THE ROLE OF REAL-WORLD EVIDENCE IN HEALTH TECHNOLOGY ASSESSMENT:

A CASE STUDY OF DIRECT ORAL ANTICOAGULANTS IN THE ATRIAL FIBRILLATION POPULATION

Giorgio Ciminata

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy (PhD)

Health Economics and Health Technology Assessment (HEHTA)

Institute of Health & Wellbeing

University of Glasgow

April 2019

Acknowledgments

I am extremely grateful to my supervisors Professor Olivia Wu, Dr Claudia Geue and Professor Peter Langhorne for their constant support and advice throughout this extraordinary journey.

Many thanks to all my colleagues at HEHTA for encouragement, offering an intellectually stimulating environment, and a wonderful playground for team bonding and personal development.

Many thanks also to the pharmacoepidemiology group based at the University of Strathclyde directed by Professor Marion Bennie for providing an ongoing collaborative professional platform.

I would also like to thank the eData Research and Innovation Service (eDRIS) group for constant technical assistance for data access.

Most importantly, I would like to thank Dr Manuela Deidda, colleague and life companion, for supporting me professionally and emotionally.

Publications, working papers and presentations

Working papers

Ciminata G, Geue C, Wu O, Deidda M, Kreif N, Langhorne P. Comparative-effectiveness of non-vitamin K antagonist oral anticoagulants and warfarin in the Scottish population: the value of real-world evidence. Health Economists' Study Group, Aberdeen, June 2017.

Ciminata G, Geue C, Langhorne P, Wu O. Inpatient, Outpatient, Prescribing and Care Home Costs associated with Atrial Fibrillation. BMJ Open, January 2019.

Ciminata G, Geue C, Wu O, Quinn TJ, Bennie M, Langhorne P. Comparative-effectiveness and safety of direct oral anticoagulants and warfarin in patients with atrial fibrillation: a national cohort study. (To be submitted).

Conference presentations

Inpatient, Outpatient, Prescribing and Care Home Costs associated with Atrial Fibrillation. European Health Economics Association conference, Hamburg, July 2016.

The Prescribing, Inpatient, Outpatient and Social Care Costs associated with Atrial Fibrillation in Scotland: a record linkage study, International Population Data Linkage Conference Data, Swansea, August 2016.

The Prescribing, Inpatient, Outpatient and Social Care Costs associated with Atrial Fibrillation in Scotland: a record linkage study, International Society of Pharmacoeconomics and Outcome Research (ISPOR), Vienna, November 2016.

The Prescribing, Inpatient, Outpatient and Social Care Costs associated with Atrial Fibrillation in Scotland: a record linkage study, UK Stroke Forum Conference, Liverpool, November 2016.

The Prescribing, Inpatient, Outpatient and Social Care Costs associated with Atrial Fibrillation in Scotland: a record linkage study, Informatics for Health, Manchester, April 2017.

Comparative-effectiveness and safety of direct oral anticoagulants and warfarin in patients with atrial fibrillation: a national cohort study, European Stroke Organisation Conference, Prague, Czech Republic, May 2017.

Comparative-effectiveness and safety of direct oral anticoagulants and warfarin in patients with atrial fibrillation: a national cohort study, International Society of Pharmacoeconomics and Outcome Research (ISPOR), Glasgow, November 2017.

Comparative-effectiveness and safety of direct oral anticoagulants and warfarin in patients with atrial fibrillation: a national cohort study, European Drug Utilisation Research Group (EuroDURG) Conference, Glasgow, November 2017.

Abstract

Real-World Evidence (RWE) refers to any "data used for decision-making that are not collected in conventional Randomised Controlled Trials (RCTs) and is increasingly used in Health Technology Assessment (HTA) as an adjustment to the evidence coming from Randomised Controlled Trials (RCTs).

RWE can provide additional evidence concerning treatment safety and effectiveness, facilitate the identification of relevant subpopulations, and permit the inclusion and analysis of clinical endpoints not expected in RCTs but observed in real life. However, the use RWE in the context of HTA is still limited.

The aim of this thesis is to explore the role of RWE in economic evaluation by exploring methods to use with observational data and the role of RWE for a case study of direct oral anticoagulants (DOACs); a class of drugs, including apixaban, dabigatran and rivaroxaban, used for the prevention of stroke in the population affected by atrial fibrillation (AF). In addition to quantifying resource use and associated healthcare expenditure for the AF population in Scotland and evaluating propensity score methods for estimating the Average Treatment Effect (ATE), specific objectives are assessing cost as well as effectiveness and safety of DOACs using Scottish linked data.

Two cohorts, one consisting of patients with a diagnosis of AF or atrial flutter, and the other of patients on any oral anticoagulant (OAC) were identified from inpatient hospital records and prescribing data for the 1997 – 2015 study period. These data were complemented by outpatient attendances, the care home census and mortality records using individual patient data linkage.

As a first step, this thesis assessed the predictors of costs and estimated inpatient, outpatient, prescribing and care-home costs associated with AF, using population-based individual-level linked data. Inpatient admissions accounted for the majority of total costs and these were the main cost driver across all age groups.

Overall, inpatient cost contributions (~75 %) were constant across age groups. This is offset by increasing care-home cost contributions. The inclusion of all available cost components is crucial for establishing overall costs, as these often extend beyond hospitalisation. Most importantly, the thesis found that patients' age has a limited impact on overall AF-related cost, and therefore may not be the main driver of future growth of AF-related costs in an ageing Scottish population.

In order to identify an appropriate method for the comparative-effectiveness analysis, propensity score (PS) based method, such as PS matching, covariate adjustment including PS as covariate, and a series of Inverse Probability Weighting (IPW) methods were tested. A cohort of patients were followed from their first oral anticoagulant prescription to first clinical event (stroke and major bleeding) or death, and censoring was applied to treatment switching or discontinuation. In this methodological chapter, the approach that uses propensity scores (PS) as a covariate was identified as the most robust method to be used in the more comprehensive comparative-effectiveness analysis.

The comparative-effectiveness analysis, including additional clinical outcomes that were also used in the pivotal RCTs assessing the efficacy of DOACs versus warfarin in the AF population, found no statistically significant differences in risk of stroke for apixaban, dabigatran and rivaroxaban compared with warfarin. There were however, concerns over safety aspects of rivaroxaban, as it was associated with increased risk of all-cause mortality.

The hazard ratios estimated from the comparative-effectiveness analysis were used to populate a Markov model to evaluate the lifetime cost- effectiveness of DOACs compared to warfarin; one-way and probabilistic sensitivity analyses were carried out to assess the uncertainty around the findings and identify key drivers. At the £20,000 threshold, apixaban and dabigatran were found to be cost-effective in AF patients who are 50 years old when starting anticoagulation. Rivaroxaban, being the least effective intervention, was dominated by warfarin, being less costly but more effective than rivaroxaban.

This thesis shows the potential of RWE in general and within the Scottish healthcare setting. The findings highlight the importance of taking into account resource utilisation beyond hospital care, and assessing several comparative-effectiveness methods to understand strengths and limitation of each. Most importantly, the findings from this thesis have the potential to inform future research, prescribing patterns and provide real-world evidence for other healthcare settings, especially where rivaroxaban is the DOAC most widely prescribed.

Finally, this thesis shows that RWE generated from routinely collected linked data in Scotland, may well support the reassessment of prescription drugs accepted conditionally by the Scottish Medicine Consortium (SMC), an independent organisation that advises the NHS Health Boards about medicines, and would therefore support the SMC in making the final acceptance decision.

Table of contents

List of Tables

List of Figures

Abbreviations

AF: Atrial Fibrillation

AIC: Akaike Information Criterion

ARISTOTLE: Apixaban for Reduction In Stroke and Other Thromboembolic Events in Atrial Fibrillation

ATE: Average Treatment Effect

BNF: British National Formulary

CCI: Charlson Comorbidity Index

CEAC: Cost-Effectiveness Acceptability Curves

CHEERS: Consolidated Health Economic Evaluation Reporting Standards

CHI: Community Health Index

CIS: Continuous Inpatient Stay

CUA: Cost Utility Analysis

DES: Discrete Event Simulation

DOACs: Direct Oral Anticoagulants

DSU: Decision Support Unit

ERG: Evidence Review Group

GI: Gastro Intestinal

GLM: Generalised Linear Model

HRGs: Healthcare Resource Groups

HRQoL: Health Related Quality of Life

HTA: Health Technology Assessment

ICD-10: International Classification of Diseases, Tenth Revision

ICER: Incremental Cost-Effectiveness Ratio

ICH: Intracranial Haemorrhage

INMB: Incremental Net Monetary Benefit

INR: International Normalised Ratio

IPCW: Inverse Probability of Censoring Weighting

IPTW: Inverse Probability of Treatment Weighting

IPW: Inverse Probability Weighting

ISD: Information Services Division

ITT: Intention To Treat

MI: Myocardial Infarction

NHS: National Health Service

NICE: National Institute for Health and Care Excellence

NMA: Network Meta-Analysis

NRS: National Records of Scotland

NVAF: Non-Valvular Atrial Fibrillation

OAC: Oral Anticoagulant

OLS: Ordinary Least Square

OPCS-4: Office of Population Censuses and Surveys, Fourth Revision

PCT: Pragmatic Controlled Trials

PIS: Prescribing Information System

PS: Propensity Score

PSA: Probability Sensitivity Analysis

QALY: Quality Adjusted Life Years

QHES: Quality of Health Economics Analyses

QOF: Quality and Outcomes Framework

RCTs: Randomised Control Trials

RE-LY: Randomized Evaluation of Long-Term Anticoagulation Therapy

ROCKET-AF: Rivaroxaban Once Daily Oral Direct Factor Xa Inhibition Compared with Vitamin K Antagonism for Prevention of Stroke and Embolism Trial in Atrial Fibrillation

RWE: Real Word Evidence

SE: Systemic Embolism

SIMD: Scottish Index of Multiple Deprivation

SMC: Scottish Medicine Consortium

SMR00: Outpatient Attendance Scottish Morbidity Records

SMR01: General Acute Inpatient and Day Case Scottish Morbidity Records

SNT: Scottish National Tariff

SSCA: Scottish Stroke Care Audit

TIA: Transient Ischaemic Attack

VKA: Vitamin K Antagonist

VTE: Venous Thromboembolism

WHO: World Health Organisation

WTP: Willingness To Pay

Author's declaration

I declare that, except where explicit reference is made to contribution of others, that this dissertation is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Signature:

Printed name: Giorgio Ciminata

Chapter 1 Introduction

1.1 Real-world evidence

Real word evidence (RWE) refers to any "data used for decision-making that are not collected in conventional randomised control trials (RCTs)" [1]. Typically, RWE data are collected from a wide range of sources including Pragmatic Controlled Trials (PCT), registries, administrative data, health surveys and electronic health records. Pragmatic Controlled Trials measure treatment effectiveness in routine clinical practice.

The design of PCTs should reflect variations in patient characteristics that would occur in clinical practice; thus, PCTs should represent the patients to whom the treatment will be applied [1, 2]. When randomisation is introduced in the design, PCTs are referred as pragmatic RCTs. A recent example of this is the Salford Lung Study investigating safety and efficacy of a once-daily dry powder inhaler reducing asthma and chronic obstructive pulmonary disease exacerbations. In this study, patients were randomised to either receive the once-daily dry powder inhaler or continue their optimised usual care based on the assessment of the asthma control test [3].

Registries collect patient level data that focus on a specific diagnosis or condition, while administrative data, generally collected at a national level, are primarily collected for administrative purposes. RWE may also be obtained from health surveys, carried out to collect data that are representative of the health of a given population, and electronic health records providing patient's clinical information in a digital format [2, 4].

The evidence from real-world scenarios is usually drawn from observational prospective and retrospective studies. A study design utilising a prospective approach investigates over time whether a given cohort of patients experience the outcome under scrutiny. The retrospective approach, typically adopted in case-control studies, looks back at exposure to risk factors leading to the actual disease state; thus, the outcome of interest has already occurred at the time of study initiation [5].

Historically, RWE has been used for clinical epidemiologic investigations, quality improvement, and safety surveillance [2].Evidence from real-world scenarios is increasingly used in health technology assessment (HTA), a process that systematically evaluates evidence concerning safety, clinical and cost-effectiveness of health technologies; in some instances social, legal and ethical aspects are factored in to the multidisciplinary evaluation process [2, 6]. The main purpose of HTA is to inform reimbursement and coverage decisions, thus RWE may play a crucial role in the pre- and post-marketing authorisation process that grant access of new health technologies to the market.

In particular, in the pre-authorisation process RWE may facilitate the identification of subpopulations that may enable better understanding of underlying prevalence of disease. In the post-marketing authorisation process, RWE may provide a more comprehensive understanding of treatment safety and effectiveness by allowing for the inclusion and analysis of clinical endpoints not measured in RCTs but observed in real life. For instance, a given side effect to a treatment may not be contemplated in a RCT but may be experienced by patients outside of the controlled RCT environment, where drugs are being used in a large population and beyond the specified RCT follow-up time [2].

Within this context, RWE may also improve the understanding of treatment sequences, treatment pathways in real life, and allow for comparisons of treatments against relevant comparators not employed in registration trials. Treatment adherence, and interactions with concomitant treatments, are other important aspects that can be investigated in RWE studies [2, 7].

In the UK, RWE is increasingly used as supplementary evidence to Phase II and Phase III of RCTs, and in "accelerated market access" where initial decisions are conditional on additional randomised and non-randomised evidence generated over time [8]. However, it could prove problematic generating additional evidence from further RCTs; as patients already having access to a given treatment, may be reluctant to taking part in any other clinical study. Although, randomised evidence could be generated from RCT, carried out in different countries, there may be generalisability issues for the study results due to differences in the target population and treatment standards [2, 4].

Despite the advantage of using RWE as a supplement to the evidence coming from RCTs in the context of HTA, its use is still limited. This is certified by a study indicating that in the UK, up to 2016, non-randomised evidence has only been used in 36% of the appraisals carried out by the National Institute for health and Care Excellence (NICE): an advisory body responsible for issuing guidelines concerning the use of new health technologies within the National Health Service (NHS) in England [9]. Nevertheless, thanks to methodological advances and increased data sources and data availability, the scope of RWE is broadening, and its use is increasing.

Still, the use of RWE provides challenges and limitations. Real-world evidence may be inconsistently collected, and missing elements may lead to reduced statistically validity or may limit the potential for answering research questions. Most importantly, RWE studies are subject to bias (e.g. confounding factors) [2]. In RCTs, these issues are controlled for by randomisation. Confounding, in general, is a distortion that modifies an association between an exposure and an outcome [10]. The main source of confounding in newly marketed medications is confounding by indication occurring when the prognostic factors, such as disease severity, used for treatment selection also affect the outcome. For instance, patients with more severe conditions are likely to receive more intense treatments. However, when comparing treatment effectiveness, the more intensive treatment may produce poorer outcomes [11].

Confounding, if not controlled for, may give rise to bias, a systematic error resulting from incorrect estimate of the true effect of an exposure on a given outcome [10]. A common type of bias using RWE is selection bias, occurring when therapies are differently prescribed according to disease severity and patient characteristics. Other types of bias are information bias, potentially caused by misclassification of data, and detection bias occurring when an outcome is more likely to be detected in one treatment group than another [12].

1.2 Real-world evidence in Scotland

In Scotland, the Farr Institute (now part of the Health Data Research UK), is involved in the use of linked electronic data to generate RWE that supports a better understanding of health at the patient and population level. The institute is based on the collaboration between Scottish universities (including the University of Glasgow, Strathclyde, Dundee, Edinburgh, St Andrews, and Aberdeen), and NHS Scotland.

The Farr institute was formerly the Scottish health informatics programme (SHIP) experience, a “Scotland-wide research platform for the collation, management, dissemination and analysis of Electronic Patient Records” funded by the Wellcome Trust, the Medical Research Council and the Economic and Social Research Council [13].

Scotland has long been at the forefront of research where routinely collected real-world data linkage is involved. While in in the late 1960s patient level data on hospitalisation, death and cancer were already available, only the advance of computing in the late 1980s has allowed for an advanced process of data linkage [14].

Health care data are gathered as Scottish Morbidity Records (SMR) reflecting the status and type of healthcare that patients receive. For instance, General Acute Inpatient and Day Case Scottish Morbidity Records (SMR01) contains all general acute admissions, categorized as inpatients or day cases, discharged from non-obstetric and non-psychiatric specialties; Outpatient Attendance Scottish Morbidity Records (SMR00) records include, at patient level, information on new and follow-up appointments at outpatient clinics for any clinical specialty. More detail on SMR01 and SMR00 are provided in Chapter 3. Other examples include the Scottish cancer registry SMR06, and records on maternity inpatient and day case collected in SMR02.

Patient level records are submitted by hospitals and health boards to the Information Services Division (ISD), a division of National Service Scotland, part of NHS Scotland. The information services division supports NHS planning, quality improvement activities and decision making by providing services in the field of health information, health intelligence and statistical services [15].

1.3 Case study

Direct Oral Anticoagulants (DOACs) including apixaban, dabigatran and rivaroxaban, have all been accepted for use in Scotland with an indication for the prevention of stroke in the Atrial Fibrillation (AF) population [16-18]. However, evidence from clinical practice on the effectiveness and safety of DOACs for AF patients is still limited. Further, as records from primary care are not available, the sample size of the AF population is potentially underestimated; therefore, this thesis focuses on patients who have a diagnostic code of AF from hospital codes. Based on these challenges, DOACs were identified a good case study for exploring the data in HTA.

1.4 Epidemiology of Atrial Fibrillation

Atrial fibrillation is the most common form of arrhythmia, increasing significantly with age, and linked to the development of highly debilitating conditions such as congestive heart failure, ischemic heart disease and stroke [19]. The risk of these conditions is significantly greater in individuals affected by AF. Therefore, AF, linked to an ageing population especially across high-income countries, is likely to have a substantial impact on the economic burden of any healthcare system [19, 20]. As per the World Health Organisation (WHO) definition, stroke is "rapidly developing clinical signs of focal (or global) disturbance of cerebral function, with symptoms lasting 24 hours or longer or leading to death, with no apparent cause other than of vascular origin" [21]. A stroke, also referred to as cerebrovascular accident, usually occurs when a blood vessel bursts or a clot causes the interruption of blood supply to the brain, thus causing damage to the brain tissue deprived of oxygen and nutrients. This in turn affects the part of the body controlled by the portion of the brain injured [21].

Stroke is typically classified as ischemic or haemorrhagic. The first typology, accounting for the majority of strokes, is caused by atherosclerotic obstruction within a blood vessel supplying blood to the brain. Strokes caused by temporal obstruction are classified as "mini strokes" or TIAs. While patients experiencing TIA may recover within an hour, symptoms of TIA should be regarded as a warning sign for a future major stroke. By contrast, haemorrhagic stroke is caused by the burst of a blood vessel generally happening because of uncontrolled hypertension. Although less common, the burst may also be caused by aneurysms or arteriovenous malformations [21].

A systematic review of studies on the epidemiology of AF in Europe, published between 2005 and 2014, estimated an AF prevalence of 2%, and an incidence rate ranging between 0.23 and 0.41 cases per 1,000 person/years [20]. Of these individuals, only 50% would develop permanent AF, and the remaining may experience paroxysmal AF (typically stop within 48 hours without any treatment) or persistent AF (each episode may last for more than seven days, or less if treated) [20]. The same study, estimated that over a population of about 500 million individuals, approximately 10 million people may experience AF [20].

Atrial fibrillation, occurring mostly in males, varies substantially according to age. While only 0.12% - 0.16% of the European population younger than 49 years is affected, for the age groups 60 - 70 and 80 years or older, the proportion rises noticeably to 3.7% - 4.2% and 10% - 17% respectively. Thus, in an ageing population the prevalence of AF is expected to grow significantly, though the main contributor is the number of elderly patients hospitalised with a secondary diagnosis of AF. Indeed AF, representing the main risk factor for ischaemic stroke, often exists with comorbidities [22].

The increase in the number of new cases, however, is not only attributable to ageing, but also to improved diagnostic capabilities allowing for a more accurate detection of AF, and improved outcomes in the treatment of AF comorbidities [22]. This, in combination with the European population growth of 0.2% - 0.3% per year on average, would bring the prevalence of AF to 2.7% - 3.3% by the year 2030 and over a population of 516 – 525 million individuals. Hence, according to Berisso and colleagues, by 2030 in Europe there will be about 14 - 17 million individuals affected by any form of AF [20].

However, an important distinction has to be made between individuals with a known diagnosis of AF, and those who are potentially at risk of AF, but are not captured because of a series of factors such as data miscoding and lack of primary care data. For instance Public Health England, reported for the year 2015/2016 a prevalence of 1.7% reflecting the proportion of patients with a diagnosis of AF over the total English population [23]. The prevalence was derived from the outputs of the Quality and Outcomes Framework (QOF), based on a pay for performance mechanism and designed to incentivise good practice in primary care [24]. However, Public Health England acknowledged that out of an estimated 1.4 million individuals at risk of AF in England, only about 985,000 were diagnosed with AF between 2015 and 2016. Therefore, the rather conservative prevalence of 1.7% was more likely to be in the proximity of 2.5% indicating that AF was undiagnosed for over 425,000 individuals across England [23]. A more recent study reported a very similar estimate (2.7%) confirming that AF is largely underdiagnosed [24].

Similarly, QOF data showed a prevalence of 1.8% in Scotland, indicating that about 96,000 individuals were diagnosed with AF in a Scottish population of about 5.5 million people in 2016. However, a report from the Cross Party Group on Heart Disease and Stroke, investigating issues concerning diagnosis, treatment and care of patients with AF in Scotland, indicated a more plausible prevalence in the proximity of 2.6%.

This would suggest that in 2016 there were 145,000 AF diagnosis, and that almost 50,000 individuals with undiagnosed AF were potentially at a greater risk of experiencing a stroke [25].

Treatments, such as beta-blockers and statins, used for the primary prevention of AF exist. However, evidence on the effectiveness of first line pharmacological treatments, and the knowledge on the pathophysiology of AF are still limited [26].Globally, at present, the main approach is focused on secondary prevention by means of pharmacological and non-pharmacological interventions aimed at controlling rhythm and rate of the heart. More specifically, anticoagulants, which are the main pharmacological intervention, aim at preventing complications, such as stroke and cardiovascular conditions [27]. Non-pharmacological interventions include electrical cardioversion used to restore the heart to its normal rhythm. Other non-pharmacological approaches include catheter and surgical ablation employed to stop abnormal electrical impulses responsible for AF [28].

1.5 AF management pathway

The management of AF is a crucial part in the prevention of stroke. Clinical guidelines have highlighted the steps that should be carried out in the management of AF, and under which circumstances anticoagulation should be administered. The full NICE AF management pathway for the prevention of stroke is presented in Figure 1.1 [29].

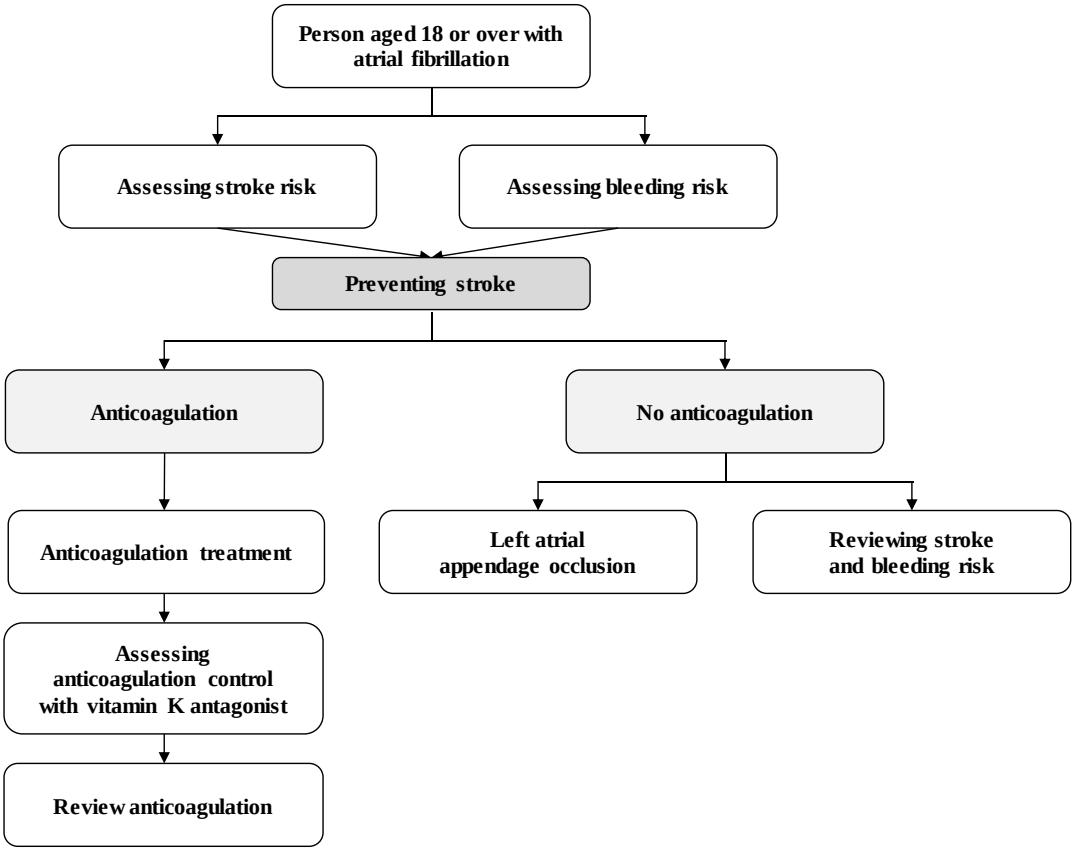


Figure 1.1 Management pathway to prevent stroke in people with atrial fibrillation
Source: Adapted from, National Institute for Health and Care Excellence. *Preventing stroke in people with atrial fibrillation*, 2015.

Firstly, in patients who are 18 years or older and with a known diagnosis of AF, the risk of ischaemic stroke and major bleeding are assessed with the use of CHA_2DS_2-VASc [30] and HAS-BLED [31] respectively. CHA_2DS_2-VASc and HAS-BLED are bespoke tools, accounting for a series of clinical and patient characteristics, used to estimate the probability of having a stroke and major bleeding in patients affected by AF.

According to the International Society on Thrombosis and Haemostasis criteria, major bleeding is defined as a clinical fatal event occurring in non-surgical patients [32].

The event should be caused by one of the following: bleeding in a critical area or organ, (e.g. intracranial, intraspinal, intraocular, retroperitoneal, intra-articular or pericardial, or intramuscular with compartment syndrome) bleeding causing a reduction in haemoglobin level of ≥2g/dL and requiring two or more blood transfusions [32].

In particular, CHA_2DS_2-VASc should be used in people with any of the following: symptomatic or asymptomatic paroxysmal, persistent or permanent atrial fibrillation, atrial flutter and a continuing risk of arrhythmia recurrence after cardioversion back to sinus rhythm [29]. In general, antiplatelet or anticoagulation should be considered for patients with CHA_2DS_2-VASc score of 1 indicating a low to moderate risk of stroke, also a different dose regimen should be considered for patients who are 80 years or older. Anticoagulation is recommended in patients with a CHA_2DS_2-VASc score of 2 or greater and reflecting moderate to high risk of stroke [30]. Nevertheless, for those patients eligible for anticoagulation, risk bleeding scores as those obtained from HAS-BLED, should be used to determine the bleeding risk associated with patients characteristics. Typically, anticoagulation should be considered for patients with HAS-BLED score ranging from 0 to 2, which reflects a relatively low to moderate risk of major bleeding [31]. More details on how these scores were calculated are provided in Chapter 6.

Whenever the use of anticoagulants is recommended, the need for anticoagulation should be reviewed annually, or more frequently, to identify the presence of any clinically relevant events for which anticoagulation is contraindicated. Anticoagulation control is generally assessed by time in therapeutic range, indicating how long International Normalised Ratio (INR) values stay within a desired range [29].

International normalised ratio is a measurement indicating the time the blood needs to clot after adding a tissue factor, and given by the ratio between the normal mean prothrombin time and patient's prothrombin time; INR is therefore used to establish the optimal anticoagulant dose [33]. Anticoagulation should be reassessed in patients showing poor anticoagulation control; generally determined by factors such as therapy adherence, anticoagulants interaction with other drugs or diet, cognitive function and alcohol consumption.

In case of poor anticoagulation control persistence, alternative stroke prevention strategies should be explored. For people on warfarin, risk and benefit associated with switching to DOACs should be taken into consideration. Compliance should also be considered when switching, poor compliance might expose patients to a greater risk of experiencing stroke or any other relevant clinical event [16-18].

As indicated by NICE guidelines, anticoagulation should not be offered to AF patients aged 65 years or younger with a low risk of stroke. For these people, the risk of stroke should be reviewed past the age of 65, or whenever any of the following clinical events are experienced: diabetes, heart failure, peripheral arterial disease, coronary heart disease, TIA or SE. Risk of stroke and bleeding should be assessed annually in patients who are on anticoagulants because of bleeding risk or other factors. As an alternative to anticoagulation, left atrial appendage occlusion should be considered; however, this treatment option should be offered whenever anticoagulants are contraindicated or not tolerated [29].

1.6 Clinical and pharmacological aspects of anticoagulants

Warfarin, to date the main pharmacological treatment for the management of AF as a preventive measure for stroke, is a Vitamin K Antagonist (VKA) working by limiting the availability of vitamin K, which in turn inhibits the synthesis of vitamin K dependent factors essential for the blood coagulation process [28]. By contrast, the DOACs, including apixaban, dabigatran, rivaroxaban and edoxaban, work directly on the coagulation factors. More specifically, apixaban and rivaroxaban directly inhibit the formation of factor Xa responsible for the conversion of prothrombin to thrombin, thus preventing the formation of fibrin clots.

In a similar mechanism of action where the drug is selective for one specific vitamin K dependent factor, dabigatran works directly on thrombin, where the inhibition disrupts the coagulation cascade and consequently prevents the formation of clots [27, 34].

Aside from the different mechanism of action, DOACs offer important advantages over warfarin and in general over any VKA drug. Compared to VKAs, the number of drug interactions of DOACs with any other drug is relatively small. However, DOACs cannot be used in concomitance with inhibitors or inducers of P-glycoprotein, a transporter responsible for the excretion of drugs from the body [28]. Similarly, the combination of DOACs with inhibitors or inducers of CYP34A (an enzyme helping metabolise drugs in the body) should be avoided. In particular, CYP3A4 inhibitors may increase serum concentration leading to increased risk of bleeding. On the other hand, CYP3A4 inducers could increase the DOACs metabolism affecting their anticoagulation effect.

As pointed out, VKAs work by limiting the availability of vitamin K, therefore food containing this vitamin may interfere with VKAs mechanism of action disrupting the anticoagulation effect. Because DOACs act directly on coagulation factors, patients have no dietary restriction as DOACs do not interact with food [27, 28].

The predictability of pharmacodynamics and pharmacokinetics is perceived to be another distinct advantage of DOAC over VKAs. While pharmacodynamics assesses the biochemical and physiological effect of the drug on the body, pharmacokinetics investigates how a drug is metabolised in the body. Because data on pharmacodynamics and pharmacokinetics have shown to be independent of body weight, age, and sex, DOACs could be used at a fixed dose without requirement for routine anticoagulant monitoring [27, 28, 34].However, anticoagulation assessment is typically carried out if urgent surgical intervention or parenteral administration are needed.

Furthermore, the rapid onset of action, following oral administration, is one of the main assets of DOACs. Equally important is the rapid offset of action, which is critical in cases where surgical intervention is required [34]. Nevertheless, important contraindications and limitations are associated with the use DOACs. Notably, DOACs may not be appropriate for patients with chronic kidney conditions as a high proportion is eliminated from the body through the kidneys. This is particularly true for dabigatran where over two thirds are expelled as an active drug [34].

Apixaban and rivaroxaban may be used cautiously at reduced doses in patients with moderate or severe renal insufficiency, with levels of creatinine clearance between 30 - 50 mL/min and 10 - 30 mL/min respectively, combined with one of the following: 80 years of age or over, body weight greater than 60 kg [16, 17, 35, 36]. The use of DOACs is also contraindicated in patients with a severe chronic liver disease. However, if hepatic conditions are mild or moderate, DOACs may still be used prudently at reduced doses. In addition, DOACs may not be suitable for patients with mechanical mitral valve issues or malignant disease, as they may trigger bleeding complications [16-18, 35-37]. Another important limitation of DOACs, with an exception for dabigatran, is the availability of an antidote to reverse coagulation in case of overdose or an urgent surgical intervention is needed [27].

Compliance is another important aspect that poses some limitation to the use of DOACs making VKAs the treatment of choice for some patients. In cases of a poor compliance, due to the short half-life of DOACs, patients will be exposed to a greater risk of experiencing stroke or any other relevant clinical event [27, 38].

1.7 Direct oral anticoagulants in Scotland

Direct oral anticoagulants have been approved throughout the UK for the prevention of stroke in the AF population [16-18, 35-37, 39]. Edoxaban, approved in Scotland in 2015, by the Scottish Medicine Consortium (SMC), an independent organisation that advises the NHS Health Boards about medicines in Scotland [40], is currently in use for the prevention of stroke in the AF population [39]. However, data availability from clinical practice is still limited. Therefore, the focus on this thesis will be on apixaban, dabigatran and rivaroxaban. DOACs are currently used as an alternative to warfarin, to date the main oral thromboprophylactic treatment for the management of AF as a preventive measure for stroke [33].

In line with other parts of the world, warfarin has been for decades the main oral anticoagulation treatment used to prevent stroke in the AF population in Scotland.
In 2011, dabigatran was the first DOAC accepted by the SMC, with an indication for the prevention of stroke and Systemic Embolism (SE) in patients with a diagnosis of Non-Valvular Atrial Fibrillation (NVAF) and associated risk factors: history of stroke Transient Ischaemic Attack (TIA) or SE, left ventricular ejection fraction (<40%), symptomatic heart failure and age ≥ 75 years. Patients aged 65 years and older were considered at risk if one of the following was present: diabetes mellitus, coronary artery disease or hypertension. A reduced twice-daily dose of 110 mg, as an alternative to the standard 150 mg, was recommended for patients with a high risk of bleeding, and in particular for patients aged 80 years or older [18].

Rivaroxaban was initially accepted for use by the SMC and indicated for the prevention of Venous Thromboembolism (VTE) in adult patients who had hip or knee replacement surgery. In 2012, rivaroxaban was accepted for use by the SMC, with a different indication aimed at preventing SE in patients with a diagnosis of NVAF plus one of the following risk factors: history of stroke or TIA, age ≥75 years, hypertension, diabetes mellitus and congestive heart failure, was obtained with a second HTA evaluation. However, the use of rivaroxaban was restricted only to patients with poor INR control or intolerance to warfarin. In addition to the 20 mg once daily standard dose, a reduced dose of 15 mg was recommended for patients with moderate to severe renal impairment

As for rivaroxaban, the –acceptance of apixaban by the SMC, for the prevention of stroke and SE in patients with a NVAF and one of the following risk factors: history of stroke or TIA, age ≥75 years, hypertension, diabetes mellitus or symptomatic heart failure, was obtained with a second HTA evaluation in 2013. In addition to a standard daily dose of 5 mg, a reduced dose of 2.5 mg was recommended for patients with a combination of the following characteristics: aged 80 years or older, body weight ≤60kg and serum creatinine level≥1.5mg/dL indicating abnormal renal function [16].

1.8 Current evidence from randomised control trials and real-world evidence

1.8.1 Randomised control trails

The efficacy of DOACs was estimated from clinical studies with respect to anticoagulation by looking primarily at ischaemic or haemorrhagic stroke and SE as the primary outcomes [41-43]. The pivotal RCTs measuring the efficacy of DOACs are described in the next paragraphs.

Apixaban

The main evidence supporting the indication of apixaban for the prevention of stroke or SE came from the "Apixaban for Reduction In Stroke and Other Thromboembolic Events in Atrial Fibrillation" (ARISTOTLE) clinical trial. The population of 18,201 patients with AF and risk factors for stroke had a median age of 70 years, with the male being greater than the female population. Overall patients were classified as having a moderate to high risk of stroke according to the $CHAD_2$ score of 2.1. The $CHADS_2$ is the precursor of the most widely used CHA_2DS_2-VASc risk calculator. The details of this tool, used to predict the potential risk of stroke in patients with AF, will be discussed in Chapter 6. Approximately 19% of patients had a history of stroke, TIA or SE. Hypertension was the risk factor for stroke, observed in over 80% for both treatment groups. In the ARISTOTLE study, both first time warfarin users and on-going users were enrolled with the objective of including a minimum of 40% first time users.

The study, using a non-inferiority design, evaluated the efficacy of apixaban (9,120 patients) compared to warfarin (9,081 patients) for ischemic or haemorrhagic stroke or SE. Patients were randomised to receive apixaban standard 5 mg twice daily standard dose (or reduced dose as previously discussed), or dose-adjusted warfarin to maintain the INR between a range of 2.0 and 3.0. The follow-up analyses were conducted following an Intention To Treat (ITT) approach, where all randomised patients are included, regardless of their adherence to protocol.

After a median follow-up of 1.8 years, apixaban 5 mg showed significant reduction in the rate of stroke or SE (HR 0.79 (95% CI 0.66, 0.95)) [42]. Apixaban was also associated with a significant reduction in risk of the primary safety outcome of major bleeding (HR 0.69 (95% CI 0.60, 0.80)) compared to warfarin [42].

Full details on baseline characteristics, events rates and HRs are presented in Table 1.1, Table 1.2 and Table 1.3 respectively.

Dabigatran

The main evidence supporting the indication of dabigatran for the prevention of stroke or SE came from the "Randomized Evaluation of Long-Term Anticoagulation Therapy" (RE-LY) clinical trial. The population of 18,113 patients with AF and an increased risk of stroke had a mean age of 71 years; males represented the majority in both treatment groups.

The study, using a non-inferiority design, evaluated the efficacy of apixaban (6,015 patients for the 110 mg reduced dose, and 6,076 for the 150 mg standard dose) compared to warfarin (6,022 patients) for ischemic or haemorrhagic stroke or SE. Similarly, RE-LY, in line with the inclusion criteria outlined in the ARISTOTLE study, enrolled on-going and first-time warfarin users. Patients were randomised to receive either 110 mg or reduced 150 mg dose of dabigatran as an alternative to dose-adjusted warfarin to maintain the INR between a range of 2.0 and 3.0. Follow-up analyses were conducted following the ITT approach [41].

After a median follow-up of 2 years, patients receiving dabigatran had a significantly reduced risk of stroke or SE (HR 0.65 (95% CI 0.52, 0.81)) when compared with patients on warfarin; but no risk difference was observed for the 110 mg reduced dose (HR 0.90 (95% CI 0.74, 1.10)). The rate of Myocardial Infarction (MI) was found to be higher in dabigatran (at any dose) than with warfarin. This may be due to a greater protection against coronary ischaemic events that warfarin may have compared to dabigatran [41]. Full details on baseline characteristics, events rates and HRs are presented in Table 1.1, Table 1.2 and Table 1.3 respectively.

Rivaroxaban

The main evidence supporting the indication of rivaroxaban for the prevention of stroke or SE came from the "Rivaroxaban Once Daily Oral Direct Factor Xa Inhibition Compared with Vitamin K Antagonism for Prevention of Stroke and Embolism Trial in Atrial Fibrillation" (ROCKET-AF) clinical trial. The population of 14,624 patients with AF and moderate-to-high risk of stroke had a mean age of 73 years; males represented the majority in both treatment groups.

The study, using a non-inferiority design, evaluated the efficacy of rivaroxaban (7,131 patients) compared to warfarin (7,133 patients) for ischemic or haemorrhagic stroke or SE. The study evaluated a 20 mg oral dose of rivaroxaban (or reduced dose in case of abnormal renal function) as an alternative to dose-adjusted warfarin to maintain the INR between a range of 2.0 and 3.0.

After a median follow-up of 1.9 years, patients receiving rivaroxaban had a significantly reduced risk of stroke or SE compared to patients on warfarin in the safety on treatment population (HR 0.79 (95% CI 0.65, 0.95)) where individuals are grouped for the analysis based on the treatment received, rather than being allocated to treatments by means of randomisation. The same level of risk reduction was observed in the per protocol population (HR 0.79 (95% CI 0.66, 0.96)) where the analysis is carried out on patients who did not discontinue or switch from the treatment they were originally allocated. Nevertheless, rivaroxaban was shown to be non-inferior to warfarin in the ITT analysis (HR 0.88 (95% CI 0.75–1.03)) [43]. The risk of stroke or SE in the ROCKET-AF was also estimated with the safety on treatment and the ITT approach. The risk for the primary outcome in the ARISTOTLE and RELY was solely estimated with the ITT approach. Full details on baseline characteristics, events rates and HRs are presented in Table 1.1, Table 1.2 and Table 1.3 respectively.

Table 1.1 Baseline characteristics

	ARISTOTLE		RE-LY			ROCKET-AF	
Baseline characteristics	Apixaban	Warfarin	Dabigatran (110mg)	Dabigatran (150mg)	Warfarin	Rivaroxaban	Warfarin
	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
Patients	9,120	9,081	6,015	6,076	6,022	7,131	7,133
Age median*(IQR) **(SD)	70 *(63 - 76)	70 *(63-76)	71.4 **(8.6)	71.5 **(8.8)	71.6 **(8.6)	73 *(65-78)	73 *(65-78)
Female	3,234 (35.5)	3,182 (35.0)	2,150 (35.7)	2,236 (36.8)	2,213 (36.7)	2,831 (39.7)	2,832 (39.7)
Age ≥75 year (SD)	2,850 (31.3)	2,829 (31.1)					
$CHAD_2$ score mean (SD)	2.1 (1.1)	2.1 (1.0)					
0 or 1	3,100 (34.0)	3,083 (34.0)	1,958 (32.6)	1,958 (32.2)	1,859 (30.9)		
2	3,262 (35.8)	3,254 (35.8)	2,088 (34.7)	2,137 (35.2)	2,230 (37.0)	925 (13.0)	934 (13.1)
≥3	2,758 (30.2)	2,744 (30.2)	1,968 (32.7)	1,981 (32.6)	1,933 (32.1)	6,205 (87.0)	6,197 (86.9)
Prior myocardial infarction	1,319 (14.5)	1,266 (13.9)	1,008 (16.8)	1,029 (16.9)	968 (16.1)	1,182 (16.6)	1,286 (18.0)
Prior clinically relevant or spontaneous bleeding	1,525 (16.7)	1,515 (16.7)					
Prior stroke, TIA, or SE	1,748 (19.2)	1,790 (19.7)	1,195 (19.9)	1,233 (20.3)	1,195 (19.8)	3,916 (54.9)	3,895 (54.6)
Heart failure			1,937 (32.2)	1,934 (31.8)	1,922 (31.9)	4,467 (62.6)	4,441 (62.3)
Peripheral vascular disease						401 (5.6)	438 (6.1)
Ventricular disease	3,235 (35.5)	3,216 (35.4)					
Diabetes	2,284 (25.0)	2,263 (24.9)	1,409 (23.4)	1,402 (23.1)	1,410 (23.4)	2,878 (40.4)	2,817 (39.5)
Hypertension	7,962 (87.3)	7,954 (87.6)	4,738 (78.8)	4,795 (78.9)	4,750 (78.9)	6,436 (90.3)	6,474 (90.8)
Drugs causing bleeding	3,781 (41.5)	3,709 (40.8)	2,404(40.0)	2,352 (38.7)	2,442 (40.6)	2,586 (36.3)	2,619 (36.7)
Renal function, creatinine clearance (IQR)						67 (52-88)	67 (52-86)
Normal, >80 ml/min	3,761 (41.2)	3,757 (41.4)					
Mild impairment, >50 to 80 ml/min	3,817 (41.9)	3,770 (41.5)					
Moderate impairment (>30 to 50 ml/min)	1,365 (15.0)	1,382 (15.2)					
Severe impairment (≤30 ml/min)	137 (1.5)	133 (1.5)					

Abbreviations: TIA= transient ischaemic attack, SE=systemic embolism.
Source : Granger et al (2011), Connolly et al. (2009), Patel et al. (2011).

