportion of events patients with decreased prediction probability; [d] proportion of non-events patients with decrease prediction probability- proportion of non-events patients with increased prediction probability

11.5.5 Any bleeding; integrated discrimination improvement and net reclassification improvement estimates

Measure of INR control	IDI – events [a]	IDI – non-events [b]	IDI (95% CI)	NRI - events (95% CI) [c]	NRI – non-events (95% CI) [d]	NRI (95% CI)
TTR group	0.0005	0	0.0005 (0.0001, 0.0009)	0.39 (0.34, 0.43)	-0.29 (-0.31, -0.28)	0.09 (0.04, 0.14)
Time in therapeutic range	0.0008	0.0001	0.0009 (0.0003, 0.0014)	-0.17 (-0.22, -0.12)	0.25 (0.23, 0.26)	0.07 (0.02, 0.13)
Time above therapeutic range	0.0006	0	0.0006 (0.0002, 0.0011)	-0.10 (-0.15, -0.05)	0.18 (0.17, 0.20)	0.08 (0.02, 0.13)
Time below therapeutic range	0	0	0.0000 (-0.0001, 0.0001)	-0.11 (-0.17, -0.06)	0.13 (0.11, 0.14)	0.01 (-0.04, 0.07)
Mean INR	0.0002	0	0.0003 (-0.0000, 0.0006)	-0.27 (-0.32, -0.22)	0.31 (0.30, 0.33)	0.05 (-0.01, 0.10)
INR standard deviation	0.0008	0.0001	0.0008 (0.0003, 0.0014)	-0.07 (-0.12, -0.02)	0.20 (0.19, 0.22)	0.13 (0.08, 0.18)
Minimum INR value	0.0001	0	0.0001 (-0.0001, 0.0003)	0.01 (-0.04, 0.06)	0.02 (0.00, 0.03)	0.03 (-0.03, 0.08)
Maximum INR value	0.0007	0	0.0007 (0.0002, 0.0013)	-0.08 (-0.14, -0.03)	0.16 (0.15, 0.17)	0.08 (0.02, 0.13)
First INR value	0.0003	0	0.0003 (0.0001, 0.0006)	-0.08 (-0.13, -0.03)	0.09 (0.08, 0.11)	0.01 (-0.04, 0.07)
Last INR value	0.002	0.0002	0.0021 (0.0013, 0.0030)	-0.33 (-0.38, -0.28)	0.45 (0.44, 0.46)	0.12 (0.07, 0.17)
INR trajectory	0.0005	0	0.0006 (0.0002, 0.0010)	-0.54 (-0.59, -0.50)	0.57 (0.56, 0.59)	0.03 (-0.01, 0.08)

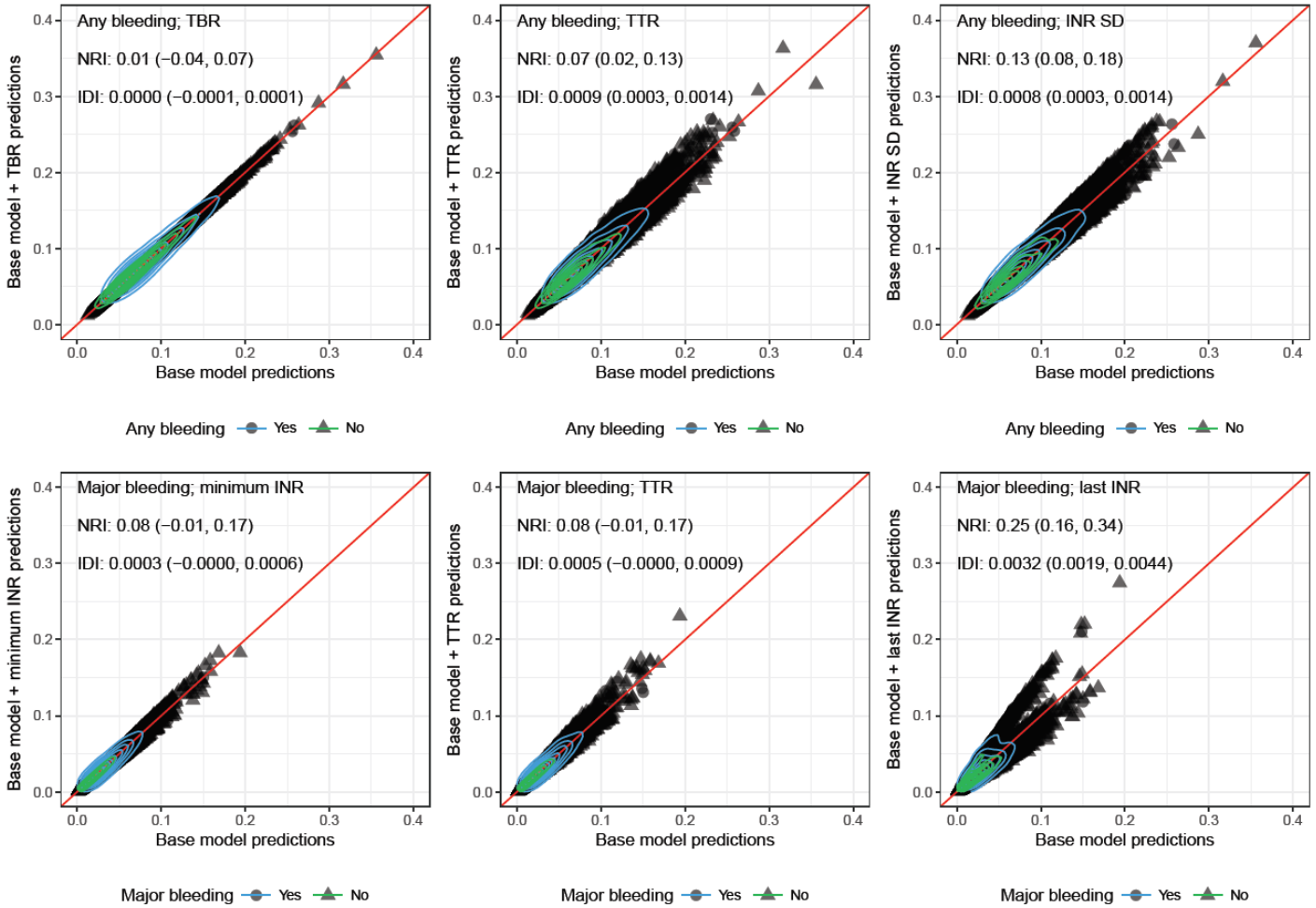
INR= international normalised ratio; IDI = Integrated discrimination improvement; NRI= net reclassification improvement; CI= confidence interval; [a] Mean probability increase for events patients; [b] Mean probability decrease for non-events patients; [c] proportion of events patients with increased prediction probability- proportion of events patients with decreased prediction probability; [d] proportion of non-events patients with decrease prediction probability- proportion of non-events patients with increased prediction probability

11.5.6 Major bleeding; integrated discrimination improvement and net reclassification improvement estimates

Measure of INR control	IDI – events [a]	IDI – non-events [b]	IDI (95% CI)	NRI - events (95% CI) [c]	NRI – non-events (95% CI) [d]	NRI (95% CI)
TTR group	0.0009	0	0.0010 (0.0003, 0.0017)	0.40 (0.32, 0.48)	-0.24 (-0.26, -0.23)	0.16 (0.07, 0.24)
Time in therapeutic range	0.0005	0	0.0005 (-0.0000, 0.0009)	-0.05 (-0.14, 0.04)	0.13 (0.11, 0.14)	0.08 (-0.01, 0.17)
Time above therapeutic range	0.0012	0	0.0012 (0.0004, 0.0020)	-0.03 (-0.12, 0.06)	0.19 (0.18, 0.21)	0.16 (0.07, 0.25)
Time below therapeutic range	0.0001	0	0.0001 (-0.0002, 0.0004)	0.00 (-0.09, 0.09)	0.08 (0.06, 0.09)	0.07 (-0.02, 0.16)
Mean INR	0.0003	0	0.0003 (-0.0001, 0.0008)	-0.14 (-0.23, -0.05)	0.23 (0.22, 0.24)	0.09 (0.00, 0.18)
INR standard deviation	0.0008	0	0.0008 (0.0001, 0.0016)	-0.03 (-0.12, 0.06)	0.22 (0.21, 0.24)	0.19 (0.10, 0.28)
Minimum INR value	0.0003	0	0.0003 (-0.0000, 0.0006)	0.04 (-0.05, 0.13)	0.04 (0.02, 0.05)	0.08 (-0.01, 0.17)
Maximum INR value	0.0014	0	0.0014 (0.0005, 0.0023)	0.06 (-0.03, 0.15)	0.06 (0.05, 0.08)	0.12 (0.03, 0.21)
First INR value	0.0005	0	0.0005 (0.0001, 0.0009)	-0.11 (-0.20, -0.02)	0.21 (0.19, 0.22)	0.10 (0.01, 0.19)
Last INR value	0.0031	0.0001	0.0032 (0.0019, 0.0044)	-0.03 (-0.12, 0.06)	0.28 (0.27, 0.29)	0.25 (0.16, 0.34)
INR trajectory	0.0006	0	0.0006 (0.0001, 0.0011)	-0.62 (-0.69, -0.54)	0.70 (0.69, 0.71)	0.08 (0.01, 0.16)