Table 1.2 RCT clinical outcomes - event rates (per 100 person years)

	ARISTOTLE		RE-LY			ROCKET-AF	
Outcome	Apixaban Events and (event rates)	Warfarin Events and (event rates)	Dabigatran (110mg) Events and (event rates)	Dabigatran (150mg) Events (event rates)	Warfarin Events (event rates)	Rivaroxaban Events (event rates)	Warfarin Events (event rates)
Stroke or SE (ITT)	212 (1.27)	265 (1.60)	182 (1.53)	134 (1.11)	199 (1.69)	269 (2.1)	306 (2.4)
Stroke or SE (per-protocol population)						188 (1.7)	241 (2.2)
Stroke or SE (safety on treatment population)						189 (1.7)	243 (2.2)
Stroke or SE or major bleeding	521 (3.17)	666 (4.11)					
Stroke or SE or major bleeding or mortality (all-cause)	1,009 (6.13)	1,168 (7.2)					
Stroke or SE or mortality (cardiovascular)						346 (3.11)	410 (3.63)
Stroke or SE or mortality (cardiovascular) or MI						433 (3.91)	519 (4.62)
Stroke	199 (1.19)	250 (1.51)	171 (1.44)	122 (1.01)	185 (1.57)	184 (1.65)	221 (1.96)
Ischaemic or unspecified stroke	162 (0.97)	175 (1.05)	159 (1.34)	111 (0.92)	142 (1.20)	149 (1.34)	161 (1.42)
Haemorrhagic stroke	40 (0.24)	78 (0.47)	14 (0.12)	12 (0.1)	45 (0.38)	29 (0.26)	50 (0.44)
SE	15 (0.09)	17 (0.1)	15 (0.13)	13 (0.11)	21 (0.18)	5 (0.04)	22 (0.19)
ICH	52 (0.33)	122 (0.8)	27 (0.23)	36 (0.3)	87 (0.74)	55 (0.5)	84 (0.7)
MI	810 (4.85)	906 (5.49)	86 (0.72)	89 (0.74)	63 (0.53)	101 (0.91)	126 (1.12)
GI bleeding	105 (0.76)	119 (0.86)	133 (1.12)	182 (1.51)	120 (1.02)		
Major bleeding	327 (2.13)	462 (3.09)	322 (2.71)	375 (3.11)	397 (3.36)	395 (3.6)	386 (3.4)
Major or clinically non-major bleeding	613 (4.07)	877 (6.01)				1,475 (14.9)	1,449 (14.5)
Any bleeding	2,356 (18.1)	3,060 (25.8)					386 (3.4)
Critical bleeding			145 (1.22)	175 (1.45)	212 (1.8)	91 (0.8)	133 (1.2)
Fatal bleeding						27 (0.2)	55 (0.5)
Mortality (all-cause)			446 (3.75)	438 (3.64)	487 (4.13)	208 (1.87)	250 (2.21)
Mortality (cardiovascular)			289 (2.43)	274 (2.228)	317 (2.69)	170 (1.53)	193 (1.71)

Abbreviations: ITT= intention to treat, SE=systemic embolism, ICH=intracranial haemorrhage, MI=myocardial infarction, GI=gastrointestinal.
Source : Granger et al (2011), Connolly et al. (2009), Patel et al. (2011).

Table 1.3 RCT clinical outcomes - Hazard Ratios

Outcome	ARISTOTLE Apixaban vs. warfarin HR 95% CI	RE-LY Dabigatran (110 mg) vs. warfarin HR 95% CI	RE-LY Dabigatran (150 mg) vs. warfarin HR 95% CI	ROCKET-AF Rivaroxaban vs. warfarin HR 95% CI
Stroke or SE (ITT)	0.79 (0.66, 0.95)	0.90 (0.74, 1.10)	0.65 (0.52, 0.81)	0.88 (0.75, 1.03)
Stroke or SE (per-protocol population)				0.79 (0.66, 0.96)
Stroke or SE (safety on treatment population)				0.79 (0.65, 0.95)
Stroke or SE or major bleeding	0.77 (0.69, 0.86)			
Stroke or SE or major bleeding or mortality (all-cause)	0.85 (0.78, 0.92)			
Stroke or SE or mortality (cardiovascular)				0.86 (0.74, 0.99)
Stroke or SE or mortality (cardiovascular) or MI				0.85 (0.74, 0.96)
Stroke	0.79 (0.65, 0.95)	0.91 (0.74, 1.12)	0.64 (0.51, 0.81)	0.85 (0.70, 1.03)
Ischaemic or unspecified stroke	0.92 (0.74, 1.13)	1.11 (0.88, 1.39)	0.76 (0.59, 0.97)	0.94 (0.75, 1.17)
Haemorrhagic stroke	0.51 (0.35, 0.75)	0.31 (0.17, 0.56)	0.26 (0.14, 0.49)	0.59 (0.37, 0.93)
SE	0.87 (0.44, 1.75)	0.71 (0.37, 1.38)	0.61 (0.30, 1.21)	0.23 (0.09, 0.61)
ICH	0.42 (0.30, 0.58)	0.30 (0.19, 0.45)	0.41 (0.28, 0.60)	0.67 (0.47, 0.93)
MI	0.88 (0.80, 0.97)	1.35 (0.98, 1.87)	1.38 (1.00, 1.91)	0.81 (0.63, 1.06)
GI bleeding	0.89 (0.70, 1.15)	1.10 (0.86, 1.41)	1.50 (1.19, 1.89)	
Major bleeding	0.69 (0.60, 0.80)	0.80 (0.69, 0.93)	0.93 (0.81, 1.07)	1.04 (0.90, 1.20)
Major or clinically non-major bleeding	0.68 (0.61, 0.75)			1.03 (0.96, 1.11)
Any bleeding	0.71 (0.68, 0.75)			
Critical bleeding		0.67 (0.54, 0.82)	0.67 (0.54, 0.82)	0.69 (0.53, 0.91)
Fatal bleeding				0.50 (0.31, 0.79)
Mortality (all-cause)	0.89 (0.80, 1.00)	0.91 (0.80, 1.03)	0.88 (0.77, 1.00)	0.85 (0.70, 1.02)
Mortality (cardiovascular)		0.90 (0.77, 1.06)	0.90 (0.77, 1.06)	0.89 (0.73, 1.10)

Abbreviations: SE=systemic embolism, ICH=intracranial haemorrhage, MI=myocardial infarction, GI=gastrointestinal
Source : Granger et al (2011), Connolly et al. (2009), Patel et al. (2011).

Overall, heterogeneity of patient populations between studies, and definition of outcomes are the key limitations of the pivotal RCTs. In particular, while the outcomes used for defining efficacy are comparable across the studies, the definition of bleeding events differs considerably [41-43]. For instance, the RE-LY study defined bleeding in terms of minor bleeding, while ROCKET-AF and ARISTOTLE defined bleeding in terms of non-major clinically relevant bleeding. This was reflected in the bleeding rates being evidently higher in ROCKET-AF and RE-LY than in the ARISTOTLE study [41-43].

Further, important baseline characteristics of patients included in the ROCKET-AF trial differed substantially from those reported in the ARISTOTLE and RE-LY trials. Although, rivaroxaban and warfarin baseline characteristics were balanced, over 80 % of patients in the ROCKET-AF study were at higher risk of stroke (CHAD2 score ≥ 3) compared to less than 33% in the ARISTOTLE and RE-LY clinical studies. Similarly, the proportion of patients with previous stroke, TIA or SE in the RE-LY was almost threefold higher than in patients included in the other two trials. A discrepancy was also observed in the proportion of patients with heart failure and diabetes, being almost double in the ROCKET-AF compared to those in the ARISTOTLE and RE-LY clinical trials [41-43].

In addition, as Al-Katib and colleagues indicated, there is uncertainty regarding the effect of apixaban compared to warfarin according to type and variation of AF. In particular, it is uncertain to what extent apixaban reduces the risk of stroke or SE, bleeding and mortality among patients with paroxysmal versus permanent AF [44].

1.8.2 Evidence from network meta-analysis

In addition to pivotal RCTs discussed in the previous sections (ARISTOTLE, RE-LY and ROCKET-AF), ten additional clinical studies [45-54] were analysed in a Network Meta-Analysis (NMA) to produce further evidence on the efficacy of DOACs [55]. Of those, five studies assessed the safety and the efficacy of either apixaban, dabigatran or rivaroxaban [47-50, 54] In NMA, multiple treatments are compared by means of direct and indirect comparison. In the latter case, comparison across trials is based on a common comparator [56]. For example, in most cases, the efficacy of DOACs in preventing stroke in the AF population was generated by comparing each DOAC against warfarin. Because warfarin is the common comparator, NMA allows conducting a series of indirect comparisons where, for instance, dabigatran or rivaroxaban could be compared against apixaban.

Only Phase II and Phase III trials evaluating the effectiveness of DOACs, VKA or antiplatelet drugs for the prevention of stroke in the AF population were included. Phase II are typically smaller than Phase III trials but still provide evidence on the efficacy of new treatments whenever findings from Phase III studies are not available. The network meta-analysis focused on five main oral anticoagulants namely: apixaban, dabigatran, rivaroxaban, edoxaban and betrixaban; but reporting the findings on the efficacy of betrixaban compared to warfarin was limited by insufficient evidence. Other anticoagulants (eribaxaban, otamixaban, darexaban, letaxaban and ximelagatran) were excluded from the analysis because administered parenterally or any other issue such as discontinuation or lack of info on the stage of clinical development [55].

Phase II studies included in the Network Meta-Analysis

Three out of seven Phase II studies included in the NMA compared either apixaban dabigatran or rivaroxaban with warfarin [48, 50, 54]

The ARISTOTLE-J study explored safety and efficacy of apixaban (2.5 and 5 mg) against warfarin in Japanese patients with non-valvular AF. The study population consisted of 222 patients aged 20 years or older with AF and at least one of the following risk factors for stroke: age 75 years or older, congestive heart failure, hypertension, diabetes and history of cerebral infarction or TIA. As in the ARISTOTLE study, both first time warfarin users and on-going users were enrolled [50]. The partially blinded study, (double blinded for apixaban, but open-label for warfarin), evaluated the safety of apixaban compared to warfarin for major and clinically relevant non-major bleeding. The efficacy was evaluated for the composites of stroke including SE and all-cause mortality, and the composite of MI or all-cause mortality. In the study, patients were randomised to receive apixaban 2.5 mg (74 patients), apixaban 5 mg (74 patients) or dose adjusted warfarin to obtain an INR <2.0 prior to entering the 12-week treatment period (74 patients). After a follow-up of 12 weeks, the occurrence of major and clinically relevant non-major bleeding was lower in the apixaban group (14% for both 2.5mg and 5 mg) compared to the warfarin group (5.3%). Apixaban also showed as better “safety profile” (a term generally used to refer to the safety of a drug) compared to warfarin, as no events of stroke, SE, MI, or all-cause mortality were observed in either apixaban group. On the contrary, two events of ischemic stroke and one event of subarachnoid haemorrhage were observed in the warfarin group [50].

The PETRO study assessed the safety of dabigatran, with or without concomitant aspirin, compared to warfarin in patients with non-valvular AF to identify the optimal doses for the Phase III clinical study. The study population consisted of 502 patients with AF at high risk for thromboembolic events and at least one of the following risk factors for stroke: age 75 years or older, symptomatic heart failure or left ventricular dysfunction, hypertension, diabetes and history of stroke or TIA [48]. The partially blinded study (double blinded for apixaban, but open-label for warfarin and concomitant aspirin) evaluated the safety of dabigatran (with or without aspirin) compared to warfarin for the primary outcome of major or clinically relevant bleeding. In the study, patients were randomised to receive twice-daily dabigatran 50 mg (105 patients), dabigatran 150 mg (166 patients), dabigatran 300 mg (161 patients), aspirin 81 or 325 mg (alone or in concomitance with dabigatran) or dose adjusted warfarin to maintain the INR between a range of 2.0 and 3.0 (70 patients) [48].

After a follow-up of 12 weeks, major bleeding events were experienced only by patients on dabigatran 300 mg plus aspirin (2.4%). The rate of the major bleeding, irrespective of the aspirin assignment, was lower in the dabigatran 50 mg group (1.9%), but higher in the 150 mg (7.7%) and 300mg (10%) groups respectively, compared to the warfarin group (5.7%). However, when aspirin assignment was taken into account, there was a significant difference in the rate of major or clinically relevant bleeding in the dabigatran 50 mg plus aspirin group (4.2%), compared to the group of patients treated with dabigatran 50 mg only, where no primary outcome events were observed [48]. Similar patterns were observed between patients on the dabigatran 300 mg plus aspirin and those who were on dabigatran 300 mg only, with the rate of the primary outcome being higher for the former group (17.2% versus 5.7%). However, the patterns were inverted for the dabigatran 150 mg plus aspirin group (5.8%) compared with the group on dabigatran 150 mg only (9%).

The overall rate of bleeding, irrespective of the aspirin assignment, was again lower in the dabigatran 50 mg group (6.5%), but higher in the 150 mg (17.8%) and 300mg (23.1%) groups respectively, compared to the warfarin group (17.1%). Further, stroke events were only experienced by patients on dabigatran 50 mg (1.9%), and the concomitance with aspirin did not seem to have a significant effect [48].

The third Phase II study, comparing a DOAC with warfarin, explored the dose response of dabigatran in patients with non-valvular AF in comparison with warfarin. The study population consisted of 166 patients 20 years or older with AF and at least one of the following risk factors for stroke: age 75 years or older, left side heart, hypertension, diabetes, history of coronary artery diseases, and history of stroke or TIA [54]. In the exploratory dose response study, patients were randomised to receive dabigatran 110 mg twice daily (46 patients), dabigatran 150 mg twice daily (58 patients) or dose-adjusted warfarin to maintain the INR between a range of 2.0 and 3.0 (or between 1.6 and 2.6 for 70 years or older patients) (62 patients) [54]. After a follow-up of 12 weeks, major bleeding events were only experienced by patients on dabigatran 150 mg (1.7%) and warfarin (3.2%). These patients were also receiving aspirin. The rate of major or clinically relevant bleeding was lower for the dabigatran plus aspirin patients (4.3% for the 110 mg and 8.6% for the 150 mg dabigatran) than in the warfarin treatment group (11.3%). The overall rate of bleeding was lower in the dabigatran 110 mg group (21.7%), but higher in the 150 mg (34.5%), compared to the warfarin group (24.2%). Further, stroke events were only experienced by patients on warfarin (3.2%) [54].

Phase III studies included in the Network Meta-Analysis

In addition to the main RCTs and Phase II studies, two additional Phase III studies comparing apixaban with aspirin and dabigatran with warfarin met the inclusion criteria of the NMA.

The J-ROCKET AF assessed the efficacy of rivaroxaban compared to warfarin in the AF Japanese population [49]. The study population consisted of 1,280 patients aged 20 years or older with AF, previous history of ischaemic stroke, TIA or SE or more than 2 of the following thromboembolism risk factors: age 75 years or older, congestive heart failure, and/or left ventricular ejection fraction, hypertension, diabetes. The study, using a non-inferiority design evaluated the efficacy of rivaroxaban (637 patients) compared to warfarin (637 patients) for the primary safety outcome of major non-major clinical bleeding and the primary efficacy composite outcome including stroke and SE. Patients were randomised to receive rivaroxaban 15 mg once daily dose, or dose adjusted warfarin to maintain the INR between a range of 2.0 and 3.0 (in patients younger than 70 years) or 1.6 and 2.6 (in patients 70 years or older) according to Japanese guidelines. Within the pre-specified maximum exposure period and the expected study duration of 2.5 years, no risk difference of the primary safety outcome was observed (HR 1.11 (95% CI 0.87, 1.42)); thus, the non-inferiority of rivaroxaban to warfarin in the safety on treatment population was confirmed. Rivaroxaban was also associated with a significant reduction in risk of the primary efficacy composite outcome (HR 0.49 (95% CI 0.24, 1.00)) compared to warfarin [49].

The AVERROES clinical study investigated the efficacy of apixaban compared to acetylsalicylic acid (Aspirin) in preventing stroke in the AF population unsuitable for vitamin K antagonist treatment [47]. The study population consisted of 5,599 patients aged 50 years or older with AF and at least one of the following stroke risk factors: age 75 years or older, symptomatic heart failure, left ventricular dysfunction or peripheral artery disease, hypertension, diabetes and history of stroke or TIA. The study, using a superiority design evaluated the efficacy of apixaban (2,808 patients) compared to aspirin (2,791 patients) for the primary outcome of stroke or SE. Patients were randomised to receive apixaban 5 mg twice-daily dose, or aspirin 81 to 324 mg once daily [47]. After a median follow-up of 1.1 years, apixaban 5mg showed a significant reduction in the primary outcome (HR 0.45 (95% CI 0.32, 0.62)) compared to aspirin. However, no risk difference of major bleeding (HR 1.13 (95% CI 0.74, 1.75)) and mortality (HR 0.79 (95% CI 0.62, 1.02)) was found for apixaban compared with aspirin [47].

The network meta-analysis evaluated the effectiveness of standard dose DOACs in different head-to-head comparisons, where dabigatran and rivaroxaban were compared with apixaban, and rivaroxaban with dabigatran. Based upon their consistency across studies, the amount of data available and the clinical relevance, the outcomes selected for the network meta-analysis were the following: stroke or SE, ischaemic stroke, MI, major bleeding, ICH, GI bleeding, clinically relevant bleeding and all-cause mortality [55].

Overall, the study found no risk difference of stroke or SE when dabigatran and rivaroxaban were compared with apixaban. However, rivaroxaban (OR 1.35 (95% CI 1.03, 1.78)) was associated with an increased risk of stroke and SE when compared against dabigatran. No difference in risk of MI was found across the DOAC to DOAC comparisons.

For what concerns the safety aspect, a significant risk increase of major bleeding compared with apixaban was found for dabigatran (OR 1.33 (95% CI 1.09, 1.62)) and rivaroxaban (OR 1.45 (95% CI 1.19, 2.33)). While there was no risk difference in ICH across all DOAC to DOAC comparisons, dabigatran (OR 1.71 (95% CI 1.21, 2.43)) and rivaroxaban, (OR 1.66 (95% CI 1.19, 2.33)) showed an increased risk of GI bleeding compared to apixaban. No risk difference in mortality was observed between DOACs. The NMA also looked at efficacy of Edoxaban, however as the focus of this thesis is on apixaban, dabigatran and rivaroxaban only, the findings on Edoxaban are not reported [55]. All findings on the efficacy of DOACs are reported in Tables 1.4

Table 1.4 Odd ratios from network-meta analysis comparing DOACs to DOACs

Outcome	Dabigatran (150 mg) vs. apixaban (5 mg) Odd ratio 95% CI	Rivaroxaban (20 mg) vs. apixaban (5 mg) Odd ratio 95% CI	Rivaroxaban (20 mg) vs. dabigatran (150 mg) Odd ratio 95% CI
Stroke or SE	0.82 (0.62, 1.08)	1.11 (0.87, 1.41)	1.35 (1.03, 1.78)
Ischaemic stroke	0.83 (0.59, 1.16)	1.01 (0.74, 1.38)	1.22 (0.87, 1.73)
MI	1.48 (0.98, 2.22)	0.92 (0.63, 1.34)	0.62 (0.41, 0.93)
Major bleeding	1.33 (1.09, 1.62)	1.45 (1.19, 1.78)	1.10 (0.90, 1.34)
ICH	0.96 (0.58, 1.60)	1.55 (0.97, 2.49)	1.61 (0.96, 2.72)
GI bleeding	1.71 (1.21, 2.43)	1.66 (1.19, 2.33)	0.97 (0.71, 1.33)
Clinically relevant bleeding	2.32 (0.74, 8.63)	1.53 (1.33, 1.75)	0.66 (0.18, 2.07)
Mortality (all-cause)	1.00 (0.84, 1.19)	0.94 (0.76, 1.17)	0.94 (0.74, 1.18)

Abbreviations: SE=systemic embolism, MI=myocardial infarction, ICH=intracranial haemorrhage, GI=gastrointestinal.
Source: Lopez et al (2017).

Standard dose apixaban was ranked as the most effective intervention for stroke, SE, MI and all-cause mortality. On the contrary, rivaroxaban was ranked the least effective. Apixaban was also ranked as the safest DOAC associated with the lowest incidence of major and GI bleeding [55]. These important findings provide further evidence that is added to the existing body of evidence.

The main limitations of the network meta-analysis were inherent to the primary data available and the assumptions made for the analysis. For instance, as few comparisons were present in more than one trial, it was not possible to fit a random effect model, thus allowing for heterogeneity in treatment effect determined by differences between treatment groups such patient characteristics or follow-up [55, 57]. Further, the reporting of the outcomes extracted for the review was incomplete; this not only reduces precision but also threatens the validity of the results of the NMA. However, the complete reporting of the relevant outcomes (stroke or SE, ischaemic stroke, MI, all-cause mortality, major bleeding, intracranial bleeding, and gastrointestinal bleeding) in the pivotal RCTs [41-43], gave some reassurance that bias due to selective reporting of outcomes was unlikely to affect the findings of the NMA [55]. Nevertheless, clinical relevant bleeding was not reported by the RE-LY trial [41]; this outcome however could be identified in real-world data, as RWE allows for the inclusion and analysis of clinical endpoints not measured in RCTs but observed in real life.

1.8.3 Real-world evidence

The health outcomes measured in a controlled environment, as in a RCT, may not reflect what happens in clinical practice, where the random allocation of treatments cannot be scientifically planned. While an RCT is typically the first step for evaluating the efficacy and the safety of a given intervention, estimating the effectiveness for the same intervention in a real-world scenario is equally important. RWE may provide a more comprehensive understanding of treatment safety and effectiveness by allowing for the inclusion and analysis of clinical endpoints not measured in RCTs but observed in real life.

In addition to the pivotal RCTs of DOACs, several RWE studies have been carried out. An existing systematic review (Ntaios, 2017) [58] identified 28 RWE studies evaluating safety and effectiveness of DOACs compared to any VKAs treatment (mainly warfarin) in the AF population. Only studies, identified in PubMed and the Web of Science until the 7th of January 2017, comparing more than one DOAC against VKA, and adequately addressing confounding by indication were included in the meta-analysis carried out in the existing systematic review [58]. Whenever two or more studies employed data coming from the same source, only the study with the longest study period was selected.

The outcomes assessed were ischaemic stroke, ischemic stroke or SE, all stroke (including ischaemic and haemorrhagic) SE, MI, ICH, GI bleeding, major bleeding and mortality. All DOACs were associated with a reduction in risk of ischaemic stroke or SE, but only the risk for the apixaban versus warfarin comparison was statistically significant. The statistically significant reduction in risk of ICH was observed across all treatment groups compared to VKA. While the reduction in risk for ICH confirms the findings from RCTs, no risk difference of ischaemic stroke or SE observed between apixaban and warfarin group diverged from the risk reduction reported in the ARISTOTE trial [42].

In the rivaroxaban treatment group, the risk of stroke or SE is matching that reported in the ROCKET-AF study only for the risk estimated with the ITT approach. Nevertheless, the meta-analysis supports the clinical non-inferiority for all DOACs in reducing the risk of stroke and composite compared to VKAs. The effect on reducing the risk of MI was not reported for apixaban, and no reduction in risk was found for dabigatran and rivaroxaban compared to VKAs. The finding provides further evidence against the supposed correlation between dabigatran and MI as observed in the RE-LY trial [58].

When evaluating the safety aspect of DOACs, the reduction and increase in risk of GI bleeding associated with apixaban and dabigatran respectively, compared with VKAs, was in line with the risk reported in the ARISTOTLE and RE-LY (dabigatran 150 mg) RCTs. The risk of major bleeding also matched the risk reported in the clinical studies, where a significant reduction in major bleeding is observed for apixaban, whereas dabigatran and rivaroxaban compared to VKAs are associated with a similar bleeding risk [41-43]. Similarly, the results on the risk of mortality are consistent with those reported in the RCTs. In the meta-analysis, apixaban and dabigatran were found to have a lower risk of mortality, but no statistically significant risk reduction was observed for rivaroxaban compared to VKAs [58]. Likewise, the ARISTOTLE and RE-LY (dabigatran 150 mg) trials reported a mortality risk reduction of 11% and 10% respectively; however, rivaroxaban was not associated with a reduction in the risk of mortality in the ROCKET-AF clinical study. Overall, the findings from the meta-analysis, reported on Table 5.1, suggest that DOACs are at least as safe as VKA in the AF population [58].

Table 1.5 Meta-analysis results

Outcome	Apixaban vs. warfarin HR 95% CI	Dabigatran vs. warfarin HR 95% CI	Rivaroxaban vs. warfarin HR 95% CI
Studies (N)	7	24	14
Ischaemic stroke	0.95 (0.75, 1.19)	0.96 (0.80, 1.16)	0.89 (0.76, 1.04)
Ischaemic stroke or SE	1.07 (0.87, 1.31)	1.17 (0.92, 1.50)	0.73 (0.52, 1.04)
All stroke or SE	0.67 (0.46, 0.98)	0.93 (0.77, 1.14)	0.87 (0.71, 1.07)
MI	not reported	0.96 (0.77, 1.21)	1.02 (0.54, 1.89)
ICH	0.45 (0.31, 0.63)	0.42 (0.37, 0.49)	0.64 (0.47, 0.86)
GI bleeding	0.63 (0.42, 0.95)	1.20 (1.06, 1.36)	1.24 (1.08, 1.41)
Major bleeding	0.55 (0.48, 0.63)	0.83 (0.65, 1.05)	1.00 (0.92, 1.08)
Mortality	0.65 (0.56, 0.75)	0.63 (0.52, 0.76)	0.67 (0.35, 1.30)

Abbreviations: SE=systemic embolism, MI=myocardial infarction, ICH=intracranial haemorrhage, GI=gastrointestinal.
Source: Ntaios et al. (2017)

Important strengths in the systematic review and meta-analysis, such as large sample size and the inclusion of studies adequately addressing confounding by indication, are acknowledged. However, several limitations inherent to the nature of administrative data such as residual confounding, a distortion that persist after controlling for confounding, still persist [58, 59].

In addition, missing information on patient characteristics, patient adherence and compliance to treatment, and details on the therapeutic range for warfarin users may undermine the robustness of the reported findings [58]. Another important limitation is the lack of distinction between different dose regimens in the analysis. As observed in in the RE-LY trial reduced dose of dabigatran did not have the same effect as the standard dose in reducing the risk of stroke [41]. Presumably, the risk of several outcomes associated with reduced doses may vary substantially in clinical practice.

Despite the limitations, the main results from Ntaios's systematic review and meta-analysis, support the main conclusions of the pivotal RCTs [41-43], and therefore strengthen their validity [58]. While, further evidence from RWE will still be likely to be limited by potential residual confounding, issues such as treatment adherence and compliance should be investigated, as poor compliance might expose patients to a greater risk of experiencing stroke or any other relevant clinical event [16-18]. In addition, info regarding the time spent within therapeutic range for patients on warfarin would allow for a more accurate comparative effectiveness analysis of DOACs versus warfarin in real life. In RCTs, the effect of DOACs may have been overestimated, as the therapeutic INR between a range of 2.0 and 3.0, was not maintained for some patients randomised to warfarin [41-43]. Further, estimating the treatment effect of DOACs according to different dose regimens may provide additional insights on how the risks of standard and reduced dose of DOACs, compared to warfarin, differ in clinical practice.

The effect of DOACs reduced dose in real-world settings was evaluated in two RWE studies, Gorst Rasmussen (2016) [60]and Nielsen (2017) [61]. Differentiating dose regimen in DOACs may be important as some evidence suggests a dose dependent effect. This was shown in the RE-LY trial assessing the efficacy of dabigatran standard (150mg) and reduced (110mg) dose [41].

The two Danish studies [60, 61], obtained the study data by merging three nationwide Danish registries: the Danish Civil Registration System covering demographic info, the Danish National Prescription Registry covering all prescription purchased since 1995, and the Danish National Patient Register covering all hospitalisation records since 1976 [60, 61]. Although these studies share many similarities in terms of study design and research questions, Gorst-Rasmussen (2016) evaluated the effectiveness of rivaroxaban against warfarin and dabigatran between February 2012 and August 2014 [60]. By contrast, Nielsen (2017) explored the effectiveness of any DOACs between August 2011 and February 2016, therefore assessing effectiveness in clinical practice within a much broader time-frame [61].

Gorst-Rasmussen (2016) found that at 2.5-year follow-up, rivaroxaban 15 mg, compared against warfarin, was associated with risk reduction for the composite of stroke or SE or TIA; but no risk difference was observed between warfarin and rivaroxaban standard dose. When evaluating the safety aspect by means of any bleeding, a risk difference compared to warfarin was not observed for either the standard or the reduced rivaroxaban dose. However, an important difference was seen between drug regimens with respect to the risk of all-cause mortality. While no difference in the risk of mortality was observed between rivaroxaban standard dose and warfarin, rivaroxaban 15 mg showed a statistically significant increase, Table 1.5 [60]. These findings were supported by the additional evidence coming from Nielsen’s study (2017), also reported in Table 1.5 [61].

Unlike, the two Danish studies [60, 61], Vinogradova (2018) [62], assessed the effectiveness of DOACs (excluding edoxaban) for different indications, and analysed the effectiveness of DOACs standard and reduced doses in patients with or without AF between 2011 and 2016, using UK primary care databases QResearch and Clinical Practice Research Datalink [62]. In the standard dose analyses comparing individual DOACs against warfarin, no difference in risk was found across the outcomes assessed. However, when assessing at the reduced doses, a reduction in the risk of ICH was observed for each DOAC. Further, the effect of apixaban on reducing the risk of major bleeding was reported but no reduction in risk was found for dabigatran compared to warfarin; however, patients on reduced dose rivaroxaban were found to be at greater risk of experiencing major bleeding than those on warfarin. Nevertheless, both reduced dose apixaban and rivaroxaban were associated with an increased risk of all-cause mortality [62].

While the findings may rise some concerns over the use of rivaroxaban, Larsen (2016) and Nielsen (2017) acknowledged that despite having controlled for confounding, residual confounding from unobserved factors may persist. Therefore, further research will be needed to access whether the association between the use of rivaroxaban and the risk of mortality is a true association [61, 63].

Table 1.6 Hazard ratios from additional RWE observational studies – DOACs vs. warfarin

Study	Outcome	Apixaban (5 mg) HR 95% CI	Apixaban (2.5 mg) HR 95% CI	Dabigatran (150 mg) HR 95% CI	Dabigatran (110 mg) HR 95% CI	Rivaroxaban (20 mg) HR 95% CI	Rivaroxaban (15 mg) HR 95% CI
Nielsen, 2017 (Denmark)	Ischaemic stroke		1.14 (0.89 - 1.46)		1.07 (0.92 - 1.23)		1.02 (0.77 - 1.34)
Vinogradova, 2018 (UK)	Ischaemic stroke	1.07 (0.79 -1.46)	1.16 (0.81 - 1.67)	1.37 (0.92 - 2.05)	1.06 (0.76 - 1.48)	0.97 (0.78 - 1.20)	1.27 (0.92 - 1.75)
Nielsen, 2017 (Denmark)	Ischaemic stroke or SE		1.15 (0.9 - 1.47)		1.07 (0.93 - 1.23)		0.99 (0.76 - 1.3)
Gorst-Rasmussen, 2016 (Denmark)	Stroke or SE or TIA					0.72 (0.51 - 1.01)	0.46 (0.26 - 0.82)
Nielsen, 2017 (Denmark)	Haemorrhagic stroke		0.76 (0.4 - 1.42)		0.60 (0.42 - 0.84)		1.25 (0.67 - 2.36)
Vinogradova, 2018 (UK)	ICH	0.41 (0.22 - 0.76)	0.44 (0.23 - 0.82)		0.48 (0.26 - 0.91)	0.85 (0.63 - 1.14)	0.86 (0.52 - 1.42)
Vinogradova, 2018 (UK)	GI bleeding	0.81 (0.58 - 1.13)	0.70 (0.45 - 1.09)	0.86 (0.51 - 1.44)	1.28 (0.94 - 1.74)	1.14 (0.93 - 1.40)	1.34 (0.98 - 1.83)
Gorst-Rasmussen, 2016 (Denmark)	Any bleeding					1.18 (0.90 - 1.55)	0.90 (0.59 - 1.35)
Nielsen, 2017 (Denmark)	Major bleeding		1.02 (0.74 - 1.41)		0.93 (0.80 - 1.08)		1.18 (0.93 - 1.5)
Vinogradova, 2018 (UK)	Major bleeding	0.62 (0.49 - 0.79)	0.68 (0.52 - 0.90)	0.79 (0.56 - 1.10)	0.93 (0.74 - 1.17)	1.06 (0.92 - 1.21)	1.25 (1.01 - 1.55)
Gorst-Rasmussen, 2016 (Denmark)	Mortality (all-cause)					0.93 (0.75 - 1.16)	1.47 (1.19 - 1.82)
Nielsen, 2017 (Denmark)	Mortality (all-cause)		1.61 (1.43 - 1.82)		1.05 (0.98 - 1.13)		1.43 (1.27 - 1.61)
Vinogradova, 2018 (UK)	Mortality (all-cause)	0.98 (0.83 - 1.15)	1.27 (1.12 - 1.45)	0.90 (0.68 - 1.19)	1.01 (0.87 - 1.18)	1.10 (0.99 - 1.21)	1.29 (1.14 - 1.45)

Note: only outcomes relevant to atrial fibrillation are reported.
SE=systemic embolism, TIA= transient ischaemic attack, ICH=intracranial haemorrhage, GI=gastrointestinal.

Additional evidence on the effectiveness of standard and reduced dose apixaban in real-world settings was provided by Proietti and colleagues (2018) [64]. The meta-analysis and systematic review identified 16 RWE studies evaluating the safety and effectiveness of apixaban compared to other DOACs and warfarin in patients with established AF diagnosis. Only studies, identified in PubMed and the Scopus database until the 6th of March 2017, comparing apixaban with warfarin or other DOACs were included in the analysis. The study evaluated the effectiveness of apixaban for stroke, major bleeding, any bleeding and ICH [64].

Although the study compared the effectiveness of apixaban with warfarin and other DOACs, because of the focus of this thesis, only results of the apixaban versus warfarin comparison are reported. In particular, Proietti's study found that apixaban was as effective as warfarin in reducing stroke; this was the case for both standard (OR 0.84 (95% CI 0.69, 1.01)) and reduced dose (OR 0.63 (95% CI 0.14, 2.81)) apixaban. Further, the risk of major bleeding was significantly lower for patients on standard (OR 0.64 (95% CI 0.51, 0.80)) and reduced dose apixaban (OR 0.55 (95% CI 0.36, 0.86)) than warfarin. Similarly, patients on apixaban had a lower risk of experiencing any bleeding with standard (OR 0.58 (95% CI 0.42, 0.81)) and reduced dose (OR 0.66 (95% CI 0.58, 0.76)) than those on warfarin [64].

1.9 Conclusion

This Chapter has discussed the role of RWE in the context of HTA; and has highlighted strength and limitations of RWE, which should be a seen as supplement to RCTs, where treatment efficacy is measured under specific conditions. Hence, RWE should be used for validating the robust evidence coming from RCTs; however, conflicting evidence may sometimes emerge. For instance, in Ntaio's study (2017) no risk difference of stroke or SE between apixaban and warfarin was observed, however the ARISTOTLE trial reported a statistically significant risk reduction [42]. Similarly, Gorst-Rasmussen (2016), Nielsen (2017), and Vinogradova (2018) reported an increased risk of all-cause mortality associated with reduced dose rivaroxaban. This finding diverged substantially from the risk reported in ROCKET-AF trial, where no risk difference of mortality was found for rivaroxaban compared to warfarin [43].

In these cases, clinicians should balance the evidence and assess whether RCTs and RWE studies look at the same population, patients characteristics and outcomes. In particular, clinicians may be interested the case-mix, treatment adherence, interactions with concomitant treatments [2, 7] or, as in the case of DOACs, the effect of standard and reduced dose apixaban and rivaroxaban compared to warfarin not investigated in the pivotal RCTs [41-43]. Thus, although residual confounding will always be a threat to the validity of RWE, important findings such as the increased risk of mortality with rivaroxaban should not automatically be discarded.

This Chapter has provided some background information propaedeutic for generating additional effectiveness evidence as pointed out in the research questions postulated in the next Chapter.

Chapter 2 Aim and objective

2.1 Aim of the thesis

This thesis aims to explore opportunities and challenges in using RWE to support HTA, and in particular, to explore the feasibility of using Scottish link health data in supporting HTA decision-making in Scotland. This will be achieved by introducing a case study of DOACs used for the prevention of stroke in the AF population.

This thesis intends to meet the following objectives:

Objective 1: Estimate and examine the composition of direct and indirect medical costs in AF using Scottish linked health data.

Objective 2: Explore, with a focus on PS based methods and the DOACs case study, methodological challenges in using RWE to estimate comparative-effectiveness.

Objective 3: Estimate the comparative-effectiveness of DOACs to understand their effectiveness in Scottish clinical practice.

Objective 4: Update, with RWE data, existing cost-effectiveness analysis of DOACs for the prevention of stroke in patients with AF in Scotland. The introduction offers an overview of all available evidence on DOACs in preventing stroke and other relevant clinical events in the AF population. The remaining thesis is structured as follows:

Chapter 3: introduces typical data linkage processes. In particular, the data used in the different analyses of this thesis are presented and data characteristics, strength and limitations are discussed.

Chapter 4: describes different regression models typically employed in healthcare costing studies, these are then empirically tested and compared. The AF cost analysis, carried out using the best fitting model identified, assesses how AF related healthcare costs vary according to demographics, comorbidities, geography and socio-economic status. The analysis also estimates how different cost components contribute to the overall healthcare cost.

Chapter 5: describes the different propensity score methods typically used in health research to balance patients' baseline characteristics between treatment groups, and estimate the treatment effect. The choice of one approach over another should not be arbitrary and all possible options should be tested. Thus, different methods are compared and applied to the linked Scottish data; the most robust method is selected for addressing objectives of Chapter 6.

Chapter 6: evaluates effectiveness and safety of DOACs in a time-to-event analysis at 2- and 6-year follow-up. In addition, assumptions for the validity of the model are tested, and alternatives to the chosen model discussed. The findings are compared against the existing evidence reported in other observational studies and pivotal RCTs.

Chapter 7: focuses on the economic evaluation of DOACs, where DOACs are compared pairwise and incrementally against warfarin, to date the main thromboprophylactic treatment for the prevention of stroke in patients affected by AF. Firstly, a review of existing cost-effectiveness models allows for the identification of a cost-effectiveness model reflecting clinical, costs and utility assumptions pertinent to Scotland. Then, the economic evaluation assesses, with the identified cost-effectiveness model, whether DOACs improve the quality of life of patients; and if so, to what extent and at what additional cost.

Chapter 8: summarises, with a focus on strengths and limitations, the main findings of this thesis. It discusses the possible direction of future research, and the implication these results may have on healthcare policies and clinical guidelines related to AF.

Chapter 3 Data source

3.1 Introduction

The data used for this thesis were obtained from ISD, and all the analyses were carried out in the National-Safe-Haven platform, a secure environment used to maintain the privacy and confidentiality of the personal information held [15]. This Chapter will describe the different ISD databases and cohorts used in this thesis, and the different steps carried out in the data cleaning process.

3.2. Data linkage

Data linkage is a complex technique used for linking records from one or more datasets by means of personal identifiers such as medical record number and demographic data including name, sex, and date of birth. In the data linkage process, the personal identifiers are matched, deterministically or probabilistically, to a population spine representing the core dataset to which other datasets can be linked [65, 66]. ISD datasets are accurately linked by means of the Community Health Index (CHI), used in Scotland to uniquely identify a person for healthcare purposes, and through well-established probability matching techniques based on Howard Newcombe principles [14, 67].

With the deterministic approach, unique personal identifiers such as social security or electoral number are regarded as truly unique, thus generating definite matches between any two records. However, due to potential coding errors, some true matches may be missed; this may happen for instance, when a unique identifier common to two datasets is used, but coding errors in one dataset may prevent records matching. The more commonly used probabilistic approach employs non-unique identifiers such as name and date of birth, in a weighting system indicating the likelihood that two records belong to the same individual [65, 66].

3.3 Data source

3.3.1 Inpatient admissions

SMR01 records contain all general acute admissions, categorized as inpatients or day cases, discharged from non-obstetric and non-psychiatric specialties. While inpatient admission implies a hospital stay overnight, day cases refer to a planned attendance to a specialty for clinical care and generally, it does not require patients to stay in the hospital overnight. Upon completion of a hospital episode defined from date of admission to the date of discharge, and regardless of whether it is inpatient or day case, an SMR01 record is generated [68]. Each episode includes non-clinical data reflecting demographic and episode management details describing date, reason, type of admission, and structures where patients were admitted from or transferred to. In particular, the type of admission would indicate whether a patient was admitted as a planned (elective) or with an emergency admission (non-elective). The details on admission/transfer would give an indication on the type of location, such as private residence, institution, same or different clinical specialty, from which an individual came from prior to hospital admission.
Each episode would also include the date, type of discharge and from where patients were discharged or transferred to. The discharge type specifies whether discharge from an inpatient or day case episode was regular or resulted from self-discharge or death [68].
In addition, the discharge/transfer aspect of an episode indicates, as for admission, type of location to which a patient is discharged or transferred to following an episode of care, or whether a patient has died. Further, for every episode the diagnostic code is recorded using the International Classification of Diseases, Tenth Revision (ICD-10) developed in 1992 by the WHO and implemented in Scotland in 1996. ICD-10 is an index of diseases and injuries used to compare conditions for epidemiological and health management purposes.

Within the SMR01 context, ICD-10 codes are reported for the main diagnosed condition (primary diagnostic position) followed by up to five additional diagnostic codes, which can describe co-morbidities [15, 68]. While ICD-10 seems to be an accurate coding system, it is argued that the increase from 17,000 codes in the previous version ICD-9 to 141,000 codes may have introduced some unnecessary complexity [69, 70]. Similarly, operations, procedures and interventions are recorded using the Office of Population Censuses and Surveys, fourth revision (OPCS-4) index. Maintained and developed by the Health and Social Care Information Centre classification service in 1987, OPCS-4, now in its sixth revision, was implemented in NHS Scotland in 2014 [71]. Further, each episode in SMR01 includes details on patients' demographics and socio-economic status such as the Scottish Index of Multiple Deprivation (SIMD). Details on health boards and geographical location expressed as urban rural classifications are also included. SMR01 also comprises Healthcare Resource Groups (HRGs) used for costing each hospital episode.

3.3.2 Outpatient attendance

SMR00 records include, at patient level, information on new and follow-up appointments at outpatient clinics for any clinical specialty. Outpatient clinics are generally located within hospital outpatient departments or a health centre [72]. As opposed to inpatient admissions, outpatient attendances do not imply a hospital overnight stay. This however differs from day cases where, although staying in hospital overnight is not required, the use of a bed is needed. Further, each episode in SMR00 includes two typologies of data: non-clinical and clinical. While the first typology includes appointment management data, attendance status, and as in SMR01, patients' demographics, socio-economic status, health board area and geographical location; the second typology indicates the referral reason and operations/ procedures recorded using the ICD-10 and OPCS-4 systems.

Although, there has been provision to record outpatient procedures since 1997, only from the 1st of April 2003 recording became mandatory. However, at the present time, the poor level of completeness of CHI numbers or other patient identifiers, still makes linking SMR00 with other datasets rather challenging. In addition, because not all fields are mandatory, such as length of waiting list, patients' ethnicity and referral reason, the information present in SMR00 is still limited. Nevertheless, improving the procedure recording system aimed at reducing the inconsistency in recording practice between health boards is on-going [72].

3.3.3 Prescribing information system

The PIS database includes prescribing records for all medicines and their associated costs, which are prescribed and dispensed by community pharmacies, dispensing doctors and a small number of specialist appliance suppliers [73]. Although the PIS database includes information starting from 1993, it is only possible to carry out analyses involving CHI related data since April 2009. In addition to the CHI number, PIS also includes prescription drugs related information such as cost, dosage, formulation code and strength, as well as details on prescribers and dispensers. Prescriptions issued in hospitals and dispensed in the community are also recorded in PIS, prescriptions dispensed in hospitals however are not included [73]. Another important aspect of the database is the inclusion of data indicating whether a prescription was prescribed and dispensed. While prescribing authorizes the use of prescriptions, dispensing indicates the actual number of prescriptions dispensed. The quality of PIS data is guaranteed by an electronic system, which has eliminated errors linked to manual data entry processes, and by several stages of record quality checking before and after they are submitted to PIS [74].

Further, the reimbursement system, which allows dispensers to be reimbursed only by submitting prescriptions for payment, guarantees a high level of individual level data completeness and availability. Nevertheless, some differences exist depending on the type of prescribers and prescriptions. For instance, for the year 2014, 98.7% of prescription data coming from general practitioners were complete and available, while those coming from dentists stayed at 1.6%. Similarly, and for the same year, the proportion of complete prescription data for cardiovascular drugs was 98.6%, comparatively for immunological products and vaccines was 71.5% [74]. In terms of coverage, PIS includes all age and sex groups, as well as socioeconomic status and geographical locations.

Although PIS population-level longitudinal data, are essential for answering research questions related to drug utilization and health outcomes, weaknesses regarding data not captured and their availability, may limit some applications. For instance, prescriptions administered to patients during inpatient stays and short-term post discharge periods, are not captured by PIS. The same applies to some specialist drugs such as growth hormone therapy and biologics administered for chronic conditions, or outpatient supplies produced by the hospital service. In addition, records on vaccines are limited, as the majority of the information may be captured by the childhood immunization database. Also, PIS does not hold information on drugs indications; for instance, DOACs may either be prescribed to AF patients or to those affected by deep vein thrombosis. Further, the low percentage of patient identifiers captured before 2009, poses limitations for carrying out longitudinal studies with PIS individual-level data for any time-period between 1993 and 2008 [73, 74].

3.3.4 Care home census

The Scottish Care Home Census is collected on an annual basis, and covers all adult care home institutions registered with the Care Inspectorate, which is responsible for regulating several social services across Scotland. The census, firstly issued by the Scottish Government in 2003, combines the former Residential Care Home Census (run by the Scottish Government) and the Private Nursing Homes Census (run by ISD Scotland) [75]. This came as a result of integrating health and social care services in Scotland according to the Public Bodies (Joint Working) (Scotland) Act 2014. The integration should guarantee a consistent provision of quality and sustainable care services in particular for the increasing number of individuals in Scotland with long-term conditions [76]. Typical items collected for the care home census are discharge dates to care home residency such as NHS and private nursing home, as well as indication on whether nursing care is required.

3.3.5 Mortality records

Mortality records for deaths occurring between 1997 and 2015 were obtained from the NRS, a database including date of death, and information on: demography, civil status. Each death in Scotland is registered with an allocated ICD-10 code.

In 2015, systems were introduced for improving the quality of data collection through independent checks on the accuracy of medical certificates of the cause of death, and there is already evidence that NRS data are of higher quality compared to NHS data [77].

3.4 Cohort identification

The data for cohort identification and the analyses addressing the objectives outlined in Chapter 2 were obtained from ISD of NHS Scotland. The data extraction process carried out by ISD is presented graphically in Figure 3.1.

Two cohorts where specified, one consisting of patients with a diagnosis of AF or atrial flutter (ICD10 code I48), the other consisting of patients on any Oral Anticoagulant (OAC) according to British National Formulary (BNF) section 2.8.2.

The initial AF cohort consisting of 279,883 patients with a known diagnosis of AF or atrial flutter were identified from the general acute inpatient and day case Scottish morbidity records 01 (SMR01) for the 1997 – 2015 study period [68]. The rationale for starting data extraction from 1997 was determined, not only by the necessity of gathering information on patients' prior medical history, but also by coding consistency issues. For instance, coding and record formatting of ICD-10 codes in SMR01 for diagnostic purposes, implemented in Scotland in 1996, have started to be consistent since 1997 [15].

A cohort of patients receiving any OAC consisting of 166,182 individuals was identified from the Scottish Prescribing Information System (PIS) for the 2009 – 2015 study period [72]. Of those on OAC, 160,630 had hospital records; this indicates that 5,552 patients who have been on OAC have never been admitted to hospital during the study period. Further, of the OAC patients with a hospital record, 92,916 patients were admitted to hospital with AF in any diagnostic position. The ISD, by means of CHI numbers (a unique 10-character numeric identifier allocated to each patient) the specified AF and OAC cohorts to PIS, SMR00, SMR01, the care home census and the National Records of Scotland (NRS) [75].

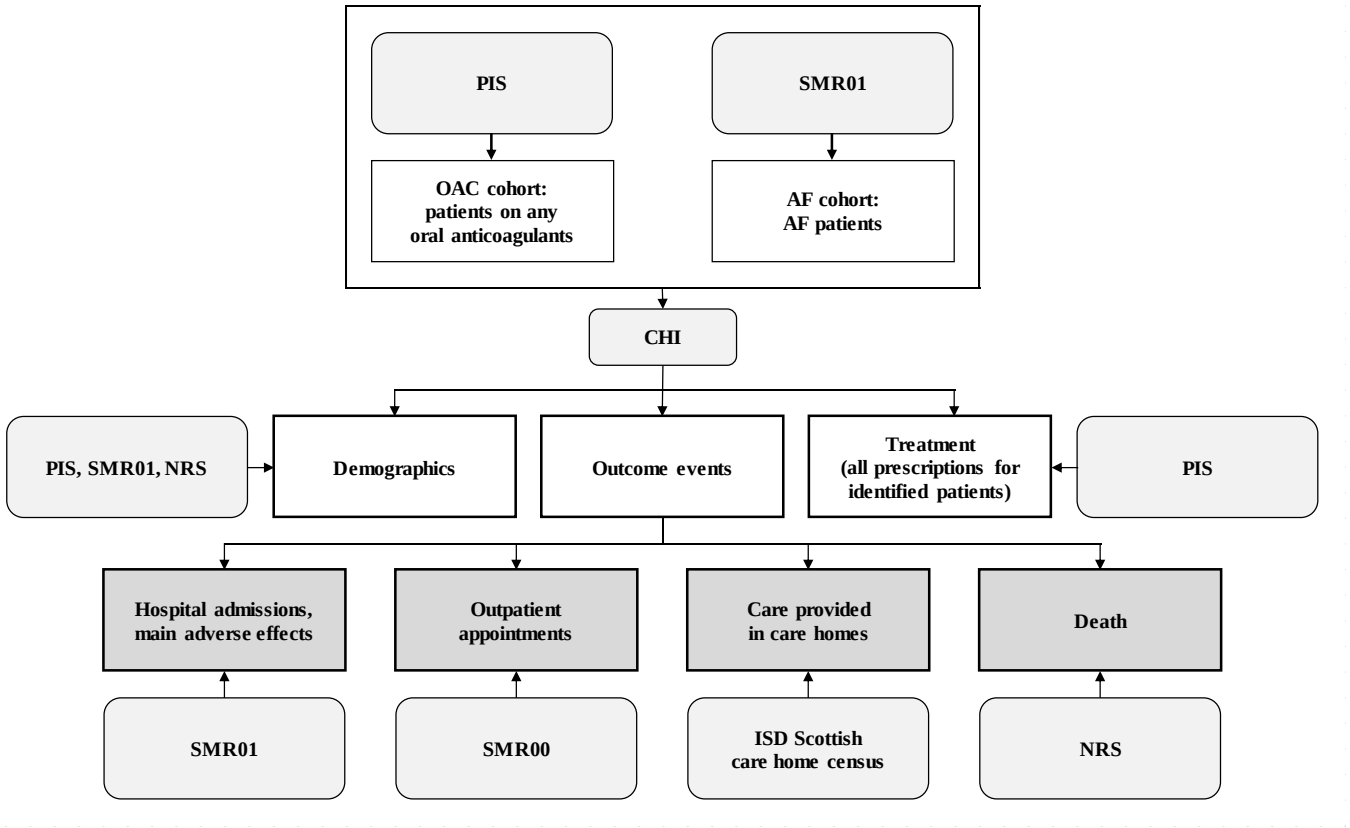


Figure 3.1 Cohort identification and data extraction

Note: Patients on any OAC and patients with AF or atrial flutter are identified from PIS and SMR01. Patients are then linked, by CHI number, to PIS, SMR00, SMR01, care home census and NRS to obtain info on demographics, outcomes events and prescribing.
Abbreviations: OAC= oral anticoagulant, PIS=prescribing information system, SMR01=Scottish morbidity records - general/acute inpatient day case dataset, SMR00=Scottish morbidity records - outpatient appointments and attendances, NRS=national health records.

3.5 Data cleaning

Prior to the analyses, the morbidity records (SMR01 and SMR00) where checked for quality and consistency. Records from health boards other than Scotland were removed. Precisely, 6,949 and 6,840 records were removed from SMR01 and SMR00 respectively. Further, duplicates were removed from SMR01 if date of admission, date of discharge, name of specialty, and ICD code for the first diagnosis were the same when comparing two or more episodes for the same patient. Similarly, SMR00 records were removed if clinic attendance date and specialty of the outpatient clinic attended of two or more episodes were identical. Including referral reason could have also been useful in identifying duplicates; however, as entering this variable is not compulsory when recording outpatient attendances, referral reason was missing for the majority of patients. Following quality control, the final number of hospital records was 4,223,817 representing 345,881 patients. While the total number of outpatient records was 7,532,181 for 337,600 patients (Figure 3.2). These cohorts were instrumental for identifying the relevant patients for this thesis and were further refined for analyses in the next Chapters.

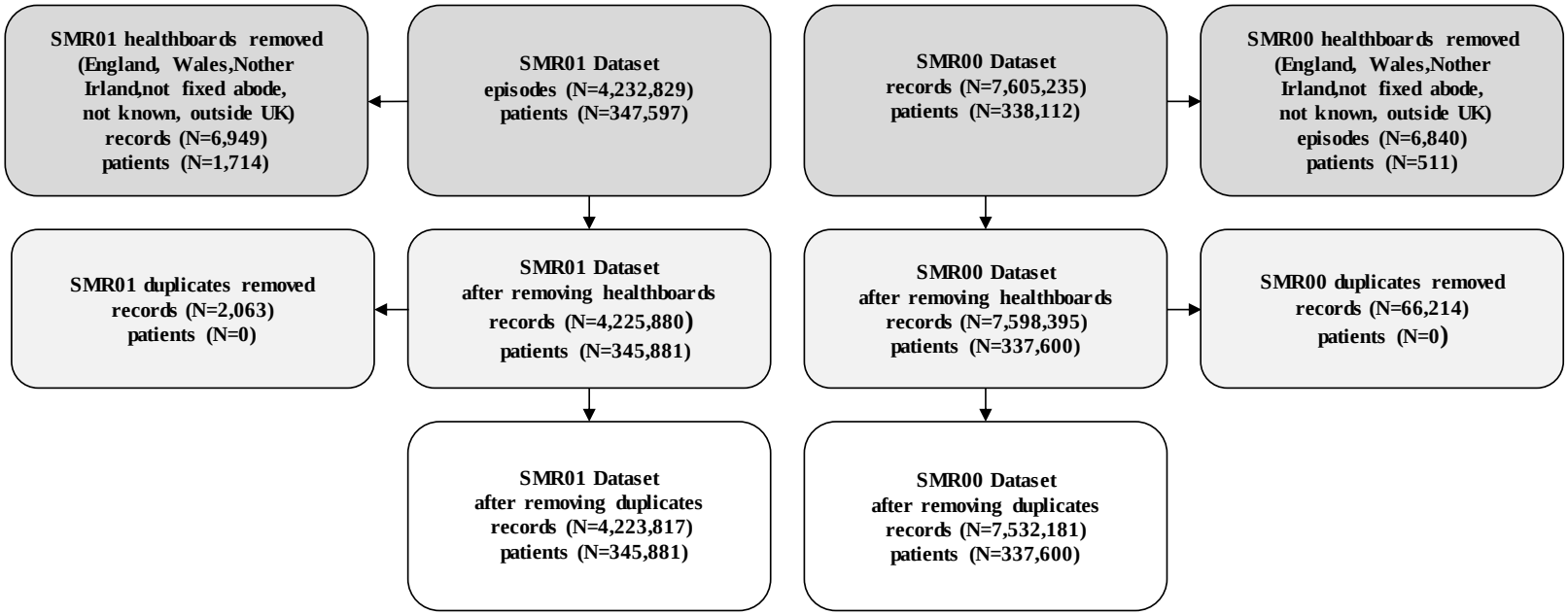


Figure 3.2 Data cleaning

Abbreviations: SMR01=Scottish morbidity records - general/acute inpatient day case dataset, SMR00=Scottish morbidity records - outpatient appointments and attendances.

3.6 Cohort descriptive

SMR01 – inpatient admissions

Cohort characteristics of patients who have been admitted to hospital are presented in Table 3.1. Of the 160,405 patients on any OAC attending outpatient clinics and with a mean age of 68 years (SD 14.6), the majority were identified in the two largest health board areas accounting for 20.7% and 13.7% respectively. This is also reflected in the categorisation of geographical areas, where large urban represented 35.8% and other urban areas represented 29.8% of the total OAC cohort. No clear distinction was observed in the proportion of patients living in the most or the least deprived areas; however, there seem to be a lesser concentration of patients in the least deprived area. Among the OAC cohort, general medicine (32.1%) followed by cardiology (9.1%) were the most frequently attended hospital specialties.

As recorded in the data, most patients were on OAC treatment as a result of being diagnosed with AF (278,286); therefore, similar baseline characteristics were observed. This was particularly evident for health boards and the categorisation of geographical areas. However, the AF cohort was on average older than the OAC cohort. Comparatively to what is observed for the OAC cohort, geriatric medicine (10.3) was the second most attended hospital specialty. Further, a clear distinction between the proportions of AF patients living in the most and least deprived was observed. The greatest proportion live in areas belonging to the most deprived category 22.5%, while those living in the least deprived areas account for 16.6% of the total AF cohort.

In addition, the distribution suggest that the comorbidities are more severe in the AF cohort than in the OAC cohort. The comorbidities were assessed by means of the Charlson Comorbidity Index (CCI), and categorised, according to clinical advise, in a way to indicate absence of comorbidities, presence of only a single comorbidity, and the presence of more than one comorbidity [78]. More details on the CCI is presented in the next chapter.

Table 3.1 Baseline characteristics of OAC and AF patients - SMR01 (1997-2015)

Baseline characteristics and outcomes (at the time of first inpatient admission)	Oral anticoagulants N (%)	Atrial fibrillation N (%)
Number of patients	160,405	278,286
BASELINE CHARACTERISTICS		
Mean age *(SD)**(range)	68 *(14.6) **(0-105)	74 *(11.9) **(0-105)
Sex		
Male	86,318 (53.8)	139,928 (50.3)
Female	74,087 (46.2)	138,358 (49.7)
Health Boards		
Greater Glasgow & Clyde	33,250 (20.7)	61,822 (22.2)
Lothian	21,478 (13.4)	41,169 (14.8)
Lanarkshire	18,529 (11.6)	31,049 (11.2)
Grampian	16,677 (10.4)	25,728 (9.3)
Tayside	14,930 (9.3)	22,003 (7.9)
Ayrshire & Arran	11,816 (7.4)	25,003 (9.0)
Highland	11,808 (7.4)	17,954 (6.5)
Fife	10,242 (6.4)	18,929 (6.9)
Forth valley	8,458 (5.3)	13,664 (4.9)
Dumfries & Galloway	6,152 (3.8)	9,798 (3.5)
Borders	4,333 (2.7)	7,222 (2.6)
Western isles	1,241 (0.8)	1,868 (0.7)
Shetland	764 (0.48)	1,036 (0.4)
Orkney	727 (0.5)	1,041 (0.4)
Geography		
Large/urban	57,368 (35.8)	106,868 (38.4)
Other/urban	47,845 (29.8)	82,601 (29.7)
Accessible small towns	14,366 (9.0)	24,938 (9.0)
Remote small towns	4,919 (3.1)	8,272 (3.0)
Very remote small towns	2,499 (1.6)	3,828 (1.4)
Accessible rural	19,506 (12.2)	30,826 (11.1)
Remote rural	6,606 (4.1)	10,371 (3.7)
Very remote rural	6,945 (4.3)	10,087 (3.6)
SIMD (Scottish index of multiple deprivation)		
1	32,213 (20.1)	62,730 (22.5)
2	33,765 (21.1)	62,632 (22.5)
3	33,163 (20.7)	55,943 (20.1)
4	31,926 (20.0)	50,691 (18.2)
5	29,337 (18.3)	46,279 (16.6)
Comorbidity		
no comorbidity	45,629 (28.5)	40,502 (14.6)
1 comorbidity	32,919 (20.5)	53,651 (19.3)
>1 comorbidity	81,857 (51.0)	184,133 (66.2)
OUTCOMES		
Median length of stay	3.6 (2.3 - 5.8)	5.4 (3.1 - 9.6)
Five most frequent specialties attended		
General medicine	630,112 (32.1)	1,207,536 (34.7)
Cardiology	178,656 (9.1)	280,577 (8.1)
General surgery (no vascular)	132,567 (6.7)	203,767 (5.9)
General surgery	120,116 (6.1)	187,530 (5.4)
Geriatric medicine	115,218 (5.9)	359,027 (10.3)

SMR00 – outpatient attendances

The baseline characteristics for patients on any OAC or with a known diagnose of AF and that have been attending outpatient clinics are presented in Table 3.2. Similar patient characteristics were observed between those who had been hospitalised, and those who had also been attending outpatient clinics. Of the 163,654 patients on any OAC attending outpatient clinics and with a mean age of 66 years (SD 14.3), the majority were identified in the two largest health board areas accounting for 20.7% and 13.6% respectively. When looking at geographical areas, large urban and other urban areas represented respectively 36.6% and 29.3% of the total OAC cohort attending outpatient clinics. No clear distinction was observed in the proportion of patients living in the most or the least deprived areas. Cardiology represented the second most attended outpatient clinic (10.3%) preceded by trauma and orthopaedic surgery (13.5%).

Baseline characteristics similarities and differences, between the OAC and AF cohorts, are similar to the ones observed in the SMR01 inpatient admissions. While this is true for age, health boards, geography and SIMD, there were differences concerning the most frequent outpatient clinic attended. Comparatively to what is observed for the OAC cohort, ophthalmology (13.1) was the second most attended outpatient clinic.

Table 3.2 Baseline characteristics of OAC and AF patients - SMR00 (1997-2015)

Baseline characteristics and outcomes (at the time of first outpatient attendance)	Oral anticoagulants N (%)	Atrial fibrillation N (%)
Number of patients	163,654	266,122
BASELINE CHARACTERISTICS		
Mean age *(SD)**(range)	66*(14.2) **(0 -105)	72*(11.8) **(0 - 108)
Sex		
Male	88,315 (54.0)	134,967 (50.7)
Female	75,339 (46.0)	131,155 (49.3)
Health Boards		
Greater Glasgow & Clyde	33,891 (20.7)	59,926 (22.5)
Lothian	22,328 (13.6)	40,026 (15.0)
Lanarkshire	18,712 (11.4)	29,492 (11.1)
Grampian	17,044 (10.4)	24,743 (9.3)
Tayside	15,144 (9.3)	20,943 (7.9)
Ayrshire & Arran	12113 (7.4)	17,248 (6.5)
Highland	11,889 (7.3)	23,235 (8.7)
Fife	10,542 (6.4)	18,138 (6.8)
Forth valley	8,609 (5.3)	12,665 (4.8)
Dumfries & Galloway	6,230 (3.8)	9,160 (3.4)
Borders	4,437 (2.7)	6,992 (2.6)
Western isles	1,234 (0.8)	1,648 (0.6)
Shetland	761 (0.5)	952 (0.4)
Orkney	720 (0.4)	954 (0.4)
Geography		
Large/urban	59,863 (36.6)	104,732 (39.4)
Other/urban	48,009 (29.3)	77,679 (29.2)
Accessible small towns	14,256 (8.7)	23,223 (8.7)
Remote small towns	4,803 (2.9)	7,494 (2.8)
Very remote small towns	2,465 (1.5)	3,510 (1.3)
Accessible rural	19,977 (12.2)	29,451 (11.1)
Remote rural	6,853 (4.2)	9,982 (3.8)
Very remote rural	7,103 (4.3)	9,587 (3.6)
SIMD (Scottish index of multiple deprivation)		
1	32,969 (20.2)	60,124 (22.6)
2	34,485 (21.1)	59,938 (22.5)
3	33,832 (20.7)	53,675 (20.8)
4	32,525 (19.9)	48,194 (18.1)
5	29,843 (18.2)	44,186 (16.6)
OUTCOMES		
Outpatient attendance per year (IQR)	5.7 (3.3 -7.6)	3.4 (2.1 - 5.5)
Five most frequent outpatient clinic attended		
Trauma & orthopaedic surgery	22,059 (13.5)	32,140 (12.1)
Cardiology	16,773 (10.3)	25,837 (9.7)
Ophthalmology	15,769 (9.6)	34,775 (13.1)
General medicine	13,847 (8.5)	23,805 (9.0)
General surgery	11097 (6.8)	n/a
Ear & nose & throat (ENT)	n/a	18,326 (6.9)

Chapter 4 The inpatient, outpatient, prescribing and care home costs associated with Atrial Fibrillation

4.1 Introduction

As highlighted in Chapter 1, AF is the most common form of arrhythmia affecting 1.8% of the adult Scottish population and rising to 6% in individuals aged over 65 years [79]. The risk of stroke, a highly debilitating condition, is significantly greater in individual affected by AF. Therefore, AF, in an ageing population, is likely to have an important impact on the economic burden of the healthcare system. A number of cost analyses on estimating the economic burden of AF exist. The majority of these studies are of selective cohorts of AF patients, based on data sourced from administrative database [80-82], health insurance databases [80, 83-85], hospital records [86, 87] and surveys [88]. Direct medical costs related to inpatient admissions, outpatient visits, as well as prescriptions have been included in these estimates; [80-88] indirect costs related to loss of productivity have been estimated among patients who were at working ages [84, 85].

Many of these studies included relatively young patients – those aged 18-20 years or older [80, 82-84, 86-88], or those under the age of 65 years [85]. However, the prevalence of AF increases significantly with age, and in particular in the 50 years or older age groups; in patients under the age of 50, AF is often associated with structural heart disease, hyperthyroidism, or alcohol excess [89]. Hence, inclusion of younger patients and the exclusion of older patients may result in imprecise cost estimates. There is a lack of generalisable studies based on large national population datasets that examine the total and the distribution of costs associated with AF [90].

To address the first Objective formulated in Chapter 2, the cost analysis carried out in this Chapter quantifies the inpatient, outpatient, prescribing and care home costs associated with AF over a five-year period. In addition, the composition of costs that are attributable to AF are examined using record-linkage of national datasets from Scotland described in Chapter 3.

4.2 Existing evidence on cost of AF

Different cost off illness studies exist that have used different approaches to estimate the cost of AF. This section describes the identified studies according to the following: population, data source, covariates, cost components and statistical method used. A summary is presented in Table 4.1.

4.2.1 Population

In the majority of the studies on cost estimation, the population comprises AF patients who are 18 years or older [80, 82-84, 86, 88]; while the remaining studies included the over 65 age group [81, 85, 87] or the entire AF population [91]. Further, among those studies patients were included according to certain characteristics. For instance, Ghate (2011) included AF patients with and without bleeding events for their analysis of healthcare costs related to warfarin-associated bleeding [83]. In an additional American study, the cohort included individuals diagnosed with AF and at least one pharmacy claim for warfarin [82].

4.2.2 Data

Across the different country specific studies, costs were primarily estimated with health insurance and hospital based administrative data. For instance, Bennel and collegues (2015) analysed the Ontario Health Insurance Program for extracting data on physician services. Data on prescription drugs were obtained from the Ontario Drug Benefit database [80]. Ghate (2011) extracted data from Thomson Reuters Medstat MarketScan Commercial Claims & Encounters and Medicare Supplemental & Coordination of Benefits database. These databases include inpatient, outpatient and prescription drug services patient level data obtained from health plans, government and public organisation [83, 92]. Four other additional US studies obtained data from different sources [82, 84, 85, 87]. In detail, Nelson and colleagues (2015), extracted the data from the Veterans Health Administration database, the largest integrated healthcare system in the USA, providing care for veterans across the country [82]. Reynolds (2017) obtained data from an AF bespoke database covering costs, therapies, adverse events and lifestyle related to AF [87]; while data in Rohrbacker (2010) and Wu (2005) studies were obtained from privately insured administrative databases [84, 85].

Amongst the European studies, Hallinen and colleagues (2006), in a Finnish study, estimated AF related costs using health service administrative data [81], while Holstenson (2011) employed hospital based administrative data from different European countries [86], and Jonsson patient and physician surveys [88].Stewart's study (2004) [91], also obtained data from different sources. In particular, data on prevalence, costs and healthcare utilisation were extracted from epidemiological studies and government datasets. More specifically, costs regarding community-based healthcare, hospital based healthcare and long-term care were used for estimating the overall cost of AF for the years 1995/1996.

For the community-based healthcare, Scottish healthcare costs were used to estimate the cost of a visit to a hospital outpatient clinic. Further, the costs for prescription drugs were obtained from Intercontinental Medical Statistics Ltd and crosschecked with government prescription data. The average cost of a visit to an anticoagulation clinic and any additional cost for monitoring of INR) for each visit were extracted from the Unit Costs of Health, and Social Care 1998 report produced by the Personal Social Services Research Unit [91].

4.2.3 Covariates

The entirety of the cost estimation models [80-88], aside from Stewart‘s model [91], were adjusted for demographics and baseline comorbidities and risk factors such as hypertension, diabetes, history of stroke, and any underlying heart disease. Wu and colleagues (2005) also looked at the insured population but focused more on the employment status for estimating healthcare related costs taking into account, in addition to demographic and comorbidities, indirect costs associated with unemployment.

4.2.4 Statistical methods

The majority of the studies used a Generalised Linear Model (GLM) model with a log link distribution [80-84, 88]. While different methods are used for estimating cost data, GLM take into account the characteristics of healthcare cost data, which in general do not follow a normal distribution and often present a skewed distribution to the right. This suggests that a small number of patients are making use of healthcare resources with very high costs [93].

Hallinen and colleagues (2006) tested GLM among OLS and log-transformed OLS, and found the latter as the best fitting cost model for their data [81]. OLS was also the model of choice for the European study predicting the related to cardiovascular disease in patients with AF [86], and the American study estimating the healthcare resource utilisation and costs associated with recurrent episodes of AF [87]. By contrast, Wu (2005) employed a multivariate two-part regression model. The study, did not specify whether a probit or a logit model was employed for estimating the probability of using a healthcare service, or whether a OLS or GLM was fitted in the second part of the model for estimating costs conditional on having incurred positive costs [85]. The technical aspects of these methods will be further discussed in the next section.

Table 4.1 Costing methodologies

Study	Population	Data	Covariates	Cost components	Statistical method
Bennell, 2015 (Canada)	AF patients (20 years or older)	Health insurance program, administrative data, Drug Benefit database	Demographics and comorbidities	Inpatient and outpatient visits and services, prescribing, long-term care services	GLM with gamma distribution and log link
Ghate, 2011 (USA)	AF patients (18 years or older) with and without bleeding events	Medstat Market Scan Commercial Claims, Medicare Supplemental & Coordination of Benefits database	Demographics, insurance status, comorbidities	Inpatient and outpatient visits and services, prescribing	GLM with gamma distribution and log link
Hallinen, 2006 (Finland)	AF patients (65 years or older)	Finnish administrative data	Age and gender and comorbidities	Inpatient and outpatient visits and services, prescribing, nursing staff, traveling by the patient	GLM, OLS with and without logarithmic transformation
Holstenson, 2011 (Europe)	AF patients (18 years or older)	Hospital based administrative data from different European countries	Age, gender and risk factors	Inpatient and outpatient visits and services, prescribing	OLS
Jonsson,2010 (Sweden and Germany)	AF patients (18 years or older)	Patient and physician surveys	Age, medical history, treatment, medical and non-medical resource use, employment status	Inpatient, outpatient, prescribing, direct non-medical costs, indirect costs	GLM, with log normal distribution and log link
Nelson, 2015 (USA)	AF patients (18 years or older) with at least one pharmacy claim for warfarin	VHA administrative data	Baseline demographics, clinical characteristics for warfarin	Inpatient and outpatient visits and services, prescribing	GLM with gamma distribution and log link
Reynolds, 2007 (USA)	AF patients (mean age 65 years)	Fibrillation Registry Assessing Costs, Therapies, Adverse events, and Lifestyle registry	Demographics, cardiac and non-cardiac comorbidities, initial medical therapies, number of AF recurrences	Inpatient, outpatient, prescribing	OLS
Rohrbacker, 2010 (USA)	AF patients (18 years or older)	Human Capital Management Services, Research Reference Database	Age, gender, marital status, race, exempt status, full-time/part-time status, salary, location, Charlson Comorbidity Index	Inpatient, outpatient, prescribing, short and long-term disability costs, productivity loss, workers' compensation costs	GLM with gamma distribution and log link
Stewart, 2004 (UK)	AF patients	Commercial and government databases	Not applicable	Inpatient and outpatient visits and services, prescribing, long-term care services	Not stated
Wu, 2005 (USA)	AF patients (65 years or younger)	Privately insured administrative database and Medicare (for direct costs)	Demographics, employment status, comorbidities	Inpatient and outpatient visits and services, prescribing, productivity loss	Multivariate two-part regression model

4.2.5 Cost components and results

Inpatient and outpatient visits and services costs as well as prescribing costs were included in the majority of the models estimating the healthcare cost associated with AF [80-83, 85, 86, 92, 94].

Bennell (2015) identified the CHA_2DS_2-VASc , as the strongest predictor of costs for AF, estimated per patient until March 2012. In particular, the CHA_2DS_2-VASc, a clinical risk score used to predict the potential risk of stroke in patients with AF [30], was used to classify patients as having a moderate to high risk of stroke. In addition, the study identified three distinct phases in which AF related costs are incurred at different levels. In the acute phase following emergency department visits, patients are likely to incur high costs, which in turn stabilise in the long-term stable/chronic phase, and increase in the phase preceding death. In particular, it was found that the phase preceding death ($8,050 (95% CI 7,666, 8,434)), compared with post emergency department visits ($1,876 (95% CI 1,822, 1,931)), and stable/chronic phase ($640 (95% CI 624, 655)), was the phase with the highest AF related costs. In the phase preceding death, hospitalisation accounted for the majority of the incurred cost (72%), whereas outpatient costs and services, referred to as physician services, contributed to the overall cost with a modest 32% [80].

Ghate (2011) indicated that patients with Intracranial Haemorrhage (ICH) or Gastro Intestinal (GI) bleeding, in the 12 month after warfarin starting date, were more likely to incur AF related healthcare costs; being respectively 64.4% and 49.0% higher than for patients who did not experience bleeding events [83].

Compared to the other studies, Hallinen (2006) adopted a slightly different approach by exploring costs associated with the use of warfarin. In addition to prescribing, inpatient and outpatient visits and services, the study included costs of nursing staff and those related to patients traveling to and from clinics. The estimated total cost per patient, at 12 month follow up and reported for the year 2002, was €589.82 (95% CI 586.68, 591.99) with the log-transformed Ordinary Least Square (OLS), and €616.00 (95% CI 579.98, 652.96) with the OLS-model [81].

For the majority of the studies, the approach of costing consisted in estimating the cost of comorbidities associated with AF, rather than estimating the cost of AF per se. Compared to the other studies, Holstenson (2011) focused only on cost related to cardiovascular disease, rather than all potential comorbidities, in patients affected by AF across different European countries. The estimated direct annual costs per patient were €933 in Greece, €1,383 in Italy, €698 in Poland, €1,316 in Spain, and €1,544 in the Netherlands [86].

The cost of illness study carried out in Sweden and Germany found the cost incurred by AF patients to be driven by the consequences of AF, such as coronary heart and cerebrovascular disease leading to hospitalisation, rather than by the treatment of AF itself. The total annual costs per patient, €7,241 in Sweden and €5,586 in Germany estimated for the year 2005, was found to increase with age; however, the AF related total cost was lower for the age groups 65 years or older than in younger patients likely to incur indirect costs associated with productivity-loss [88].

Similarly, Nelson (2015) adopted a somewhat different costing approach compared to other studies, by estimating and comparing healthcare costs between newly anticoagulated patients with out-of-range (INR < 2, INR > 3) or in range ($2 \leq INR \leq 3$) INR values. Those with out-of-range INR incurred the highest overall healthcare cost, ranging from $3,419 to $5,126 and including hospital, outpatient, and pharmacy costs, than those with the optimal INR range [82].

In Reynolds`s study (2007) the cost, associated with recurrent episodes of AF at 3 and 6 months intervals, were estimated for patients in whom the first AF episode became permanent, and for those with a sinus rhythm of either 0, 1–2, or ≥3 indicating the recurrence of AF. The mean annual costs per patient, for year 2002, was $2,372, $3,385, $6,331, and $10,312 for each group respectively. Further, the recurrence of AF was found to increase the overall healthcare cost by $1,600 annually, and hospital cost being the cost element with greatest variation [87].

Rohrbacker (2010) estimating the cost of AF and cardiac arrhythmias in an employed population, found the annual costs being higher by $3,958 in the AF group compared to employees not affected by AF. Similarly, the cost for patients with cardiac arrhythmias exceeded, by $3,958, the cost for those without the condition. In addition, the number of absence days due to sick leave and short-term disability were considerably higher in the AF and cardiac arrhythmias groups than in the condition free employees [84].

Stewart (2004), evaluating the cost of AF for the whole AF population in the UK in 1995, identified hospitalisation and drug prescriptions, accounting for 50% and 20% respectively, as the cost components contributing the most to the overall AF direct cost of £244 million. Long-term nursing home care following hospital admission, accounted for an additional £46.4 million. Stewart's study also predicted the overall direct and nursing home care cost to increase respectively by an additional £46.4 and £111 million for the year 2000 [91].

In Wu's study (2005), the total AF related cost was estimated per patient, for the study period 1999-2002, for the whole AF population and for a sub set of patients with unemployment data available. For the latter group, to reflect the employer perspective, direct and indirect costs were estimated. For the whole AF population, the estimated annual total direct cost was ($15,553), significantly higher than the cost estimated for non-AF patients ($3,204). For the subgroup with employment status, the estimated total cost was ($18,454) and ($3,461) for the AF and non-AF patients respectively. Indirect costs were higher in the non-AF patients accounting for 20% of the total cost, than in the AF patients where indirect costs represented only the 15% of the overall estimated healthcare cost [85].

4.3. Regression models for healthcare expenditure

OLS and GLM are the two types of regression models that are routinely used in the analysis of costs within a healthcare context to explain variation in costs. In regression analysis, the residuals, often referred to as the error terms, indicate the difference between the actual value of the independent variable and the value predicted by the regression equation. For a regression model to be valid, the error terms must be constant, not following a specific pattern and have a mean of zero.

Generally, the distribution of the error term, in patient level healthcare costs, is characterised by a high degree of heteroscedasticity indicating important differences in the size of the residuals across values of a given independent variable. In other words, the residuals of healthcare costs are not randomly distributed around zero, and therefore the error variance may not be constant. It follows that heteroscedasticity implies the presence of a non-constant error variance violating the OLS assumption of constant variance. In the presence of heteroscedasticity, although the resulting coefficients will be unbiased, the estimates will be inefficient; hence, OLS cannot be used as the Best Linear Unbiased Estimator. An efficient and unbiased estimator will have a minimum variance and a difference between its expected value and the true value being estimated equal to zero. Adopting a log transformation approach for the OLS model may minimise the issues related to heteroscedasticity and skewness of the data. However, with this method, the error term is additive and a log scale variance proportional to the square of the mean indicates that the variance is not constant. Hence, in a log transformed OLS model, the mean needs to be retransformed to the original scale of the data. Because retransformation returns the geometric mean, which is smaller than the mean of the raw cost data, a smearing factor needs to be applied to obtain estimates of the arithmetic mean. As defined by Duan (1983), the smearing factor is a "nonparametric estimate of the expected response on the untransformed scale after fitting a linear regression model on a transformed scale"[95]. However, the selection of a smearing factor is arbitrary and mainly depends on type of distribution of the log cost [93, 96]. For instance, when the distribution is normal and there is no difference between log variances, what is generally called a "common smearing factor" would be used. Alternatively, if the log variances differ, a "subgroup – specific smearing factor" may be more suitable. OLS is part of the GLM family, and can also be represented as GLM regression with Gauss family and identity link. Nevertheless, comparatively to log OLS, in a GLM regression model, the mean of the covariates rather than the mean of the cost variable is transformed.

This is achieved by specifying a link function indicating the relationship between the mean and the linear predictor given by the explanatory variables, and the distributional family reflecting the relationship between the mean and the variance [93, 96].

Equation 1 shows the structure of the GLM model

$$E(Y) = \mu = g\text{-}1\ (X\beta) \qquad \text{(Equation 4.1)}$$

Where E(Y) is the expected value of Y (cost), (Xβ) the linear predictor and g the link function. In a GLM regression, characterised by a multiplicative error where the variance is proportional to the mean, predictions are made on the raw cost scale; consequentially no retransformation is required [93]. The most common distributional families reflecting the mean-variance relationship and link functions are reported in Table 4.2. Regression models for healthcare expenditure, such as GLM, OLS and log transformed OLS, can be compared by means of Akaike Information Criterion (AIC), which measures goodness of fit. Then, distributional families and link function can be tested as described in the methods section of this Chapter, to identify the best fitting ones for the set of data used in this cost analysis.

Table 4.2 Distributional families

Distributional families and Link functions		
Distributional families	Gaussian Poisson Gamma Inverse Gaussian	Constant variance Variance proportional to mean Gamma Variance proportional to cube of mean
Link function	Identity Square root Log Reciprocal	g(u)=xiβi g(u)=xiβi² g(u)=exp(xiβi) g(u)=1/(xiβi)

4.4 Methods

The present work extends the existing British study [91] in the use of individual-level linked data and by estimating healthcare costs that are not directly related to AF, but incurred by patients with AF in the 5 years following their first AF event. Therefore, any subsequent costs incurred by AF patients within the period described are included. In addition, while the existing study estimated the AF cost for the entire British population in 1995, in this study the average AF related healthcare costs per person is estimated for a longer time-period ranging from 1997 to 2015. Further, in the present study, costs were estimated with an incidence-based approach. Cost analyses or cost of illness studies typically adopt either a prevalence or incidence based approach [97]. In the context of AF, the prevalence-based approach determines costs attributable to all cases of AF in a given year, while the incidence based approach estimates the new cases of AF in a given year, and then applies a lifetime cost estimate to the new cases [98].

In addition to a prevalence or incidence-based approach, a top-down or bottom-up strategy is typically adopted in cost of illness studies. Top-down is generally suitable when costing homogenous services such as nursery and long-term care, as it assumes an equal distribution of resources between patients. With this approach, typically focused on national average costs, total expenditure is divided by total units of activity to derive a unit cost. This differs from the bottom-up approach where the calculation of unit costs is based on the resource utilisation measured at the patient level [98].

A further distinction between costing analyses is between the medicalized and the global comprehensive approaches. In the first case, only expenditures directly attributable to a particular disease are used for estimating the overall costs. While the medicalized approach can be used to identify highly specific expenditures, it may also lead to underestimation or overestimation of the economic burden of a given disease; this may happen when cost estimation is not adequately adjusted for confounders highly correlated with the disease of interest. Conversely, the global comprehensive approach, used in this analysis, includes all the expenditures incurred by a population with a particular disease [99]. These expenditures are not necessarily related to the disease of interest; for instance, expenditures related to orthopaedics surgery or cancer treatment incurred by a patient with AF, will count towards the global comprehensive cost of AF.

4.4.1 Data preparation

As described in the previous Chapter, after checking for data entry errors and removal of duplicate records, an initial AF cohort consisting of 278,286 individuals hospitalised with a diagnosis of AF or atrial flutter was identified. Based on clinical advice and the evidence that prevalence and incidence of AF typically increase exponentially from 50 years onwards [89], the analysis including individuals 50 years or older would be inclusive of all patients potentially at risk of AF. The choice on the age cut-off for the AF cohort was also based on the indication of oral anticoagulants for the AF population. Most AF patients in the cohort are also on direct oral anticoagulants, and patients who are 50 years or older are likely to be on anticoagulants only because of AF, while patients younger than 50 (only about 3% of AF patients in the cohort) could be on anticoagulants for reasons other than AF.

The final AF cohort as identified from SMR01 records for analysis consisted of 272,716 patients. Individual-level data merging was then carried out with SMR00, PIS, care home census and mortality records as depicted in Figure 4.1. Incident AF events (ICD10 code I48) were identified using all six diagnostic positions in SMR01, with a look back period of five years to minimise double counting. Patients were followed up for five years following incident AF event in terms of their healthcare resource use, care home admissions and mortality. Since AF is often a precursor of stroke and cardiovascular conditions, an estimation of costs for a period of five years post AF event would allow us to fully capture costs associated with an AF patient.

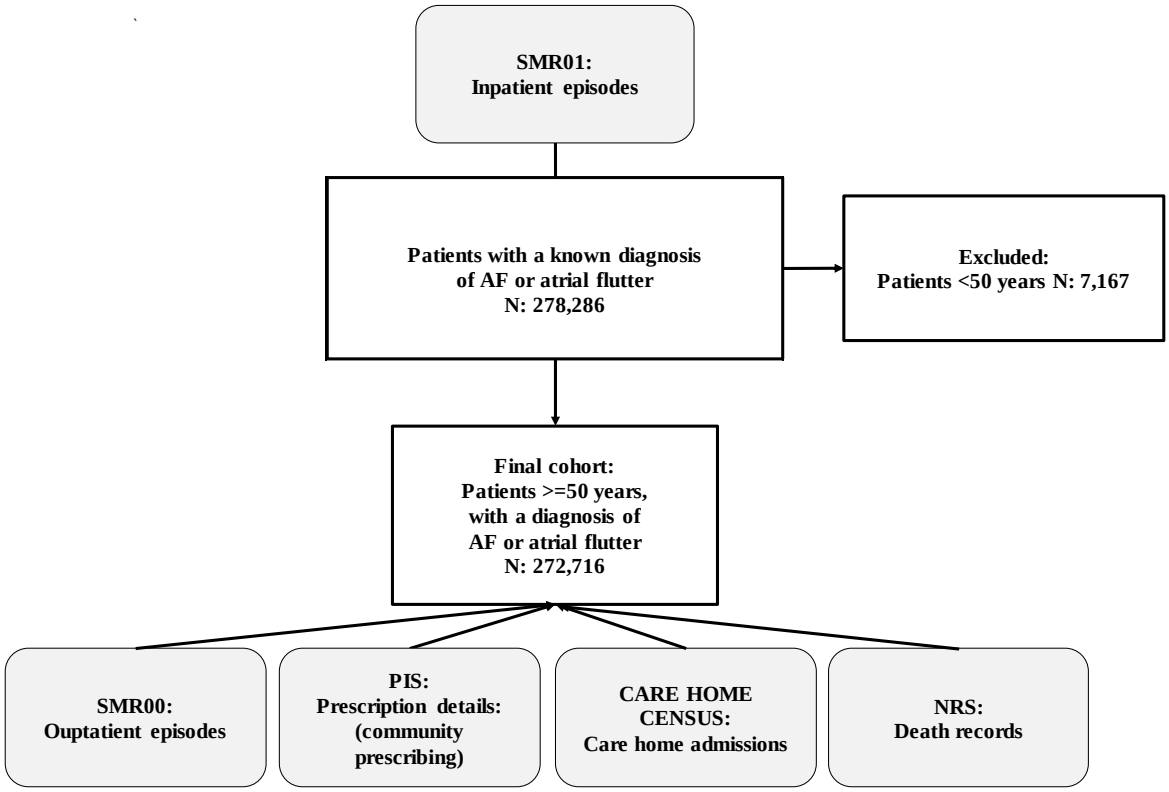


Figure 4.1 Data preparation

Abbreviations: PIS=prescription information system, SMR01=Scottish morbidity records - general/acute inpatient day case dataset, SMR00=Scottish morbidity records - outpatient appointments and attendances, NRS=national health records.

4.4.2 Costing

The cost analysis carried out in this Chapter quantifies and examines the distribution of inpatient, outpatient, prescribing and care home costs associated with AF over a five-year period. In this section, the methods of assigning unit costs to each cost component are in turn described.

Inpatient care costs

Inpatient care costs were assigned utilising the latest (2013/2014) Scottish National Tariff (SNT), a list of standard average prices based on the HRGs system [100]. The SNT uses HRG4 for grouping clinically similar treatments that use similar levels of healthcare resources. More specifically, hospital admissions are grouped according to their corresponding patient operative procedures and diagnoses; this allows for the identification of groups and subgroups of patients, and the treatment they receive [101].

For instance, in a case where a patient aged 58 is admitted to hospital with a primary diagnosis of cerebral infarction due to thrombosis of cerebral arteries (ICD-10 163.3) and the dominant procedure received is computed tomography scan (OPCS-4 U05.1) , the HRG4 code AA35A indicating stroke with severe complications and comorbidities will be used for the hospital admission. Secondary diagnoses and comorbidities may also be used for further differentiation of patients and identify any additional resource used [101]. In Scotland, the HRG based tariff was developed to support NHS health boards with what is known as “cross boundary flow activity “where boards can charge one another for individuals receiving treatment in one health board but residing in another. This system overcomes the impracticality of using average costs within specialties; indeed, costs may vary considerably depending on the treatment performed and the severity of the admission.

The SNT is published by ISD in the Scottish Health Service Costs, commonly referred to as the "Cost Book', and contains cost information on hospital and primary care services; therefore, allowing healthcare expenditure to be identified by specialty, type of admission and location. Scottish healthcare costs are available at specialty level only [100].