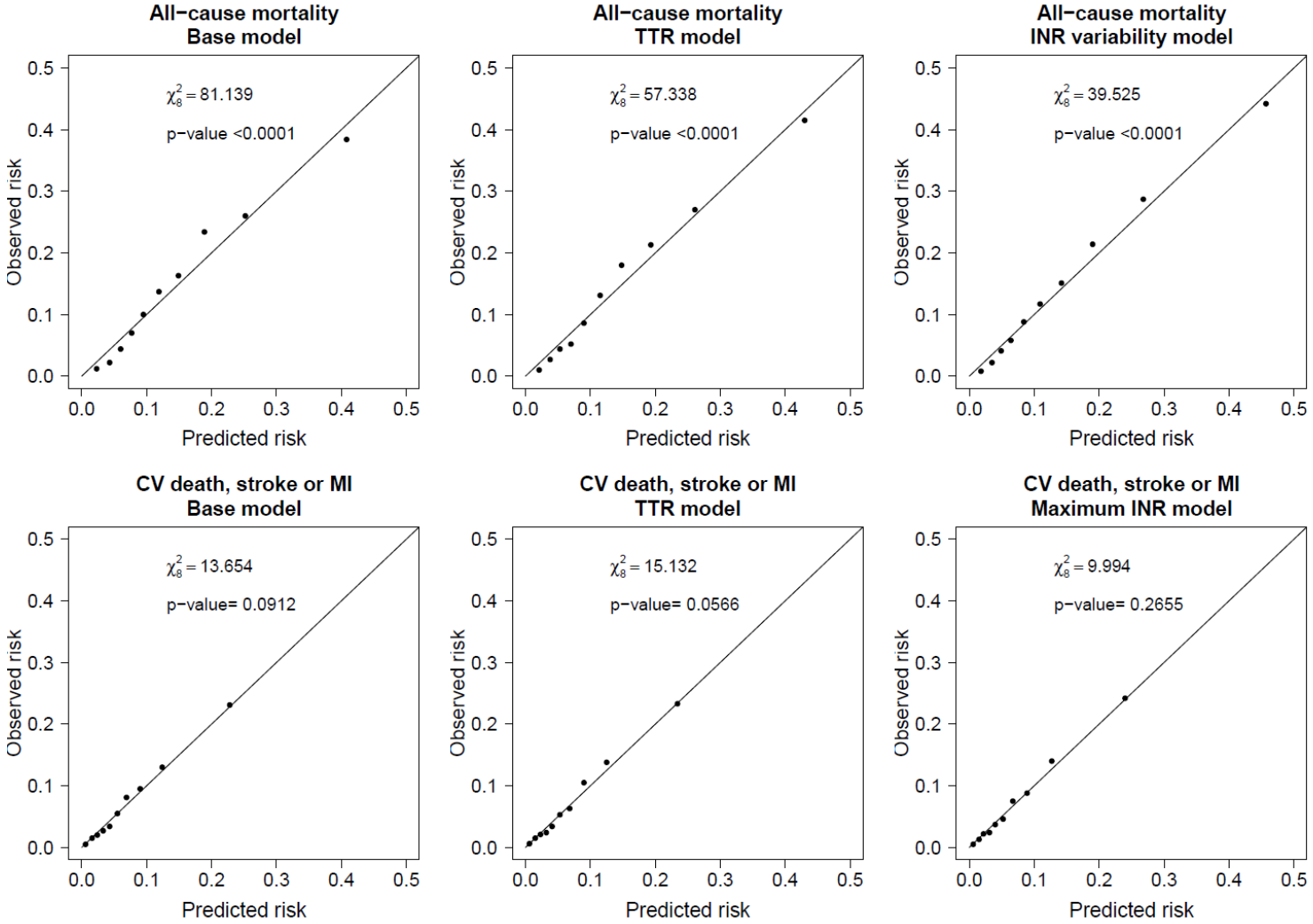
INR= international normalised ratio; IDI = Integrated discrimination improvement; NRI= net reclassification improvement; CI= confidence interval; [a] Mean probability increase for events patients; [b] Mean probability decrease for non-events patients; [c] proportion of events patients with increased prediction probability- proportion of events patients with decreased prediction probability; [d] proportion of non-events patients with decrease prediction probability- proportion of non-events patients with increased prediction probability

11.5.7 Predicted any bleeding and major bleeding; comparing base models with models including measures of INR control

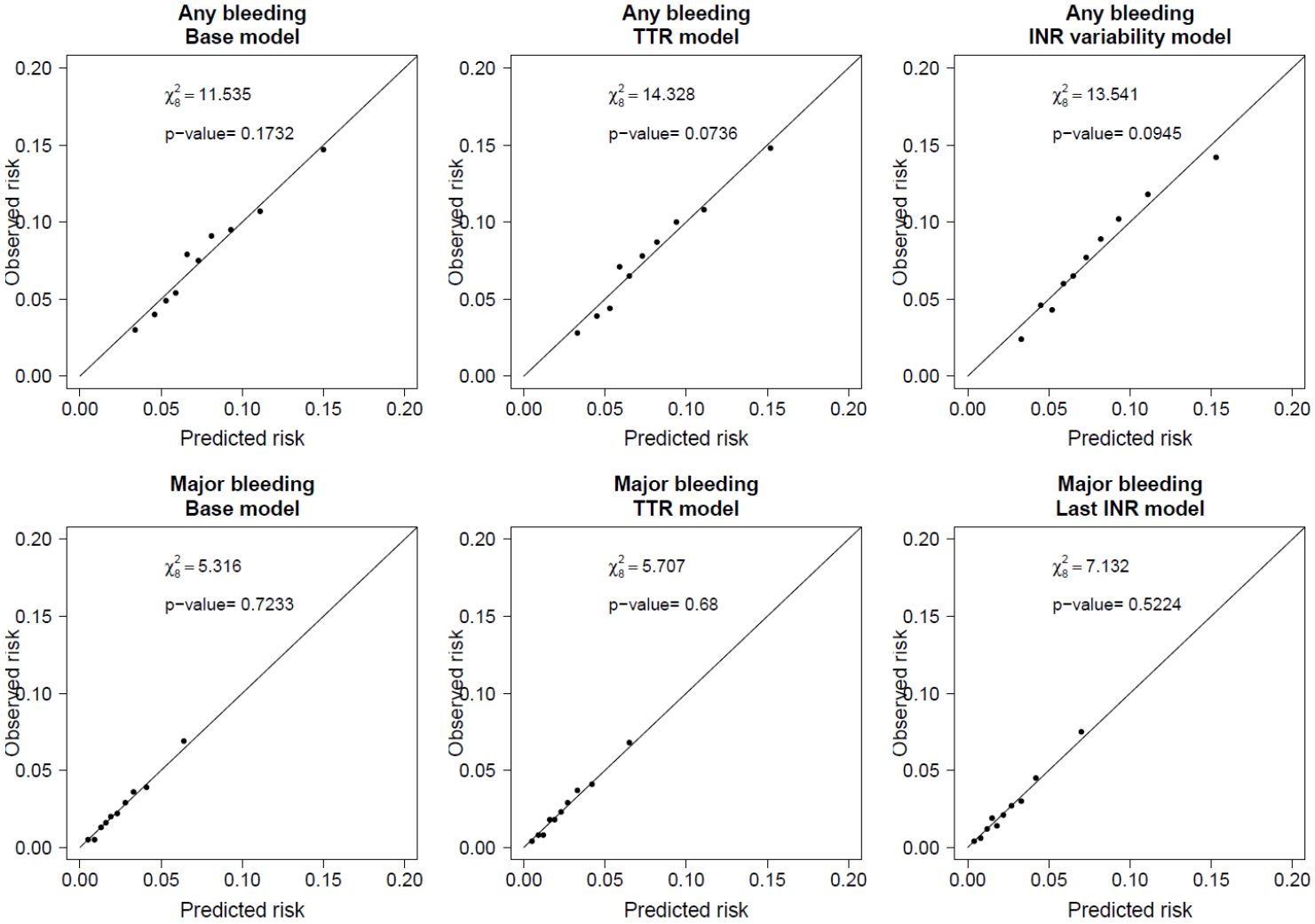


TBR= time below therapeutic range; NRI= net reclassification improvement; IDI= integrated discrimination improvement; TTR= time in therapeutic range; SD= standard deviation