However, a cost weighting system allows estimating costs at more specific HRG level, where the relative cost weights obtained from English NHS reference costs are applied to the Scottish data. The underlying assumption for this mechanism is that the resource differential when comparing two procedures or conditions is the same in Scotland as in England. The SNT includes elective costs incurred during planned procedures, as well as non-elective costs attributed to emergency treatments. Because two typologies of costs are present within the SNT, the choice on whether an elective or non-elective cost should be attributed to a single episode is based on the admission type. For instance, if a patient is admitted and the admission type indicates that this is an emergency (e.g. patient injury-road traffic accident) a non-elective cost should be used. Comparatively, in the case of a routine admission where a waiting list is planned, the corresponding elective cost for that specific episode will be applied [100].

Once the total cost per episode is defined, the total cost for a Continuous Inpatient Stay (CIS) should also be determined. In Scotland, a CIS describes the entire duration of an inpatient stay in a hospital. A CIS can consist of several episodes, and is defined from the date of admission to the date of discharge and can span several specialties. However, the SNT is based on unit costs per spell rather than individual hospital episodes or a CIS [100]. A spell differs from a CIS as it combines all hospital episodes associated with a stay in hospital within the same specialty. Hence, the CIS needs to be partitioned into spells whenever a change in specialty occurs.

For instance, if a patient is admitted to a cardiology department and then transferred to thoracic surgery, two different spells will be observed and costed. If within a CIS, two or more episodes are linked to the same specialty, only the highest incurred cost is taken into account, and the remaining episodes will be replaced with a zero cost. The process explaining how a cost per CIS using episode costing was obtained is graphically presented in Figure 4.2.

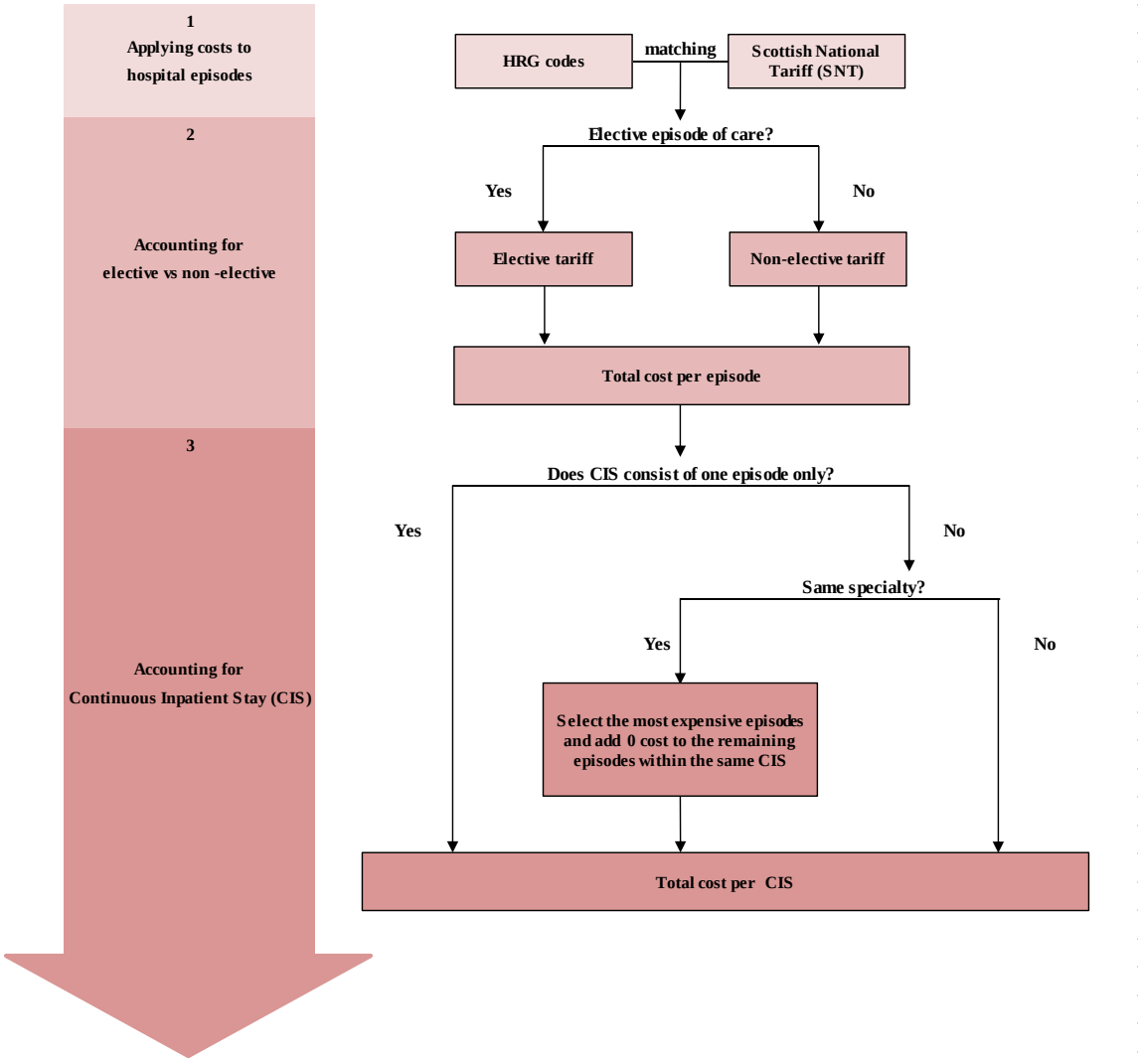


Figure 4.2 Deriving a cost per CIS using per episode costing

Care home costs

Care home costs, obtained from the care home census and based on length of stay or residency, reflect the average weekly charge for residents in care homes according to source of founding (publicly funded residents and self-funders) for the period 31st of March 2007- 31st of March 2015. Care home residency was established from care home census records, reporting admission to a care home like structure. An average of care home charges for long stay residents was calculated using information on whether nursing care was provided or not (charge with nursing £694, without nursing £614). The average weekly care home charge was expressed per day (£93), so that only the effective days spent in a care home were costed [75].

Cost for outpatient attendances

Costs for outpatient attendances were estimated by merging SMR00 with outpatient specialty costs published by ISD in the Scottish Health Service Cost [72]. The assignment of unit costs to outpatient clinics depended on whether this took place in an outpatient consultant or nurse led clinic. The first typology of clinics led by consultants (SMR00 clinic type code 1 and 2) are mainly for consultation and embrace minor as well as complex and expensive treatments. The second typology of clinic is managed by nurses (SMR00 clinic type code 3) but can sometimes be supervised by consultants. The services provided by the nurse led clinic, may however not be costed separately, but rather be included in the consultant clinic report [72].

Prescription costs

The cost of each prescription dispensed per patient was obtained from PIS [74]. Firstly, the price per unit was obtained by dividing the item price by the pack size. Secondly, the total number of items dispensed was obtained by multiplying the number of items dispensed by the number of instalments.

In Scotland prescriptions that are addictive or cause any significant harm if taken in high doses, are administered to patients in instalments. While this may not apply to the DOACs drugs prescribed for preventing stroke in patients with AF, addictive drugs such as methadone are likely to be prescribed in instalments. As the objective of the cost analysis is to estimate healthcare costs related to AF, all drugs prescribed for the AF cohort were included. This implies that aside from DOACs and warfarin, the cost of other non-AF related drugs, such as antidepressants or oncological treatments, were taken into account when estimating the total cost of prescriptions per patient in any given year between 2009 and 2015. AF-related costs per patient where estimated per year. Hence, the costs of two or more CIS within one year were aggregated to reflect yearly hospital cost. Similarly, the total yearly outpatient costs were obtained by summing the costs of individual outpatient episodes occurring within one year. The same method was applied to charges for care home stays and prescription costs.

4.4.3 Econometric model

Healthcare expenditure data are typically characterised by: i) a significant proportion of zero-cost observations for individuals who have not utilised any healthcare resources in a given time period, and ii) a skewed distribution for positive costs. In the AF costing analysis, patients were followed up for 5 years from their incident AF event (the cost were estimated only for those patients that were still alive after the 5 years post AF incident event) but during this time not every patient incurs costs in each follow-up year. Therefore, a two-part model was used for estimating first, the probability of incurring any cost in a given year and second, the mean cost conditional on having incurred positive costs [93, 96].

The first part of the model estimates the probability of using a healthcare service in a given time period as described in the following equation:

$$\Pr(HCE_{it} > 0) = \alpha + \beta_1 A_{it} + \beta_2 G_i + \beta_3 Y_i + \beta_4 \sum_{s=2}^{5} F_i + \beta_5 \sum_{s=2}^{5} S_i + \beta_6 \sum_{u=2}^{8} U_i +$$

$$\beta_7 \sum_{h=2}^{14} H_i + \beta_8 \sum_{c=2}^{3} C_{it} + \left(\beta_9 \sum_{c=2}^{3} C_{it} * A_{it} \right) + u_i$$

(Equation 4.2)

Where: A is age at the time of admission (reference category: 50 -54 age group); G is sex (reference category: male); Y is year of admission; F is follow-up year (reference category: 1st year); S is SIMD categories (reference category: most deprived (1)); U is the urban/rural classification (reference category: large urban area); H is health board of inpatient admission (reference category: Greater Glasgow & Clyde); C is the Charlson comorbidity index (reference category: no comorbidities);; u_i is the error term for patient *i* at time *t*.

The same explanatory variables were used in the second part of the model, with a gamma distribution and log link, estimating costs conditional on having incurred positive costs (Equation 4.3).

$$E[HCE] = g(x\beta)$$

(Equation 4.3)

Where xβ is the linear predictor for HCE

Mean costs per patient per year following their incident AF event were then calculated by multiplying first and second modelling part (Equation 4.4).

$$E[HCE|X] = \Pr(HCE > 0|X) * E[HCE|HCE > 0, X]$$

(Equation 4.4)

A user-written STATA programme “glmdiagnostic.do” [96], performing four different tests simultaneously, was used to identify the most appropriate distributional family and link function. In particular, the modified Park test indicates the most appropriate family (Poisson, Gamma, Gaussian NLLS, Inverse Gaussian or Wald) for a given link function. The other three tests indicate the most plausible link function for a given distribution. More specifically, the Pregibon link test will check for the linearity of response, whereas, the Pearson correlation and the Modified Hosmer-Lemeshow tests will look at systematic bias in fit on the raw scale [96].

4.4.4 Econometric model covariates

The two-part model adjusted for age, sex, year of inpatient admission, socio-economic status, urban-rural classification, health board, comorbidities and mortality. These covariates are considered to be the main confounders that have an effect on costs incurred by an AF population, and were selected according to existing literature and clinical advice. The same set of covariates explain part one (estimates the probability of incurring any cost in a given year) and part two (estimates the mean cost conditional on having incurred positive costs) of the model.

Age

The model controlled for age because AF and associated comorbidities are age-related conditions, and may have an impact on the overall costs expected to increase as the AF cohort ages. Although hospital, outpatient and prescribing costs are expected to increase marginally with age, the cost for care home residency is assumed to increase substantially. As mentioned, only patients 50 years and older were included in the analysis. Patients’ age at admission is measured using categories, where the youngest (50-54 years) served as the reference group.

Sex

To improve the accuracy of the costs estimates, the cost analysis focused on sex rather than gender; therefore the "not known" and the "non-specified" categories were excluded from the analysis. Further, it was assumed that costs vary between males and females, in particular those for care home residency.

Socioeconomic status

Variation in healthcare utilisation and associated costs and care home residency by socio-economic status is controlled for using the SIMD. The SIMD reflects areas of multiple deprivation ranked from the most to the least deprived and expressed as categories defined by fifths of the distribution of deprivation. where the most and the least deprived areas are represented by 1 and 5 respectively [102]. The index is used in Scotland for shaping policies aimed at addressing issues related to areas with high levels of deprivation.
In the econometric model, the least deprived category is used as a reference category for cost estimation, and any increase or decrease in AF cost is compared against this category. In particular, individuals in the most deprived areas may be more likely to be in hospital, as they may not have support from relatives at home, and therefore be more likely to have a longer length of stay. These patients may also incur higher hospital AF related cost than those in the least deprived area.

Health Board Area

In Scotland, there are 14 regional health boards responsible for the provision of healthcare [103]. Hence, potential differences in healthcare utilisation and prescribing costs may reflect variation in clinical practice and prescribing behaviour rather than the ability of patients to access care. Greater Glasgow & Clyde representing the largest health board is used as the reference category for estimating costs.

Geography

Patients living in urban areas may have easier access to care compared to patients living in more remote areas, which is controlled for by including the 8-fold classification measuring rurality [104] . In the 8-fold classification, the large urban category serves as the reference category to compare against. According to Scottish Government definition, a large urban area is characterised by settlements of 125,000 or more people, as opposed to very remote rural areas, which embraces a population of less than 3,000 people and a driving time of over 60 minutes to a settlement of 10,000 people or more [104]. Similar to health board areas, variations in costs for hospital, outpatient and care home utilisation are expected to be seen in for different geographical areas.

Comorbidity index

Patients with one or more comorbidities are expected to incur significantly higher costs than those with none. This was accounted for by including the CCI, which defines the severity of comorbidities, within a set of 17 pre-specified conditions, by assigning a weight from 1 to 6 to each category based on the adjusted risk of mortality. A single comorbidity score for each patient is obtained by summing all weights. While a higher score may suggest an increased risk of mortality, a score of zero indicates no comorbidities [78]. To assess whether including CCI may have improved the predictive ability of the model, the regression model described in this paragraph was compared against the same model adjusted for comorbidities, and model fit was compared using the AIC. A difference of 0.04 suggested that there is no evidence of one being better than the other model. Nevertheless, the lowest AIC indicated the model including CCI as the model with the best fit for the given set of data. The effect of any of the covariates described may depend on the particular level or value of any other explanatory variable, giving rise to what is known as interaction [105].

Interaction

Based on expert clinical advice, one interaction term assuming an interaction of age with comorbidities was included in the econometric model. An interaction effect occurs if the effect of a given covariate depends on the value of another covariate. Intuitively, a linear relationship between age and comorbidities suggests that the level of comorbidities increases, as patients get older. [106].

4.5 Results

4.5.1 Cohort characteristics

At the time of first hospitalisation, outpatient clinic attendance or care home admission, of the 272,716 AF patients - (mean age of 71 years (SD 10.6)), the majority were identified in the two largest urban health board areas (Greater Glasgow & Clyde and Lothian), accounting for 22.3% and 14.9% respectively. This is also reflected in the categorisation of geographical areas, where large urban represented 38.5% and other urban areas represented 29.7% of the total AF cohort.

A greater proportion of patients live in areas belonging to the most deprived category compared with those living in the least deprived areas – SIMD category 1 and category 5 representing 22.6% and 16.4% of the AF cohort respectively.

Among those with inpatient admission, general medicine (34.7%) followed by geriatric medicine (10.7%) were the most frequently attended hospital specialties. Most AF patients, who had been hospitalised, had also been attending outpatient clinics; therefore, similar pattern were observed across the different characteristics. Cardiology represented the second most attended outpatient clinic (10.84%) preceded by ophthalmology (13.5%).

Of the 272,716 AF patients, 4,331 were admitted to a care home, of which 22.3% were readmitted to hospital with a median length of stay of four days (IQR 1.0-16.0).

In the 5 years following the first AF incident event, the distribution of health board, geography and SIMD remained unchanged. On the other hand, the proportion of AF patients dying in the 5 years follow-up changed considerably. Within the first year, about 25% AF patients died; but much reduced percentages of mortality were observed for the remaining years of the 5-year follow-up. An inverse trend was observed for the increase of number of patients with more than one comorbidity.

In the 5 years following the incident AF event, the distribution of health board, geography and SIMD remained unchanged. During the first year about 25% of AF patients died; but lower rates of mortality were observed for years two to five. An inverse trend was observed for the increase of number of patients with more than one comorbidity.

Furthermore, over the 5-year follow-up the composition of the most frequent specialties attended remained fairly unchanged. Similar trends were observed for the composition of the most frequently attended outpatient clinics; however, haematology was no longer listed among the top five.

The proportion of patients admitted to a care home increased from 1.6% (time 0) to 2.0% (year 5), but the proportion of those admitted to hospital from a care home and median length of stay remained unchanged. Cohort characteristics and outcomes of patients who have either been admitted to hospital, a care home like structure, or attending outpatient clinics are presented in Table 4.3.

Table 4.3 Baseline characteristics and outcomes of AF patients 50 years or older

Baseline characteristics and outcomes	time 0	Follow-up year 1	Follow-up year 2	Follow-up year 3	Follow-up year 4	Follow-up year 5
Number of patients	272,716	205,043	188,066	173,605	161,931	152,361
Mean age (SD) (range)	71 (10.6) (50-108)	70 (10.4) (50-103)	70 (10.4) (50-103)	69 (10.4) (50 - 103)	68 (10.5) (50 - 101)	68 (10.5) (50 - 101)
Sex						
Male	135,683 (49.8)	104,970 (51.2)	96,798 (51.5)	89,984 (51.8)	84,461 (52.2)	79,881 (52.4)
Female	137,033 (50.2)	100,073 (48.8)	91,268 (48.5)	83,621 (48.2)	77,470 (47.8)	72,480 (47.6)
Number of deaths	n/a	67,673 (24.8)	16,977 (8.3)	14,461 (7.7)	11,674 (6.7)	9,570 (5.9)
Health boards						
Greater Glasgow & Clyde	60,774 (22.3)	45,292 (22.1)	41,269 (21.9)	37,769 (21.8)	34,968 (21.6)	32,683 (21.5)
Lothian	40,498 (14.9)	30,334 (14.8)	27,788 (14.8)	25,655 (14.8)	23,902 (14.8)	22,500 (14.8)
Lanarkshire	30,105 (11.0)	22,842 (11.1)	20,881 (11.1)	19,309 (11.1)	18,064 (11.2)	17,061 (11.2)
Grampian	25,208 (9.2)	19,440 (9.5)	17,960 (9.5)	16,653 (9.6)	15,609 (9.6)	14,709 (9.7)
Tayside	24,468 (9.0)	17,502 (8.5)	15,951 (8.5)	14,700 (8.5)	13,616 (8.4)	12,753 (8.4)
Ayrshire & Arran	21,543 (7.9)	16,295 (7.9)	14,989 (8.0)	13,859 (8.0)	12,981 (8.0)	12,286 (8.1)
Highland	18,584 (6.8)	13,578 (6.6)	12,481 (6.6)	11,502 (6.6)	10,705 (6.6)	10,065 (6.6)
Fife	17,612 (6.5)	13,855 (6.8)	12,891 (6.9)	11,998 (6.9)	11,259 (7.0)	10,667 (6.0)
Forth valley	13,308 (4.9)	10,018 (4.9)	9,215 (4.9)	8,551 (4.9)	8,041 (5.0)	7,593 (4.0)
Dumfries & Galloway	9,645 (3.5)	7,364 (3.6)	6,780 (3.6)	6,314 (3.6)	5,911 (3.7)	5,545 (3.6)
Borders	7,148 (2.6)	5,480 (2.7)	5,068 (2.7)	4,700 (2.7)	4,435 (2.7)	4,203 (2.8)
Western isles	1,812 (0.7)	1,467 (0.7)	1,337 (0.7)	1,245 (0.7)	1,159 (0.7)	1,084 (0.7)
Shetland	1,009 (0.4)	813 (0.4)	750 (0.4)	698 (0.4)	665 (0.4)	626 (0.4)
Orkney	1,002 (0.4)	763 (0.4)	706 (0.4)	652 (0.4)	616 (0.4)	586 (0.4)

Table 4.3 Baseline characteristics and outcomes of AF patients 50 years or older (continued a)

Baseline characteristics and outcomes	time 0	Follow-up year 1	Follow-up year 2	Follow-up year 3	Follow-up year 4	Follow-up year 5
Geography						
Large/urban	104,841 (38.5)	78,225 (38.2)	71,326 (37.9)	65,426 (37.7)	60,681 (37.5)	56,801 (37.3)
Other/urban	80,794 (29.7)	60,342 (29.4)	55,264 (29.4)	51,043 (29.4)	47,644 (29.4)	44,856 (29.4)
Accessible small towns	24,492 (9.0)	18,251 (8.9)	16,770 (8.9)	15,475 (8.9)	14,436 (8.9)	13,572 (8.9)
Remote small towns	8,126 (3.0)	6,055 (3.0)	5,536 (2.9)	5,136 (2.0)	4,775 (2.9)	4,524 (3.0)
Very remote small towns	3,712 (1.4)	2,882 (1.4)	2,642 (1.4)	2,449 (1.4)	2,301 (1.4)	2,179 (1.4)
Accessible rural	30,122 (11.1)	23,102 (11.3)	21,447 (11.4)	19,996 (11.5)	18,824 (11.6)	17,896 (11.7)
Remote rural	10,277 (3.8)	7,882 (3.8)	7,357 (3.9)	6,860 (3.0)	6,467 (4.0)	6,110 (4.0)
Very remote rural	9,908 (3.6)	7,900 (3.9)	7,343 (3.9)	6,858 (3.0)	6,457 (4.0)	6,084 (4.0)
SIMD						
1	61,686 (22.6)	44,494 (21.7)	40,225 (21.4)	36,668 (21.1)	33,759 (20.8)	31,355 (20.6)
2	61,704 (22.6)	45,255 (22.1)	41,154 (21.9)	37,734 (21.7)	35,038 (21.6)	32,769 (21.5)
3	54,937 (20.1)	41,610 (20.3)	38,285 (20.4)	35,408 (20.4)	33,024 (20.4)	31,120 (20.4)
4	49,448 (18.1)	38,196 (18.6)	35,385 (18.8)	32,928 (18.0)	30,984 (19.1)	29,392 (19.3)
5	44,933 (16.4)	35,486 (17.3)	33,015 (17.6)	30,865 (17.8)	29,124 (18.0)	27,723 (18.2)
Comorbidity						
no comorbidity	64,758 (23.8)	56,815 (27.7)	55,952 (29.8)	55,258 (31.8)	54,758 (33.8)	54,241 (35.6)
1 comorbidity	68,368 (25.1)	50,779 (24.8)	47,982 (25.5)	45,679 (26.3)	43,871 (27.1)	42,370 (27.8)
>1 comorbidities	139,590 (51.2)	97,449 (47.5)	84,132 (44.7)	72,668 (41.9)	63,302 (39.1)	55,750 (36.6)

Table 4.3 Baseline characteristics and outcomes of AF patients 50 years or older (continued b)

Baseline characteristics and outcomes	time 0	Follow-up year 1	Follow-up year 2	Follow-up year 3	Follow-up year 4	Follow-up year 5
INPATIENT						
Length of stay	5.5 (3.2 -9.8)	4.8 (2.3-9.0)	4.5 (2.5-8.5)	4.2 (2.4-8.0)	4.0 (2.3 - 7.5)	4.0 (2.2 - 7.0)
Most frequent specialty attended						
General medicine	1,162,296 (34.7)	512,240 (39.1)	456,474 (39.0)	402,890 (38.9)	353,143 (38.8)	311,109 (38.7)
Geriatric medicine	358,641 (10.7)	162,068 (12.4)	137,258 (11.7)	113,977 (11.1)	93,520 (10.3)	77,040 (9.6)
Cardiology	263,414 (7.9)	129,737 (9.9)	121,278 (10.4)	113,374 (11.0)	105,769 (11.6)	98,982 (12.3)
General surgery (no vascular)	195,416 (5.8)	62,356 (4.8)	57,055 (4.9)	51,614 (5.0)	46,400 (5.1)	41,835 (5.2)
General surgery	179,613 (5.4)	51,971 (4.0)	47,247 (4.0)	42,298 (4.1)	37,893 (4.2)	33,761 (4.2)
OUTPATIENT						
Median outpatient attendance (IQR)	2.0 (1.0 - 4.0)	2.0 (1.0 - 4.0)	2.0 (1.0 - 4.0)	2.0 (1.0 - 4.0)	2.0 (1.0 - 4.0)	2.0 (1.0 - 4.0)
Most frequent outpatient clinic attended						
Ophthalmology	732,525 (13.5)	1,975 (8.0)	1,975 (8.0)	1,975 (8.0)	1,975 (8.0)	1,975 (8.0)
Cardiology	588,174 (10.84)	3,406 (13.7)	3,406 (13.7)	3,406 (13.7)	3,406 (13.7)	3,406 (13.7)
Trauma and orthopaedic surgery	452,052 (8.33)	1,859 (7.5)	1,859 (7.5)	1,859 (7.5)	1,859 (7.5)	1,859 (7.5)
Haematology	335,872 (6.19)	n/a	n/a	n/a	n/a	n/a
General surgery (excluded vascularisation)	n/a	1,408 (5.7)	1,408 (5.7)	1,408 (5.7)	1,408 (5.7)	1,408 (5.7)
Dermatology	304,168 (5.61)	1,304 (5.3)	1,304 (5.3)	1,304 (5.3)	1,304 (5.3)	1,304 (5.3)
Care home						
Number of patients admitted	4,331 (1.6)	3,486 (1.7)	3,197 (1.7)	3,298 (1.9)	3,077 (1.9)	3047 (2.0)
Number of hospitalisations from care home	966 (22.3)	736 (21.1)	716 (22.4)	765 (23.2)	729 (23.7)	731 (24.0)
Hospitalisations from care home, median length of stay (IQR)	4.0 (1.0 - 16.0)	4.0 (1.3 - 16.0)	3.9 (1.1 - 13.0)	3.9 (1.2 - 12.8)	3.8 (1.1 - 12.3)	3.8 (1.3- 12.5)

4.5.2 Econometric modelling diagnostic tests

Based on the review of existing AF costing studies, GLM models were compared against OLS regression and log transformed OLS by means of AIC, which measures goodness of fit. When comparing the different models, GLM reported the lowest AIC (18.021) indicating the best fit for the given set of data, compared to OLS regression (AIC 20.025) and log transformed OLS (AIC 20.179). Following the statistical test conducted with the user-written STATA programme "glmdiagnostic.do" to identify the most appropriate distributional family and link function, a Gamma distribution was identified as the best fitting distribution for the log link GLM model (Appendix I).

4.5.3 Econometric modelling results

Regression results for both modelling parts are presented in Table 4.4. However, the coefficients from the first modelling part have been converted into ORs (Appendix II, Table I) for ease of interpretation of the likelihood of utilising any health or social care services. Overall, an inversely U-shaped association between age and the likelihood of utilising any health or social care services was observed – a gradual increment in the likelihood in resource use with advancing age up to 80 years, when compared with the reference group (50-54 years), while patients 80 years or older showing a decreased probability of utilising healthcare services. However, this association was not observed in the second modelling part, estimating costs conditional on having incurred positive costs, where a gradient between age and costs indicated increasing costs as the cohort ages.

The use of health or social care services and associated costs decreases gradually in the first 5 years following the first AF incident event. Further, a marginal increase in health or social care services utilisation was observed for patients living in the most deprived areas, when compared with patients living in areas with the lowest level of deprivation. For patients with comorbidities, the odds of utilising healthcare services were 1.74 (one comorbidity) and 2.20 (two or more comorbidities) times higher compared to those with no comorbidities.

Table 4.4 Regression results: probability of healthcare resources utilisation and cost estimation

Covariates	Probability (1st modelling part) Coefficient (95%CI)	Std. Err	Probability (2nd modelling part) Coefficient (95%CI)	Std. Err
Age group (years)				
50-54	Reference			
55-59	0.297 (0.221, 0.374)	0.039	0.078 (0.023, 0.133)	0.028
60-64	0.428 (0.353, 0.502)	0.038	0.150 (0.099, 0.202)	0.026
65-69	0.497 (0.424, 0.569)	0.037	0.224 (0.173, 0.274)	0.026
70-74	0.542 (0.470, 0.613)	0.037	0.344 (0.295, 0.392)	0.025
75-79	0.650 (0.579, 0.722)	0.037	0.418 (0.370, 0.466)	0.024
80-84	0.567 (0.494, 0.639)	0.037	0.546 (0.496, 0.597)	0.026
85-89	0.510 (0.433, 0.588)	0.040	0.738 (0.683, 0.793)	0.028
90-max	0.285 (0.196, 0.373)	0.045	0.975 (0.907, 1.043)	0.035
Sex				
Male	Reference			
Female	0.023 (0.002, 0.044)	0.011	0.047 (0.033, 0.060)	0.007
Date of admission	0.206 (0.203, 0.208)	0.001	-0.057 (-0.059, -0.056)	0.001
Follow-up years				
1	Reference			
2	-1.630 (-1.656, -1.603)	0.014	-0.177 (-0.188, -0.165)	0.006
3	-1.859 (-1.887, -1.831)	0.014	-0.142 (-0.154, -0.129)	0.007
4	-1.944 (-1.974, -1.915)	0.015	-0.030 (-0.043, -0.016)	0.007
5	-2.048 (-2.081, -2.014)	0.017	-0.032 (-0.049, -0.014)	0.009
SIMD category				
1	Reference			
2	0.019 (-0.014, 0.051)	0.017	-0.050 (-0.070, -0.030)	0.010
3	-0.027 (-0.061, 0.007)	0.017	-0.077 (-0.099, -0.056)	0.011
4	-0.058 (-0.093, -0.023)	0.018	-0.110 (-0.132, -0.088)	0.011
5	-0.062 (-0.096, -0.027)	0.018	-0.157 (-0.179, -0.135)	0.011
Geography				
Large urban	Reference			
Other urban	-0.170 (-0.201, -0.139)	0.016	-0.031 (-0.051, -0.012)	0.010
Accessible small towns	-0.199 (-0.241, -0.157)	0.021	-0.051 (-0.077, -0.025)	0.013
Accessible rural	-0.252 (-0.292, -0.212)	0.020	-0.063 (-0.089, -0.037)	0.013
Remote small towns	-0.168 (-0.232, -0.104)	0.033	-0.015 (-0.057, 0.028)	0.022
Remote rural	-0.361 (-0.418, -0.304)	0.029	-0.068 (-0.106, -0.030)	0.020
Very remote small towns	-0.437 (-0.534, -0.341)	0.049	-0.097 (-0.161, -0.032)	0.033
Very remote rural	-0.413 (-0.488, -0.339)	0.038	-0.085 (-0.139, -0.032)	0.027
Health boards				
Great Glasgow and Clyde	Reference			
Lothian	-0.053 (-0.092, -0.015)	0.020	-0.054 (-0.076, -0.032)	0.011
Lanarkshire	0.006 (-0.034, 0.047)	0.021	-0.077 (-0.102, -0.053)	0.013
Ayrshire and Arran	-0.427 (-0.471, -0.382)	0.023	-0.065 (-0.094, -0.035)	0.015
Grampian	0.022 (-0.024, 0.068)	0.023	-0.052 (-0.078, -0.025)	0.013
Tayside	-0.498 (-0.540, -0.456)	0.021	-0.100 (-0.128, -0.072)	0.014
Fife	-0.094 (-0.146, -0.042)	0.026	-0.025 (-0.059, 0.008)	0.017
Highland	-0.196 (-0.258, -0.134)	0.032	-0.041 (-0.080, -0.002)	0.020
Forth Valley	-0.546 (-0.597, -0.495)	0.026	-0.115 (-0.149, -0.080)	0.018
Dumfries and Galloway	-0.339 (-0.400, -0.278)	0.031	-0.174 (-0.214, -0.133)	0.021
Borders	-0.573 (-0.638, -0.509)	0.033	-0.107 (-0.154, -0.061)	0.024
Western Isles	-1.295 (-1.409, -1.181)	0.058	0.145 (0.061, 0.229)	0.043
Orkney	-0.487 (-0.641, -0.334)	0.078	0.001 (-0.118, 0.121)	0.061
Shetland	-0.681 (-0.831, -0.532)	0.076	-0.064 (-0.200, 0.073)	0.070
Comorbidity				
no comorbidities	Reference			
1 comorbidity	0.552 (0.425, 0.678)	0.065	0.423 (0.339, 0.507)	0.043
>1 comorbidities	0.789 (0.606, 0.971)	0.093	0.917 (0.832, 1.002)	0.043

4.5.4 Cost estimates

The estimated mean annual cost per AF patient was £3,114 (95% CI: £3,093, £3,215). Inpatient cost was £3,032 (95% CI: £3,006, £3,058) in the first year following the first AF incident event but dropped to £2,276 (95% CI: £2,233, £2,319) in the fifth year. Similarly, outpatient cost was £363 (95% CI: £360, £366) in the first year following the first AF incident event but dropped to £286 (95% CI: £282, £290) in the fifth year. Conversely, the cost of care home was £131 (95% CI: £121, £142) in the first year following the first AF incident event but dropped to £242 (95% CI: £220, £263) in the fifth year.

As shown in Table 4.5, compared against the other health boards, patients residing in Great Glasgow and Clyde incurred the highest estimated cost of £3,370 (95% CI: £3,318, £3,423). Comparatively, patients living in Forth Valley incurred the lowest estimated cost of £2,742 (95% CI: £2,657, £2,826). AF patients from large urban areas incurred the highest estimated cost of £3,273 (95% CI: £3,230, £3,315). Comparatively, the lowest estimated cost of £2,764 (95% CI: £2,590, £2,937) was incurred by AF patients living in “very remote rural areas”. Further, patients from the most deprived areas seemed to incur higher overall costs than those from the least deprived areas. A clear gradient of estimated costs decreasing from the most to the least deprived areas was observed.

Table 4.5 Average annual cost per AF patient by health board, geography and SIMD

Category	Mean cost (£) (95% CI)
Health board	
Great Glasgow and Clyde	3,370 (3,318, 3,423)
Lothian	3,169 (3,111, 3,227)
Lanarkshire	3,123 (3,057, 3,189)
Ayrshire and Arran	2,949 (2,875, 3,022)
Grampian	3,210 (3,140, 3,281)
Tayside	2,808 (2,741, 2,875)
Fife	3,241 (3,144, 3,338)
Highland	3,140 (3,032, 3,248)
Forth Valley	2,742 (2,657, 2,826)
Dumfries and Galloway	2,686 (2,585, 2,786)
Borders	2,747 (2,626, 2,869)
Western Isles	2,979 (2,722, 3,236)
Orkney	3,115 (2,739, 3,490)
Shetland	2,808 (2,421, 3,195)
Geography	
Large urban	3,273 (3,230, 3,315)
Other urban	3,090 (3,048, 3,132)
Accessible small towns	3,016 (2,947, 3,085)
Accessible rural	2,953 (2,889, 3,018)
Remote small towns	3,142 (3,015, 3,270)
Remote rural	2,883 (2,781, 2,986)
Very remote small towns	2,764 (2,590, 2,937)
Very remote rural	2,807 (2,666, 2,949)
SIMD (Scottish index of multiple deprivation)	
1	3,359 (3,309, 3,408)
2	3,205 (3,159, 3,250)
3	3,095 (3,046, 3,144)
4	2,981 (2,933, 3,029)
5	2,843 (2,795, 2,891)

Costs incurred by female AF patients are higher than the costs observed in the male group. Table 4.6 shows males and females with an estimated cost of £3,036 (95% CI: £3,007, £3,064) and £3,192 (95% CI: £3,162, £3,223) respectively. While there was little difference between the total costs and the distribution of costs for inpatient, outpatient and prescription costs, the difference seemed more pronounced when comparing the care home component of costs (5.0% of total costs among males versus 7.3% of total costs among females).

Table 4.6 Average annual cost per AF patient by sex and healthcare or care home sector

Category	Cost estimates	
	Mean cost (£)	(95%CI)
Male		
Inpatient (%)	2,273 (74.8)	(2,251, 2,294)
Outpatient (%)	323 (10.6)	(320, 326)
Care home (%)	153 (5.0)	(139, 168)
PIS (%)	263 (8.7)	(260, 266)
Total	3,036	(3,007, 3,064)
Female		
Inpatient (%)	2,323 (72.8)	(2,301, 2,345)
Outpatient (%)	325 (10.2)	(322, 328)
Care home (%)	233 (7.3)	(218, 248)
PIS (%)	280 (8.8)	(277, 283)
Total	3,192	(3,162, 3,223)

The average annual cost per AF patient by age and for each health or care home sector is presented in Table 4.7. Considering the individual contribution of each cost component to the overall costs, inpatient costs were found to be the main cost driver across all age groups. While inpatient cost contribution remained constant with an average contribution of about 75% to the overall cost for patients aged between 50 and 84 years, it decreased for patients over 85 years of age. Similar patterns were observed for outpatient and prescribing costs. On the contrary, the contribution of care home costs to the overall costs increased with age (0.6% for patients aged 50-54 years and approximately 19.7% for patients who are 90 years or older).

The contribution of each setting to the total health and care home costs by the number of existing comorbidities is presented in Table 4.8. While inpatient and total costs varied considerably with the number of comorbidities, outpatient and care home contributions remained fairly constant.

Table 4.7 Average annual cost per AF patient by age and healthcare or care home sector

Age groups	Hospital total cost (£)		Outpatient l total cost (£)		Care home total cost (£)		PIS total cost (£)	
	Mean (95%CI)	%	Mean (95%CI)	%	Mean (95%CI)	%	Mean (95%CI)	%
50-54	1,857 (1,765, 1,948)	75	337 (324, 350)	13.6	16 (9, 41)	0.6	267 (250, 284)	10.8
55-59	2,065 (1,996, 2,135)	75.9	352 (342, 361)	12.9	17 (1, 36)	0.6	286 (274, 298)	10.5
60-64	2,091 (2,041, 2,141)	75.7	352 (345, 359)	12.7	26 (14, 39)	0.9	295 (286, 304)	10.7
65-69	2,152 (2,109, 2,195)	76.1	353 (348, 358)	12.5	33 (21, 45)	1.2	290 (284, 296)	10.3
70-74	2,247 (2,211, 2,283)	76.1	353 (349, 358)	12	63 (49, 77)	2.1	290 (285, 295)	9.8
75-79	2,344 (2,312, 2,376)	76.7	348 (343, 352)	11.4	84 (71, 97)	2.7	281 (277, 285)	9.2
80-84	2,469 (2,432, 2,506)	75.1	317 (313, 321)	9.6	235 (211, 258)	7.1	268 (265, 272)	8.1
85-89	2,665 (2,618, 2,712)	72.7	276 (271, 281)	7.5	472 (431, 513)	12.9	255 (250, 259)	7
90-max	2,765 (2,694, 2,836)	69.8	194 (187, 200)	4.9	779 (706, 852)	19.7	226 (220, 232)	5.7
Total	2,297 (2,282, 2,313)	74.3	324 (322, 326)	10.5	201 (190, 212)	6.5	271 (269, 273)	8.8

Table 4.8 Average annual cost per AF patient by CCI and healthcare or care home sector

Cost component	No comorbidity cost (£)		1 comorbidity cost (£)		>1 comorbidity cost (£)	
	Mean (95%CI)	%	Mean (95%CI)	%	Mean (95%CI)	%
Hospital	1,786 (1,767, 1,805)	74.9	2,253 (2,226, 2,280)	73.5	3,148 (3,110, 3,186)	73.4
Outpatient	285 (282, 288)	12.0	307 (304, 311)	10.0	409 (405, 414)	9.5
Care home	131 (117, 146)	5.5	219 (199, 239)	7.1	256 (235, 276)	6.0
PIS	185 (183, 188)	7.8	269 (266, 273)	8.8	386 (381, 391)	9.0
Total	2,384 (2,357, 2,412)	100	3,065 (3,026, 3,104)	100	4,289 (4,237, 4,340)	100

4.6 Discussion

4.6.1 Cohort characteristics

The descriptive analysis does not only provide an overview of patient characteristics, but also reveals important aspects related to AF. The mean age of 71 years reported for the cohort affected by AF in SMR01 is in line with the mean age suggested in the literature. As expected, general medicine and cardiology were identified as the specialties patients were admitted to most frequently over time. The high number of hospital episodes in geriatric medicine, is in line with the average age of patients. The highest number of hospitalisations was also expected in large affluent urban areas such as the Greater Glasgow & Clyde and Lothian with reasonable access to hospitals, and in areas with the highest index of deprivation for the reasons discussed in the previous section.

Ophthalmology and cardiology were respectively the first and the second most attended outpatient clinic over time. Indeed, these specialties deal with age related conditions, which generally have a high prevalence in elderly patients. The attendance of the haematology outpatient clinic is also consistent with expectations, as patients on warfarin, to date the main oral anticoagulant treatment for the management of AF as a preventive measure for stroke, require constant monitoring of blood clotting rate to determine the effects of oral anticoagulants. While this is true at time of first attendance, in the 5-year follow-up haematology is no longer among the five most frequently attended outpatient clinic attended, presumably due to discontinuation of warfarin or switching to another OAC. Across health boards, geographical locations and socio-economic status, the proportion of outpatient clinics attended reflects the proportion of hospital episodes, indicating consistency between hospitalisation and follow-up visits.

4.6.2 Estimated cost of AF

A greater proportion of AF patients was found in the most deprived areas, this combined with the likelihood of people living in the most deprived category having longer inpatient stays, may explain the difference in inpatient care utilisation between patients from the most and the least deprived areas, with associated costs being higher for the former group. As AF is more likely to affect the elderly [19], AF related costs were expected to increase with age. As health deteriorates with age, older age groups are assumed to make greater use of healthcare services, and therefore incur higher costs than younger age groups. However, age was found to have a modest impact on overall healthcare costs, being fairly consistent across age groups. This finding is in line with existing evidence indicating that healthcare expenditure depends not only on patients' calendar age, but is also significantly associated with remaining lifetime [107].

On the other hand, comorbidity had a considerable effect on the overall cost, increasing significantly in patients with more than one comorbidity compared to patients in other CCI categories.. Decreasing inpatient and outpatient costs for the oldest patients were offset by increasing care home costs, in particular for women. Indeed, the main cause for higher overall costs incurred by women is attributable to the higher likelihood for elderly women to reside in care homes. Interestingly, care home contribution to the overall costs was noticeably lower for patients with multiple comorbidities than for those with one comorbidity. This may suggest that sicker patients are more likely to be in hospital than in a care home.

To date, only one single study published in 2004 has estimated the cost of AF in Scotland; the authors estimated the cost of AF in 1995/1996, and projected these to the year 2000 [91]. Previous work has focussed on a 12-months follow-up, which seems limited in order to capture all healthcare resource utilisation for AF patients. This analysis offers a longer follow-up and a contemporary estimate of healthcare costs related to AF including all relevant care settings. This analysis offers a distinct advantage over previous work as costs, rather than being based on extrapolated rates using a prevalence-based approach [91], are estimated with an incidence-based method using patient-level morbidity records. Using an incidence-based approach to costing and a broad perspective to capture the majority of costs associated with AF, several routinely collected administrative datasets from Scotland were combined, including care home utilisation. A robust Scottish record linkage system, where administrative health data are routinely collected, allows Scotland to be at the forefront of research where patient data linkage is involved [14, 108].

Existing analyses, including this one, regardless of econometric model choice and covariates used, show that costs due to inpatient admission are the main contributor to overall healthcare cost for AF patients; hospitalisation as the main cost driver was also identified by the previous Scottish study published in 2004 and using patient level data [91]. This is a pertinent finding that may well support future policies on opportunistic screening in the population at risk of AF, and in particular in Scotland where 1 in 3 patients with AF are currently undiagnosed [25]. The European AF management guidelines and the Scottish Cross-Party Group ‘Heart Disease and Stroke’, recently recommended that people who are 65 years or older and at risk of AF and associated comorbidities such as cardiovascular disease, diabetes or respiratory disease should be screened opportunistically in primary care, pharmacies or community settings. With rigorous screening and appropriate treatment, hospitalisations could be avoided and costs reduced [25, 109].

Although capturing most healthcare sectors and related costs, it was not possible to obtain data on general practice consultations, as those data are currently not routinely available in Scotland. This, coupled with the potential risk of AF going undiagnosed and clinical miscoding of morbidity records, may lead to an underestimation of the size of the AF cohort and associated costs. Other limitations are mostly inherent to the nature of administrative data, such as missing records or incomplete data. Prescribing and care home data were only available from 2009 and 2012, so their contribution to overall AF related costs might be underestimated. Despite these limitations, it was possible to harness high quality patient-level linked data to identify a cohort of AF patients and to estimate AF related costs in Scotland.

While the choice of a given costing method should be driven by the research question and the nature of the data, it is accepted that OLS or GLM, with different combinations as discussed in previous sections of this Chapter, are the two main approaches that should be employed in the costing of health care expenditure data. Clearly, the issues and the limitations inherent to the costing approaches mentioned should be taken into account when selecting the most appropriate regression model.

4.7 Conclusions

The inclusion of all available cost components is crucial for establishing overall costs, as these often extend beyond hospitalisation. The study identified hospitalisation as the main cost driver that may well support the implementation of AF screening policies that could potentially reduce costs for the most expensive healthcare setting, i.e. inpatient care.

Most importantly, the study concludes that patients' age has a limited impact on the overall AF related cost, and therefore may contribute much less to future growth of AF related cost in an ageing Scottish population.

This Chapter, combined with Chapter 1, provides an overview of the landscape of AF in Scotland in terms of epidemiology, disease management, treatment availability and overall costs. The next Chapter will offer an overview of the methods used in comparative-effectiveness research, with a particular focus on PS based methods that will be tested and compared.

Chapter 5 Methods for comparative-effectiveness analysis

5.1 Introduction

Comparative-effectiveness research is a rapidly evolving field that provides clinicians, patients and policy makers with clinical evidence needed to make informed decisions concerning healthcare related issues. In particular, comparative-effectiveness research evaluates the relative effectiveness of different treatment options for a specific medical condition and in a selected population, and aims to reduce the gap between clinical research and clinical practice l [11, 110]. In this context, both RCTs and non-randomised studies (RWE) contribute to generating clinical evidence for the decision-making.

Randomised controlled trials typically recognised as “gold standard” for providing evidence on efficacy, are designed for measuring efficacy in a controlled environment and in a selected population where the treatment under examination is randomly assigned. Randomisation ensures that differences in patient characteristics such as age, sex, comorbidities and disease severity, are similarly distributed between treatment groups; thus, the observed effect in the study population can be attributable to the difference in outcomes between treatment groups [41, 42, 111].

In non-randomised studies (RWE), evidence on effectiveness is generated in an uncontrolled environment. The absence of randomisation does not allow for an unbiased comparison between patients who are exposed and those who are not exposed to the treatment under study. Hence, the observed differences in health outcomes between the groups may be influenced by the population characteristics or other additional factors rather than by the treatment effect. Nevertheless, RWE may provide additional insight concerning treatment safety and effectiveness of a treatment and in some cases may be the only available source of evidence if randomised data are not available.

In particular, RWE captures natural heterogeneity that reflects variation in patient population. In RCTs, patients are recruited according to certain demographic and clinical characteristics, and variation is reduced by applying exclusion criteria; this however contributes to the limited generalizability of RCTs. Further, unlike RCTs where evidence is generated in a controlled environment, RWE may capture inevitable differences in clinical practice [112].

As stated in Chapter 4, a lack of randomisation in RWE studies gives rise to confounding by indication, occurring when the prognostic factors, such as disease severity, used for treatment selection also affect the outcome. For instance, if patients with more severe conditions receive more intense treatments. However, when comparing treatment effectiveness, the more intensive treatment may produce poorer outcomes [11]. A more practical example is provided by Kyriacou et al (2016) who identified potential confounding by indication in a clinical setting where the association between tracheal intubation during paediatric in-hospital cardiac arrest and survival were investigated. In particular, children with more severe conditions and lower probability of survival were more likely to be intubated. However, this likelihood was driven by the fact that having severe conditions is a strong predictor of both, mortality and clinical decision to intubate [113].

Historically, regression adjustment has been used to address confounding in RWE; but over the last decade, there has been an increasing interest in the application of Propensity Score based (PS) methods, such as matching and inverse probability weighting (IPW), within the use of observational data in medical research. Propensity score methods attempt to mimic the RCTs process or randomisation by estimating the probability of treatment assignment conditional on observed baseline characteristics [11, 59].

Propensity score methods offer, as it will be discussed in the next section, several advantages over conventional regression methods [59]. However, while PS methods may reduce the bias due to observable confounders such as age, sex and existing previous history of stroke, other unobserved confounding, such as patients' tolerability and access to healthcare may still bias the PS estimates. Propensity score methods can address observed confounding if the assumption of 'no unobserved confounding' is reasonable [114]. Therefore, in order to address both observed and unobserved confounding, different statistical methods such as instrumental variable, difference in differences and regression discontinuity should be used as an alternative to PS methods [115, 116]. An overview of these methods will be also be provided in the next section.

In order to address the second Objective defined in Chapter 2, this Chapter explores, within the comparative-effectiveness framework and the DOAC case study, different PS methods for estimating the treatment effect comparing warfarin and DOACs, and assesses whether results for AF-related clinical outcomes differ depending on the PS method used. The motivation for concentrating on PS methods is that the main focus will be on observed confounders. Further, for this analysis, to increase the overall sample size, a cohort termed "combined DOAC" was created to include users of either apixaban, dabigatran or rivaroxaban.

The PS method, identified as the most robust, will inform the comparative-effectiveness model, in the subsequent Chapter where differences in effectiveness between DOACs and warfarin for several outcomes are assessed. In the following paragraphs, a brief description of all the instruments available to address observed and unobserved confounding will be provided.

5.2 Propensity score based techniques

5.2.1 Propensity score

Propensity score methods attempt to mimic the RCTs process or randomisation by estimating the probability of treatment assignment conditional on observed baseline characteristics [11, 59]. Typical PS methods are matching, IPW and the use of covariate adjustment with PS. In general, the underlying assumption for these methods is that balancing observed baseline characteristics would reduce bias [117]. The PS is generally estimated using a logistic regression model, where treatment status, being the dependent variable coded as 0/1, is regressed against observed baseline characteristics. For the PS to be valid, two conditions must hold. Firstly, treatment assignment must be independent of potential outcomes conditional on the observed baseline characteristics. Secondly, patients must have a positive probability of receiving either treatment [118].

The PS estimation involves several steps; to start with, covariates to be included should be selected according to certain characteristics. In general these variables should be measured baseline covariates associated with treatment assignment (e.g. blood pressure, weight or glucose level), covariates that are potential confounders affecting the outcome (e.g. age and sex) and variables that are true confounders affecting treatment assignment and the outcome of interest (e.g. comorbidity co-medication or life style factors such as smoking and alcohol consumptions) [59].

As a second step, the PS distribution should be inspected graphically to ascertain whether an adequate overlapping distribution has been achieved. Distributions of the predicted probabilities between treatment groups should overlap to indicate that covariates for control and treatment groups are comparable [59, 117].

Then, it should be established whether any of the covariates included in the model are satisfactory balanced. This is typically done by means of standardised differences comparing the mean of continuous and binary variables between treatment groups. In detail, for a continuous covariate, the standardised difference is obtained as indicated in Equation 5.2 [59].

$$d = \frac{(\bar{x}_{treatment} - \bar{x}_{control})}{\sqrt{((s^2_{treatment} + s^2_{control})/2)}}$$

Equation (5.2)

Where $\bar{x}$ *treatment* and $\bar{x}$ *control* are the sample mean of the treated and untreated group respectively; S^2*treatment* and S^2*control* are the sample variance of the treated and untreated group respectively.

For dichotomous variables, the standardised difference is obtained as described in Equation 5.3 [59].

$$d = \frac{(\hat{p}_{treatment} - \hat{p}_{control})}{\sqrt{(((\hat{p}_{treatment}(1 - \hat{p}_{treatment})) + (\hat{p}_{control}(1 - \hat{p}_{control})))/2)}}$$

Equation (5.3)

Where $\hat{p}$ *treatment* and $\hat{p}$ *control* are the prevalence or mean of the dichotomous variable of the treated and untreated group respectively.

While the approach described for calculating standardised differences works for PS matching and IPW methods, for PS covariate adjustment an approach calculating the conditional standardised difference is used instead (Appendix III, Equation III-1 – III-5).

For any of the PS method tested, differences in the means of covariates are considered negligible if the value is below the threshold of 0.1 standard deviation. Although there is no universal agreement on what the threshold for the standardised difference should be, the threshold of 0.1 is now considered by researchers as an adequate measure for diagnostic purposes assessing covariates balance and imbalance [119]. Upon assessment of PS distribution and covariate balance/imbalance, the most appropriate matching or weighting method should be selected based on the best fitting model where covariates show the lowest standardised difference. Table 5.1 shows a summary of assumptions and advantages and disadvantages for the PS methods typically used: PS matching, IPW and covariate adjustment with PS.

Table 5.1 PS methods

Method	Assumptions	Advantage	Disadvantage
PS matching	Groups are matched according to their propensity score	Only propensity model specification is required.	Propensity model specification is required. Significant proportion of individuals may be omitted when matching samples
IPW	A weight reflecting the probability of being exposed to a given treatment is given to each patient	Only propensity model specification is required. Analysis is conducted on the entire cohort	Propensity model specification is required
Covariate adjustment with PS	Outcome variable is regressed to the treatment status and the estimated propensity score	Analysis is conducted on entire cohort	Propensity and regression model specification is required

Propensity score methods offers several advantages over the regression adjustment methods and differ in a number of ways; however, the latter has long been used to address confounding in RWE studies. Regression adjustment involves the use of a two-step estimator. Firstly, two separate regression equations for the treatment and the control group are estimated; then the differences estimated from the outcomes of the two groups are averaged across all individuals [59, 114].

Regression adjustment is often used in its simpler form, known as covariate adjustment, offering more flexibility by allowing for all relevant patient characteristics from the treatment and control group to be included into a single equation [114]. The main distinction between PS methods and regression or covariate adjustment methods consist in the following: the former model the probability of receiving treatment, the latter model the outcome of interest. Other differences between PS and regression methods consists in how the treatment effect is estimated. Regression or covariate adjustment estimates the conditional treatment effect ("treatment effect on the individual"); here the coefficient from the regression is interpreted as an estimated change in outcome while keeping all other variables constant. PS based methods estimate the marginal treatment effect ("treatment effect on the population"), where the outcome rates from matched or weighted population are just compared without referring to each patient's characteristics. In practice, both conditional and marginal effects should lead to similar conclusions, but the latter will be closer to the treatment effect estimated in RCTs [11, 59].

The main advantages of PS methods over regression adjustment consist in minimising the risk of model overfitting and offering ways for assessing model specification. In particular, the risk of overfitting is minimised by allowing for balanced comparison when several factors are associated with exposure. Further, propensity score methods allows for the inclusion of all observed covariates into a single propensity score; this is particularly relevant when the number of observed covariates is large in proportion to the number of patients or outcome events [116] .This is summarised in the following equation:

$$e_i = P_r(Z_i = 1|X_i)$$

(Equation 5.1)

Where PS (e_i) is a balancing score that reflects the probability (P_r) of patients being exposed to a given treatment (Z_i) conditional on the observed covariates (X_i)

Evaluation of model specification, allowed by available diagnostics test, can indicate whether a PS model is correctly specified by assessing the distribution of the predicted probabilities and comparing the mean of continuous and binary variables between treatment groups. On the contrary, goodness-of-fit test for regression and covariate adjustment do not indicate whether the model is correctly specified [120].

Despite the differences and advantages of PS methods, a double robust approach combines the PS estimators with regression or covariate adjustment; where the latter would control for residual imbalances between the treatment and control group in the PS models. The advantage of using a double robust approach consists in reducing the risk of model misspecification, as either the PS or the outcome regression model needs to be correctly specified [59, 121]. There are several possible causes of model misspecification in PS models, such as overfitting, omitting higher order terms (i.e. ignoring non-linearity), and incorrect functional form. These might refer to either PS equation, outcome equation, or both. As described by Austin (2007), PS model misspecification affects the covariate balance and result in a downward bias in the treatment effect estimate, due to the "non-collapsibility" of marginal versus conditional causal effect [122].

When regression adjustment and IPW are combined in a regression model, augmented IPW, using an augmentation term, is typically used as the estimator to correct for misspecification in the treatment model. However, this estimator may result in unstable scenarios where the treatment probabilities are close to zero; in these cases, a regression adjustment estimator using the inverse probability weights is used to adjust for the outcome model misspecification [117, 123]. Double robust is also used with matching, where the regression estimator adjusts for residual imbalances between the treatment and control group in the matched sample.

As reported by Kreif et al. (2014), this approach would produce less bias than the double robust methods using unstable IPW determining outcome and the PS models misspecification [121]. Further, Elze et al. (2017) used a double robust approach by using PS as a covariate in the outcome model, along with other covariates reflecting patient characteristics and the treatment status [120].

5.2.2 Matching

Propensity score matching creates a sub-sample of treated and untreated groups sharing a similar PS value. This allows outcomes for the treated and the untreated to be directly compared [59]. In particular, when the outcome is continuous, such as a measurement on a numerical scale, the treatment effect is given by the difference in mean outcome between the treated and the untreated. On the other hand, when the outcome is dichotomous, indicating for instance those who experience a clinical event and those who do not, the treatment effect is given by the difference in proportion of patients who experience the event in the treated and the untreated groups within the matched sample [59, 118].

In its simplest form, PS matching uses the nearest neighbour matching approach where, as the name implies, an individual from the treatment group is matched with another individual from the control group with the closest PS to that of the treated individual. However, if several individuals from the control group have PSs that are equally close to that of the treated individual, the individual from the control group is selected randomly [59, 117, 118, 124-126]. With this approach, the matching may be further improved by applying a caliper distance that specifies a threshold under which the PSs of matched individuals must stay. The issue with the caliper is that in the literature there is no agreement on what the maximal acceptable distance should be [127].

A key aspect of this PS method is selecting whether matching should be done with or without replacement. In the first case, any patient from the control group can be used several times for more than one treated individual. Replacement is particularly useful in settings were the treatment significantly outnumbers the control group. By contrast, matching without replacement allows patients from the control group to be matched against those in treatment group only once [59, 124-126].

A further distinction is between greedy and optimal matching. The greedy approach is based on an algorithm that makes best possible matching between treated and untreated individuals with close PS at each stage. With the greedy approach, a matched pair is not reused for further matching. On the contrary, with optimal approach a matched pair could be used several times, thus reducing PS differences within matched pairs [59, 117].

With the PS matching method, a correct specification of the PS model may suffice [59, 125-127]. Nevertheless, as mentioned, with a double robust approach, incorporating relevant covariates in both the PS matching model and the outcome regression model, will compensate for any potential covariate imbalance, and will further reduce the bias caused by residual differences in observed baseline covariates. However, while PS matching has largely been used in comparative-effectiveness research, there are limitations in its application. The generalizability of results may be an important issue when using the matching method as a significant proportion of individuals will be omitted when creating the matched sub-sample.

5.2.3 Inverse Probability Weighting

An alternative PS method used to address observed confounding is IPW. The most common form for this method is Inverse Probability of Treatment Weighting (IPTW). With this approach, a weight, reflecting the probability of being exposed to a given treatment and equal to the reciprocal of the PS, is assigned to each patient in the treatment group, whereas a weight equal to the reciprocal of one minus the PS is assigned to patients in the control group [59, 128]:

$$\omega_i = \frac{1}{e_i} + \frac{1}{1 - e_i}$$

(Equation 5.4)

Unlike the PS matching method, IPTW analysis is carried out on the entire cohort. Nevertheless, IPTW offers, along with matching, an important advantage over the covariate adjustment with PS approach, requiring only the PS model specification for a correct ATE estimation [123, 128, 129]. However, with poor PS overlap, the resulting extreme weights directly derived from PS may undermine the robustness of the model. Extreme weights, caused by a PS close to 0 for the treated or 1 for the control group, indicate that patients may have potentially been assigned to treatment regardless of their baseline characteristics. For instance, patients with high PS score maybe patients that, despite being eligible because of their characteristics to a given treatment, do not get the treatment. Conversely, patients with low propensity scores maybe patients that are on a given treatment despite not being the appropriate candidates [130]. In these circumstances, trimming is generally performed to remove individuals with extreme PS values from the analysis; observations with extreme PSs are removed if <0.1 or >0.9 for the treatment and the control group respectively [127, 131]. Alternatively, large weights can be truncated to the desired threshold. While there is no agreement on the optimal threshold, as reported in some studies, any weight greater than 10 is considered a large weight [128, 132].

While trimming and truncating may reduce any residual bias, excluding all individuals with extreme values may increase the covariate imbalance between treatment and control group [59, 120, 125, 126].

IPW is typically used in the context of IPTW, but it has increasingly been employed as IPCW to address treatment switching or discontinuation issues within RCTs and with observational data [123, 128, 129, 133]. IPCW was firstly developed for RCTs to address treatment crossover and informative censoring. Thus, IPCW is particularly relevant in RCTs assessing new oncological treatments, where treatment switching typically occurs following disease progression. In this scenario, patients, being censored because starting oncological treatment before the protocol-defined progression, will have a different risk of treatment failure than those who stay on treatment [134]. Despite, its applications in RCTs, IPWC could also be used in the context of RWE where treatment switching is more likely to happen. In detail, with IPCW, the probability of being censored is used for estimating weights. In this setting, censoring is applied to those individuals who switch treatment, and the estimator assigns extra weights to individuals who are not censored but share similar characteristics with the switchers. The weights from IPCW are then added to those obtained from IPTW to obtain the overall weight reflecting ATE and censoring [128, 135, 136].

5.2.4 Covariate adjustment with propensity score

In the covariate adjustment approach, the treatment effect is obtained from the regression coefficient of a fitted regression model where the outcome variable is regressed on the treatment status and the estimated PS [59]. Depending on whether the outcome is continuous or dichotomous, a regression model would be selected accordingly.

Covariate adjustment with PS overcomes the potential limitations encountered with PS matching by allowing for conducting analyses on the entire cohort rather than on the PS matched sample; so no observations are lost [59, 120]. However, as with IPW methods, extreme PSs maybe removed, but this, while reducing any residual bias, may reduce the sample size and increase the covariate imbalance between treatment and control group [59, 120, 125, 126]. Other issues arising from the covariate adjustment approach include the correct specification of the regression model once the PS model is appropriately specified. In regression adjustment, the propensity score is included in the outcome equation; however, its relationship with the outcome is unknown and can be only assumed. Also, being the propensity score itself a function of the covariates, the inclusion of the propensity score as an additional covariate implies establishing a relation between the outcome and the covariates which is affected by the PS functional model, and might be different from the true functional form [137]. Because of the need for correctly specifying both the PS and the outcome model, this method is generally regarded as inefficient compared to other PS approaches. Nevertheless, including PS as an additional covariate, along with other covariates in the outcome model, thus resulting in a double robust approach, will increase the efficiency of the method [59, 114, 120].

5.3 Unobserved confounding

Different methods, such as regression adjustment and those based on PS estimation are typically used to reduce observed confounding inherent to observational data, and as discussed, PS based methods are particularly efficient in addressing confounding by indication. However, if in addition to observed confounders, unobserved confounders or other factors also affect the outcomes, different statistical methods may be used to correct for potential bias and estimate the treatment effect in the context of RWE. These methods are based on the use of instrumental variables, difference in differences and regression discontinuity. A brief description of these methods is provided below.

Instrumental variable

The instrumental variable approach, allows the estimation of the causal effect of treatment on the outcome in a non-randomised setting in the presence of unobserved confounding. Therefore, instrumental variable can be considered as a method that resembles the treatment assignment process in RCTs where patients could randomly be separated into two distinctive subgroups according to, for instance, policy change or geographical differences [11, 138]. In a clinical context, different values of the instrumental variable for two groups may reflect different probabilities of receiving a treatment. In particular, if the treatment rate differs, but the outcome in the two groups does not, then the treatment cannot be indicated as the cause for causing the outcome. On the contrary, if the outcomes between the two groups differ, then the treatment could be indicated as the cause for the outcome; here, comparing outcomes for two groups with different values of instrumental variable would be the same as comparing groups that have been randomised for treatment allocation.

For the instrumental variable approach to be an effective method to mitigate unobserved confounding, different assumptions must be tested and satisfied. In particular, the instrumental variable must be associated with exposure to the intervention, thus determining the treatment assignment; but at the same time, it must not be associated with the outcome variable. Finally, the instrumental variable must be associated to the outcome only because of the use of the intervention. However, in many cases it is very difficult to identify an instrument that satisfies the assumptions described [115]. In addition, it should be considered that instrumental variable estimates the local average treatment effect (LATE) that is the average treatment effect (ATE) estimated only for a specific population.

An instrumental variable could also, as suggested by Brookhart (2010) [138], be used to assess the safety and effectiveness of a new treatment using hospital formularies and administrative data. In the example provided, a new treatment is introduced for the treatment of acute MI. While existing clinical evidence suggests that the new treatment is more effective than existing therapies, the new treatment is expensive and may cause side effects in certain patients; thus, the new treatment would only be added to the formulary of some hospitals. This would create a system where patients are randomly allocated to exposure; where patients are not randomized to hospitals, but formulary status is randomized to patients, in the sense that patients are not aware of hospitals' formularies [120].

Difference in differences

The difference in differences approach, controlling for unobserved differences, estimates the treatment effect by comparing changes over time in the outcome of interest between treatment and control group. This approach relies on the two central assumptions: "common time effects across groups" and "no composition changes within groups". More specifically, this method assumes that unobserved characteristics of an individual are fixed over time. Hence, calculating the difference in outcomes in the same way as if neither group was exposed to the treatment. In addition to unobserved differences, this approach also controls for observed differences; thus minimising the risk of omitted variable bias caused by unmeasured confounders or measurement error [139]. Historically, difference in differences has been employed in evaluation of health-care policies; however, it could be used for estimating the treatment effect whenever data are available for the period preceding treatment initiation and after, and for treatment and comparison groups.
The main limitation of this method, however, consists in finding similar study groups, where treatment exposure is the only difference [140].

Regression discontinuity

Comparatively, with regression discontinuity, the treatment effect is estimated by comparing treatment exposure above and below a certain threshold of a given continuous variable for both, control and treatment group. Additional terms may be added to the model so that the slope can vary above and below a given threshold. As in the instrumental variable, regression discontinuity estimates the LATE rather than the ATE [116].

Further, depending on whether the threshold affects treatment effect deterministically or probabilistically, regression discontinuity may have a sharp or fuzzy design. In the first case, a given threshold makes a clear distinction between those who receive the treatment (e.g. those above the given threshold) and those who do not (e.g. those above the given threshold). However, in reality, this is hardly the case, as there may be individuals in treatment group who do not receive the treatment, and individuals in the control group who receive the treatment. The fuzzy design addresses this issue with a probabilistic approach where the probability of receiving the treatment is higher above or below the given threshold.

Because of its characteristics, regression discontinuity is generally regarded as the best approximation of an observational study to an RCT [115, 116, 141]. An example of regression discontinuity is represented by CD4 glycoproteins counts to establish eligibility for antiretroviral therapy among HIV-positive patient. Exogenous factors such as smoking and exercise will cause variability in the CD4 count measurement; this variability mimicking a random variable is then used for assessing whether observations are above or below a certain threshold (in this case 200 cells/μL CD4 count); this will allow for estimating treatment effect and establishing treatment assignment [142].

5.4 Methods

5.4.1 Data sources and cohort

Initially, all patients treated with either warfarin or DOACs were identified from PIS. Individual-level data linkage was then carried out with SMR01 and mortality records to identify AF patients, clinical events and calculate mortality rates. The clinical outcomes were identified from SMR01 according to ICD-10 and OPCS-4 codes.
In this Chapter, the focus is on stroke-all (including ischaemic and haemorrhagic), major bleeding and all-cause mortality; details on the ICD-10 and OPCS-4 codes used for identifying clinical outcomes and mortality are provided in Chapter 6.

In order to represent a typical cohort of anticoagulant users, the analysis was limited to patients who were 50 years or older at the time of the first OAC prescription. PIS records are available from 2009 onwards; therefore, to establish a cohort of patients with a first prescription of warfarin or DOACs, and no exposure to anticoagulation within one year prior to the index date, only patients starting anticoagulation from 2010 onwards were included in the analysis.

Further, to ensure that only patients that were likely to have received OACs because of an AF diagnosis were included (ICD10 code I48X), any patients with a diagnosis other than AF were excluded from the analysis. Clinical codes for inclusion and exclusion criteria are presented in Table 5.2. In addition, a cohort including users of any DOAC "combined DOAC" was created to increase the overall treatment sample size and assess whether any of the PS approaches tested were sensitive to sample size.

Table 5.2 Inclusion and exclusion criteria

Outcome	Diagnostic, procedure and drug codes
Inclusion criteria	
Atrial fibrillation	ICD-10: I48
Oral anticoagulant	BNF: 02.08.02
Exclusion criteria	
Mitral stenosis, valvular disease or heart valve replacement	ICD-10: I05 – I08, I34 – I37, Q22, Q23, Z95.2 – Z95.4 OPCS-4: K02.3, K25.3, K25.4, K25.8, K25.9, K26.3, K26.4, K29.1 – K29.4, K30.1, K31.1
VTE	ICD-10: I26, I63.6, I67.6, I80.1-I80.9, I81, I82.2- I82.9

5.4.2 Propensity score model

Different PS models comparing the effectiveness of DOACs against warfarin for three major AF related clinical outcomes (stroke, major bleeding and mortality) were produced. This was done in three major steps. Firstly, the PSs were estimated for each treatment, and then with aid of the NICE DSU document on the use of observational data to inform estimates of treatment effectiveness in technology appraisal, the PS methods to be tested were identified. In detail, based on the initial assessment of PS overlap between the treatment and the control group, the NICE DSU document indicates what method should be used to address observed confounding, as long as the assumption of no unobserved confounding is reasonable [114].

The document suggests that with a good PS overlap, and by assuming that "a regression model is a good approximation of the effect of covariates on outcome" (best specification possible in terms of functional form and covariates), regression adjustment, on its own or as double robust combined with IPW, should be used. Alternatively, in cases where there is not a good approximation, the methods that should be used are IPW, on its own or as double robust combined with regression adjustment, or matching [114].

With a poor overlap, the guidelines suggest the use of matching methods to improve the overlap; then if a good or a moderate balance is obtained, regression methods should be applied on the matched sample. If poor overlap is still present following matching, trimming the sample may improve the overlap. With this approach, however, there is the risk of removing a large portion of individuals from the analysis, causing the estimates to refer to the treatment effect on the treated rather than the ATE. If after trimming, a poor balance is still persistent, it may not be possible to conduct the analysis assuming selection on observables [114]. Finally, PS model specification assessment was carried out to identify the most robust PS based method among the ones initially selected. In this step PS overlap was assessed for each of the methods identified and for each treatment group [114].

PS estimation

The PS estimation was carried out for each individual DOAC (apixaban, dabigatran and rivaroxaban), warfarin and the combined DOAC cohorts, resulting in five different PS models. Propensity scores were estimated with a logit model accounting for the following baseline characteristics of first time OAC users: age at the time of the first prescription, sex, SIMD, CHA_2DS_2-VASc and HAS-BLED score, ischaemic stroke or SE or TIA, vascular disease, hypertension, diabetes, cancer, prescription predisposing bleeding, and comorbidity. The PS in each of the five different PS models was estimated according to the full set of covariates listed above. The risk scores were calculated for each patient for the 5-year periods prior to their first anticoagulation prescription.

The CHA_2DS_2-VASc is an extension of the previously developed $CHADS_2$. The risk score of both tools depends on patients' previous history of stroke, TIA, congestive heart failure, hypertension and diabetes identified in SMR01 according to their corresponding ICD10 codes as shown in Table 5.3. Patient age, if over 65 years, is also included in the risk score equation. For the CHA_2DS_2-VASc, the risk score is calculated by also taking into account patients' sex and age if greater than 75 years, and whether there is a prior history of vascular disease including MI, peripheral arterial disease, or aortic plaque. These are important risk predictors that may allow for a more accurate calculation of the potential risk of stroke in patients affected by AF [30, 143]. In the CHA_2DS_2-VASc tool, a point is given for each variable as described in Table 5.3; two points are given if a patient has a previous history of stroke or TIA [143].

Although CHA_2DS_2-VASc is widely used, there were limitations in the development of the risk-predicting tool that could have an impact on predicting thromboembolic events. Firstly, information on the occurrence of thromboembolic events during the first year from the baseline survey was only available for 69% of patients. Secondly, the primary analysis was based on 1,084 patients that were not anticoagulated at baseline, and during the first year follow up, only 18% started anticoagulation [143]. In the PS models, CHA_2DS_2-VASc was entered as a categorical variable where the risk of stroke is represented by the following risk categories: low to moderate (0-1), moderate to high (2-3) and high (>=4).

Table 5.3 CHA2DS2-VASc - ICD-10 and OPCS-4 codes

Variable	Diagnostic, procedure and drug codes	Points if present
Stroke or TIA	ICD-10: I26, I63, I64, G45.8, G45.9, G46.3 - G46.7	2
Congestive heart failure	ICD-10: I11.0, I13.0, I13.2, I50	1
Hypertension	ICD-10: I10 - I15	1
Vascular disease	ICD-10: I20 - I22, I70, I73.1, I73.8, I73.9, I74	1
Diabetes mellitus	ICD-10: E10, E11, E13, E14, G59.0, G63.2, H28.0, H36.0, M14.2, N08.3, O24.0, O24.1, O24.3	1
Sex Category (female)	not applicable	1
Age ≥65 years	not applicable	1
Age ≥75 years	not applicable	1

The HAS-BLED scoring system, (Table 5.4), is similar to the one described for CHA_2DS_2-VASc. The risk score is calculated according to age, history of hypertension, renal and liver disease, stroke, TIA and major bleeding [31, 144]. Alcohol intake and the use of drugs predisposing to bleeding such as aspirin, Clopidogrel, or non-steroidal anti-inflammatory drugs, are two other important variables determining the potential risk of bleeding in AF patients. Alcohol intake was defined according to alcohol abuse related hospital admissions, and according to prescription of drugs used for alcohol dependence defined by BNF subsection 04.10.01 and identified from PIS records.
Similarly, the use of drugs promoting bleeding was established according to BNF section 02.09 and BNF subsection 10.01.01.

As, laboratory data were not part of the administrative health datasets, labile INR indicating quality of anticoagulation by means of unstable or high INRs or poor time in therapeutic range was not included in determining the potential risk of bleeding. However, as Poli et al. suggest, the redefined version of HAS-BLED, renamed HAS-BED and not taking into account the labile INR item, is still able to predict a high risk of bleeding [144].

The full list of variables and codes required for calculating the HAS-BLED risk score is reported in Table 5.4. In the PS models, HAS-BLED was entered as a categorical variable where the risk of bleeding is represented by the following risk categories: low to moderate (0-2), and moderate to high (>=4).

Table 5.4 HAS-BLED - ICD-10 and OPCS-4 and BNF codes

Variable	Diagnostic, procedure and drug codes	Points if present
Hypertension	ICD-10: I10 - I15	1
Renal disease	ICD-10: I12, I13, N00 - N05, N07, N11, N14, N17 - N19, Q61	1
Liver disease	ICD-10: B15.0, B16.0, B16.2, B19.0, K70 - K76	1
Stroke or TIA	ICD-10: I26, I63, I64, G45.8, G45.9, G46.3 - G46.7	1
Major bleeding	ICD-10: D62, H11.3, H35.6, H43.1, I60 – I62, J94.2, K25.0, K25.2, K25.4, K25.6, K26.0, K26.2, K26.4, K26.6, K27.0, K27.2, K27.4, K27.6, K28.0, K28.2, K28.4, K28.6, K29.0, K62.5, K92.0-92.2, N02, N95.0, R04, R31, R58	1
Alcohol intake	ICD-10: E52, F10, G31.2, G62.1, G72.1, I42.6, K29.2, K70, K86.0, O35.4, T51, Z71.4, Z72.1 BNF: 04.10.01	1
Drugs predisposing to bleeding	BNF: 02.09, 10.01.01	1
Age ≥65 years	not applicable	1

The previous history of stroke or SE or TIA, vascular disease, hypertension, diabetes, prior to OAC initiation, was ascertained from inpatient records. Further, previous history of cancer was included as it may be a reason for anticoagulation. All variables mentioned above and reflecting baseline characteristics that may influence treatment assignment were included in the PS models, as the variance inflation factor did not show evidence of collinearity. Because the PS method identified as the most robust will inform the comparative-effectiveness model in the subsequent Chapter, collinearity was also tested for standard and reduced DOACs (Appendix IV, Table IV-1).

PSs were estimated with the following equation:

$$(0 < \text{PS} \leq 1) = \alpha + \beta_1 A_i + \beta_2 G_i + \sum_{s=2}^{5} \beta_3 S_i + \sum_{s=2}^{3} \beta_4 CH_i + \beta_5 B_i$$

$$+ \sum_{c=2}^{3} \beta_6 C_i + \beta_7 S_i + \beta_8 V_i + \beta_9 H_i + \beta_{10} D_i + \beta_{11} CA_i + \beta_{12} P_i + u_i$$

(Equation 5.5)

Where A is age at the time of admission (reference category: 50 -54 age group); G is sex (reference category: male); S is SIMD category (reference category: most deprived category (1)); CH is the CHA_2DS_2-VASc score (reference category: 0-1); B is the HAS-BLED score (reference category: 0-2); C is Charlson comorbidity index (reference category: no comorbidities); S,V,H,D,CA are stroke or SE or TIA, vascular disease, hypertension, diabetes and cancer (dummy variables 0=no event, 1=event); P is prescription predisposing bleeding (reference category: no prescription); u_i is the error term for patient *i*

PS models to be tested

Assuming that the assumption of no unobserved confounding is reasonable, and in line with the objective of this Chapter, only PS based methods were tested in this section. In particular, PS matching, covariate adjustment including PS as covariate and a series of IPW methods were tested as shown in Figure 5.1. In this process, the PS distribution was assessed for each model to assess whether the cohorts were adequately balanced and to identify potentially extreme weights. As the use of DOACs for the prevention of stroke in the AF population is relatively new compared to warfarin, it was assumed that warfarin (the control) outnumbered the treatment group (DOACs); this was however tested and presented in the Results section.

Thus, the PS matching without replacement was selected as the most suitable PS matching method. In the covariate adjustment method, the only step required in the PS model was PS estimation.

The IPW methods tested were IPTW and IPTW combined with IPCW. In the IPTW method, a weight, reflecting the probability of being exposed to the combined DOAC cohort and equal to the reciprocal of the PS, was assigned to each patient in the treatment group; the same process was carried out for each DOAC individually. A weight equal to the reciprocal of one minus the PS was assigned to patients in the warfarin group. In the trimmed IPTW, individuals with extreme PS values were removed from the analysis; these were PSs <0.1 for the DOAC cohorts and PSs >0.9 for the warfarin group.
Typically, 5% of patients with the most extreme PS are excluded from the analysis; however, this seems to be arbitrary and clear guidelines on this matter are not currently available [120].

In the IPW method combining IPTW with IPCW, two different sets of weights were estimated. The weights for IPTW were estimated as discussed. Those for IPCW were estimated by censoring patients who switch treatment (from warfarin to DOACs, from DOACs to warfarin, or between DOACs), and assigning weights to individuals who were not censored but shared similar characteristics with the switchers. Then, IPTW and IPCW weights were added to obtain the overall weight reflecting ATE and censoring.

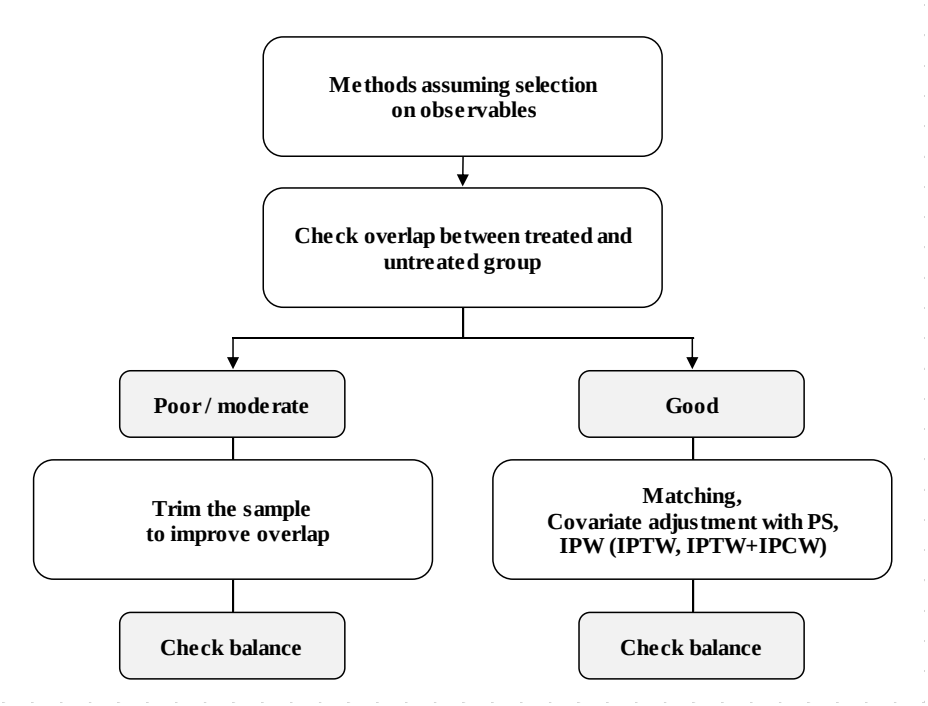


Figure 5.1 Propensity score model selection

Abbreviations: PS=propensity score, IPW=inverse probability weighting, IPTW=inverse probability of treatment weighting, IPCW= inverse probability of censoring weighting.
Source: Adapted from, Faria et al. (2015).

PS model specification assessment

Following a selection of PS based models, adequacy of model specification was assessed graphically and by means of standardized difference, a measure generally used to compare the mean of variables between treatment groups. Differences in the means of covariates were considered negligible if below the threshold of 0.1 standard deviation [127].

5.4.3 Outcome model

Cox proportional hazards regression (time-to-event analysis) was used to compare risks between treatment groups for three major AF related clinical events: stroke-all (including haemorrhagic and ischaemic), major bleeding and all-cause mortality. Following a double robust approach, some of the covariates used in the PS models were used in the outcome models; these were age, sex, comorbidity, socio economic status, and year of prescription. This was done for all PS based methods tested.

Following a continuous treatment approach [63], patients were censored if they switched or discontinued treatment; for each method, the risks of stroke, major bleeding or death (for patients exposed to either DOACs or warfarin) were estimated from anticoagulation initiation to the time of clinical event or death during a 2-year follow-up period. However, descriptive statics on treatment switching are reported at 2-year and 6-year follow-up. The first clinical event for each treatment was determined within a competing risk framework. In this analysis, treatment discontinuation, i.e. temporal gaps between consecutive prescriptions, was considered to be occurring if the gap exceeded a 28 days threshold, and the supply of the penultimate prescription did not fill the gap. The threshold was identified in a drug utilisation study using the same patient-level data utilised in this thesis [27].

For the IPW method combining IPTW with IPCW, censoring was specifically modelled in the PS model.

In this Chapter, the Cox model was used as a proof of concept, characteristics and assumptions of the Cox model will be discussed in more detail in Chapter 6. In addition to comparing PS models in terms of robustness by measuring the standardised differences for each covariate, the ATE, estimated with the outcome model for each of the clinical outcomes selected, was compared across method to assess whether and how it differs depending on the PS method used. The ATE estimates the treatment effect on the cohort of interest but could also selectively estimate the treatment effect only for those patients, that within the same cohort, are treated (treatment effect on the treated -ATT). Because the focus of this analysis was estimating the treatment effect on the whole AF population, regardless of whether patients where on DOACs or warfarin, the ATE was estimated.

5.5 Results

5.5.1 Cohort characteristics

From the initial cohort of 166,182 patients on any OAC identified from the PIS between April 2009 and December 2015, a cohort of 165,627 patients on either warfarin or DOACs were identified. The final cohort resulted in 33,965 treatment naïve patients, of which 26,387 were on warfarin, 3706 on apixaban, 484 on dabigatran and 3,388 on rivaroxaban. The diagram showing how the cohort for this analysis was obtained is presented in Figure 5.2

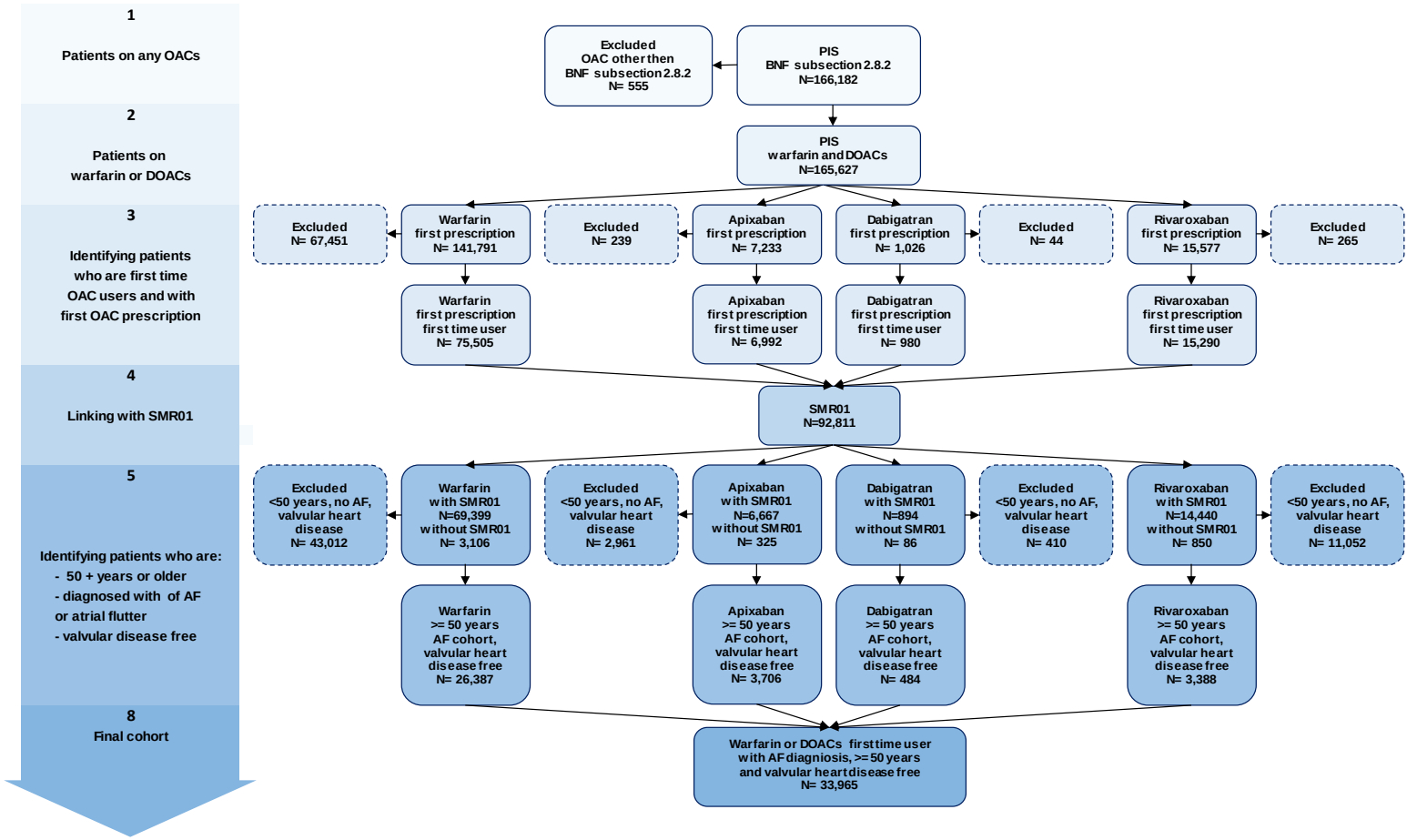


Figure 5.2 Study population

The initial study population, firstly identified from PIS, was linked to inpatients and mortality data to identify patients that were likely to have received oral anticoagulants because of an AF or atrial flutter diagnosis according to the ICD10 code I48X. Only patients 50 years or older and naïve to anticoagulation were included in the analysis, those with a history of valvular heart disease or venous thromboembolism were excluded.