11.5.8 Calibration of base models and models including measures of INR control for all-cause mortality and cardiovascular death, stroke or MI



11.5.9 Calibration of base models and models including measures of INR control for any bleeding and major bleeding



12 References

1. Bhatnagar P, Wickramasinghe K, Williams J, Rayner M, Townsend N. The epidemiology of cardiovascular disease in the UK 2014. *Heart.* 2015;101(15):1182-1189.
2. Fitzmaurice DA, Blann AD, Lip GY. Bleeding risks of antithrombotic therapy. *BMJ.* 2002;325(7368):828-831.
3. Mehran R, Rao SV, Bhatt DL*, et al.* Standardized Bleeding Definitions for Cardiovascular Clinical Trials: A Consensus Report From the Bleeding Academic Research Consortium. *Circulation.* 2011;123(23):2736-2747.
4. National Institute for Health and Care Excellence. Atrial fibrillation: the management of atrial fibrillation (Clinical guideline 180). 2014; http://guidance.nice.org.uk/CG180.
5. Kirchhof P, Benussi S, Kotecha D*, et al.* 2016 ESC Guidelines for the management of atrial fibrillation developed in collaboration with EACTS The Task Force for the management of atrial fibrillation of the European Society of Cardiology (ESC) Developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESCEndorsed by the European Stroke Organisation (ESO). *European journal of cardio-thoracic surgery.* 2016:ezw313.
6. Lip GY, Nieuwlaat R, Pisters R, Lane DA, Crijns HJ. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the euro heart survey on atrial fibrillation. *Chest.* 2010;137(2):263-272.
7. Pisters R, Lane DA, Nieuwlaat R*, et al.* A novel user-friendly score (HAS-BLED) to assess 1-year risk of major bleeding in patients with atrial fibrillation: the Euro Heart Survey. *Chest.* 2010;138(5):1093-1100.
8. Denaxas SC, George J, Herrett E*, et al.* Data resource profile: cardiovascular disease research using linked bespoke studies and electronic health records (CALIBER). *Int J Epidemiol.* 2012;41(6):1625-1638.
9. Roth GA, Forouzanfar MH, Moran AE*, et al.* Demographic and Epidemiologic Drivers of Global Cardiovascular Mortality. *New England Journal of Medicine.* 2015;372(14):1333-1341.
10. Bauer KA. Pros and cons of new oral anticoagulants. *ASH Education Program Book.* 2013;2013(1):464-470.
11. Connolly SJ, Ezekowitz MD, Yusuf S*, et al.* Dabigatran versus warfarin in patients with atrial fibrillation. *N Engl J Med.* 2009;361(12):1139-1151.
12. Mega JL, Braunwald E, Wiviott SD*, et al.* Rivaroxaban in Patients with a Recent Acute Coronary Syndrome. *New England Journal of Medicine.* 2012;366(1):9-19.
13. Patel MR, Mahaffey KW, Garg J*, et al.* Rivaroxaban versus Warfarin in Nonvalvular Atrial Fibrillation. *New England Journal of Medicine.* 2011;365(10):883-891.
14. Granger CB, Alexander JH, McMurray JJV*, et al.* Apixaban versus Warfarin in Patients with Atrial Fibrillation. *New England Journal of Medicine.* 2011;365(11):981-992.
15. Alexander JH, Lopes RD, James S*, et al.* Apixaban with Antiplatelet Therapy after Acute Coronary Syndrome. *New England Journal of Medicine.* 2011;365(8):699-708.
16. Connolly SJ, Eikelboom J, Joyner C*, et al.* Apixaban in Patients with Atrial Fibrillation. *New England Journal of Medicine.* 2011;364(9):806-817.
17. Bonaca MP, Bhatt DL, Cohen M*, et al.* Long-term use of ticagrelor in patients with prior myocardial infarction. *N Engl J Med.* 2015;372(19):1791-1800.
18. Montalescot G, Sechtem U, Achenbach S*, et al.* 2013 ESC guidelines on the management of stable coronary artery disease The Task Force on the management of stable coronary artery disease of the European Society of Cardiology. *Eur Heart J.* 2013;34(38):2949-3003.
19. Lip GY, Kakar P, Watson T. Atrial fibrillation—the growing epidemic. *Heart.* 2007;93(5):542.

20. Gage BF, van Walraven C, Pearce L*, et al.* Selecting Patients With Atrial Fibrillation for Anticoagulation: Stroke Risk Stratification in Patients Taking Aspirin. *Circulation.* 2004;110(16):2287-2292.
21. Camm AJ, Kirchhof P, Lip GY*, et al.* Guidelines for the management of atrial fibrillation The Task Force for the Management of Atrial Fibrillation of the European Society of Cardiology (ESC). *Eur Heart J.* 2010;31(19):2369-2429.
22. Camm AJ, Lip GYH, De Caterina R*, et al.* 2012 focused update of the ESC Guidelines for the management of atrial fibrillation. *European Heart Journal.* 2012;33(21):2719-2747.
23. Steg PG, James SK, Atar D*, et al.* ESC Guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation. *European Heart Journal.* 2012;33(20):2569-2619.
24. NICE. Myocardial infarction with ST-segment elevation: The acute management of myocardial infarction with ST-segment elevation; Nice clinical guideline 167. 2013; https://www.nice.org.uk/guidance/cg167.
25. Hamm CW, Bassand J-P, Agewall S*, et al.* ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation The Task Force for the management of acute coronary syndromes (ACS) in patients presenting without persistent ST-segment elevation of the European Society of Cardiology (ESC). *Eur Heart J.* 2011;32(23):2999-3054.
26. NICE. Unstable angina and NSTEMI: The early management of unstable angina and non-ST-segment-elevation myocardial infarction; Clinical guideline 94. 2010; https://www.nice.org.uk/guidance/cg94.
27. Subherwal S, Bach RG, Chen AY*, et al.* Baseline Risk of Major Bleeding in Non–ST-Segment–Elevation Myocardial Infarction: The CRUSADE (Can Rapid risk stratification of Unstable angina patients Suppress ADverse outcomes with Early implementation of the ACC/AHA guidelines) Bleeding Score. *Circulation.* 2009;119(14):1873-1882.
28. Mehran R, Pocock SJ, Nikolsky E*, et al.* A Risk Score to Predict Bleeding in Patients With Acute Coronary Syndromes. *Journal of the American College of Cardiology.* 2010;55(23):2556-2566.
29. NICE. Management of stable angina; NICE clinical guideline 126. 2011; https://www.nice.org.uk/guidance/cg126.
30. Bhatt DL, Flather MD, Hacke W*, et al.* Patients with prior myocardial infarction, stroke, or symptomatic peripheral arterial disease in the CHARISMA trial. *Journal of the American College of Cardiology.* 2007;49(19):1982-1988.
31. Tendera M, Aboyans V, Bartelink M-L*, et al.* ESC Guidelines on the diagnosis and treatment of peripheral artery diseases. *European Heart Journal.* 2011;32(22):2851-2906.
32. Perk J, De Backer G, Gohlke H*, et al.* European Guidelines on cardiovascular disease prevention in clinical practice (version 2012). *European Heart Journal.* 2012;33(13):1635-1701.
33. National Institute for Health and Care Excellence. Venous thromboembolic disease: diagnosis, management and thrombophilia testing (Clinical guideline 144). 2012; http://nice.org.uk/guidance/cg144.
34. Konstantinides SV, Torbicki A, Agnelli G*, et al.* 2014 ESC guidelines on the diagnosis and management of acute pulmonary embolism. *Eur Heart J.* 2014;35(43):3033-3069, 3069a-3069k.
35. 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.
36. 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.