Overall, mean age of patients at the time of the first prescription was similar across all treatment groups. Across all treatments, patients with the highest risk of stroke measured using the CHA_2DS_2-VASc score represented the majority. This is clearly visible in the histogram (Figure V-1) presented in Appendix IV, indicating that most patients had a CHA_2DS_2-VASc score of 4 points (laying within the >=4 risk category) across all treatment groups. Patients with the lowest risk of bleeding measured using the HAS-BLED score represented the majority. Again, this is clearly visible in the histogram (Figure V-2) presented in Appendix III, indicating that most patients had a HAS-BLED score of 2 points (laying within the >=2 risk category) across all treatments; however, the proportion of patients with the lowest risk of stroke was higher in the dabigatran group than any other treatment group. Similarly, while most patients had no comorbidities across all treatment groups, those on dabigatran represented the biggest proportion (Appendix V, Figure V-3). Further, the proportion of patients with a history of stroke or SE or TIA was lower in the dabigatran group than any other treatment group. Over one third of patients on anticoagulation had hypertension, which is an important risk factor for stroke. In addition, the majority of patients across all treatment groups were also on medication predisposing to bleeding such as aspirin and non-steroidal anti-inflammatory drugs than patients on any other DOACs. Patients' baseline characteristics are reported in Table 5.5

Table 5.5 Baseline characteristics

Characteristics	Apixaban N(%)	Dabigatran N(%)	Rivaroxaban N(%)	Warfarin N(%)
Cohort	3,706	484	3,388	26,387
Sex				
Men	1,983 (53.51)	304 (62.81)	1,805 (53.28)	14,437 (54.71)
Women	1,723 (46.49)	180 (37.19)	1,583 (46.72)	11,951 (45.29)
Mean age (SD)	75 (10.26)	72 (10.44)	75 (10.23)	74 (9.23)
SIMD (Scottish index of multiple deprivation)				
1	845 (22.80)	44 (9.09)	524 (15.47)	5,019 (19.02)
2	737 (19.89)	89 (18.39)	613 (18.09)	5,637 (21.36)
3	615 (16.59)	125 (25.83)	749 (22.11)	5,581 (21.15)
4	642 (17.32)	132 (27.27)	828 (24.44)	5,305 (20.10)
5	867 (23.39)	94 (19.42)	674 (19.89)	4,845 (18.36)
CHA_2DS_2-VASc score				
0-1	826 (22.29)	164 (33.88)	697 (20.57)	5,547 (21.02)
2-3	1,179 (31.81)	144 (29.76)	1,078 (31.82)	9,084 (34.43)
>=4	1,701 (45.90)	176 (36.36)	1,613 (47.61)	11,756 (44.55)
HAS-BLED				
0-2	2,338 (63.09)	347 (71.69)	2,178 (64.29)	16,878 (63.96)
>=3	1,368 (36.91)	137 (28.31)	1,210 (35.71)	9,509 (36.04)
Comorbidity				
no comorbidity	1,814 (48.95)	317 (65.50)	1,714 (50.59)	13,831 (52.42)
1 comorbidity	883 (23.83)	83 (17.15)	769 (22.70)	5,974 (22.64)
>1 comorbidities	1,009 (27.23)	84 (17.36)	905 (26.71)	6,582 (24.94)
Stroke or TIA	543 (14.65)	43 (8.88)	502 (14.82)	3,490 (13.23)
Vascular disease	633 (17.08)	70 (14.46)	601 (17.74)	4,778 (18.11)
Hypertension	1,307 (35.27)	154 (31.82)	1,237 (36.51)	9,653 (36.58)
Diabetes mellitus	528 (14.25)	62 (12.81)	476 (14.05)	3,558 (13.48)
Cancer	321 (8.66)	41 (8.47)	290 (8.56)	1,982 (7.51)
Drug causing bleeding	2267 (61.17)	282 (58.26)	1,974 (58.26)	17,373 (65.84)

At 2-year follow-up, patients initiating anticoagulation with warfarin or dabigatran were more likely to discontinue compared to patients on other anticoagulation treatments. On the contrary, the majority of patients starting anticoagulation with apixaban where more likely to stay on treatment. Patients initiating anticoagulation with warfarin were also more likely to switch to rivaroxaban than any other DOAC. A much smaller proportion of patients also switched from DOAC to warfarin and from DOAC to DOAC. Findings for discontinuation and switching at 6 years were in line with those observed at 2 years follow-up. The full description on patients switching and discontinuing is presented in Table 5.6.

Table 5.6 Treatment discontinuation or switching

Discontinuing/switching	2-year follow-up N (%)	6-year follow-up N (%)
All patients		
Discontinue	11,366 (33.46)	13,239 (38.98)
Switch warfarin to DOAC	2,146 (6.32)	3,568 (10.50)
Switch DOAC to warfarin	150 (0.44)	164 (0.48)
Switch DOAC to DOAC	196 (0.58)	214 (0.63)
Warfarin		
Discontinue	9,980 (37.82)	11,502 (43.59)
Switch to apixaban	746 (2.83)	1,360 (5.15)
Switch to dabigatran	226 (0.86)	301 (1.14)
Switch to rivaroxaban	1,174 (4.45)	1,907 (7.23)
Apixaban		
Discontinue	348 (9.39)	410 (11.06)
Switch to warfarin	30 (0.81)	31 (0.84)
Switch to dabigatran	9 (0.24)	9 (0.24)
Switch to rivaroxaban	35 (0.94)	35 (0.94)
Dabigatran		
Discontinue	200 (41.32)	259 (53.51)
Switch to warfarin	31 (6.40)	37 (7.64)
Switch to apixaban	16 (3.31)	25 (5.17)
Switch to rivaroxaban	37 (7.64)	42 (8.68)
Rivaroxaban		
Discontinue	838 (24.73)	1,068 (31.52)
Switch to warfarin	89 (2.63)	96 (2.83)
Switch to apixaban	79 (2.33)	83 (2.45)
Switch to dabigatran	20 (0.59)	20 (0.59)

5.5.2 Propensity score distribution

The PSs for the DOACs versus warfarin comparison showed an adequate overlapping distribution, and no extreme PSs are observed (Figure 5.3). By contrast, when comparing DOACs individually against warfarin different scenarios emerged. The PSs generated from the apixaban versus warfarin analysis, graphically showed an adequate overlap, however extreme weights were observed for the apixaban group. In particular, 760 patients amongst the apixaban users had PSs <0.1(Figure 5.4). To avoid trimming a large proportion of patients (20% of individuals on apixaban) with extreme PS and causing further covariate imbalance, only 5% of patients with the most extreme PS were excluded from the analysis.

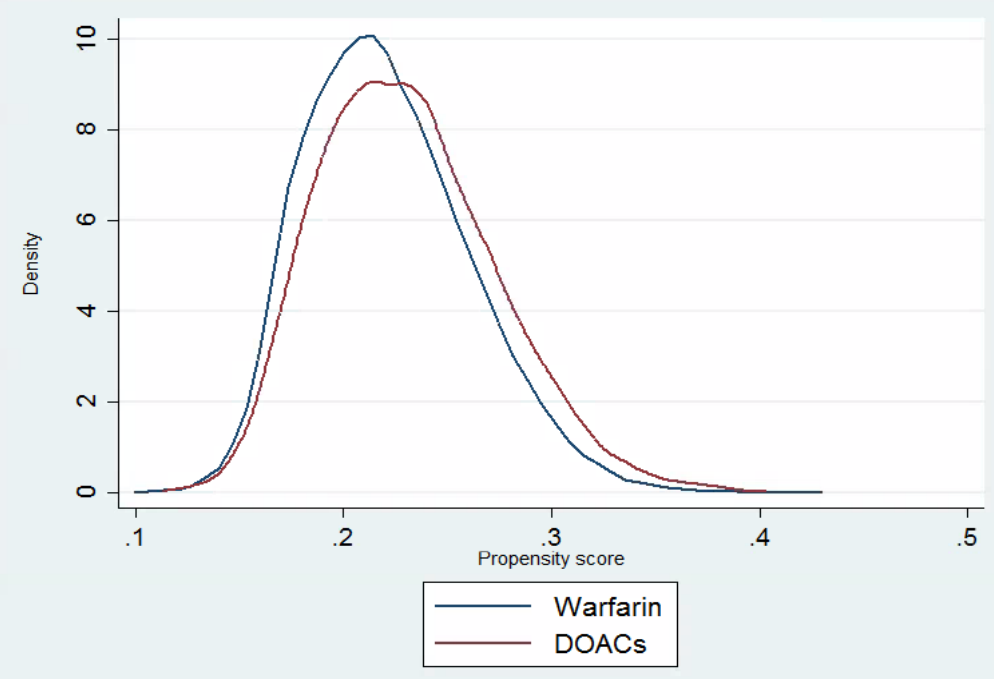


Figure 5.3 PS distribution for combined DOACs and warfarin

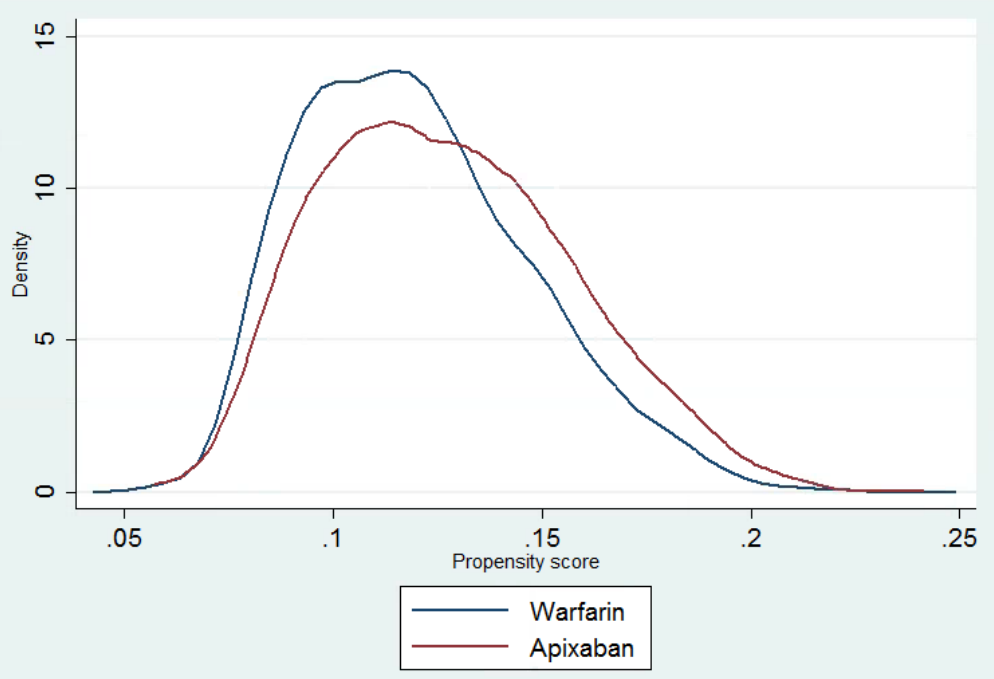


Figure 5.4 PS distribution for apixaban and warfarin

The totality of the PS generated from the dabigatran versus warfarin comparison were extreme and had a poor overlap (Figure 5.5). In these cases, applying PS trimming or extreme weights truncation is clearly not feasible, and any PS method may give incorrect ATE estimates.

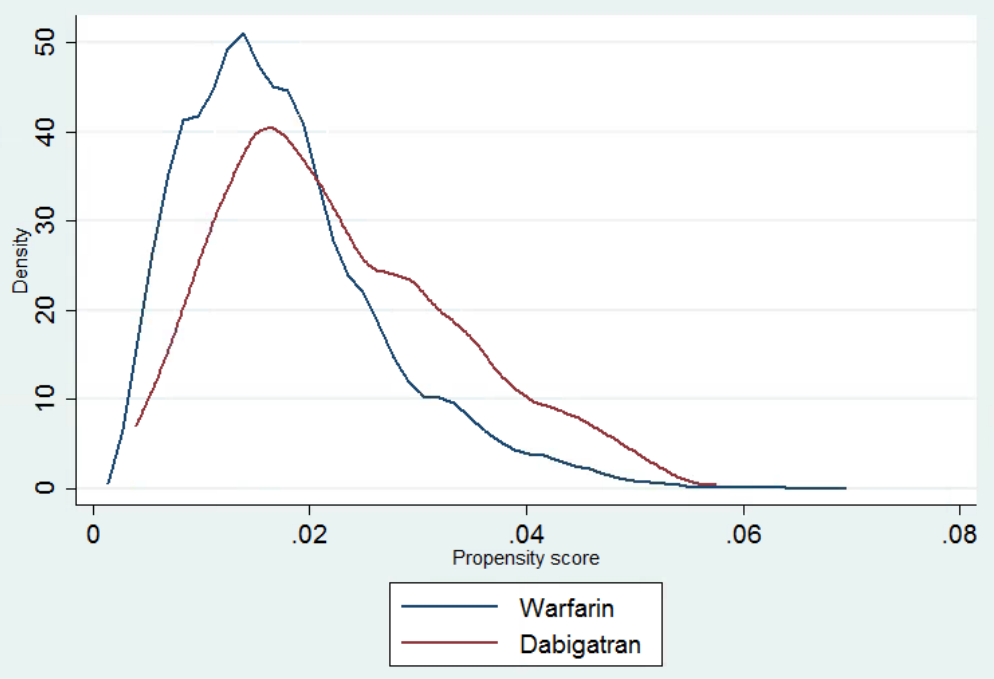


Figure 5.5 PS distribution for dabigatran and warfarin

Despite an acceptable overlap of the PS distribution, 1,021 patients on rivaroxaban had extreme PS (Figure 5.6). As with apixaban, to avoid additional imbalance only 5% of patients with the most extreme PSs were excluded from the analysis.

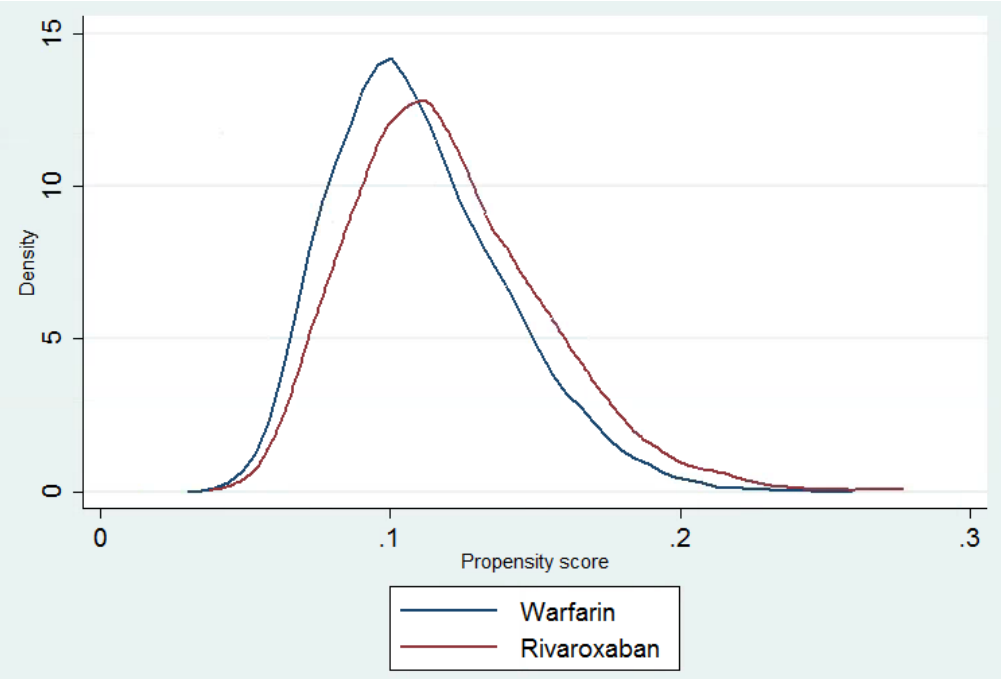


Figure 5.6 PS distribution for rivaroxaban and warfarin

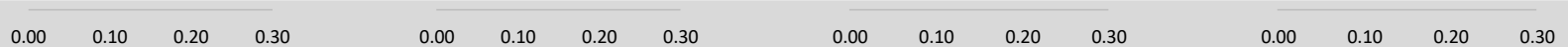
5.5.3 Covariate balance assessment

Combined DOACs versus warfarin

Following the first graphical assessment on the PS models specification, the distribution of baseline covariates between treatment groups was assessed by means of standardized difference. As shown in Table 5.7, the unadjusted standardized differences indicated an adequate starting balance for the baseline characteristics of patients on any DOAC (combined DOACs) or on warfarin, with differences in the means of covariates below the threshold of 0.1 standard deviation. Overall, a good balance of patients' covariates was obtained with every PS model tested. However, the standardised difference for CHA_2DS_2-VASc score >=4, HAS-BLED score >=3 did not improve with the PS matching method. Nevertheless, the standardised difference for these patient characteristics was still below the threshold regardless of being adjusted or unadjusted. The same issue was encountered when adopting the IPW approach were no balance improvement was observed for the HAS-BLED score >=3 and having one comorbidity.

Table 5.7 Covariate imbalance assessment for combined DOACs vs. warfarin

Characteristics	Unadjusted	PSM adjusted	PS covariate adjusted	IPTW adjusted
Women	0.014	0.008	0.003	0.010
Mean age (SD)	0.086	0.034	0.008	0.013
SIMD quintile 2	0.059	0.003	0.003	0.005
SIMD quintile 3	0.037	0.006	0.003	0.005
SIMD quintile 4	0.026	0.000	0.003	0.005
SIMD quintile 5	0.080	0.012	0.003	0.005
CHA2DS2-VASc score 2-3	0.058	0.004	0.005	0.004
CHA2DS2-VASc score >=4	0.030	0.034	0.005	0.004
HAS-BLED score >=3	0.004	0.023	0.003	0.012
Comorbidity 1 comorbidity	0.006	0.001	0.002	0.007
Comorbidity >1 comorbidity	0.033	0.031	0.002	0.007
Stroke or TIA	0.033	0.027	0.003	0.010
Vascular disease	0.024	0.008	0.001	0.002
Hypertension	0.020	0.015	0.003	0.015
Diabetes mellitus	0.017	0.041	0.003	0.010
Cancer	0.040	0.020	0.000	0.005
Drugs causing bleeding	0.128	0.019	0.001	0.013



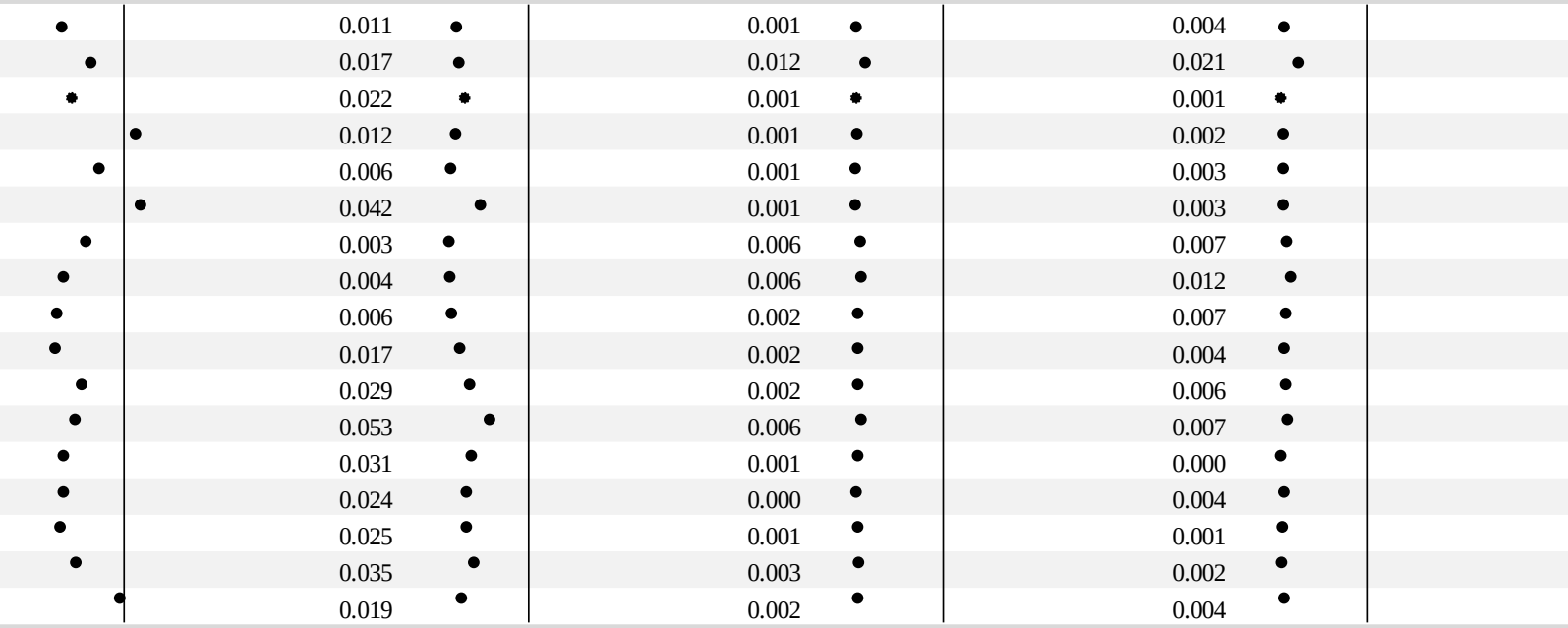
Abbreviations: SIMD=Scottish index of multiple deprivation, TIA=transient ischaemic attack.

Apixaban versus warfarin

An overall adequate balance for baseline characteristics was also observed between apixaban and warfarin, where the standardised difference was further reduced for every covariate with the PS covariate adjustment and IPW method. However as previously observed, balance of stroke or TIA, vascular disease and diabetes mellitus, did not improve when using the PS matching method (Table 5.8).

Table 5.8 Covariate imbalance assessment for apixaban vs. warfarin

Characteristics	Unadjusted	PSM adjusted	PS covariate adjusted	IPTW adjusted
Women	0.024	0.011	0.001	0.004
Mean age (SD)	0.061	0.017	0.012	0.021
SIMD quintile 2	0.036	0.022	0.001	0.001
SIMD quintile 3	0.117	0.012	0.001	0.002
SIMD quintile 4	0.071	0.006	0.001	0.003
SIMD quintile 5	0.124	0.042	0.001	0.003
CHA2DS2-VASc score 2-3	0.056	0.003	0.006	0.007
CHA2DS2-VASc score >=4	0.027	0.004	0.006	0.012
HAS-BLED score >=3	0.018	0.006	0.002	0.007
Comorbidity 1 comorbidity	0.017	0.017	0.002	0.004
Comorbidity >1 comorbidity	0.049	0.029	0.002	0.006
Stroke or TIA	0.041	0.053	0.006	0.007
Vascular disease	0.027	0.031	0.001	0.000
Hypertension	0.027	0.024	0.000	0.004
Diabetes mellitus	0.022	0.025	0.001	0.001
Cancer	0.042	0.035	0.003	0.002
Drugs causing bleeding	0.097	0.019	0.002	0.004

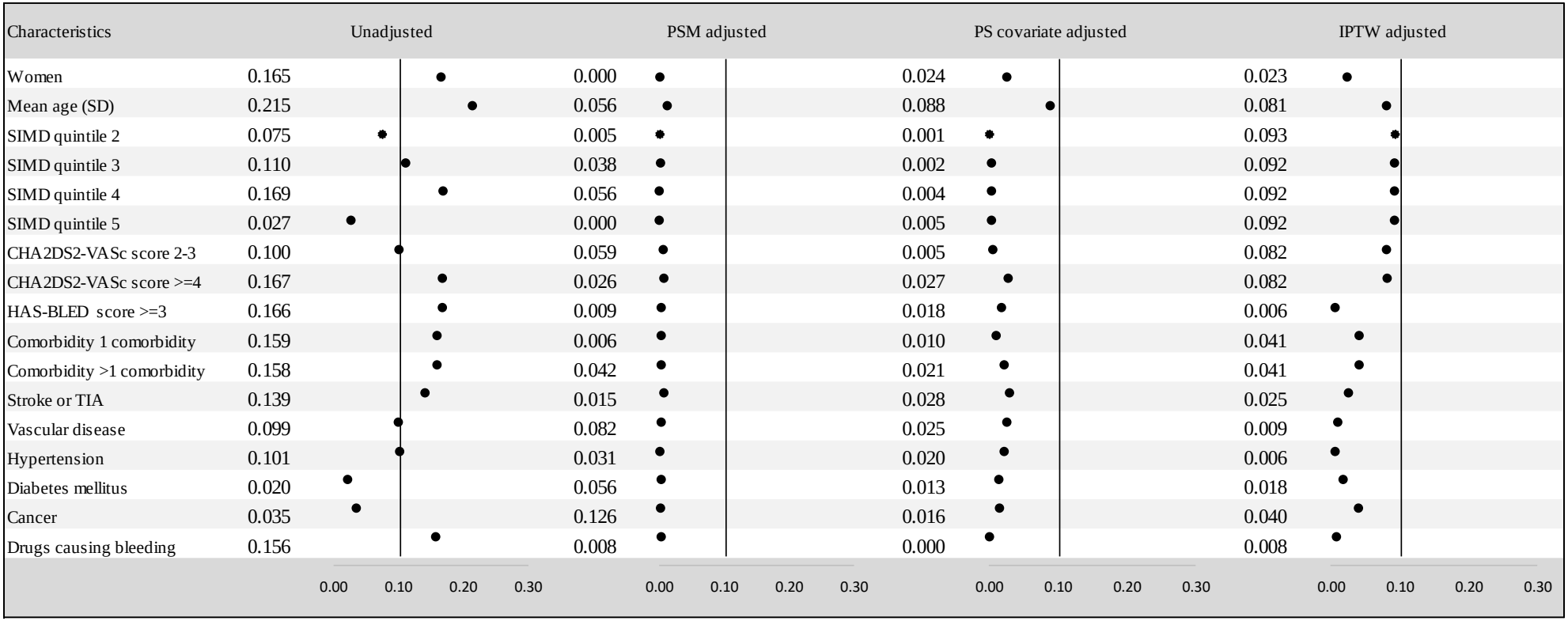


Abbreviations: SIMD=Scottish index of multiple deprivation, TIA=transient ischaemic attack.

Dabigatran versus warfarin

The standardized difference above the threshold reflected a substantial difference in the starting baseline characteristics of dabigatran and warfarin users. An adequate balance was however achieved for all covariates with the PS covariate adjustment method. Again, PS matching failed to provide an optimal balance in terms of patient characteristics between the dabigatran and warfarin groups. Similarly, improved balance was not achieved for every covariate when using the IPW approach. In particular, the balance for socio demographic characteristics captured by SIMD (category 2 and 5) and the covariate indicating the diagnosis of cancer, although still below the threshold, was suboptimal compared to the unadjusted initial baseline characteristics balance (Table 5.9).

Table 5.9 Covariate imbalance assessment for dabigatran vs. warfarin



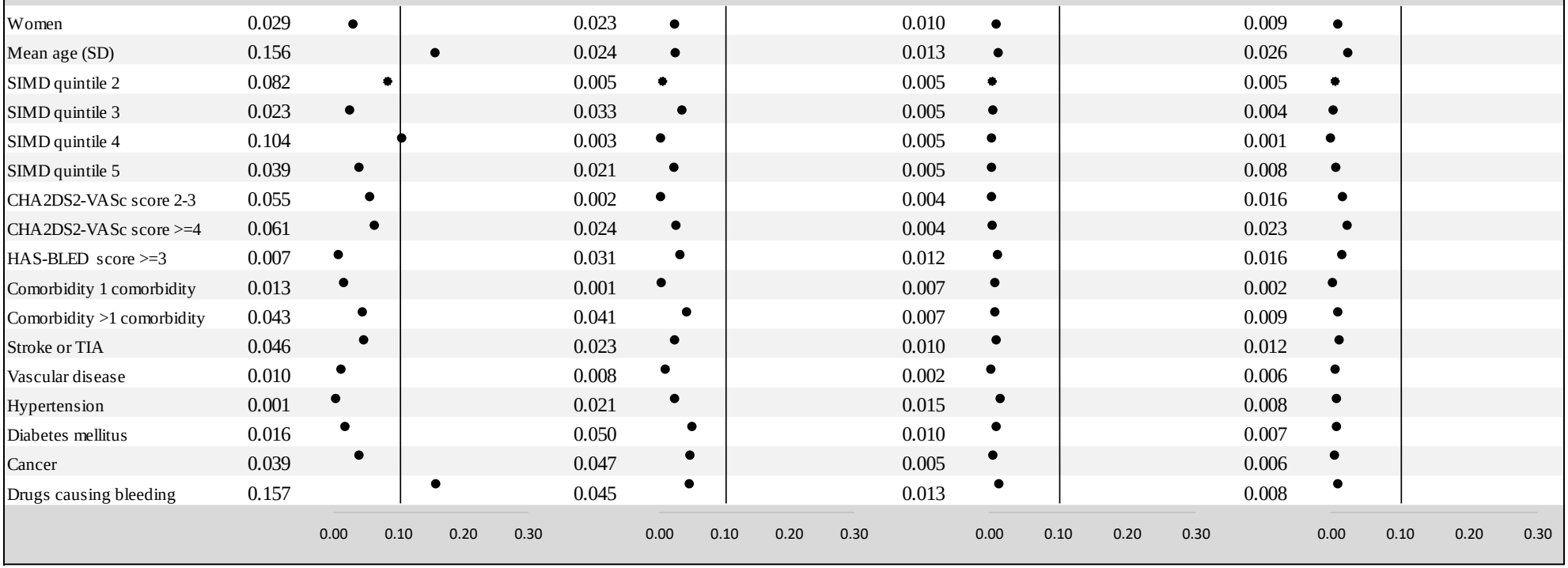
Abbreviations: SIMD=Scottish index of multiple deprivation, TIA=transient ischaemic attack.

Rivaroxaban versus warfarin

An overall good starting balance was also observed for the baseline characteristics of rivaroxaban and warfarin users. However, the mean age of patients at the time of the first prescription, SIMD (category 3) and the use of drugs causing bleeding, seemed to some extent differ between the two treatment groups. While, the PS methods brought the balance for those covariates below the 0.1 threshold, the balance for HAS-BLED score >=3 and hypertension did not improve with any of the methods tested (Table 5.10).

Table 5.10 Covariate imbalance assessment for rivaroxaban vs. warfarin

Characteristics	Unadjusted	PSM adjusted	PS covariate adjusted	IPTW adjusted
Women	0.029	0.023	0.010	0.009
Mean age (SD)	0.156	0.024	0.013	0.026
SIMD quintile 2	0.082	0.005	0.005	0.005
SIMD quintile 3	0.023	0.033	0.005	0.004
SIMD quintile 4	0.104	0.003	0.005	0.001
SIMD quintile 5	0.039	0.021	0.005	0.008
CHA2DS2-VASc score 2-3	0.055	0.002	0.004	0.016
CHA2DS2-VASc score >=4	0.061	0.024	0.004	0.023
HAS-BLED score >=3	0.007	0.031	0.012	0.016
Comorbidity 1 comorbidity	0.013	0.001	0.007	0.002
Comorbidity >1 comorbidity	0.043	0.041	0.007	0.009
Stroke or TIA	0.046	0.023	0.010	0.012
Vascular disease	0.010	0.008	0.002	0.006
Hypertension	0.001	0.021	0.015	0.008
Diabetes mellitus	0.016	0.050	0.010	0.007
Cancer	0.039	0.047	0.005	0.006
Drugs causing bleeding	0.157	0.045	0.013	0.008



Abbreviations: SIMD=Scottish index of multiple deprivation, TIA=transient ischaemic attack

5.6 Discussion

In clinical practice, population case mix may diverge substantially, making a comparison of safety and effectiveness of two health interventions difficult. PS methods allow for reducing any potential imbalance between covariates and obtaining more homogenous and comparable treatment groups [59, 125]. When comparing combined DOACs, apixaban and rivaroxaban against warfarin, with the exception of the dabigatran versus warfarin analysis, a reasonable balance of baseline characteristics was observed even before adjusting with PS estimates. However, in some cases the standardised difference indicated that the balance of certain baseline characteristics between treatments did not improve after PS adjustment. This occurrence is reported in the literature, and it seems to be common with the PS matching method [131].

Overall patients on dabigatran were younger, with a low risk of stroke and with fewer comorbidities compared to patients on any other DOAC or warfarin. This seems to suggest that dabigatran was selectively prescribed to patients with lower risk of stroke and in general healthier than patients on either warfarin, apixaban or rivaroxaban. Evidence of selective prescribing of dabigatran in younger patients with lower risk of stroke has been reported in the literature [63]. Further, as expected, patients on either reduced dose apixaban or reduced dose rivaroxaban, where at greater risk or stroke, bleeding and had more comorbidities than patients on apixaban and rivaroxaban standard dose. As per clinical guidelines reduced dose apixaban and rivaroxaban are recommended for patients aged 80 years or older [16, 18].

Among the PS methods tested, PS covariate adjustment, less sensitive to sample size, was shown to be the most robust method. As discussed, this is one of the main advantages over the PS matching approach. From a sole sample size perspective, PS covariate adjustment and IPW methods offer the same advantage, but trimming of PS or truncation of extreme weights may reduce the sample size further.

Theoretically, PS covariate adjustment is less robust than PS matching and IPW methods, as it is more sensitive to distributional assumptions and PS specification, therefore not reflecting the true treatment effect [123, 128, 129]. Nevertheless, PS covariate adjustment was found to be a valid option to adjust for confounding by indication and in some instances outperformed the other methods reporting much reduced standardised differences. When censoring was incorporated in an IPTW-IPCW model rather than being modelled explicitly, the results were in line with those generated with other methods. Moreover, PS methods may not necessarily perform better than conventional standard regression. In particular, Elze and colleagues (2017) found that in the presence of substantial covariate imbalance with individuals with very large weights, IPW methods give inaccurate treatment effect estimates. In the case studies evaluated, after truncation, the estimated treatment effect moved towards the crude treatment effect, indicating the inadequacy of these methods in adjusting for covariate imbalance in the presence of heavy weights. On the other hand, the performance of PS matching and standard covariate adjustment were comparable, although PS matching gave less accurate estimates in some instances [104].

5.7 Limitations

The analysis carried out provides an overview of the PS based methods used to address confounding by indication; however, there were a number of limitations inherent to the nature of RWE and PS based methods. Firstly, the relatively small size of the cohorts did not allow the analysis to test for PS by stratification, a method involving the stratification of individuals into mutually exclusive subgroups according to their estimated PS [59]. In this method, individuals are typically divided into five distinct subgroups using the quintiles of the estimated PS; this will consent the treated and the untreated groups to have similar PS values and comparable distributions of measured baseline covariates. However, PS by stratification does not seem to perform optimally in datasets with few outcomes [59, 120].

A further constraint in this analysis, concerns the limitation of PS methods of addressing unmeasured confounding which may still bias the estimates. In particular, it is recognised that confounding by indication is the main source of confounding in newly marketed medications where early adopters are most likely to prescribe new drugs when they become available, whereas other prescribers may prefer to opt for existing drugs with proved and established clinical effectiveness [145]. While PS methods can address confounding by indication, there may still be unobserved confounders that are difficult to measure.

5.8 Conclusions

Comparative-effectiveness research is rapidly evolving; in this context, RWE has a crucial role in supporting decision-making by generating clinical evidence for clinicians, patients and policy makers. Propensity score methods are powerful tools that mimic RCTs and address confounding by indication typical in RWE studies. However, PS methods have strengths and limitations that should be acknowledged. Propensity score matching and IPW methods are considered theoretically superior to PS covariate adjustment as the latter may be more prone to model misspecification.

Nevertheless, in this study, PS covariate adjustment was found to be the most robust method, PS matching and IPW methods also performed well, but were excluded to avoid further sample size reduction. Therefore, the use of a single best method for reducing bias due to confounding by indication should be avoided, and the choice of adequate PS methods may vary according to the characteristics of the data. It follows that, as long as assumptions such as no unobserved confounding hold, several methods should be identified and tested.

For the comparative-effectiveness of DOACs in the next Chapter, the choice of any PS model over another is unlikely to lead to divergent conclusions. Nevertheless, for the reasons discussed, PS used as a covariate is the approach that will be used for the comparative and cost-effectiveness analyses in the Chapters 6 and 7.

Chapter 6 Comparative-effectiveness analysis

6.1 Introduction

This Chapter builds on the work carried out in the previous Chapter where covariate adjustment with PS was identified as the best fitting PS model for the DOACs case study. As mentioned, PS methods are typically used in observational studies to address confounding by indication [59]. While the PS model described in Chapter 5 only focused on stroke, major bleeding and all-cause mortality, in this Chapter a more comprehensive comparative-effectiveness analysis is carried out including all primary outcomes reported in the pivotal RCTs and existing observational studies assessing efficacy and effectiveness of DOACs in the AF population [41-43, 58, 60, 61, 63]. In addition, although a Cox model was used as a proof of concept in Chapter 5 for comparing risks of selected clinical outcomes, in this Chapter a more rigorous approach (e.g. assessing proportionality violation) is adopted to identify the most appropriate method for the time-to-event analysis. Evidence on effectiveness of DOACs in clinical practice and in the long term is limited; thus, to address the third Objective of this thesis, this Chapter aims at providing further evidence to inform clinical practice in Scotland. Results estimated in this Chapter will be used to populate the cost-effectiveness model in Chapter 7.

6.2 Methods – data

Prior to executing the comparative-effectiveness analysis, three main preparatory steps were carried out. Firstly, the clinical outcomes were identified from SMR01 according to ICD-10 and OPCS-4 codes informed by the literature [37, 81, 112-114]. Then CHA_2DS_2-VASc and HAS-BLED scores were calculated. As discussed, these are bespoke tools used to estimate the probability of having a stroke and bleeding in patients affected by AF according to pre-existing conditions and patient characteristics. Finally, as described in Chapter 5, PSs were calculated according to a series of patients' baseline characteristics.

These steps are described in more detail in the following sections. The comparative-effectiveness analysis was carried out on the same study population described in the previous Chapter (Figure 5.2).

6.2.1 Clinical outcomes

In order to define the full effectiveness and safety profile of DOACs, several additional clinical outcomes were included in the analysis. In particular, the effectiveness of DOACs was assessed with respect to stroke and various composites (stroke or SE, stroke or SE or TIA, stroke or SE or all-cause mortality) as shown in Table 6.1. The clinical outcome of stroke was defined by the inclusion of both ischaemic and haemorrhagic stroke. Other relevant clinical outcomes were MI, ICH and mortality due to all-cause, stroke or cardiovascular events. The safety profile was defined by assessing the risk of GI and major bleeding (including haemorrhagic stroke, GI and other major bleeding events). ICD-10 and OPCS-4 codes from SMR01 used to identify clinical outcomes, obtained from a recent study assessing the effectiveness of DOACs using the same patient level data utilised in this thesis, are presented in Table 6.1 [146].

Table 6.1 Clinical outcomes - ICD-10 and OPCS-4 codes

Outcome	Diagnostic, procedure and drug codes
Effectiveness	
Stroke-all	ICD-10 for Ischaemic stroke: I63, I64, G46.3-G46.7 OPCS-4 for Ischaemic stroke: U54.3 ICD-10 for Haemorrhagic stroke: I60-61
Stroke or SE	Same as Stroke-all plus ICD-10 for SE: I74
Stroke or SE or TIA	Same as Stroke or SE plus ICD-10 for TIA: G45.8, G45.9
Stroke or SE or mortality (all-cause)	Same as Stroke or SE plus death identified form NRS
MI	ICD-10: I21, I22 OPCS-4: K50.2, K50.3
Mortality (all-cause)	Death identified from NRS
Mortality (stroke)	death identified form NRS plus ICD-10 and OPCS-4 codes used for Stroke-all
Mortality (cardiovascular)	death identified form NRS plus ICD-10: I11, I13, I20-I26, I46, I47, I49, I50, I60, I61, I63, I64, I67, I73, I74
Safety	
ICH	ICD-10: I61
GI bleeding	ICD-10: K25.0, K25.2, K25.4, K25.6, K26.0, K26.2, K26.4, K26.6, K27.0, K27.2, K27.4, K27.6, K28.0, K28.2, K28.4, K28.6, K29.0, K62.5, K92.0-92.2
Major bleeding	Including Haemorrhagic stroke, GI bleeding plus ICD-10 codes for Other major bleeds: D62, H11.3, H35.6, H43.1, I62, J94.2, N02, R04, R31, R58, N95.0

6.2.3 PS model

As discussed in Chapter 5, after establishing adequacy of the PS model specification by graphical inspection and by means of standardized difference, the PS model using the PS as a covariate resulted the most robust method compared to PS matching and various IPW approaches. The model using PS as a covariate was then selected as the model for the comparative-effectiveness analysis carried out in this Chapter. As described in Chapter 5 and shown in Equation 5.5, PS were estimated using a logistic regression model according to the following baseline characteristics: age at the time of the first prescription, sex, SIMD, prescription predisposing bleeding, comorbidity CHA_2DS_2-VASc and HAS-BLED scores, previous history of ischaemic stroke or SE or TIA, vascular disease, hypertension, diabetes, and cancer.

6.3 Methods – model

Prior to conducting the comparative-effectiveness analysis, in addition to identifying clinical outcomes and calculating scores, an assessment of the proportionality assumption for the Cox proportional hazards model was carried out; proportionality is the main assumption that has to be satisfied for a semi-parametric or a parametric model to be valid [147, 148]. In general, the Cox proportional hazards model is a semi-parametric model indicating no association between estimated parameters and time, and assuming proportional hazards. The general hazard model can be written as follows:

$$h(t, X) = h_0(t) * \exp(\beta X)$$

(Equation 6.1)

Where $h_0(t)$ is the baseline hazard function when all explanatory variables X are set to zero, and (βX) is the linear predictor. The hazard rate (h) at time (t) is then obtained by multiplying the baseline hazard function by the exponential form of the linear predictor [149].

While a semi-parametric model only provides estimates at a fixed time point, a parametric model is generally used to extrapolate the survival curves beyond the observed follow-up times in order to match a specific time horizon. Thus, parametric models are particularly useful in economic evaluation [148, 149]. Parametric models that can be parametrised as a proportional hazards model follow distributions such as Exponential, Weibull and Gompertz, which are monotonic with time; thus, the distributional assumption implies that the hazard has a specific shape [150, 151].

In particular, in the Exponential model, the simplest parametric model, the hazard function is constant over time and can be written as follows [150]:

$$h(t) = \lambda \; for \; 0 \leq t < \infty$$

Equation (6.2)

Where λ is a positive constant and *t* is time.

In the Weibull distribution depending on the shape and scale parameter, the hazard rate is assumed to increase ($\gamma > 1$) or decrease ($\gamma < 1$) monotonically over time, thus offering more flexibility than the exponential model. This is also a characteristic of the Gompertz distribution. However, while the Weibull distribution is linear with respect to the log of time, the Gompertz distribution has a log-hazard function, which is linear with respect to time [150, 151]. Approaches for assessing proportionality and addressing proportionality violation, whenever present, will be elucidated in the next sections. In addition, details on the sensitivity analysis, carried out to assess the risk of an event associated with apixaban and rivaroxaban, compared to warfarin, when administered at their reduced doses, will be discussed.

6.3.1Proportionality assumption assessment

Following the NICE Decision Support Unit (DSU) guidelines on how to select the most suitable survival model for economic evaluation, log-cumulative hazard plots were produced for assessing the type of hazard in the observed patient-level data, and establish whether proportional hazards can be assumed [150]. In other terms, the ratio between hazards for patients in different treatment groups must be constant over time [148].Violation of the proportional hazard assumption may lead to overestimating the effect of predictor variables associated with increased hazard ratios over time.

Similarly, the effect of covariates linked to decreased hazard ratios over time may be underestimated [148]. The log-cumulative hazard plot was obtained for all clinical outcomes by plotting the log of the negative log of the estimated survival density function versus the log of time. For the proportionality assumption to be satisfied, the two lines, one for each treatment, should be fairly parallel and not crossing [148].

6.3.2 Methods for addressing proportionality violation

Non-parallel or crossing log-cumulative hazard plots may suggest a lack of proportionality and different methods may be used to address proportionality violation. With the counting process approach, time is split into time intervals to identify the time-periods in which proportionality is met. While this approach to some extent could be a valid alternative to a conventional semi-parametric model, such as Cox proportional hazard model, a sufficient sample size for all time-periods is required [148]. Another option for dealing with proportionality violation is stratification, where the whole sample is stratified into subgroups according to any categorical variable interacting with time. Hence, a good stratification variable is often age. With this approach, the baseline hazard function would differ between specified subgroups [148].

Accelerated failure time models are other potential alternatives for the analysis of time-to-event data whenever proportionality is not met. With this approach, the predictors act multiplicatively on the time rather than on the hazard scale, therefore providing a measure of event time ratio rather than hazard ratio. Further, proportionality violation could be addressed by including interactions between time and predictive variables that are time dependent.

If the interactions are not statistically significant, proportionality cannot be rejected, and it is assumed that there is no violation of the proportionality assumption. Testing time dependent variables is comparable to testing for scaled Schoenfeld residuals on functions of time. In particular, Schoenfeld residuals corresponds to the difference between the observed and expected value for the covariate of interest. In this test, non-proportionality is assumed if the slope of the scaled residuals is equal to zero, and the plotted residuals against time show a uniform pattern [147, 148].

According to the DSU guidelines, when proportionality is not met, other approaches may be considered. In particular, if plots do not produce parallel lines, a piecewise model may be used. In piecewise parametric models, generally used when variable hazards are observed over time, different models are fitted to different time-periods. As exponential models are typically used with piecewise parametric models, the different time periods will have constant hazard rates [150]. Other parametric models, such as Weibull and Gompertz, may also be considered, but the hazard will be assumed to be varying over time. Parametric models are typically used for extrapolating the hazard rates beyond the observed data; however piecewise parametric models may not be particularly informative when applied to the extrapolated portion of the survival curve as hazards are not observed [148, 150].