37. García Rodríguez LA, Pérez Gutthann S. Use of the UK General Practice Research Database for pharmacoepidemiology. *Br J Clin Pharmacol.* 1998;45(5):419-425.
38. Morley KI, Wallace J, Denaxas SC*, et al.* Defining disease phenotypes using national linked electronic health records: a case study of atrial fibrillation. *PLOS ONE.* 2014;9(11):e110900.
39. Pujades-Rodriguez M, Timmis A, Stogiannis D*, et al.* Socioeconomic deprivation and the incidence of 12 cardiovascular diseases in 1.9 million women and men: implications for risk prediction and prevention. *PLOS ONE.* 2014;9(8):e104671.
40. 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(9932):1899-1911.
41. George J, Rapsomaniki E, Pujades-Rodriguez M*, et al.* How Does Cardiovascular Disease First Present in Women and Men? Incidence of 12 Cardiovascular Diseases in a Contemporary Cohort of 1,937,360 People. *Circulation.* 2015:CIRCULATIONAHA. 114.013797.
42. 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(2):105-113.
43. Pujades-Rodriguez M, Duyx B, Thomas SL*, et al.* Rheumatoid Arthritis and Incidence of Twelve Initial Presentations of Cardiovascular Disease: A Population Record-Linkage Cohort Study in England. *PLOS ONE.* 2016;11(3):e0151245.
44. Bell S, Daskalopoulou M, Rapsomaniki E*, et al.* Association between clinically recorded alcohol consumption and initial presentation of 12 cardiovascular diseases: population based cohort study using linked health records. *BMJ.* 2017;356:j909.
45. Shah AD, Denaxas S, Nicholas O, Hingorani AD, Hemingway H. Neutrophil Counts and Initial Presentation of 12 Cardiovascular Diseases: A CALIBER Cohort Study. *J Am Coll Cardiol.* 2017;69(9):1160-1169.
46. Schulman S, Beyth RJ, Kearon C, Levine MN. Hemorrhagic complications of anticoagulant and thrombolytic treatment: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). *Chest.* 2008;133(6 Suppl):257S-298S.
47. Rao AK, Pratt C, Berke A*, et al.* Thrombolysis in Myocardial Infarction (TIMI) Trial--phase I: hemorrhagic manifestations and changes in plasma fibrinogen and the fibrinolytic system in patients treated with recombinant tissue plasminogen activator and streptokinase. *J Am Coll Cardiol.* 1988;11(1):1-11.
48. Wiviott SD, Antman EM, Gibson CM*, et al.* Evaluation of prasugrel compared with clopidogrel in patients with acute coronary syndromes: design and rationale for the TRial to assess Improvement in Therapeutic Outcomes by optimizing platelet InhibitioN with prasugrel Thrombolysis In Myocardial Infarction 38 (TRITON-TIMI 38). *Am Heart J.* 2006;152(4):627-635.
49. Wiviott SD, Braunwald E, McCabe CH*, et al.* Prasugrel versus clopidogrel in patients with acute coronary syndromes. *N Engl J Med.* 2007;357(20):2001-2015.
50. An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction. *N Engl J Med.* 1993;329(10):673-682.
51. Stone GW, Bertrand M, Colombo A*, et al.* Acute Catheterization and Urgent Intervention Triage strategY (ACUITY) trial: study design and rationale. *Am Heart J.* 2004;148(5):764-775.
52. Stone GW, Witzenbichler B, Guagliumi G*, et al.* Bivalirudin during primary PCI in acute myocardial infarction. *N Engl J Med.* 2008;358(21):2218-2230.
53. Stone GW, McLaurin BT, Cox DA*, et al.* Bivalirudin for patients with acute coronary syndromes. *N Engl J Med.* 2006;355(21):2203-2216.

54. Yusuf S, Zhao F, Mehta SR*, et al.* Effects of clopidogrel in addition to aspirin in patients with acute coronary syndromes without ST-segment elevation. *N Engl J Med.* 2001;345(7):494-502.
55. Eikelboom JW, Mehta SR, Anand SS*, et al.* Adverse impact of bleeding on prognosis in patients with acute coronary syndromes. *Circulation.* 2006;114(8):774-782.
56. Mehta SR, Bassand JP, Chrolavicius S*, et al.* Design and rationale of CURRENT-OASIS 7: a randomized, 2 x 2 factorial trial evaluating optimal dosing strategies for clopidogrel and aspirin in patients with ST and non-ST-elevation acute coronary syndromes managed with an early invasive strategy. *Am Heart J.* 2008;156(6):1080-1088 e1081.
57. Mehta SR, Tanguay JF, Eikelboom JW*, et al.* Double-dose versus standard-dose clopidogrel and high-dose versus low-dose aspirin in individuals undergoing percutaneous coronary intervention for acute coronary syndromes (CURRENT-OASIS 7): a randomised factorial trial. *Lancet.* 2010;376(9748):1233-1243.
58. Montalescot G, White HD, Gallo R*, et al.* Enoxaparin versus unfractionated heparin in elective percutaneous coronary intervention. *N Engl J Med.* 2006;355(10):1006-1017.
59. James S, Akerblom A, Cannon CP*, et al.* Comparison of ticagrelor, the first reversible oral P2Y(12) receptor antagonist, with clopidogrel in patients with acute coronary syndromes: Rationale, design, and baseline characteristics of the PLATelet inhibition and patient Outcomes (PLATO) trial. *Am Heart J.* 2009;157(4):599-605.
60. Wallentin L, Becker RC, Budaj A*, et al.* Ticagrelor versus clopidogrel in patients with acute coronary syndromes. *N Engl J Med.* 2009;361(11):1045-1057.
61. Moscucci M, Fox KA, Cannon CP*, et al.* Predictors of major bleeding in acute coronary syndromes: the Global Registry of Acute Coronary Events (GRACE). *Eur Heart J.* 2003;24(20):1815-1823.
62. Rationale and design of the GRACE (Global Registry of Acute Coronary Events) Project: a multinational registry of patients hospitalized with acute coronary syndromes. *Am Heart J.* 2001;141(2):190-199.
63. Spencer FA, Moscucci M, Granger CB*, et al.* Does comorbidity account for the excess mortality in patients with major bleeding in acute myocardial infarction? *Circulation.* 2007;116(24):2793-2801.
64. Schulman S, Kearon C. Definition of major bleeding in clinical investigations of antihemostatic medicinal products in non-surgical patients. *J Thromb Haemost.* 2005;3(4):692-694.
65. Ndrepepa G, Schuster T, Hadamitzky M*, et al.* Validation of the Bleeding Academic Research Consortium Definition of Bleeding in Patients With Coronary Artery Disease Undergoing Percutaneous Coronary Intervention. *Circulation.* 2012;125(11):1424.
66. Hamon M, Lemesle G, Tricot O*, et al.* Incidence, source, determinants, and prognostic impact of major bleeding in outpatients with stable coronary artery disease. *J Am Coll Cardiol.* 2014;64(14):1430-1436.
67. Kikkert WJ, van Geloven N, van der Laan MH*, et al.* The prognostic value of bleeding academic research consortium (BARC)-defined bleeding complications in ST-segment elevation myocardial infarction: a comparison with the TIMI (Thrombolysis In Myocardial Infarction), GUSTO (Global Utilization of Streptokinase and Tissue Plasminogen Activator for Occluded Coronary Arteries), and ISTH (International Society on Thrombosis and Haemostasis) bleeding classifications. *J Am Coll Cardiol.* 2014;63(18):1866-1875.
68. Vranckx P, White HD, Huang Z*, et al.* Validation of BARC Bleeding Criteria in Patients With Acute Coronary Syndromes: The TRACER Trial. *J Am Coll Cardiol.* 2016;67(18):2135-2144.
69. Kaatz S, Ahmad D, Spyropoulos AC, Schulman S. Definition of clinically relevant non-major bleeding in studies of anticoagulants in atrial fibrillation and venous thromboembolic disease in non-surgical patients: communication from the SSC of the ISTH. *J Thromb Haemost.* 2015;13(11):2119-2126.

70. Raiford DS, Perez Gutthann S, Garcia Rodriguez LA. Positive predictive value of ICD-9 codes in the identification of cases of complicated peptic ulcer disease in the Saskatchewan hospital automated database. *Epidemiology.* 1996;7(1):101-104.
71. de Abajo FJ, Rodriguez LA, Montero D. Association between selective serotonin reuptake inhibitors and upper gastrointestinal bleeding: population based case-control study. *BMJ.* 1999;319(7217):1106-1109.
72. Wahl PM, Rodgers K, Schneeweiss S*, et al.* Validation of claims-based diagnostic and procedure codes for cardiovascular and gastrointestinal serious adverse events in a commercially-insured population. *Pharmacoepidemiol Drug Saf.* 2010;19(6):596-603.
73. Crooks CJ, Card TR, West J. Defining upper gastrointestinal bleeding from linked primary and secondary care data and the effect on occurrence and 28 day mortality. *BMC Health Serv Res.* 2012;12(1):392.
74. Valkhoff VE, Coloma PM, Masclee GM*, et al.* Validation study in four health-care databases: upper gastrointestinal bleeding misclassification affects precision but not magnitude of drug-related upper gastrointestinal bleeding risk. *J Clin Epidemiol.* 2014;67(8):921-931.
75. Arnason T, Wells PS, van Walraven C, Forster AJ. Accuracy of coding for possible warfarin complications in hospital discharge abstracts. *Thromb Res.* 2006;118(2):253-262.
76. Cunningham A, Stein CM, Chung CP*, et al.* An automated database case definition for serious bleeding related to oral anticoagulant use. *Pharmacoepidemiol Drug Saf.* 2011;20(6):560-566.
77. Friberg L, Skeppholm M. Usefulness of Health Registers for detection of bleeding events in outcome studies. *Thromb Haemost.* 2016;116(6):1131-1139.
78. Steffensen FH, Kristensen K, Ejlersen E, Dahlerup JF, Sorensen HT. Major haemorrhagic complications during oral anticoagulant therapy in a Danish population-based cohort. *J Intern Med.* 1997;242(6):497-503.
79. Hollowell J, Ruigomez A, Johansson S, Wallander MA, Garcia-Rodriguez LA. The incidence of bleeding complications associated with warfarin treatment in general practice in the United Kingdom. *Br J Gen Pract.* 2003;53(489):312-314.
80. Toyoda K, Yasaka M, Iwade K*, et al.* Dual antithrombotic therapy increases severe bleeding events in patients with stroke and cardiovascular disease: a prospective, multicenter, observational study. *Stroke.* 2008;39(6):1740-1745.
81. Hansen ML, Sorensen R, Clausen MT*, et al.* Risk of bleeding with single, dual, or triple therapy with warfarin, aspirin, and clopidogrel in patients with atrial fibrillation. *Arch Intern Med.* 2010;170(16):1433-1441.
82. Lamberts M, Olesen JB, Ruwald MH*, et al.* Bleeding after initiation of multiple antithrombotic drugs, including triple therapy, in atrial fibrillation patients following myocardial infarction and coronary intervention: a nationwide cohort study. *Circulation.* 2012;126(10):1185-1193.
83. Gallagher AM, van Staa TP, Murray-Thomas T*, et al.* Population-based cohort study of warfarin-treated patients with atrial fibrillation: incidence of cardiovascular and bleeding outcomes. *BMJ Open.* 2014;4(1):e003839.
84. Lovelock CE, Molyneux AJ, Rothwell PM. Change in incidence and aetiology of intracerebral haemorrhage in Oxfordshire, UK, between 1981 and 2006: a population-based study. *The Lancet Neurology.* 2007;6(6):487-493.
85. Hreinsson JP, Kalaitzakis E, Gudmundsson S, Bjornsson ES. Upper gastrointestinal bleeding: incidence, etiology and outcomes in a population-based setting. *Scand J Gastroenterol.* 2013;48(4):439-447.
86. Hallas J, Dall M, Andries A*, et al.* Use of single and combined antithrombotic therapy and risk of serious upper gastrointestinal bleeding: population based case-control study. *BMJ.* 2006;333(7571):726.

87. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ.* 2009;338:b375.
88. Royston P, Moons KGM, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. *BMJ.* 2009;338.
89. Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. *BMC Med Res Methodol.* 2013;13(1):33.
90. 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.
91. 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.
92. Hippisley-Cox J, Coupland C, Vinogradova Y*, et al.* Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. *BMJ*2007.
93. Hippisley-Cox J, Coupland C. Predicting risk of upper gastrointestinal bleed and intracranial bleed with anticoagulants: cohort study to derive and validate the QBleed scores. *BMJ.* 2014;349:g4606.
94. Hippisley-Cox J, Coupland C, Robson J, Sheikh A, Brindle P. Predicting risk of type 2 diabetes in England and Wales: prospective derivation and validation of QDScore. *BMJ.* 2009;338:b880.
95. Hippisley-Cox J, Coupland C, Brindle P. Derivation and validation of QStroke score for predicting risk of ischaemic stroke in primary care and comparison with other risk scores: a prospective open cohort study. *BMJ.* 2013;346:f2573.
96. Hippisley-Cox J, Coupland C. Predicting the risk of chronic Kidney Disease in men and women in England and Wales: prospective derivation and external validation of the QKidney Scores. *BMC Fam Pract.* 2010;11(1):49.
97. Ducrocq G, Wallace JS, Baron G*, et al.* Risk score to predict serious bleeding in stable outpatients with or at risk of atherothrombosis. *Eur Heart J.* 2010;31(10):1257-1265.
98. Gage BF, Yan Y, Milligan PE*, et al.* Clinical classification schemes for predicting hemorrhage: Results from the National Registry of Atrial Fibrillation (NRAF). *Am Heart J.* 2006;151(3):713-719.
99. Pocock SJ, Stone GW, Mehran R, Clayton TC. Individualizing treatment choices using quantitative methods. *Am Heart J.* 2014;168(5):607-610.
100. National Institute for Health and Care Excellence. Cardiovascular disease: risk assessment and reduction, including lipid modification (Clinical guideline 181). 2014; https://www.nice.org.uk/guidance/cg181.
101. Hemingway H, Croft P, Perel P*, et al.* Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. *BMJ.* 2013;346:e5595.
102. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ.* 2015;350:g7594.
103. De Caterina R, Husted S, Wallentin L*, et al.* Vitamin K antagonists in heart disease: current status and perspectives (Section III). Position paper of the ESC Working Group on Thrombosis--Task Force on Anticoagulants in Heart Disease. *Thromb Haemost.* 2013;110(6):1087-1107.
104. Keeling D, Baglin T, Tait C*, et al.* Guidelines on oral anticoagulation with warfarin - fourth edition. *Br J Haematol.* 2011;154(3):311-324.
105. Rosendaal FR, Cannegieter SC, van der Meer FJ, Briet E. A method to determine the optimal intensity of oral anticoagulant therapy. *Thromb Haemost.* 1993;69(3):236-239.
106. Kearon C, Akl EA, Ornelas J*, et al.* Antithrombotic Therapy for VTE Disease: CHEST Guideline and Expert Panel Report. *Chest.* 2016;149(2):315-352.

107. Nishimura RA, Otto CM, Bonow RO*, et al.* 2014 AHA/ACC Guideline for the Management of Patients With Valvular Heart Disease: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation.* 2014;129(23):e521-643.
108. Vahanian A, Alfieri O, Andreotti F*, et al.* Guidelines on the management of valvular heart disease (version 2012). *Eur Heart J.* 2012;33(19):2451-2496.
109. Macedo AF, Bell J, McCarron C*, et al.* Determinants of oral anticoagulation control in new warfarin patients: analysis using data from Clinical Practice Research Datalink. *Thromb Res.* 2015;136(2):250-260.
110. Rose AJ, Hylek EM, Ozonoff A*, et al.* Patient characteristics associated with oral anticoagulation control: results of the Veterans AffaiRs Study to Improve Anticoagulation (VARIA). *J Thromb Haemost.* 2010;8(10):2182-2191.
111. Witt DM, Delate T, Clark NP*, et al.* Outcomes and predictors of very stable INR control during chronic anticoagulation therapy. *Blood.* 2009;114(5):952-956.
112. Witt DM, Delate T, Clark NP*, et al.* Twelve-month outcomes and predictors of very stable INR control in prevalent warfarin users. *J Thromb Haemost.* 2010;8(4):744-749.
113. Wieloch M, Sjalander A, Frykman V*, et al.* Anticoagulation control in Sweden: reports of time in therapeutic range, major bleeding, and thrombo-embolic complications from the national quality registry AuriculA. *Eur Heart J.* 2011;32(18):2282-2289.
114. Dlott JS, George RA, Huang X*, et al.* National assessment of warfarin anticoagulation therapy for stroke prevention in atrial fibrillation. *Circulation.* 2014;129(13):1407-1414.
115. Nelson WW, Desai S, Damaraju CV*, et al.* International normalized ratio stability in warfarin-experienced patients with nonvalvular atrial fibrillation. *Am J Cardiovasc Drugs.* 2015;15(3):205-211.
116. Singer DE, Hellkamp AS, Piccini JP*, et al.* Impact of global geographic region on time in therapeutic range on warfarin anticoagulant therapy: data from the ROCKET AF clinical trial. *J Am Heart Assoc.* 2013;2(1):e000067.
117. Apostolakis S, Sullivan RM, Olshansky B, Lip GY. Factors affecting quality of anticoagulation control among patients with atrial fibrillation on warfarin: the SAMe-TT(2)R(2) score. *Chest.* 2013;144(5):1555-1563.
118. Pokorney SD, Simon DN, Thomas L*, et al.* Patients' time in therapeutic range on warfarin among US patients with atrial fibrillation: Results from ORBIT-AF registry. *Am Heart J.* 2015;170(1):141-148, 148 e141.
119. Currie CJ, McEwan P, Emmas C, Morgan CL, Peters JR. Anticoagulation in patients with non-valvular atrial fibrillation: an evaluation of stability and early factors that predict longer-term stability on warfarin in a large UK population. *Curr Med Res Opin.* 2005;21(12):1905-1913.
120. Melamed OC, Horowitz G, Elhayany A, Vinker S. Quality of anticoagulation control among patients with atrial fibrillation. *Am J Manag Care.* 2011;17(3):232-237.
121. Atas H, Sahin AA, Barutcu Atas D*, et al.* Potential Causes and Implications of Low Target Therapeutic Ratio in Warfarin-Treated Patients for Thrombosis Prophylaxis. *Clin Appl Thromb Hemost.* 2017;0(0):1076029617695484.
122. Boulanger L, Kim J, Friedman M*, et al.* Patterns of use of antithrombotic therapy and quality of anticoagulation among patients with non-valvular atrial fibrillation in clinical practice. *Int J Clin Pract.* 2006;60(3):258-264.
123. Nelson WW, Choi JC, Vanderpoel J*, et al.* Impact of co-morbidities and patient characteristics on international normalized ratio control over time in patients with nonvalvular atrial fibrillation. *American Journal of Cardiology.* 2013;112(4):509-512.
124. Szummer K, Gasparini A, Eliasson S*, et al.* Time in Therapeutic Range and Outcomes After Warfarin Initiation in Newly Diagnosed Atrial Fibrillation Patients With Renal Dysfunction. *J Am Heart Assoc.* 2017;6(3).