6.3.3 Time-to-event analysis

After establishing from the proportionality assumption assessment that the proportionality was violated across almost the entirety of the clinical outcomes, data were split into 2-year intervals to identify the point where proportionality assumptions were met. However, with time intervals, most of the events have not been estimated for the apixaban group, as most of the events where happening within the first 2 years since anticoagulation initiation. For dabigatran, the number of events was less than 10 for most of the outcomes. Whenever the number of clinical events was less than 5, due to disclosure restrictions, the sign < 5 was used instead (Appendix VI, Table VI-1). As in most time intervals events were not experienced, the HRs for several outcomes across the different treatment comparisons were not estimated or were associated with a broad confidence interval indicating small sample size and a low number of observed events (Appendix VI, Table VI-2).

While stratification could have been a valid alternative to address proportionality violation, it was not a practical option as dividing the cohort into subsamples would reduce the sample size further, making the estimation of hazards for several clinical outcomes problematic as observed when splitting the time into intervals. Missing proportionality could also be addressed by ATF models that provide a measure of event time ratio rather than hazard ratios. This approach however does not express differences in risks as the Cox model does [148, 150]; thus making it unpractical to compare the output from this analysis with those reported in other comparative-effectiveness studies of DOACs.

Hence, in the final Cox model, proportionality violation was addressed by introducing interaction terms between time and the variables violating the proportionality assumption, thus creating time varying covariates. With this approach, if the interactions are not statistically significant, proportionality cannot be rejected and it is assumed that there is no violation of the proportionality assumption; therefore, only statistically significant interactions were included in the final model [152].

As proportionality was addressed by introducing time varying covariates, the semi-parametric Cox proportional hazards regression model was then used to compare risks of the clinical outcomes described in the previous paragraphs across the different treatments. While PS methodologies account for differences in baseline characteristics, as described in Chapter 5, there might still be residual bias after applying the scores; thus, the final Cox proportional hazards model was adjusted for age, sex and comorbidities, resulting in the following equation:

$$h_i(t) = h_0(t) * \exp\left(\beta_1 I_i + \beta_2 P_i + \beta_3 A_i + \beta_3 G_i + \beta_4 \sum_{c=2}^{3} C_{it}\right)$$

(Equation 6.3)

Where I is the intervention (reference category: control), P is the propensity score, A is age at the time of first event (reference category 50-54 age group); G is sex (reference category: male); C is the Charlson comorbidity index (reference category: no comorbidities).

Note: to relax the proportionality assumption, time interactions were included for the variables violating the proportional hazard assumption.

The cohort for the time-to-event analysis, as described in Chapter 5, represents a typical cohort of anticoagulant users, who were 50 years or older at the time of the first OAC prescription. The final cohort resulted in 33,965 first time OAC users, of which 26,387 were on warfarin, 3,706 on apixaban, 484 on dabigatran and 3,388 on rivaroxaban.

Following a continuous treatment approach, the risks of AF related clinical events or death with combined dose DOACs (including standard and reduced doses) compared to warfarin were estimated from anticoagulation initiation to the time of first clinical event or death; patients, followed for 2 and 6 year, were censored if they switched or discontinued treatment. In particular, it was decided to follow patients for 2 years to match the follow-up reported in the pivotal RCTs [41-43]. The choice of following patients for 6 years, although the time intervals indicated that patients on DOACs experienced most of the clinical events within the first 2 years from anticoagulation initiation, was instead based on the PIS data availability covering the time period 2010-2015.

The rationale for censoring patients discontinuing treatment, was based on the approach taken in the pivotal RCTs [41-43] and on the evidence that discontinuation happens in real -world scenarios. To reflect what happens in clinical practice, unlike in RCTs, patients were also censored if they switched between anticoagulation treatments. For instance, a patient would be censored if started anticoagulation treatment with warfarin, experienced a stroke and then switched to apixaban. Thus, censoring switchers allowed to establish to what treatment an occurring event was attributable to.

Because patients can be at risk of more than one mutually exclusive event, time-to-event for each clinical outcome for each treatment comparison was determined within a competing risk framework, where patients can experience only one first event of the several competing first events [153].

In other words, the risk for the first event to occur is estimated in the presence of competing events; for instance, the risk of experiencing the first stroke is estimated in presence of other possible outcomes such as CHI, MI or death.

Event rates for each treatment arm were reported for 100 person-years. In detail, the number of events divided by the total exposure time (person-time), from initiation of anticoagulation to the clinical outcome of interest, is multiplied by 100 to obtain the final 100 person-years event rate.

In addition, the crude incidence of events over time, PS unadjusted and not accounting for the presence of competing risk, is presented in the form of Kaplan Meier curves. Incidence of events, PS adjusted and accounting for competing events, is depicted as cumulative incidence curves. Stata version 14 was used for the statistical analysis.

6.3.4 Sensitivity analysis

Existing observational studies have reported differences in the risk of stroke and mortality when DOACs were administered at different doses [60-62]. Therefore, sensitivity analyses discerning apixaban and rivaroxaban doses were carried out to identify any risk reduction of the main clinical outcomes. While it was possible to conduct time-to-event analyses on standard and reduced dose for the apixaban and rivaroxaban cohorts, it was not possible to perform the same analysis for the dabigatran treatment group due to sample size.

In particular, the objective of the sensitivity analysis was to establish whether apixaban and rivaroxaban are accountable for increased risk of stroke and mortality when administered as a reduced dose**.** The PS model used for the sensitivity analysis was the same that was employed for the main analysis in Chapter 5, where the model is described in detail.

Tests for collinearity in the analyses for the standard and reduced dose apixaban and rivaroxaban were also performed in Chapter 5 and the variance inflation factor did not show any evidence of collinearity (Appendix IV, Table IV-1). Thus, the PS model was again specified according to the following baseline characteristics: age at the time of the first prescription, sex, SIMD, prescription predisposing bleeding, comorbidity CHA2DS2-VASc and HAS-BLED scores, previous history of ischaemic stroke or SE or TIA, vascular disease, hypertension, diabetes, and cancer. The assessment of the proportionality assumption was carried out for standard and reduced dose apixaban and rivaroxaban. Kaplan Meier curves and cumulative incidence curves, as for the main analysis, were also be presented.

6.4 Results

As observed when assessing the proportionality assumptions for the Cox model (Appendix VI, Table VI-1), patients on DOACs, experienced most of the clinical outcomes within the first 2 years from anticoagulation initiation. Further, patients on DOACs were exposed to the treatment for a much shorter period of time, compared to those on warfarin. Thus, the findings for the time-period beyond the first 2 years from anticoagulation initiation did not provide a complete picture, but were included in order to utilise all observational data available.

6.4.1 Cohort characteristics

The comparative-effectiveness analysis in this Chapter, was carried out on the same cohort of patients described in Chapter 5. However, the sensitivity analysis was carried out on AF patients that were either on standard (apixaban 5m, rivaroxaban 20 mg), or reduced dose (apixaban 2.5mg, rivaroxaban 15 mg) DOACs.

Due to the relatively small sample size of dabigatran, it was not possible to discern standard and reduces dose regimen. Thus, in this section, baseline characteristics of patients that were on either standard or reduced dose apixaban and rivaroxaban are presented.

Patients on apixaban and rivaroxaban reduced dose were on average older at the time of the first prescription than those on standard dose. Patients on reduced dose were also at a higher risk of stroke, measured using the CHA_2DS_2-VASc score, compared to those on standard dose. This is clearly visible in the histogram (Figure VII-1) presented in Appendix VII, indicating that for scores >=4, patients on reduced dose represents the majority.

The risk of bleeding, measured with the HAS-BLED score, was also higher for patients on reduced dose apixaban and rivaroxaban, compared to those on standard dose. Again, this is clearly visible in the histogram (Figure VII-2) presented in Appendix VII, indicating that for scores >=3, patients on reduced dose represents the majority. Further, the proportion of patients with more than one comorbidity (Appendix VII, Figure VII-3), and the proportion of patients with a history of AF related conditions was higher in reduced dose groups than in patients on standard dose regimen. Patients' baseline characteristics are reported in Table 6.2.

Table 6.2 Baseline characteristics (standard and reduced dose)

Characteristics	Apixaban 5mg N (%)	Apixaban 2.5mg N (%)	Rivaroxaban 20 mg N (%)	Rivaroxaban 15mg N (%)
Cohort	2,784	922	2,735	518
Sex				
Men	1,672 (60.06)	311 (33.73)	1,533 (56.05)	210 (40.54)
Women	1,113 (39.98)	610 (66.16)	1,202 (43.95)	308 (59.46)
Mean age (SD)	72 (9.35)	84 (7.17)	73 (8.14)	83 (8.16)
SIMD (Scottish index of multiple deprivation)				
1	635 (22.81)	210 (22.78)	412 (15.06)	97 (18.73)
2	544 (19.54)	193 (20.93)	467 (17.07)	116 (22.39)
3	471 (16.92)	144 (15.62)	608 (22.23)	111 (21.43)
4	494 (17.74)	149 (16.16)	678 (24.79)	114 (22.01)
5	641 (23.02)	226 (24.51)	570 (20.84)	80 (15.44)
CHA_2DS_2-VASc score				
0-1	798 (28.66)	28 (3.04)	648 (23.69)	20 (3.86)
2-3	1,001 (35.96)	178 (19.31)	932 (34.08)	106 (20.46)
>=4	986 (35.42)	715 (77.55)	1,155 (42.23)	392 (75.68)
HAS-BLED				
0-2	1,873 (67.28)	465 (50.43)	1,839 (67.24)	246 (47.49)
>=3	912 (32.76)	456 (49.46)	896 (32.76)	272 (52.51)
Comorbidity				
no comorbidity	1,477 (53.05)	337 (36.55)	1,458 (53.31)	200 (38.61)
1 comorbidity	655 (23.53)	228 (24.73)	624 (22.82)	105 (20.27)
>1 comorbidity	652 (23.42)	357 (38.72)	653 (23.88)	213 (41.12)
Stroke or TIA	368 (13.22)	175 (18.98)	394 (14.41)	97 (18.73)
Vascular disease	446 (16.02)	187 (20.28)	460 (16.82)	123 (23.75)
Hypertension	937 (33.66)	370 (40.13)	959 (35.06)	228 (44.02)
Diabetes mellitus	410 (14.73)	118 (12.80)	367 (13.42)	93 (17.95)
Cancer	225 (8.08)	96 (10.41)	227 (8.30)	49 (9.46)
Drug causing bleeding	1,706 (61.28)	561 (60.85)	1,566 (57.26)	323 (62.36)

6.4.2 Comparative-effectiveness

The log-cumulative hazard plots for each comparison are presented and discussed in detail in Appendix VIII, Figure VIII-1–VIII-9; while the test of significance for the time-varying covariates carried out for each variable included in each Cox model, is presented in Appendix VIII, Table VIII-1. In addition, the crude estimates depicted in the Kaplan Meier curves for each comparison are presented and discussed in detail in Appendix IX, Figure IX-1–IX-9. The estimates accounting for PS and competing risk adjustments are shown graphically and discussed in detail in the cumulative incidence curves in Appendix X, Figure X-1–X-9. Although, not particularly informative, HRs for all clinical events at 6-year follow-up are presented in Appendix XI Figure XI-1 for completeness.

Apixaban versus warfarin

Patients on apixaban had higher absolute event rates than warfarin across all clinical outcomes (Table 6.3). Hazard ratios (HRs) for all clinical events are presented in Figure 6.1. No difference in risk was observed for any of the clinical outcomes when comparing efficacy and safety profiles of apixaban versus warfarin.

Dabigatran versus warfarin

Absolute clinical event rates for patients on dabigatran were generally lower than the rates observed in the warfarin group (Table 6.3). The trend was inverted for GI bleeding rates, which were higher in the dabigatran than the warfarin group. The same trend was observed for major bleeding and mortality due to stroke; however, the difference in event rates between treatments is negligible. No difference in risk was observed for most of the clinical outcomes when comparing efficacy and safety profiles of dabigatran versus warfarin (Figure 6.1). However, the risk of GI bleeding was found to be greater in the dabigatran group than in the warfarin group.

Rivaroxaban versus warfarin

Patients on rivaroxaban had higher event rates than warfarin users across all clinical outcomes (Table 6.3). There were no differences in risk of stroke and stroke composites between rivaroxaban and warfarin patients (Figure 6.1). However, when all-cause mortality was included in the stroke composite, an increased risk for the rivaroxaban treatment group was observed. When evaluated separately, patients on rivaroxaban showed an increased risk of all-cause mortality compared to those on warfarin. Similarly, the risk of mortality due to cardiovascular conditions was greater in the rivaroxaban group. When assessing the safety profile, increased risks of GI bleeding and major bleeding were observed for patients on rivaroxaban compared to the warfarin treatment group.

Table 6.3 Number of events and event rates for DOACs and warfarin

Outcome	Apixaban events and (event rates)	Dabigatran events and (event rates)	Rivaroxaban events and (event rates)	Warfarin events and (event rates)
Stroke all	54 (2.16)	<5 (<0.80)	66 (2.18)	537 (1.35)
Stroke or SE	57 (2.28)	<5 (<0.80)	68 (2.25)	575 (1.45)
Stroke or SE or TIA	71 (2.84)	5 (0.80)	80 (2.64)	718 (1.81)
Stroke or SE or mortality (all cause)	215 (8.61)	24 (3.84)	325 (10.73)	2122 (5.34)
MI	40 (1.60)	<5 (<0.80)	27 (0.89)	312 (0.79)
Major bleeding	116 (4.64)	22 (3.52)	191 (6.31)	1298 (3.27)
ICH	14 (0.56)	not estimated	16 (0.53)	83 (0.21)
GI bleeding	40 (1.60)	13 (2.08)	65 (2.15)	428 (1.08)
Mortality (all cause)	159 (6.36)	20 (3.20)	258 (8.52)	1558 (3.92)
Mortality (stroke)	9 (0.36)	<5 (<0.80)	16 (0.53)	87 (0.22)
Mortality (cardiovascular)	89 (3.56)	12 (1.92)	124 (4.09)	827 (2.08)

Note: event rates are estimated per 100 person-years. Due to disclosure restrictions, in the case of fewer than five events, "<5" was reported.
Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack, MI=myocardial infarction, ICH=intracranial haemorrhage, GI=gastrointestinal.

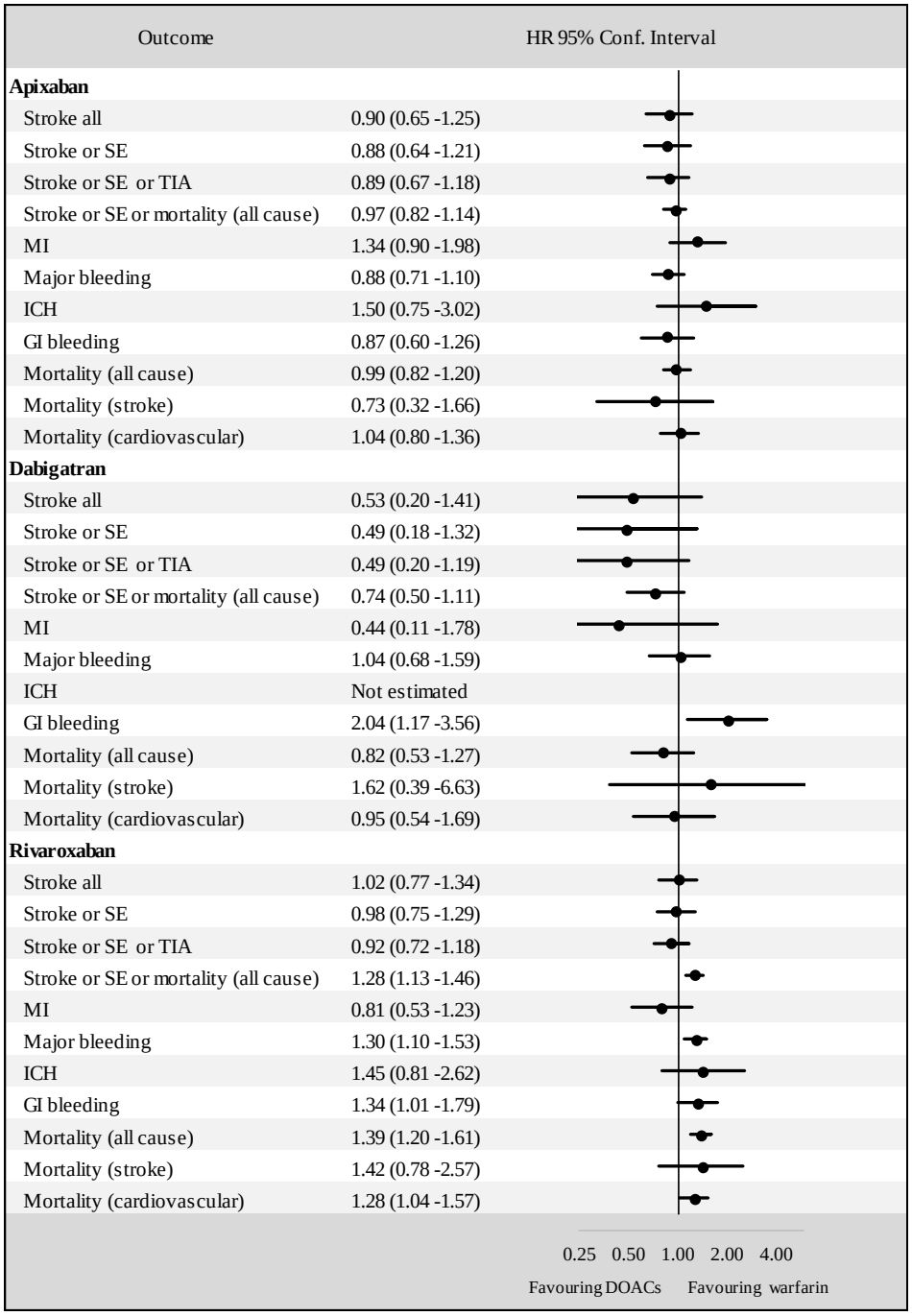


Figure 6.1 Hazard ratios since first prescription, DOACs vs. warfarin

Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack, MI=myocardial infarction, ICH=intracranial haemorrhage, GI=gastrointestinal.

6.4.3 Sensitivity analysis

The log-cumulative hazard plots for each comparison are presented and discussed in detail in Appendix XII, Figure XII-1–XII-9; while the test of significance for the time-varying covariates carried out for each of variable included in each Cox model, is presented in Appendix XII, Table XII-1. In addition, the crude estimates depicted in the Kaplan Meier curves for each comparison are presented and discussed in detail in Appendix XIII, Figure XIII-1–XIII-9. The estimates accounting for PS and competing risk adjustments are shown graphically and discussed in detail in the cumulative incidence curves in Appendix X, Figure X-1–X-9. As in the main analysis, for completeness, HRs for all clinical events at 6-year follow-up are presented in Appendix XV Figure XV-1 (apixaban) and Figure XV-2 (rivaroxaban).

Apixaban standard and reduced dose

Comparing apixaban against warfarin, the difference in event rates for any of the outcomes was considerably higher for patients on the reduced dose apixaban than it was for those on the standard dose (Table 6.4). Absolute event rates for both apixaban doses were higher than the rates in the warfarin group. No risk differences for any of the clinical outcomes were observed between apixaban and warfarin regardless of the dose administered (Figure 6.2).

Table 6.4 Number of events and event rates for 2 years since first prescription, apixaban standard (5 mg) and reduced dose (2.5 mg)

Outcome	Apixaban 5 mg event and (event rates)	Apixaban 2.5 mg event and (event rates)
Stroke all	35 (1.82)	19 (3.28)
Stroke or SE	37 (1.93)	20 (3.46)
Stroke or SE or TIA	45 (2.34)	26 (4.49)
Stroke or SE or mortality (all cause)	127 (6.61)	88 (15.21)
MI	24 (1.25)	16 (2.77)
Major bleeding	87 (4.53)	29 (5.01)
ICH	9 (0.47)	5 (0.86)
GI bleeding	29 (1.51)	11 (1.90)
Mortality (all cause)	91 (4.74)	68 (11.75)
Mortality (stroke)	5 (0.26)	4 (0.69)
Mortality (cardiovascular)	49 (2.55)	40 (6.91)

Note: event rates are estimated per 100 person-years.
Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack, MI=myocardial infarction, ICH=intracranial haemorrhage, GI=gastrointestinal.

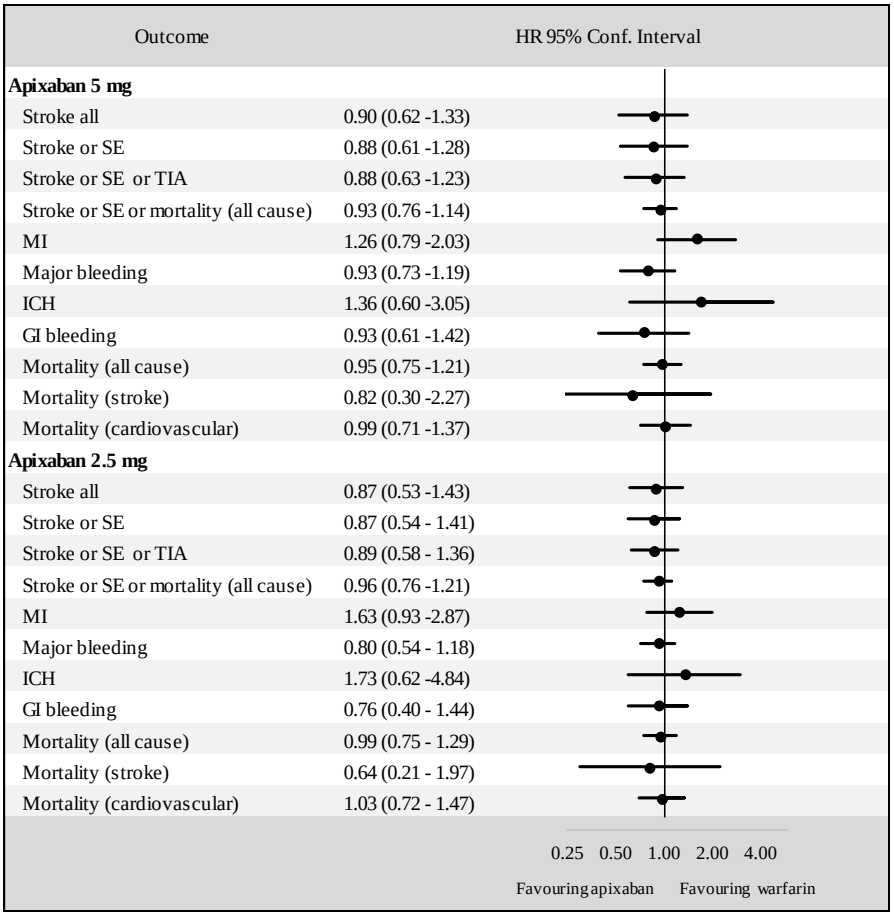


Figure 6.2 Hazard ratios since first prescription, apixaban standard (5 mg) and reduced dose (2.5 mg) vs. warfarin

Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack, MI=myocardial infarction, ICH=intracranial haemorrhage, GI=gastrointestinal.

Rivaroxaban standard and reduced dose

For patients on reduced dose rivaroxaban, the all-cause mortality rate was higher than with warfarin. (Table 6.5). Increased risk of major bleeding, all-cause mortality and mortality due to cardiovascular conditions were observed for patients on rivaroxaban standard 20 mg dose compared to those on warfarin (Figure 6.3). An increased risk of GI bleeding and any major bleeding was observed for patients on the reduced 15 mg dose compared to those on warfarin.

Table 6.5 Number of events and event rates for 2 years since first prescription, rivaroxaban standard (20 mg) and reduced dose (15 mg)

Outcome	Rivaroxaban 20 mg event and (event rates)	Rivaroxaban 15 mg event and (event rates)
Stroke all	52 (2.17)	10 (2.30)
Stroke or SE	54 (2.26)	10 (2.30)
Stroke or SE or TIA	63 (2.63)	12 (2.76)
Stroke or SE or mortality (all cause)	245 (10.23)	69 (15.86)
MI	16 (0.67)	10 (2.30)
Major bleeding	145 (6.06)	41 (9.42)
ICH	13 (0.54)	2 (0.46)
GI bleeding	43 (1.80)	20 (4.60)
Mortality (all cause)	192 (8.02)	59 (13.56)
Mortality (stroke)	12 (0.50)	3 (0.69)
Mortality (cardiovascular)	87 (3.63)	32 (7.36)

Note: event rates are estimated per 100 person-years.
Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack, MI=myocardial infarction, ICH=intracranial haemorrhage, GI=gastrointestinal.

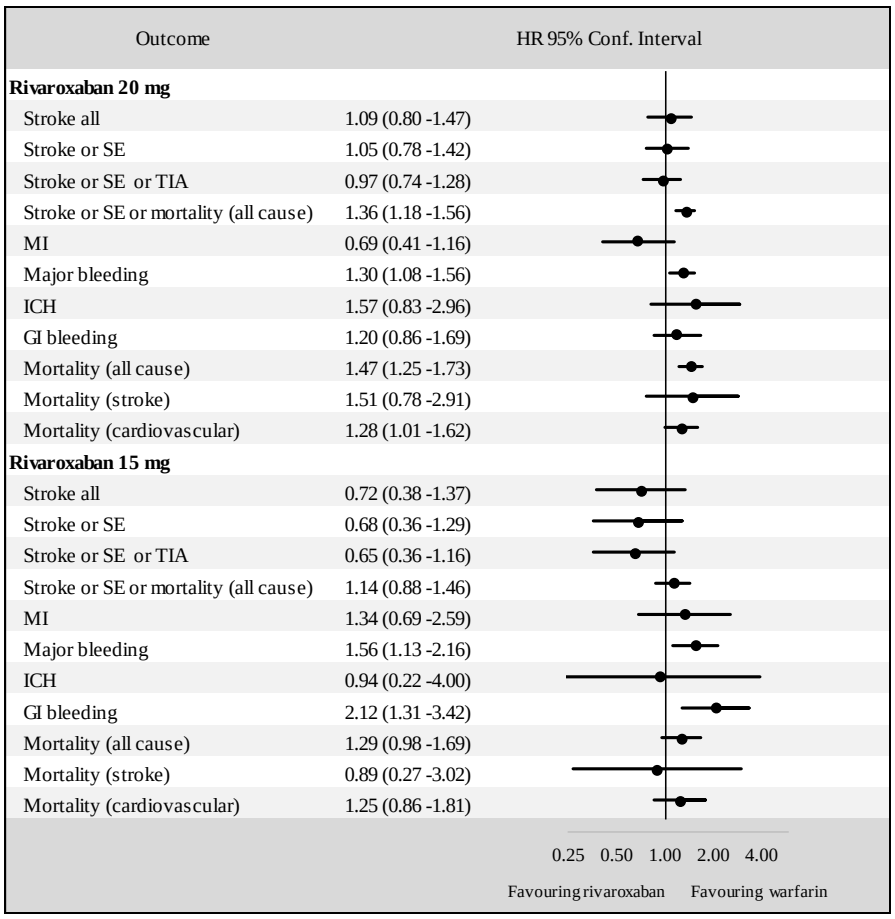


Figure 6.3 Hazard ratios since first prescription, rivaroxaban standard (20 mg) and reduced dose (15 mg) vs. warfarin

Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack, MI=myocardial infarction, ICH=intracranial haemorrhage, GI=gastrointestinal

6.5 Discussion

Evidence from clinical practice offers the opportunity to assess how well treatment works in real life under uncontrolled conditions. This type of evidence typically gathered from observational studies, poses some important challenges such as confounding and selection bias, which, as discussed in Chapter 5, can be mitigated but not eliminated. Hence, additional evidence from real-world scenarios can be used as a supplement rather than a substitute to the robust evidence coming from RCTs, where treatment efficacy is measured under specific conditions.

In this section, the evidence from Scottish clinical practice is discussed and compared against the evidence from RCTs and other observational studies discussed in Chapter 1. However, while it is essential to acknowledge the importance of the evidence generated from indirect comparison by means of NMA as discussed in Chapter 1, such evidence cannot be compared with the findings from this analysis not covering DOAC to DOAC comparison.

The descriptive statistics of patients on either standard or reduced dose DOACs, confirms, as discussed in Chapter 5, that older and sicker patients were more likely to be on reduced dose apixaban or rivaroxaban. A striking finding from this comparative-effectiveness analysis is the increased risk of mortality associated with rivaroxaban. The increased risk of mortality may be linked to selective prescribing, a mechanism determined by the possibility of DOACs being prescribed according to the patients' clinical profile [154, 155], and acknowledged in two observational studies [62, 63]. For instance, in Larsen's study (2016), dabigatran was selectively prescribed to younger patients with lower risk of stroke and less renal impairment [63]. Similarly, Nielsen (2017) reported that patients who were sicker and older were more likely to be on rivaroxaban than on warfarin [61]. Presumably, sicker and older patients may not be able to cope with the burden associated with warfarin and the constant INR monitoring. Thus, while patients on warfarin may discontinue or switch treatment prior to death, those on DOACs may be dying while on treatment, but of causes that are age related and not linked to AF [62]. While the findings may rise some concerns over the use of rivaroxaban, Larsen (2016) and Nielsen (2017) acknowledged that despite using PS methods, allowing for addressing confounding by indication, some unmeasured residual confounding and selective prescribing behaviour would persist. Therefore, further research will be needed to access whether the association between the use of rivaroxaban and the risk of mortality is a true association [61, 63].

As described by Vinogradova (2018), selective prescribing may also have an effect on the increased risk of bleeding in patients on warfarin compared to those on DOACs. This translates into bleeds being more likely to be detected in warfarin patients who are regularly monitored for their INR [62]. It is possible that, in the study dataset, patients on DOACs will be sicker than patients treated with warfarin; this is something that cannot be captured fully by the CCI. Similarities and differences with other studies are discussed in the next sections and presented in Appendix XVI (Figure XVI-1, XV-2 for standard dose and XV-3 for reduced dose).

Similarity and differences with other studies

Apixaban

The findings regarding GI bleeding in the ARISTOTLE clinical study [42] were confirmed in this comparative-effectiveness analysis where no risk difference was found between apixaban and warfarin patients at 2 years from anticoagulation initiation. The finding from this analysis, however, differed from those reported in one observational study (Ntaios (2017)) where a risk reduction of GI bleeding was found for patients on apixaban [58] .Further, the risk difference for stroke between apixaban and warfarin, although reduced, was not statistically significant. In the ARISTOLTE trial, a significant risk reduction for stroke in apixaban patients compared to those on warfarin was observed. However, when stroke was assessed according to specific subtypes, no risk difference was reported for ischaemic or unspecified stroke [42]. In contrast, Nielsen (2017) distinguished between doses [61] and found no risk differences for stroke between patients on apixaban 2.5 mg reduced dose and those on warfarin. Similar findings were reported by Vinogradova (2018) where the effectiveness of apixaban compared to warfarin was assessed at 5 mg standard and 2.5 mg reduced dose [62].

Hence, in analysis carried out in this Thesis, differences in the risk for stroke between patients on any apixaban dose and patients treated with warfarin seem to be much closer to those observed in clinical practice than those reported in the ARISTOTLE study.

Dabigatran

A full comparison of results from this comparative-effectiveness analysis with findings from the RE-LY trial was not possible, since the ability to assign patients to different dose regimen was limited due to sample size [41]. In the RE-LY study, a minimum of 18,000 patients were enrolled to obtain an 84% power ensuring that the upper bound of the 97.5% confidence interval of the risk of the primary outcome would fall below 1.46. The findings for the risk of stroke are in line with those observed in the RE-LY trial for patients receiving dabigatran at a reduced dose. Similarly, findings from RE-LY suggesting an increased risk of GI bleeding in patients treated with dabigatran 150 mg compared to those on warfarin were confirmed in this analysis and in a meta-analysis of observational studies [58].

Rivaroxaban

When comparing the risk of stroke and safety profiles (risk of bleeding) between patients on rivaroxaban and those on warfarin, the findings are comparable to those reported in the ROCKET-AF trial [43]. In the ROCKET-AF trial, the risk ratio of major bleeding was not estimated as a single clinical outcome, but event rates were reported for each body site. Major bleeding was associated with an event rate of 3.6% per 100 patient years for patients on rivaroxaban, whereas patients on warfarin had an event rate of 3.4% [43]. In the clinical study, the biggest contributor to major bleeding was GI bleeding with an event rate of 3.2% for rivaroxaban patients compared to an event rate of 2.2% for patients on warfarin [43].

The findings on the risk of stroke between patients on rivaroxaban and those on warfarin were also comparable to those reported in an observational study by Gorst-Rasmussen (2016), where no risk difference was observed between treatment group doses [60].

The sensitivity analysis showed an increase of GI bleeding in patients treated with reduced dose rivaroxaban, but no differences were observed for standard dose compared to warfarin. The findings for the standard dose are in line with those reported by Vinogradova (2018) [62], while those for the reduced dose may be due to selection bias resulting from people with greater risk of bleeding being more likely to be given rivaroxaban reduced dose. This is backed up by the baseline characteristics for DOACs standard and reduced dose previously presented in Table 6.2, and the histogram presented in Appendix VII (Figure VII-2). A different pattern was observed when looking at the risk of all-cause mortality, where it was observed an increased risk of mortality for patients on rivaroxaban standard dose but not for those on 15 mg reduced dose. While findings on the risk of mortality associated with reduced dose rivaroxaban are comparable to the risk reported in the ROCKET-AF trial [43], the increased risk of mortality observed for patients receiving the standard dose matches results reported by Nielsen (2017)[61] and Gorst-Rasmussen (2016) [60]. However, in these two observational studies, the increased risk of all-cause mortality was found for patients receiving rivaroxaban 15 mg reduced dose [60, 61].

6.6 Limitations

The comparative-effectiveness analysis carried out provides evidence on how DOACs work in clinical practice and whether this class of drugs offer a clear advantage in preventing stroke and other important pathologies in the AF population. Nevertheless, this analysis was constrained by a number of limitations.

Firstly, the relatively low number of events, as in the case of dabigatran, may undermine the robustness of the findings. The pivotal RCTs have identified the optimal number of events and participants needed to test the hypothesis of DOACs non-inferiority in preventing the primary outcome of stroke or SE. However, with observational data, it is not possible to determine the number of events needed to obtain a robust statistical power.

Although the number of events observed in the dabigatran group is clearly a limitation of this Thesis, the analyses for each treatment group were carried out to get an indication of effect of dabigatran in preventing stroke and other AF related comorbidities in a real-world setting. Nevertheless, as the evidence from this analysis suggest that patients are increasingly switching from warfarin to DOACs, additional observational data may be available in the future.

The potential risk of clinical miscoding of morbidity records is another possible limitation of the study that should be acknowledged. Further, given the nature of observational data and non-randomised evidence, when estimating the ATE, there will always be some residual bias due to observed confounders, even if PS methods are applied. Unobserved confounders, not dealt with in this analysis, such as patients' tolerability or access to healthcare, may generate additional bias.

In addition, the proportionality assumption of the Cox model should be assessed cautiously, and several approaches should be tested to address possible violation of this assumption before selecting the final model for analysis. Ultimately, depending on the scope of the research, the Cox model may still offer the most appropriate method for estimating the ATE, even when the proportionality assumption is violated.

6.7 Conclusions

This analysis provides RWE that apixaban is at least as effective as warfarin for the prevention of stroke and AF associated comorbidities, but also highlights some safety and effectiveness concerns regarding the use of rivaroxaban. The increased risk of bleeding and mortality associated with rivaroxaban is in line with findings from other observational studies. However, the difference in risk of mortality may be caused by true association or bias from selective prescribing. Methods to deal with selective prescribing and reduce confounding by indication exist, as the PS method used in this analysis, but as these issues can be mitigated, and not entirely eliminated, there will always be a risk of residual selection bias even when adequate methods are sufficiently adjusted for selective prescribing. Thus, further research will be needed to confirm the association between the use of rivaroxaban and the risk of mortality. Further, the evidence on dabigatran gives an indication of the effect in preventing stroke and AF associated comorbidities; but given the sample size and the number of events observed, the findings are less robust compared to those obtained for apixaban and rivaroxaban. Nevertheless, the evidence provided in this analysis allows for an assessment of how well treatments work in a non-randomised setting, and the additional evidence provided in this study should be used as a supplement rather than a substitute to the robust evidence coming from RCTs.

In this Chapter, event rates and risks of major clinical events associated with AF have been estimated and compared against those reported in the pivotal RCTs [41-43] and existing observational studies [60-62]; thus allowing for identifying the main differences between the findings generated with a randomised and a non-randomised approach. Most of the estimated event rates and HRs from the analysis in this Chapter will be utilised to populate the cost-effectiveness model needed for the economic evaluation of DOACs, which is carried out and discussed in Chapter 7.

Chapter 7 Economic evaluation

7.1 Introduction

Economic evaluation is the preferred method for the appraisal of treatments or healthcare programmes including the identification, measurement, valuation and comparison of costs and consequences of the alternatives being considered [156]. This Chapter focuses on the economic evaluation of DOACs using RWE of effectiveness. While several comparative-effectiveness studies have assessed the effectiveness of DOACs in clinical practice [58, 60, 61, 63, 157] RWE on the cost-effectiveness is still limited, as in most cost-effectiveness studies, the effectiveness parameters were extracted from the pivotal RCTs [41-43]. To date, no cost-effectiveness study from UK perspective has used findings from RWE studies to show whether DOACs are cost-effective in clinical practice. To meet the fourth objective described in Chapter 2, a review of existing studies assessing the cost-effectiveness of DOACs compared to warfarin in the AF population, reflecting the UK clinical practice and adopting the NHS perspective was carried out. Then the identified model was updated with patient level data to establish whether DOACs are a cost-effective option to warfarin for AF patients in Scottish clinical practice.

7.1.1 Cost utility analysis

A common objective of economic evaluation techniques, within the healthcare framework, is to evaluate the level of health benefits and effects in relation to the level of resource use [151, 156]. Cost Utility Analysis (CUA), is the preferred type of economic evaluation in HTA as it uses Quality Adjusted Life Years (QALYs) as generic outcome measure, which provides a common currency for measuring gains in life expectancy and Health Related Quality of Life (HRQoL) that reflect the effectiveness of interventions [158, 159]. One QALY represents one year in perfect health, while death is associated with zero QALYs.

Individual preferences under uncertainty, obtained from the general population expressed as health state utilities, are combined with the time spent in a given health state to calculate the number of total QALYs accumulated in that given health state. Health utilities are typically measured with generic instruments such as EQ-5D, SF-6D, Health Utilities Index or other instruments that measure condition specific health utilities [158, 160, 161].

Despite the availability of different instruments, NICE recommends the use of EQ-5D for measurement and valuation of HRQoL. Because EQ-5D, extensively researched and validated, is not disease specific, it can be used for most diseases areas, thus allowing for comparisons of interventions across disease areas[162]. The design of EQ-5D allows for capturing a broad spectrum of level of health across five different dimensions: mobility, ability to self-care, ability to undertake usual activities, pain and discomfort, anxiety and depression. Each dimension comprises 5 levels (or 3 in the first EQ-5D version) of perceived problems. For each of the possible unique health states, given by combining dimensions with level, utility scores are obtained through the application of valuation methods such as visual analogue scale, time trade-off or standard gamble [160, 163].
In particular, visual analogue scale records respondents' self-reported health on a single line ranging from 0 to 100; where the lower and upper end represent worst and best possible health state respectively [161].

Time trade-off measures with a scaling method the relative amount of time that patients would be willing to trade-off in order to survive in a given health state. Patients may choose to stay in their ill state or improve their health status with a reduced life expectancy. Standard gamble, typically used in economics to measure preferences, reflect the risk that patients would be willing to take in a two-way outcome scenario [161].

Here, patients are asked to choose between a sure outcome (living the remaining life expectancy in a particular health state) or gamble for a state of optimal health or instant death [161, 164].

In CUA, costs are generally classified as direct, indirect and intangible. While direct costs reflect expenses directly associated with a condition such as prescription drugs, medical equipment and patient transportation; indirect or productivity costs reflect the costs incurred because of productivity losses resulting from morbidity or mortality associated with a given condition. On the other hand, intangible costs are these experienced because of pain or adverse events, and cannot directly be measured in monetary terms [151, 156, 165]. In CUA, cost-effectiveness is expressed as Incremental Cost-Effectiveness Ratio (ICER), obtained by the difference in expected cost divided by the difference in QALYs of two interventions. The calculation for the ICER is presented in the following equation:

$$\text{ICER} = \frac{\Delta C\ (C_{1-}\ C_0)}{\Delta E\ (E_{1-}\ E_0)}$$

(Equation 7.1)

Where C_1 and C_0 are the cost for intervention and control group; and E_1 and E_0 represent the effectiveness measure for intervention and control group.

In the UK, interventions with an ICER below the threshold (an indication of Willingness To Pay (WTP)) of £20,000 per QALY gained are deemed cost-effective [156, 166]. However, an ICER greater than £20,000 may still be accepted for reimbursement if there are concerns that QALYs (generated by EQ-5D) may not adequately reflect the health gain as a result of an intervention for a given disease area. Further, other factors such as ethics and lack of alternative treatments are taken into account for reimbursement purposes. This is often the case with orphan or end of life drugs [162].

The ICER, generally used for presenting results from CUA, could also be expressed as Incremental Net Monetary Benefit (INMB), a summary statistics that provide information on the size or scale of the interventions being evaluated. A positive INMB indicates that an intervention is cost-effective compared to the alternatives for a given threshold.

Specifically, INMB is obtained by multiplying the incremental utility (QALY) by the threshold, and then subtracting the incremental cost [156]. The calculation for the INMB is presented in the following equation:

$$\text{INMB} = (\Delta E * WTP) - \Delta C$$

(Equation 7.2)

7.1.2 Data source

The data needed in economic evaluation are generally extracted from different evidence sources including RCTs, systematic reviews, and existing economic evaluation studies. The duration of the studies however, as in the case of most RCTs, may have a follow-up limited to a few years; therefore, a decision analytic model generating evidence that goes beyond the limited duration of a study may be required. While some of the parameters of the model can be generated, or extrapolated from the study, such as cost of the intervention and long-term survival, more often multiple external sources will be needed to complement the evidence generated within-study. Informing the model parameters using several data sources, besides the evidence generated from within-trial RCTs, is a standard practice in economic evaluation.

However, designing the health economics component from the very early stage of trial design would allow to collect relevant information related to resource use and health related quality of life, taking into account specificities related to the intervention (e.g. impact on intermediate, rather than final outcomes) or to the context (e.g. multinational trials). This would produce reliable estimates of cost-effectiveness that can subsequently be extrapolated for a long-term time horizon and incorporated into an economic model [151, 156].

7.1.3 Models in economic evaluation

Analytical models, reflecting the clinical pathway of any health condition, and depicting different consequences according to the interventions adopted, are at the core of economic evaluation. The structure of a model would normally depend on the clinical pathway. In general, decision trees, Markov models or discrete event simulation (DES) models are used. Decision trees are the simplest model structure, here distinct branches represent clinically meaningful pathways and associated health outcomes. At the point of branches intersection, often referred to as node, a probability indicating the occurrence and non-occurrence of a specific event must be specified. These probabilities are then multiplied by costs and outcomes assigned, to obtain the expected values of outcomes and associated costs for each treatment being evaluated [151, 167].

Markov models employ mutually exclusive health states that can be occupied sequentially at any given time. Markov models are stochastic models, which exploit memory-less properties, where the probability for a cohort of moving from one health state to another does not depend on previous events, but only on the present health state.

In a Markov model, moving from one health state to another depends on the probability of transitioning between states, and cycles are used to indicate discrete time-periods. Typically, in scenarios where clinical events are relatively rare, the cycle length could be one year; similarly, if clinical events occur more frequently, or the rate changes rapidly over time, the cycle length may be shorter. Costs and utilities are accumulated during each cycle according to the time spent in each state or the transition between states [151, 156]. A Markov model may not be appropriate when modelling conditions or clinical events requiring transition probabilities to change over time. However, the issue concerning time dependency of transition probabilities can be handled by having independent transition probabilities for each cycle. For example, when comparing mortality, this can be done with life tables, allowing transition probabilities to change as the cohort ages. As per Office of National Statistics definition, life tables indicate “trends for the UK and constituent countries in the average number of years people will live beyond their current age measured by "period life expectancy", analysed by age and sex” [77].

Time dependency and the memory-less nature of cohort Markov models can also be dealt with using a DES approach. This maintains the concept of health state and discrete cycles but models the system as a series of events occurring over time. The risk of these events depends on individual patient characteristics and changes over time as more events are experienced. This feature of DES is particularly useful in modelling complex pathways where future events are likely to depend on patients’ clinical history. Thus, DES may be particularly useful for conditions were clinical measures such as body mass index and blood pressure are recorded.