125. Penning-van Beest FJA, Geleijnse JM, van Meegen E*, et al.* Lifestyle and diet as risk factors for overanticoagulation. *Journal of Clinical Epidemiology.* 2002;55(4):411-417.
126. Tomita H, Kadokami T, Momii H*, et al.* Patient factors against stable control of warfarin therapy for Japanese non-valvular atrial fibrillation patients. *Thromb Res.* 2013;132(5):537-542.
127. Ciurus T, Cichocka-Radwan A, Lelonek M. Factors affecting the quality of anticoagulation with warfarin: experience of one cardiac centre. *Kardiochir Torakochirurgia Pol.* 2015;12(4):334-340.
128. Barrios V, Escobar C, Prieto L*, et al.* Anticoagulation Control in Patients With Nonvalvular Atrial Fibrillation Attended at Primary Care Centers in Spain: The PAULA Study. *Rev Esp Cardiol (Engl Ed).* 2015;68(9):769-776.
129. Anguita Sanchez M, Bertomeu Martinez V, Cequier Fillat A. Quality of Vitamin K Antagonist Anticoagulation in Spain: Prevalence of Poor Control and Associated Factors. *Rev Esp Cardiol (Engl Ed).* 2015;68(9):761-768.
130. Wypasek E, Mazur P, Bochenek M*, et al.* Factors influencing quality of anticoagulation control and warfarin dosage in patients after aortic valve replacement within the 3 months of follow up. *J Physiol Pharmacol.* 2016;67(3):385-393.
131. Mohammed S, Aljundi AH, Kasem M, Alhashemi M, El-Menyar A. Anticoagulation control among patients with nonvalvular atrial fibrillation: A single tertiary cardiac center experience. *J Adv Pharm Technol Res.* 2017;8(1):14-18.
132. Kooistra HA, Gebel M, Sahin K, Lensing AW, Meijer K. Independent predictors of poor vitamin K antagonist control in venous thromboembolism patients. Data from the EINSTEIN-DVT and PE studies. *Thromb Haemost.* 2015;114(6):1136-1143.
133. Proietti M, Lane DA, Lip GY. Chronic Kidney Disease, Time in Therapeutic Range and Adverse Clinical Outcomes in Anticoagulated Patients with Non-valvular Atrial Fibrillation: Observations from the SPORTIF Trials. *EBioMedicine.* 2016;8:309-316.
134. Ansell J, Hollowell J, Pengo V*, et al.* Descriptive analysis of the process and quality of oral anticoagulation management in real-life practice in patients with chronic non-valvular atrial fibrillation: the international study of anticoagulation management (ISAM). *J Thromb Thrombolysis.* 2007;23(2):83-91.
135. Gotsman I, Ezra O, Hirsh-Rokach B*, et al.* A Patient-Specific Tailored Intervention Improves INR Time in Therapeutic Range and INR Variability in Heart Failure Patients. *Am J Med.* 2017;10.1016/j.amjmed.2017.02.030.
136. Morgan CL, McEwan P, Tukiendorf A*, et al.* Warfarin treatment in patients with atrial fibrillation: observing outcomes associated with varying levels of INR control. *Thromb Res.* 2009;124(1):37-41.
137. Lehto M, Niiranen J, Korhonen P*, et al.* Quality of warfarin therapy and risk of stroke, bleeding, and mortality among patients with atrial fibrillation: results from the nationwide FinWAF Registry. *Pharmacoepidemiol Drug Saf.* 2017;10.1002/pds.4194.
138. Oden A, Fahlen M. Oral anticoagulation and risk of death: a medical record linkage study. *BMJ.* 2002;325(7372):1073-1075.
139. Jones M, McEwan P, Morgan CL*, et al.* Evaluation of the pattern of treatment, level of anticoagulation control, and outcome of treatment with warfarin in patients with non-valvar atrial fibrillation: a record linkage study in a large British population. *Heart.* 2005;91(4):472-477.
140. van Leeuwen Y, Rosendaal FR, Cannegieter SC. Prediction of hemorrhagic and thrombotic events in patients with mechanical heart valve prostheses treated with oral anticoagulants. *J Thromb Haemost.* 2008;6(3):451-456.
141. Lind M, Fahlen M, Kosiborod M, Eliasson B, Oden A. Variability of INR and its relationship with mortality, stroke, bleeding and hospitalisations in patients with atrial fibrillation. *Thromb Res.* 2012;129(1):32-35.

142. Ibrahim S, Jespersen J, Poller L. The clinical evaluation of International Normalized Ratio variability and control in conventional oral anticoagulant administration by use of the variance growth rate. *J Thromb Haemost.* 2013;11(8):1540-1546.
143. Razouki Z, Ozonoff A, Zhao S, Jasuja GK, Rose AJ. Improving quality measurement for anticoagulation: adding international normalized ratio variability to percent time in therapeutic range. *Circ Cardiovasc Qual Outcomes.* 2014;7(5):664-669.
144. Rose AJ, Delate T, Ozonoff A, Witt DM. Comparison of the Abilities of Summary Measures of International Normalized Ratio Control to Predict Clinically Relevant Bleeding. *Circ Cardiovasc Qual Outcomes.* 2015;8(5):524-531.
145. Van Den Ham HA, Klungel OH, Leufkens HGM, Van Staa TP. The patterns of anticoagulation control and the risk of stroke, bleeding and mortality in patients with non-valvular atrial fibrillation. *Journal of thrombosis and haemostasis.* 2013;11(1):107-115.
146. Proietti M, Senoo K, Lane DA, Lip GYH. Major Bleeding in Patients with Non-Valvular Atrial Fibrillation: Impact of Time in Therapeutic Range on Contemporary Bleeding Risk Scores. *Scientific Reports.* 2016;6:24376.
147. January CT, Wann LS, Alpert JS*, et al.* 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation. *Circulation.* 2014;130(23):e199-e267.
148. Walker GA, Heidenreich PA, Phibbs CS*, et al.* Mental Illness and Warfarin Use in Atrial Fibrillation. *American Journal of Managed Care.* 2011;17(9):617-624.
149. Rodriguez F, Hong C, Chang Y*, et al.* Limited English proficient patients and time spent in therapeutic range in a warfarin anticoagulation clinic. *J Am Heart Assoc.* 2013;2(4):e000170.
150. Razouki Z, Ozonoff A, Zhao S, Rose AJ. Pathways to poor anticoagulation control. *J Thromb Haemost.* 2014;12(5):628-634.
151. Yong C, Azarbal F, Abnousi F*, et al.* Racial Differences in Quality of Anticoagulation Therapy for Atrial Fibrillation (from the TREAT-AF Study). *American Journal of Cardiology.* 2016;117(1):61-68.
152. Grzymala-Lubanski B, Svensson PJ, Renlund H, Jeppsson A, Sjalander A. Warfarin treatment quality and prognosis in patients with mechanical heart valve prosthesis. *Heart.* 2017;103(3):198-203.
153. Hellyer JA, Azarbal F, Than CT*, et al.* Impact of Baseline Stroke Risk and Bleeding Risk on Warfarin International Normalized Ratio Control in Atrial Fibrillation (from the TREAT-AF Study). *American Journal of Cardiology.* 2017;119(2):268-274.
154. Williams B, Evans M, Honushefsky A, Berger P. A Clinical Prediction Model for Time in Therapeutic Range While on Warfarin in Newly Diagnosed Atrial Fibrillation. *Journal of the American College of Cardiology.* 2017;69(11):549.
155. Veeger NJ, Piersma-Wichers M, Tijssen JG, Hillege HL, van der Meer J. Individual time within target range in patients treated with vitamin K antagonists: main determinant of quality of anticoagulation and predictor of clinical outcome. A retrospective study of 2300 consecutive patients with venous thromboembolism. *Br J Haematol.* 2005;128(4):513-519.
156. Burton C, Isles C, Norrie J, Hanson R, Grubb E. The safety and adequacy of antithrombotic therapy for atrial fibrillation: a regional cohort study. *Br J Gen Pract.* 2006;56(530):697-702.
157. Cavallari LH, Aston JL, Momary KM*, et al.* Predictors of unstable anticoagulation in African Americans. *J Thromb Thrombolysis.* 2009;27(4):430-437.
158. Okumura K, Komatsu T, Yamashita T*, et al.* Time in the therapeutic range during warfarin therapy in Japanese patients with non-valvular atrial fibrillation. - A multicenter study of its status and infuential factors. *Circ J.* 2011;75(9):2087-2094.
159. Costa GL, Lamego RM, Colosimo EA, Valacio RA, Moreira Mda C. Identifying potential predictors of high-quality oral anticoagulation assessed by time in therapeutic

international normalized ratio range: a prospective, long-term, single-center, observational study. *Clin Ther.* 2012;34(7):1511-1520.
160. Han SY, Palmeri ST, Broderick SH*, et al.* Quality of anticoagulation with warfarin in patients with nonvalvular atrial fibrillation in the community setting. *J Electrocardiol.* 2013;46(1):45-50.
161. Bishop MA, Streiff MB. Effects of anticoagulation provider continuity on time in therapeutic range for warfarin patients. *J Thromb Thrombolysis.* 2016;42(2):283-287.
162. Alyousif SM, Alsaileek AA. Quality of anticoagulation control among patients with atrial fibrillation: An experience of a tertiary care center in Saudi Arabia. *J Saudi Heart Assoc.* 2016;28(4):239-243.
163. Rombouts EK, Rosendaal FR, Van Der Meer FJ. Daily vitamin K supplementation improves anticoagulant stability. *J Thromb Haemost.* 2007;5(10):2043-2048.
164. Sconce E, Avery P, Wynne H, Kamali F. Vitamin K supplementation can improve stability of anticoagulation for patients with unexplained variability in response to warfarin. *Blood.* 2007;109(6):2419-2423.
165. Clarkesmith DE, Pattison HM, Lip GY, Lane DA. Educational intervention improves anticoagulation control in atrial fibrillation patients: the TREAT randomised trial. *PLOS ONE.* 2013;8(9):e74037.
166. Hylek EM, Chang Y, Skates SJ, Hughes RA, Singer DE. Prospective study of the outcomes of ambulatory patients with excessive warfarin anticoagulation. *Archives of Internal Medicine.* 2000;160(11):1612-1617.
167. Freixa R, Blanch P, Ibernon M*, et al.* [Identification of factors responsible for oral over-anticoagulation in outpatients with heart disease]. *Rev Esp Cardiol.* 2003;56(1):65-72.
168. Wittkowsky AK, Devine EB. Frequency and causes of overanticoagulation and underanticoagulation in patients treated with warfarin. *Pharmacotherapy.* 2004;24(10):1311-1316.
169. Clark NP, Witt DM, Delate T*, et al.* Thromboembolic consequences of subtherapeutic anticoagulation in patients stabilized on warfarin therapy: the low INR study. *Pharmacotherapy.* 2008;28(8):960-967.
170. Wallentin L, Yusuf S, Ezekowitz MD*, et al.* Efficacy and safety of dabigatran compared with warfarin at different levels of international normalised ratio control for stroke prevention in atrial fibrillation: an analysis of the RE-LY trial. *Lancet.* 2010;376(9745):975-983.
171. Gallagher AM, Setakis E, Plumb JM, Clemens A, van Staa TP. Risks of stroke and mortality associated with suboptimal anticoagulation in atrial fibrillation patients. *Thromb Haemost.* 2011;106(5):968-977.
172. Gallagher AM, Puri S, Staa TV. Linkage of the General Practice Research Database (GPRD) with other data sources. Pharmacoepidemiol Drug Saf; 2011;20(S1):230-364;http://dx.doi.org/10.1002/pds.2206.
173. Mathur R, Bhaskaran K, Chaturvedi N*, et al.* Completeness and usability of ethnicity data in UK-based primary care and hospital databases. *J Public Health (Oxf).* 2014;36(4):684-692.
174. Herrett E, Gallagher AM, Bhaskaran K*, et al.* Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol.* 2015;44(3):827-836.
175. Herrett E, Smeeth L, Walker L, Weston C, Group obotMA. The Myocardial Ischaemia National Audit Project (MINAP). *Heart.* 2010;96(16):1264-1267.
176. Lip GY, Laroche C, Dan GA*, et al.* 'Real-world' antithrombotic treatment in atrial fibrillation: The EORP-AF pilot survey. *Am J Med.* 2014;127(6):519-529 e511.
177. Budaj A, Brieger D, Steg PG*, et al.* Global patterns of use of antithrombotic and antiplatelet therapies in patients with acute coronary syndromes: insights from the Global Registry of Acute Coronary Events (GRACE). *Am Heart J.* 2003;146(6):999-1006.
178. Filippi A, Sessa E, Trifirò G*, et al.* Oral anticoagulant therapy in Italy: prescribing prevalence and clinical reasons. *Pharmacological Research.* 2004;50(6):601-603.

179. van Staa TP, Leufkens HG, Abenhaim L*, et al.* Use of oral corticosteroids in the United Kingdom. *QJM.* 2000;93(2):105-111.
180. Petersen I, Hayward AC. Antibacterial prescribing in primary care. *J Antimicrob Chemother.* Vol 60 Suppl 1. England2007:i43-47.
181. Iwagami M, Tomlinson LA, Mansfield KE*, et al.* Prevalence, incidence, indication, and choice of antidepressants in patients with and without chronic kidney disease: a matched cohort study in UK Clinical Practice Research Datalink. *Pharmacoepidemiol Drug Saf.* 2017;26(7):792-801.
182. Gardarsdottir H, Heerdink ER, van Dijk L, Egberts AC. Indications for antidepressant drug prescribing in general practice in the Netherlands. *J Affect Disord.* Vol 98. Netherlands2007:109-115.
183. Marston L, Nazareth I, Petersen I, Walters K, Osborn DP. Prescribing of antipsychotics in UK primary care: a cohort study. *BMJ Open.* 2014;4(12):e006135.
184. Rapsomaniki E, Shah A, Perel P*, et al.* Prognostic models for stable coronary artery disease based on electronic health record cohort of 102 023 patients. *Eur Heart J.* 2014;35(13):844-852.
185. Pujades-Rodriguez M, George J, Shah AD*, et al.* Heterogeneous associations between smoking and a wide range of initial presentations of cardiovascular disease in 1 937 360 people in England: lifetime risks and implications for risk prediction. *Int J Epidemiol.* 2015;44(1):129-141.
186. Pasea L, Chung SC, Pujades-Rodriguez M*, et al.* Personalising the decision for prolonged dual antiplatelet therapy: development, validation and potential impact of prognostic models for cardiovascular events and bleeding in myocardial infarction survivors. *Eur Heart J.* 2017;38(14):1048-1055.
187. Timmis A, Rapsomaniki E, Chung SC*, et al.* Prolonged dual antiplatelet therapy in stable coronary disease: comparative observational study of benefits and harms in unselected versus trial populations. *BMJ.* 2016;353:i3163.
188. Trialists'Collaboration A. Collaborative meta-analysis of randomised trials of antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients. *BMJ.* 2002;324(7329):71-86.
189. Patrono C, Andreotti F, Arnesen H*, et al.* Antiplatelet agents for the treatment and prevention of atherothrombosis. *Eur Heart J.* 2011;32(23):2922-2932.
190. You JJ, Singer DE, Howard PA*, et al.* Antithrombotic therapy for atrial fibrillation: Antithrombotic therapy and prevention of thrombosis, 9th ed: american college of chest physicians evidence-based clinical practice guidelines. *CHEST Journal.* 2012;141(2_suppl):e531S-e575S.
191. Roffi M, Patrono C, Collet JP*, et al.* 2015 ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation: Task Force for the Management of Acute Coronary Syndromes in Patients Presenting without Persistent ST-Segment Elevation of the European Society of Cardiology (ESC). *Eur Heart J.* 2016;37(3):267-315.
192. Shah AD, Nicholas O, Timmis AD*, et al.* Threshold haemoglobin levels and the prognosis of stable coronary disease: two new cohorts and a systematic review and meta-analysis. *PLoS Med.* 2011;8(5):e1000439.
193. Steyerberg EW, Moons KG, van der Windt DA*, et al.* Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med.* 2013;10(2):e1001381.
194. Levine GN, Bates ER, Bittl JA*, et al.* 2016 ACC/AHA Guideline Focused Update on Duration of Dual Antiplatelet Therapy in Patients With Coronary Artery Disease: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines: An Update of the 2011 ACCF/AHA/SCAI Guideline for Percutaneous Coronary Intervention, 2011 ACCF/AHA Guideline for Coronary Artery Bypass Graft Surgery, 2012 ACC/AHA/ACP/AATS/PCNA/SCAI/STS Guideline for the Diagnosis and Management of Patients With Stable Ischemic Heart Disease, 2013

ACCF/AHA Guideline for the Management of ST-Elevation Myocardial Infarction, 2014 AHA/ACC Guideline for the Management of Patients With Non-ST-Elevation Acute Coronary Syndromes, and 2014 ACC/AHA Guideline on Perioperative Cardiovascular Evaluation and Management of Patients Undergoing Noncardiac Surgery. *Circulation.* 2016;134(10):e123-155.

195. Sorensen R, Hansen ML, Abildstrom SZ*, et al.* Risk of bleeding in patients with acute myocardial infarction treated with different combinations of aspirin, clopidogrel, and vitamin K antagonists in Denmark: a retrospective analysis of nationwide registry data. *Lancet.* 2009;374(9706):1967-1974.
196. Hylek EM, Evans-Molina C, Shea C, Henault LE, Regan S. Major hemorrhage and tolerability of warfarin in the first year of therapy among elderly patients with atrial fibrillation. *Circulation.* 2007;115(21):2689-2696.
197. Janssen KJ, Vergouwe Y, Donders AR*, et al.* Dealing with missing predictor values when applying clinical prediction models. *Clin Chem.* 2009;55(5):994-1001.
198. Jernberg T, Hasvold P, Henriksson M*, et al.* Cardiovascular risk in post-myocardial infarction patients: nationwide real world data demonstrate the importance of a long-term perspective. *Eur Heart J.* 2015;36(19):1163-1170.
199. Rapsomaniki E, Thuresson M, Yang E*, et al.* Using big data from health records from four countries to evaluate chronic disease outcomes: a study in 114 364 patients after myocardial infarction. *Eur Heart J Qual Care Clin Outcomes.* 2016;http://dx.doi/10.1093/ehjqcco/qcw004.
200. Mauri L, Kereiakes DJ, Yeh RW*, et al.* Twelve or 30 months of dual antiplatelet therapy after drug-eluting stents. *N Engl J Med.* 2014;371(23):2155-2166.
201. Keaney JF, Jr. Balancing the risks and benefits of dual platelet inhibition. *N Engl J Med.* 2015;372(19):1854-1856.
202. Pocock SJ, Mehran R, Clayton TC*, et al.* Prognostic modeling of individual patient risk and mortality impact of ischemic and hemorrhagic complications: assessment from the Acute Catheterization and Urgent Intervention Triage Strategy trial. *Circulation.* 2010;121(1):43-51.
203. Salisbury AC, Wang K, Cohen DJ*, et al.* Selecting antiplatelet therapy at the time of percutaneous intervention for an acute coronary syndrome: weighing the benefits and risks of prasugrel versus clopidogrel. *Circ Cardiovasc Qual Outcomes.* 2013;6(1):27-34.
204. Garg P, Galper BZ, Cohen DJ, Yeh RW, Mauri L. Balancing the risks of bleeding and stent thrombosis: a decision analytic model to compare durations of dual antiplatelet therapy after drug-eluting stents. *Am Heart J.* 2015;169(2):222-233 e225.
205. Yeh RW, Secemsky E, Kereiakes DJ*, et al.* Individualizing Treatment Duration of Dual Antiplatelet Therapy after Percutaneous Coronary Intervention: An Analysis from the DAPT Study. *Circulation.* 2015;132:2267-2285.
206. Terkelsen CJ, Lassen JF, Norgaard BL*, et al.* Mortality rates in patients with ST-elevation vs. non-ST-elevation acute myocardial infarction: observations from an unselected cohort. *Eur Heart J.* 2005;26(1):18-26.
207. Steg PG, Lopez-Sendon J, Lopez de Sa E*, et al.* External validity of clinical trials in acute myocardial infarction. *Arch Intern Med.* 2007;167(1):68-73.
208. Moons KGM, Kengne AP, Grobbee DE*, et al.* Risk prediction models: II. External validation, model updating, and impact assessment. *Heart.* 2012;98(9):691-698.
209. Muller-Nordhorn J, Binting S, Roll S, Willich SN. An update on regional variation in cardiovascular mortality within Europe. *Eur Heart J.* 2008;29(10):1316-1326.
210. Collins GS, Ogundimu EO, Cook JA, Manach YL, Altman DG. Quantifying the impact of different approaches for handling continuous predictors on the performance of a prognostic model. *Stat Med.* 2016;10.1002/sim.6986:n/a-n/a.
211. Buuren S, Groothuis-Oudshoorn K. MICE: Multivariate imputation by chained equations in R. *J Stat Softw.* 2011;45(3).
212. Cox DR. Note on Grouping. *J Am Stat Assoc.* 1957;52(280):543-547.

213. Walker SM, Asaria M, Manca A*, et al.* Long term health care use and costs in patients with stable coronary artery disease: a population based cohort using linked electronic health records (CALIBER). *Eur Heart J Qual Care Clin Outcomes.* 2016;http://dx.doi.org/10.1093/ehjqcco/qcw003.
214. Asaria M, Walker SM, Palmer SJ*, et al.* Using electronic health records to predict costs and outcomes in stable coronary artery disease. *Heart.* 2016;http://dx.doi.org/10.1136/heartjnl-2015-308850.
215. Bonaca MP, Bhatt DL, Braunwald E*, et al.* Design and rationale for the Prevention of Cardiovascular Events in Patients With Prior Heart Attack Using Ticagrelor Compared to Placebo on a Background of Aspirin-Thrombolysis in Myocardial Infarction 54 (PEGASUS-TIMI 54) trial. *Am Heart J.* 2014;167(4):437-U444.
216. Melgaard L, Gorst-Rasmussen A, Lane DA*, et al.* Assessment of the cha2ds2-vasc score in predicting ischemic stroke, thromboembolism, and death in patients with heart failure with and without atrial fibrillation. *Jama.* 2015;314(10):1030-1038.
217. Stenestrand U, Wijkman M, Fredrikson M, Nystrom FH. Association between admission supine systolic blood pressure and 1-year mortality in patients admitted to the intensive care unit for acute chest pain. *Jama.* 2010;303(12):1167-1172.
218. Henriksson M, Palmer S, Chen R*, et al.* Assessing the cost effectiveness of using prognostic biomarkers with decision models: case study in prioritising patients waiting for coronary artery surgery. *BMJ.* 2010;340:b5606.
219. Pencina MJ, D' Agostino RB, Vasan RS. Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. *Statistics in Medicine.* 2008;27(2):157-172.
220. Pencina MJ, D'Agostino RB, Sr., Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med.* 2011;30(1):11-21.
221. Leening MJ, Vedder MM, Witteman JC, Pencina MJ, Steyerberg EW. Net reclassification improvement: computation, interpretation, and controversies: a literature review and clinician's guide. *Ann Intern Med.* 2014;160(2):122-131.
222. Kerr KF, Wang Z, Janes H*, et al.* Net reclassification indices for evaluating risk prediction instruments: a critical review. *Epidemiology.* 2014;25(1):114-121.
223. Pepe MS, Fan J, Feng Z, Gerds T, Hilden J. The Net Reclassification Index (NRI): a Misleading Measure of Prediction Improvement Even with Independent Test Data Sets. *Stat Biosci.* 2015;7(2):282-295.
224. Kawai VK, Cunningham A, Vear SI*, et al.* Genotype and risk of major bleeding during warfarin treatment. *Pharmacogenomics.* 2014;15(16):1973-1983.