Nevertheless, DES implementation and execution is computationally more intensive than the more conventional Markov approach, requiring patient-level data to obtain time-to-event estimates and associated distributions [167]. Further, the computational burden limits the flexibility of a DES model to assess parameter uncertainty. This is a major concern with the Probabilistic Sensitivity Analysis (PSA), a type of sensitivity analysis based on the probability distribution of parameters and used to assess uncertainty around the findings, as it requires two levels of simulation for estimating a single expected value and for sampling from distribution. In other words, if 10,000 simulations are needed for the two level simulation, 100 million simulations would be required at the patient level; thus, compared to a Markov model, DES would require a much larger number of simulations [151, 168].

Methods to reduce the computational burden and simulation time, such as Gaussian process [169] and ANOVA methods [170] have been proposed. However, simulation studies comparing outputs from Markov and DES models have shown that these models produced very similar results; and have highlighted that differences in the results mostly depend on assumptions and available evidence, rather than the simulation method adopted [171, 172]. For instance, Brown's simulation study (2000) compared a cohort-based model versus a patient level model, and found that the former predicted higher rates of mortality and MI but fewer strokes. However, this was mainly determined by model differences in terms of model structure, assumptions and evidence used [172].

7.2 Review of economic models

To date several cost-effectiveness studies, employing a wide range of model structures and assumptions, have been developed to assess the relative value of different treatment strategies for the prevention of stroke in the AF population. This review aims at identifying, in existing studies, common model features that best reflect the UK clinical practice and perspective. The present review, is an extension of the existing Limone's systematic review (2013) assessing studies on cost-effectiveness of DOACs (including apixaban, dabigatran and rivaroxaban) for stroke prevention in the AF population [173]. An additional study, also carried out by Limone and colleagues [174], aided the identification of the best fit for purpose model for this Chapter, by highlighting common flaws in published models evaluating the cost-effectiveness of pharmacological interventions in preventing stroke in the AF population.

The existing systematic review by Limone et.al [174], searched MEDLINE, EMBASE, National Health Service Economic Evaluation Database, Health Technology Assessment and Tufts Cost-Effectiveness Analysis Registry from 2008 to 2012. Then, a quality assessment of models was conducted using the Quality of Health Economics Analyses (QHES) rating scale, a validated measure containing 16 items to evaluate the quality of conduction and reporting of cost-effectiveness analyses. The scores in the rating scale range from 0 to 100, moving from the lowest to the highest possible quality. In Limone's critical appraisal, only models obtaining a QHES score of 80 were considered of high quality [166].

In the existing review, 18 economic evaluation were identified; of these 4 assessed apixaban, 9 assessed dabigatran (150 mg, 110 mg) and 4 assessed rivaroxaban. Warfarin was the main comparator in almost all cost-effectiveness models assessed.

The majority of the models identified in Limone's review (2013) reported share a common core structure taken from the Sorensen and Gage models [175, 176], designed for evaluating the cost-effectiveness of dose adjusted warfarin. In the earliest model designed by Gage and colleagues in 1995, a Markov model was built, using a time horizon of 10 years, to evaluate the expected outcomes of three treatment alternatives (warfarin, aspirin and no therapy) for 65 years or older patients with an history of NVAF. The model included 10 health states for each of the three options (well, moderate-severe stroke, mild stroke, moderate-severe stroke, second stroke, mild ICH, moderate-severe ICH, death), and patients remained in the "well" state until TIA, stroke, haemorrhage or death occurred.

The second stroke health state was included to capture the effect of previous events on future events. Further, important anticoagulation related adverse clinical events such as MI and extra-cranial bleeds were not included in the model. The model also allowed for switching from warfarin to aspirin in case of bleed and from aspirin to warfarin if stroke was experienced [176]. The Sorensen model is an extension of the Gage model where, while sharing a very similar structure, additional health states such as ischemic stroke and SE were included [175]. In addition , almost all studies identified in Limone's review (2013) employed Markov models, only one study used a DES approach [173]. As previously discussed, the DES approach still maintains the concept of health state and discrete cycles as with the Markov approach, but models the system as a series of events occurring over time [151].

Despite common structures shared between models, Limone identified a noteworthy variation in the complexity and the assumptions made. In particular, the complexity was reflected in the number of health states included. Among the apixaban models, all 4 models included the MI state, 3 included minor bleeding, and only 1 model included SE. Similarly, among the dabigatran models, all 13 models included the MI state, 11 included minor bleeding, and only 8 models included SE. In the rivaroxaban models, MI and minor bleeding were included in all 4 models, but only 3 of those also modelled SE [173]. As Limone pointed out, the omission could have occurred because the difference in risk between treatments for the outcomes of interest was not statistically significant. For instance, the risk reduction for MI was not statistically significant in the RE-LY and ROCKET-AF trials; similarly, SE did not reach statistical significance in the ARISTOTLE and RE-LY clinical studies [41-43, 174].

Another common issue identified in most of the cost-effectiveness models critically appraised by Limone, concerns the way INR control and time in therapeutic range was dealt with. While most of the models included adjusted dose warfarin as a treatment alternative, the rate of clinical outcomes such as stroke or bleeding was not always adequately adjusted according to INR control and time in therapeutic range.
This could be an issue as the time in therapeutic range measured in a controlled environment, as in RCTs, may be higher than the one recorded in clinical practice. Thus, the lack of event rates adjustment in the cost-effectiveness models, may lead to conclusions diverging from those drawn when RWE is utilised [173, 174].

7.2.1 Review - methods

To reflect UK clinical practice, only cost-effectiveness studies adopting the UK NHS perspective were included in the review. The existing systematic review from Limone (2013) included cost-effectiveness studies published up to 2012 [173]; hence, a further search was conducted, through MEDLINE and EMBASE databases, to identify any other relevant studies carried out between 2012 and 2018. The syntax for the search strategy is reported in Appendix XVII. The following items for each study were extracted: perspective, interventions, base-case population characteristics, model type and assumptions, time horizon, discount rate, drug discontinuation or switching, results and source of uncertainty.

A clear reference to the AF population and the inclusion of costs-effectiveness evaluation were the two main selection criteria. Models evaluating the cost-effectiveness of DOACs in a population either than AF, e.g. VTE was excluded. While several studies have assessed the cost-effectiveness of DOACs adopting different country specific perspective assumptions, this review focused entirely on the studies evaluating the three main DOACs (apixaban, dabigatran and rivaroxaban) and adopting the UK perspective. A critical appraisal of the methodology and reporting of the identified models, except for government reports, was conducted using the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) checklist for the items to include when reporting economic evaluations of health interventions [177].

7.2.2 Review - results

The literature search, after removing duplicates, initially identified 227 studies (Figure 7.1). Of these, 29 were not reporting an economic study, 24 were review articles, 8 were not comparing DOACs, 3 were not in English language, leaving with 27 studies eligible for further assessment. Of these 27, a total of 5 articles assessing the cost-effectiveness of DOACs in the AF population, and adopting a UK perspective were identified [178-182].

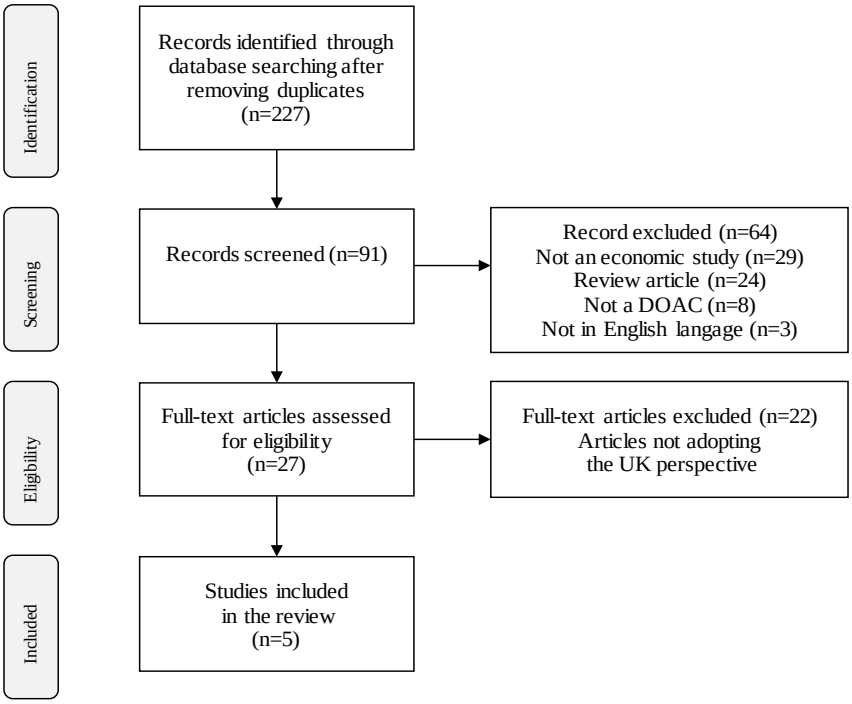


Figure 7.1 Results of literature search

A description for each study identified the literature search and included in the review is reported in Table 7.1.

Population

The base-population characteristics are generally matching those of the RCTs; where, on average patients suitable for anticoagulation therapy and with different levels of risk of stroke initiated anticoagulation treatment at 70 years of age [41-43].

Interventions

The cost-effectiveness of apixaban was assessed in the 5 studies identified where standard dose apixaban was assessed against dabigatran, rivaroxaban and warfarin [178-182]. Of those, 3 studies included aspirin as an additional intervention as an alternative to anticoagulation discontinuation due to a series of clinical events such MI, ICH or major bleeding [178, 179, 181]. The cost-effectiveness of dabigatran was investigated in 4 of 5 studies identified [179-182] where dabigatran was assessed at different doses (dabigatran 110 mg, dabigatran 150 mg and sequential dabigatran) against apixaban, rivaroxaban and warfarin. Aspirin was included as an alternative intervention in 2 of these studies [179, 181]. The cost-effectiveness of rivaroxaban was also investigated in 4 of the studies identified [179-182]; and of these, 2 studies [179, 181] included aspirin as an alternative interventions to anticoagulation treatments.

Model type

All studies identified employed Markov models. An important element that became evident when scrutinising the existing models was that the majority, as the ones identified in Limone`s review (2013) [173], share a common core structure taken from the Sorensen and Gage models [175, 176]. The models also were populated with the safety and efficacy data of the pivotal RCTs (Aristotle, RE-LY, and ROCKET-AF) [41-43]. Two of these models [178, 182] also included evidence from the AVERROES trial [47].

There were however, differences in complexity reflected by the number of health states used. In particular, Dorian's model included the following 11 health states: healthy with AF, ischaemic stroke (mild, moderate, and severe), haemorrhagic stroke (mild, moderate, and severe), SE, MI, treatment discontinuation and death [178].

The manufacturer model for apixaban reported 18 health states: health with AF, ischaemic stroke (mild, moderate, severe and fatal), haemorrhagic stroke (mild, moderate, severe and fatal), SE (fatal and non-fatal), MI, ICH (fatal and non-fatal), other major bleeds, clinically relevant non-major bleeds, cardiovascular hospitalisations, death [182]. Lip's model (2014) [179] was the same as the one presented by Dorian (2014) [178].

In Sterne's model (2017), the AF cohort enters the model in the "AF well" state, those are patients who are healthy and with AF. The cohort may then experience Stroke, MI, major bleed, ICH or die at any time. Over time, patients could again experience Stroke, MI, major bleed, ICH, but this time the risk of experiencing an event would depend on previous event history. This would result in 17 health states [180]. Verhoef's model (2014) only included 9 health states: healthy with AF, ischaemic stroke, TIA, MI, SE, ICH, extra-cranial haemorrhage, disability and death.

Cycle length

As indicated in all of the models adopting the UK perspective, and specified by the time horizon, patients were followed for their lifetime. The time horizon, indicating the time frame of the analysis, is divided into equal increments of time (Markov cycles) during which transition from one state to another can be made. The majority of the models adopted a cycle length of 6 weeks [178, 179, 182], thus reflecting the frequency of recurrent events following a first stroke or any other main clinical event associated with AF. Sterne (2017) and Verhoef (2014) adopted cycle lengths of 3 and 1 months, respectively [180, 181]. Further, the model transition probabilities indicating the probability of moving from one health state to another were derived, from all of the studies, from RCTs that assessed the efficacy and safety of the DOAC under consideration [41-43]. Additional evidence on DOACs effectiveness was generated from NMA carried out by apixaban manufacturer [182] and Sterne (2017) [180].

Assumptions- HRQOL

Several assumptions were made regarding the level of utilities between treatments and the risk of experiencing relevant clinical events between health states. In particular, Dorian (2014) [178] and Lip (2014) [179] assumed the same level of utility decrement between DOACs and aspirin (being one of the comparators in addition to warfarin), while the apixaban manufacturer model assumed a level of disutility (decrement in utility) only associated with warfarin, and same risk of MI for apixaban and dabigatran [182].

Assumptions -effectiveness

Further, Lip (2014) assumed the recurrence of ischaemic or haemorrhagic stroke and the rate of all-cause mortality being the same across treatments [179]. In addition, the apixaban manufacturer model and Sterne (2017) assumed the relative treatment effect being constant over time. Sterne, also assumed the same probability of dying for patients with a history of clinically relevant bleeding, ICH or stroke [180, 182].

Assumptions discontinuation and switching

In the all studies identified, assumptions were made around drug discontinuation and switching. Primarily, patients who discontinued anticoagulation therapy due to ICH or major bleeding events switched to aspirin or interrupted treatment. Discontinuation and switching rates were based on those observed in RCTs [41-43].

Results

In most studies, all DOACs were deemed cost-effective compared to warfarin with an ICER well below the £20,000 threshold [178-180, 182]. Overall DOACs were found to have a high probability, greater than 80%, of being cost-effective compared to warfarin [43]. However, Verhoef (2014) reported considerably lower probabilities of DOACs being cost-effective at £20,000, with less than 50% for apixaban and dabigatran standard dose and 5% for rivaroxaban 20 mg [181] .

Sensitivity analysis and key drivers

The risk of stroke and ICH and all-cause mortality were among the clinical parameters the mostly influenced the cost-effectiveness models. Other impactful parameters were the cost associated with clinical events or treatments. In particular, treatment and INR monitoring costs were identified as key drivers in the majority of the cost-effectiveness studies identified [178, 179, 181, 182].

Table 7.1 Cost-effectiveness studies identified in the review

Study	Interventions	Population	Model type (perspective, cycle length and assumptions)	Discontinuation and switching	Results	Sensitivity analysis and key drivers
Dorian, 2014	Apixaban 10 mg daily, warfarin 5 mg (average) daily, aspirin 150 mg daily	70 years mean age with AF and suitable for VKA therapy. 70 years mean age with AF and not suitable for VKA therapy.	Markov decision model, UK perspective, lifetime horizon, cycle length (6 weeks). Data source: AVERROES and ARISTOTLE trials Assumption: risk of stroke recurrence not treatment specific, same utility decrements for apixaban and aspirin, no apixaban efficacy for secondary stroke prevention, drug discontinuation and switching.	Discontinuation of apixaban and switch to Aspirin due to ICH (56% of patients), or interrupt treatment for 6 weeks (44% of patients). Discontinuation of aspirin and switch to Warfarin due to ischaemic stroke or SE. Permanent discontinuation of treatment due to haemorrhagic stroke or MI.	Apixaban vs warfarin £11,909 (97% at £20,000 threshold). Apixaban vs aspirin £7,196 (99% at £20,000 threshold).	Main drivers: apixaban risk of stroke, apixaban utility decrement, apixaban other death rate, daily cost apixaban, aspirin stroke rate.
Edwards, 2012 (manufacturer model)	Apixaban, dabigatran 150 mg/110 mg, dabigatran 110 mg, rivaroxaban, warfarin.	70 years mean age with AF and suitable for VKA therapy. 70 years mean age with AF and not suitable for VKA therapy.	Markov decision model, UK perspective, lifetime horizon, cycle length (6 weeks). Data source: AVERROES and ARISTOTLE trial Assumptions: same risk of MI and level of other-cause discontinuation for apixaban and dabigatran, constant relative treatment effect across CHADS2 score categories, disutility associated with treatment and not age adjusted, drug discontinuation and switching.	Discontinuation of anticoagulation due to MI or haemorrhagic stroke. Discontinuation of aspirin (as second line treatment) and switch to warfarin due to ischaemic stroke or SE. Switching to aspirin or no treatment due to ICH, major bleed or other causes.	Apixaban vs warfarin £11,008 (80% at £20,000 threshold). Dabigatran 110 mg dominated by dabigatran blend. Rivaroxaban extendedly dominated by apixaban. Dabigatran blend extendedly dominated by apixaban.	Main drivers: apixaban vs warfarin: warfarin disutility, risk of ICH, ischaemic stroke or other-cause mortality, INR monitoring costs, QALYs discount rate. Main drivers: apixaban vs rivaroxaban or Dabigatran, risk of ICH, ischaemic stroke or other-cause mortality, apixaban absolute risk of stroke, second-line aspirin risk of stroke.

Table 7.1 Cost-effectiveness studies identified in the review (continued a)

Study	Interventions	Population	Model type (perspective, cycle length and assumptions)	Discontinuation and switching	Results	Sensitivity analysis and key drivers
Lip, 2014	Apixaban 5 mg twice daily, dabigatran 150 mg twice daily, dabigatran 110 mg twice daily, rivaroxaban 20 mg once daily, aspirin, adjusted dose warfarin.	70 years mean age with AF and suitable for VKA therapy.	Markov decision model, UK perspective, lifetime horizon, cycle length (6 weeks). Data source: ARISTOTLE, RE-LY and ROCKET-AF trials Assumptions: recurrence of ischaemic or haemorrhagic stroke same for all treatments, case fatalities and all-cause mortality rates same for all treatments, same utility decrement for DOACs and aspirin.	Discontinuation of anticoagulation and switch to aspirin due to bleeding events.	Apixaban vs dabigatran 150 mg £9,611 (83% at £20,000 threshold). Apixaban vs dabigatran 110 mg £4,497 (98% at £20,000 threshold). Apixaban vs rivaroxaban 20 mg £5,305 (85% at £20,000 threshold).	Main drivers: Drugs prices, ICH and stroke rates, and hospitalisation rates.
Sterne, 2017	Apixaban 5 mg twice daily, dabigatran 150 mg twice daily, rivaroxaban 20 mg once daily, adjusted dose warfarin.	70 years mean age with AF and suitable for VKA therapy.	Markov decision model, UK perspective, 30 years time horizon, cycle length (3 months). Data source: ARISTOTLE, RE-LY and ROCKET-AF trials and NMA. Assumptions: future risk of death for bleeds and ICH same as stroke, events rate and relative treatment effects independent of age, similar post-ICH and post-ischaemic stroke management costs, drug discontinuation and switching.	Discontinuation of dabigatran and switch to warfarin due to MI. Switching to no treatment due to ICH or haemorrhagic stroke. Switching from DOAC to warfarin, or the reverse, due to ischaemic stroke bleeding, SE or TIA.	Apixaban 5 mg £7,533* (close to 60% at £20,000/£30,000 threshold). Dabigatran 150 mg £5,279*. Rivaroxaban 20 mg £6,365*.	Main drivers: apixaban low rates of MI, ICH and clinical relevant bleeding.

* Incremental net monetary benefit

Table 7.1 Cost-effectiveness studies identified in the review (continued b)

Study	Interventions	Population	Model type (perspective, cycle length and assumptions)	Discontinuation and switching	Results	Sensitivity analysis and key drivers
Verhoef, 2014	Apixaban 5 mg twice daily, dabigatran 150 mg twice daily, rivaroxaban 20 mg once daily, coumarin derivatives, Aspirin	70 years mean age with AF initiating anticoagulation therapy.	Markov decision model, UK perspective, lifetime horizon, cycle length (1 month) Data source: ARISTOTLE, RE-LY and ROCKET-AF trials. Assumptions: frequency of INR yearly measurements, event rates stable after 2 years, about one third of ischaemic stroke assumed TIA, drug switching.	Discontinuation of anticoagulation and switch to aspirin due to ICH.	Apixaban 5 mg €11,470 (£10,058) ** (35% at £20,000 threshold). Dabigatran 150 mg €11,171 (£9,796) ** (41% at £20,000 threshold). Rivaroxaban 20 mg €16,949 (£14,862) ** (5% at £20,000 threshold).	Apixaban 5 mg (35% at £20,000 threshold). Dabigatran 150 mg (41% at £20,000 threshold). Rivaroxaban 20 mg (5% at £20,000 threshold). Main drivers: time in range, risk of ICH for rivaroxaban, cost of apixaban.

** Cost converted from euro to British pound

7.2.3 Review – discussion

The models appraised with the CHEERS tool were in general of good quality. However, for all models, a rational and a justification for the choice of the discount rate and health outcomes used as the measure of benefit was not provided. Nevertheless, it is appreciated that most models in economic evaluation adopting the NHS perspective, typically use a 3.5% discount rate for cost and utilities, and use QALYs as the outcome measure as per NICE reference case recommendations [183]. Another common issue identified in most of the models identified was a lack of reporting methods for adjusting estimated unit cost. In particular it was not clear how and whether costs where inflated. The complete CHEERs checklists for each of the identified study is reported in Appendix XVIII (Table XVIII-1−XVIII-4).

Different models evaluating the cost-effectiveness of DOACs have been developed, and although a common structure borrowed from the Sorensen and Gage models [175, 176] is shared, they differ in terms of health state, assumptions and inputs used. These differences reflect different levels of complexity; however, the models evaluated also shared common features. For instance, all models assumed that patients started anticoagulation treatment at 70 years of age, thus reflecting the age enrolment of the pivotal RCTs [41-43].

Beyond a common core structure, the existing cost-effectiveness models, as the ones identified in the Limone's review [173], were populated with the safety and efficacy data of the pivotal RCTs (Aristotle, RE-LY, ROCKET-AF) [41-43]. Dorian's model (2014) [178] and the apixaban manufacturer model [182] also included evidence from the AVERROES trial [47] as they included aspirin as an alternative to anticoagulation discontinuation (patients discontinue anticoagulation with DOAC and switch to aspirin due to bleeding events) [178, 179, 181].

However, this implies the inclusion of a VKA unsuitable patient population, and may not be relevant for the cost-effectiveness analysis in this Chapter. The number of health states employed in each model reflected different levels of complexity. For instance, the manufacturer model for apixaban categorised ischaemic and haemorrhagic stroke according to four different levels of severity, and distinguished SE and ICH between fatal and fatal. Achieving this level of complexity may only be possible with availability of patient level data as in the RCTs. Nevertheless, even with real-world data it may not be feasible to categorise stroke according to severity [182]. Further, unlike all the other models identified in the review, Verhoef (2014) included disability as a health state; that is a disability from warfarin-associated intracranial and extra cranial haemorrhages [181].

All models identified assessed apixaban [178-182], which was found to be a cost-effective option compared to warfarin. The ICERs for those models were below the NICE recommended threshold of £20,000, and in line with the most plausible ICER of £12,757 identified by the Evidence Review Group (ERG), an external and independent academic group that reviews the manufacturer or sponsor's evidence submission on behalf of NICE [184]. In the model where the dabigatran was compared against warfarin and found to be cost-effective [181], the ICER reported was again well below the £20,000 threshold and the most plausible ICER of £18,900 calculated by the ERG group.

As mentioned, all the models identified in the review were Markov based models. Despite, the availability of patient-level data, the option of using a DES model was excluded for the reasons elucidated in the previous paragraph comparing Markov versus DES approaches. Specifically, for the DOACs case study, a DES model would generate a new set of probabilities for future events by updating patients' health profiles over time according to their characteristics, risk scores, and their experience of clinical events [185].

However, in the cost-effectiveness analysis carried out in this Chapter, patients' characteristics and risk scores for estimating risk of events were only measured at baseline; and, as it will be further discussed, the risk of future events were obtained from the literature [186]. Therefore, despite the availability of patient level data, Sterne's model seemed the best alternative for the cost-effectiveness analysis undertaken in this Chapter.

In particular, Sterne's model seemed the most fit for purpose because of the ease of implementation, with no stroke severity distinction between mild, moderate and severe [180]. Currently, there is not sufficient evidence to assume that stroke severity depends on the anticoagulation treatment received [35]. Furthermore, health states for non-clinically relevant minor bleed were not included as they were assumed to have minimal impact on cost, quality of life and future risks. Hence, no distinction was made between bleed locations such as GI bleeding or other types of bleed [180]. The model also appeared to be inclusive of all the relevant health states including TIA and SE. Unlike other models [178, 179, 182], TIA and SE health states were included but only for the acute phase. Further, TIA was used as an alternative to the reversible ischaemic neurological deficit -RIND (used to reflect recovery from a temporary stroke or TIA) to represent a non-disabling minor stroke [180].

7.3 Methods

The objective of this economic evaluation is to assess the cost-effectives of DOACs when compared to warfarin in Scottish clinical practice, from a clinical and economic perspective for the prevention of stroke in the AF population. In the base-case analysis, it was assumed that patients entered the model at 50 years of age. This assumption was based on clinical advice and the evidence that prevalence and incidence of AF typically increase exponentially from 50 years onwards [89].

Thus, the analysis including patients from the age of 50 would be inclusive of all patients potentially at risk of AF. The baseline characteristics were the same described for the AF cohort in Chapter 5 and Chapter 6. However, a further subgroup analysis was carried out on a cohort of patients 70 years or older at the time of coagulation initiation. This is line with the anticoagulation starting age adopted in most cost-effectiveness models reviewed [178-182].

In the cost-effectiveness analysis, the DOACs, including apixaban, dabigatran and rivaroxaban, currently used in Scottish clinical practice, were compared against warfarin and also incrementally against each other. In addition to the main analysis, a set of subgroup analyses was carried out to assess the cost-effectiveness of apixaban and rivaroxaban, when administered at their standard or reduced dose; subgroup analysis was not possible for dabigatran due to sample size.

Cost and outcomes were extrapolated for 50 years for a cohort of patients 50 years or older; thus, a lifetime horizon was adopted. Total costs and QALYs for each intervention were discounted at 3.5% as per NICE guidelines [184], and presented against warfarin and incrementally, according to their effectiveness, to identify dominance and extended dominance. These are concepts used in economic evaluation as described in the NICE reference case.

While dominance is established when a given intervention is less effective and more costly compared to the previous alternative, extended dominance is determined by the ICER of an intervention being higher compared to the next more effective alternative [151]. As CUA is typically used in economic evaluations, QALY was used as the outcome measure reflecting differences in health benefits between interventions.

As per NICE reference case recommendation, the study adopted the NHS perspective [183]. The NHS perspective includes cost of treatment, administration and monitoring, and those related to service utilisation such as general practitioner visits and hospital admissions. Costs associated to management of treatment side effects may also be included. However, transportation, over the-counter purchases and indirect costs are not included.

While a broader societal perspective may be relevant for this study, reflecting a series of social opportunity costs, as for instance those generated by reducing productivity loss, NICE recognise that the societal perspective may introduce bias against people not at work because of retirement age or unable to work because of health conditions.

7.3.2 Model structure

The multistate Markov model was developed based on the latest published DOAC cost-effectiveness model, which adopted the UK perspective [180]. The model used in this economic evaluation maintained the Sterne's model structure (2017) [180], but different assumptions were made. As in Sterne's model (2017) [180], the same model structure is used for each treatment strategy with variations in event probabilities and associated utilities and costs. While the main structure, using a cycle length of 3 months is maintained, assumptions on treatment switching and discontinuation were not implemented as there were directly modelled in the treatment effectiveness estimation in Chapters 5 and 6.

As shown in Figure 7.2 [180], the AF cohort enters the model in a “AF” state, patients may then experience Stroke, MI, major bleed, ICH or die at any time. Over time, patients could again experience Stroke, MI, major bleed, ICH, but this time the risk of experiencing an event would depend on previous event history. For instance, if patients move from MI to the MI+S health state, the risk of stroke will depend on the present risk of stroke, but also on the future risk of stroke for patients who had an MI. Similarly, if patients move from MI+S to the ICH+MI+S health state, the risk of ICH will depend on the present risk of ICH, but also on the future risk of ICH for patients who had an MI and a stroke.

Patients moving to the S+B+MI+ICH health state, may experience stroke, bleeding, MI or ICH; and again, the risk of moving into this health state will depend on previous event history. For patients who are moving from ICH+MI+S health state, the future risk of bleeding will depend on the previous history of ICH, MI and stroke.

Patients may also experience TIA or SE; these however are considered temporary health states having only a short-term effect in terms of costs and utilities. On the other hand, stroke, ICH, MI and bleeding are expected to have some effect in the long-term, and therefore need to be modelled.

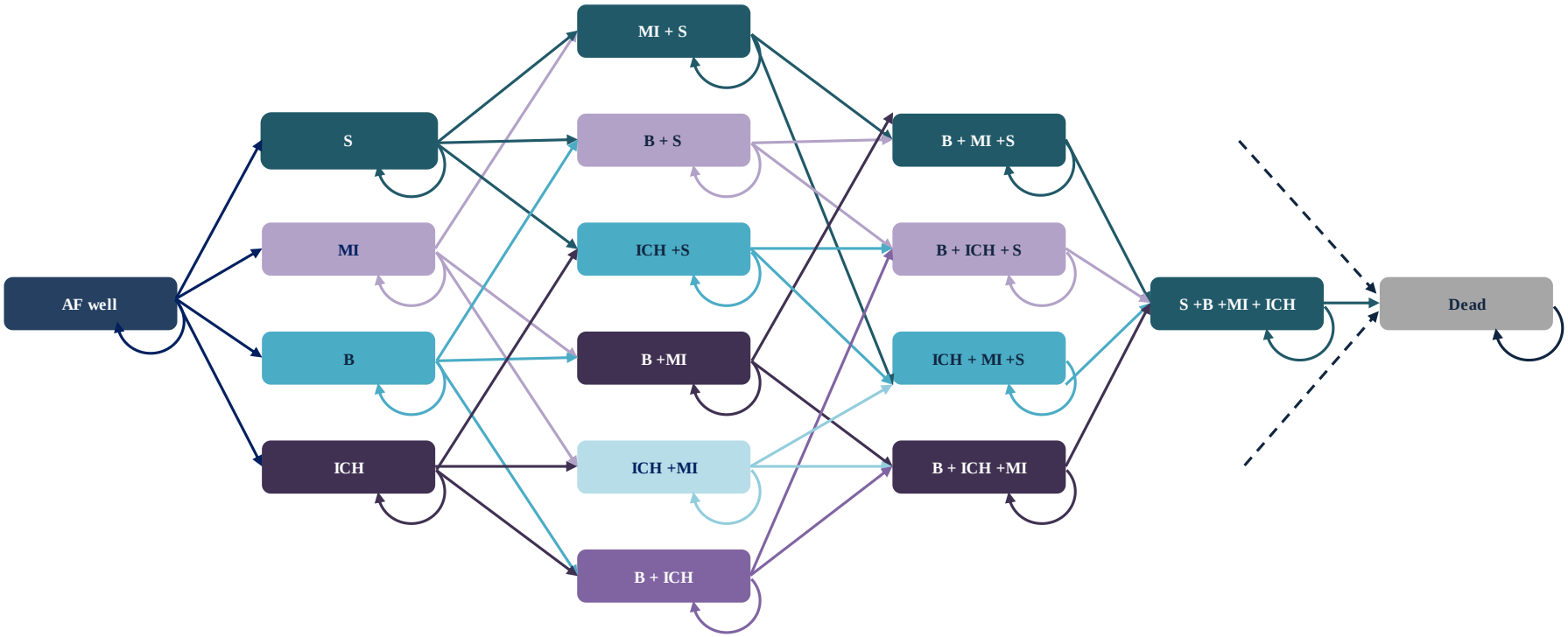


Figure 7.2 Multi-state Markov model for the prevention of stroke in the AF population

Abbreviations: S=ischaemic stroke, MI= myocardial infarction, B= major bleeding, ICH= intracranial haemorrhage
Source: Adapted from, Sterne et al. (2017).

7.3.3 Model inputs

Event rates and transition probabilities

In addition to the clinical outcomes described in Chapter 6, ischaemic stroke, TIA and SE were estimated individually to accommodate the transient role that these states have in the model. The effectiveness parameters were estimated following the same method described in Chapters 5 and 6. The cumulative incidence curves for DOACs any dose, standard dose, and reduced dose are presented in Appendix XIX (Figure XIX-1−XIX-6). While the cumulative incidence curves generated in Chapter 6 show the ATE, those presented in Appendix XIX show the relative treatment effect, and therefore one single regression (where warfarin is the reference group) was run including all four treatment options represented by a dummy variable. Nevertheless, the interpretation in terms of probability of event is the same, and the HRs presented in this section diverge marginally from those presented in Chapter 6. The ICD 10 codes used for disease identification were previously discussed and presented in Chapter 6, Table 6.1.

The model transition probabilities for each treatment and each clinical outcome were obtained by firstly converting warfarin event rates into probabilities, and then applying the HRs from the incremental time-to-event analysis to the warfarin event rates. The transition probabilities for warfarin were assumed to follow a beta distribution were α represents the number of events and β is obtained by subtracting the number of events from the number of person-years. The beta distribution, constrained on the interval 0-1, reflects the proportion of observed outcomes and represents the conjugate of the binomial distribution [151]. Warfarin event rates and transition probabilities are presented in Table 7.2.

Table 7.2 Warfarin event rates and transition probabilities

Outcome	Event rate	Beta distribution α	Beta distribution B	Transition probability (3 months cycle)
TIA	0.45	180	39,536	0.0011
SE	0.12	49	39,667	0.0003
Stroke	1.44	573	39,143	0.0036
MI	0.99	393	39,323	0.0025
Major bleeding	4.03	1,599	38,117	0.0100
ICH	0.28	113	39,603	0.0007
Mortality (all-cause)	5.30	2,104	37,612	0.0132

Abbreviations: TIA=transient ischemic attack, SE=systemic embolism, MI= myocardial infarction, ICH= intracranial haemorrhage.

Event rates were converted into transition probabilities according to the following equation [151]:

$$Probability = 1 - exp^{(-rt)}$$

(Equation 7.3)

Where r is the event rate and t the time.

To obtain the transition probabilities, the event rates presented per 100 person-years, were firstly brought back to their original values, then the transition probability equation was applied. All transition probabilities where divided by 91.3125 to convert yearly values into 3 months cycles transition probabilities for the Markov model. For instance, based on the event rate of TIA for warfarin of 0.0045 (or 0.45 per 100 person-years) the transition probability would be calculated as shown in the equation below:

$$\text{Warfarin TIA probability} = 1\text{-}\exp(-(0.0045/365.25)*91.3125) = 0.0011$$

Equation (7.4)

The transition probability for DOACs were obtained by multiplying the warfarin event rates by the HRs of DOACs clinical outcomes, then the transition probability equation (as described in Equation 7.4) was applied. For instance, based on the event rate of TIA for warfarin of 0.0045 (or 0.45 per 100 person-years) and the HR for apixaban of 0.87, the transition probability would be calculated as shown in the equation below:

$$\text{Apixaban TIA probability} = 1\text{-}\exp(-((0.0045*0.87)/365.25)*91.3125) = 0.001$$

Equation (7.5)

All HRs and transition probabilities calculated for pooled, standard and reduced dose, are presented in Table 7.3. The number of events observed and estimated HRs for the 70 years or older subgroup are reported in Appendix XX (Table I). The HRs reflecting the treatment effect were assumed to follow a lognormal distribution, where the confidence limits calculated on the log scale indicate lognormal as the most appropriate distributional form for risk parameters [151].

As in previous models, event rates and treatment effects were assumed independent of age; hence, the derived transition probabilities were homogenous throughout the model. However, life tables for Scotland were included in the model, thus allowing transition probabilities to change over time as the cohort ages [77].

The probability of experiencing a clinical event depends on previous events history. For instance, if one patient on warfarin moves from the MI to the MI+S health state, the risk of stroke will depend on the actual risk of stroke associated with warfarin (estimated with the comparative-effectiveness methodology discussed in the previous Chapter), but also on the future risk of stroke for a patient who had an MI. The risk of future events in Table 7.4 are age and treatment independent, and assumed to follow a lognormal distribution.

The risks of future events, taken from Sterne`s model, are based on a study that estimated the risk of SE, TIA, MI and bleeding according to previous event history [186] and a study that estimated the risk of previous events on mortality [187]. Further, as there is no evidence available on the effect of bleeding on ICH and mortality, Sterne assumed that the effect of bleeding and ICH on mortality would be the same as stroke.

Table 7.3 DOACs HRs and transition probabilities

Intervention (compared to warfarin)	Outcome	HR 95 % CI	Standard error	Transition probability (3 months cycle)
Apixaban				
	TIA	0.87 (0.47 -1.63)	0.275	0.0011
	SE	0.62 (0.17 -2.25)	0.415	0.0002
	Stroke	0.86 (0.6 -1.23)	0.153	0.0032
	MI	1.27 (0.86 -1.87)	0.243	0.0033
	Major bleeding	0.85 (0.69 -1.05)	0.091	0.0092
	ICH	1.54 (0.79 -3.02)	0.508	0.0011
	Mortality (all-cause)	0.99 (0.83 -1.99)	0.092	0.0129
Apixaban (standard dose)				
	TIA	0.82 (0.38 -1.78)	0.313	0.0010
	SE	0.55 (0.12 -2.51)	0.428	0.0002
	Stroke	0.91 (0.6 -1.39)	0.195	0.0034
	MI	1.23 (0.77 -1.96)	0.282	0.0032
	Major bleeding	0.91 (0.72 -1.15)	0.109	0.0099
	ICH	1.42 (0.65 -3.13)	0.567	0.0010
	Mortality (all-cause)	0.96 (0.77 -1.22)	0.113	0.0125
Apixaban (reduced dose)				
	TIA	0.98 (0.40 -2.39)	0.457	0.0012
	SE	0.85 (0.10 -6.95)	0.883	0.0003
	Stroke	0.78 (0.44 -1.38)	0.219	0.0029
	MI	1.52 (0.87 -2.66)	0.452	0.0040
	Major bleeding	0.78 (0.53 -1.14)	0.156	0.0085
	ICH	1.77 (0.64 -4.92)	0.890	0.0012
	Mortality (all-cause)	1.01 (0.77 -1.32)	0.136	0.0132
Dabigatran				
	TIA	0.44 (0.06 -3.18)	0.525	0.0005
	SE*	1.00 (0.00 -0.00)	0.000	0.0003
	Stroke	0.62 (0.23 -1.66)	0.294	0.0023
	MI	0.43 (0.11 -1.76)	0.306	0.0011
	Major bleeding	1.02 (0.67 -1.56)	0.211	0.0111
	ICH*	1.00 (0.00 -0.00)	0.000	0.0007
	Mortality (all-cause)	0.80 (0.52 -1.26)	0.176	0.0104
Rivaroxaban				
	TIA	0.65 (0.35 -1.22)	0.205	0.0008
	SE	0.44 (0.10 -1.94)	0.396	0.0001
	Stroke	0.95 (0.69 -1.30)	0.145	0.0036
	MI	0.81 (0.54 -1.24)	0.163	0.0021
	Major bleeding	1.29 (1.09 -1.52)	0.107	0.0140
	ICH	1.52 (0.85 -2.73)	0.420	0.0011
	Mortality (all-cause)	1.42 (1.23 -1.64)	0.098	0.0185
Rivaroxaban (standard dose)				
	TIA	0.66 (0.33 -1.34)	0.234	0.0008
	SE	0.59 (0.14 -2.57)	0.415	0.0002
	Stroke	1.02 (0.72 -1.45)	0.174	0.0038
	MI	0.70 (0.42 -1.18)	0.175	0.0018
	Major bleeding	1.29 (1.07 -1.55)	0.115	0.0140
	ICH	1.58 (0.84 -2.99)	0.481	0.0011
	Mortality (all-cause)	1.50 (1.27 -1.76)	0.117	0.0196
Rivaroxaban (reduced dose)				
	TIA	0.51 (0.12 -2.11)	0.373	0.0006
	SE*	1.00 (1.00 -1.00)	1.335	0.0003
	Stroke	0.71 (0.35 -1.46)	0.249	0.0027
	MI	1.40 (0.72 -2.69)	0.463	0.0037
	Major bleeding	1.57 (1.14 -2.17)	0.252	0.0171
	ICH	1.02 (0.24 -4.34)	0.740	0.0007
	Mortality (all-cause)	1.32 (1.00 -1.73)	0.182	0.0172

* Assumed no treatment effect difference between warfarin and dabigatran for SE and ICH.
Abbreviations: TIA= transient ischaemic attack, SE=systemic embolism, MI=myocardial infarction, ICH=intracranial haemorrhage

Table 7.4 Effect of previous events on future events (HRs)

Risk factor	Future stroke	Future TIA/SE	Future ICH	Future bleed	Future mortality
Stroke	4.00 (3.78, 4.22)	3.61 (3.44, 3.78)	1.64 (1.39, 1.94)	1.39 (1.27, 1.52)	1.32 (0.98. 1.77)
ICH	1.78 (1.56, 2.03)	1.82 (1.62, 2.04)	10.2 (8.59, 12.2)	2.95 (2.57, 3.39)	1.32 (0.98, 1.77)
Bleeding	1.32 (1.21, 1.44)	1.36 (1.26, 1.46)	3.54 (3.02, 4.17)	3.32 (3.06, 3.60)	1.32 (0.98, 1.77)
MI	1.24 (1.17, 1.33)	1.29 (1.22, 1.36)	0.94 (0.78, 1.12)	1.24 (1.15, 1.35)	1.03 (0.73, 1.46)

Abbreviations: ICH=intracranial haemorrhage, MI=myocardial infarction.
Source: Sterne et al. (2017).

7.3.4 Cost and utility inputs

Cost inputs

While effectiveness parameters were obtained from the survival analysis as previously described, cost and utility parameters were obtained from various sources. The cost for the clinical events occurring in the acute phase (first 3 months from anticoagulation initiation) were mainly obtained from the SNT 2013/2014 [100], and then averaged according to the level of activity identified in the NHS reference cost 2011/2012. The activity is measured according to the number of attendances, bed days, clients, episodes, tests, or other unit of activity appropriate to the service [188]. To match the clinical events from SNT and level of activity from the NHS reference cost, HRGs codes were used. However, while a more up to date version of the NHS reference cost is publicly available, the HRG codes did not match those reported in the SNT for the year 2013/2014; thus, HRG from the NHS reference cost for the year 2011/2012 were used. The cost of MI was double the amount of the reference cost, thus accounting for potential follow-up costs [180].

Acute management costs for ischaemic stroke and ICH, obtained from the Sterne model, are based a study assessing in the UK the acute and long-term costs of stroke in atrial fibrillation patients [189]. Costs for the post-acute phase were identified from the same study, and averaged according to the number of patients experiencing a non-disabling, moderately disabling or a totally disabling stroke [189].

All costs were inflated to 2017/2018 prices, using the consumer price index for medical interventions, and are presented in Table 7.5 [190]. The unit cost for the DOACs were obtained from the 2018 BNF database [191-193], and calculated for a 3-month cycle; while the average drug and monitoring costs of warfarin were obtained from a NICE costing report on AF [194] cited in different cost-effectiveness studies. The unit cost of each DOAC (Table 7.6) is known and therefore assumed to be fixed. However, the cost of warfarin and associated monitoring typically varies significantly across clinical practice, and therefore a uniform distribution ranging from 50% to 150% of the mean cost was assumed [194].

Table 7.5 Costs of clinical events

Event (Non-Elective)	Cost £ (SD)	Activity	HRG code	Average weighted cost	Inflated cost	Distribution	Source
Acute phase cost							
Ischaemic stroke	10,844 (15,733)				12,705 (18,433)	Normal (12,705, 1,448)	Sterne (2017)
ICH	10,683 (12,885)				12,517 (15,096)	Normal (12,517, 3,661)	Sterne (2017)
TIA	1,368		AA29Z		1,495	Uniform (748, 2,243)	Scottish National Tariff*, NHS reference cost**
MI	2,518***		EB10Z	5,036	5,503	Uniform (2,752, 8,255)	Scottish National Tariff*, NHS reference cost**
SE	7,254	1,545	QZ17A	3,673	4,013	Uniform (2,007, 6,020)	Scottish National Tariff*, NHS reference cost**
	3,549	10,946	QZ178				
	1,984	2,473	QZ17C				
Clinical relevant bleeding	4,783	16,116	FZ38D	3,008	3,287	Uniform (1,644, 4,931)	Scottish National Tariff*, NHS reference cost**
	3,222	19,304	FZ38E				
	625	2,806	FZ38F				
	5,101	13,336	FZ43A				
	3,369	16,491	FZ43B				
	664	1,935	FZ43C				
	159		VB07Z				
Post-acute phase cost							
Non disabling (N=66)	2,135 (3,675)			3,370 (6368)	3,948 (7,461)	Normal (3,948, 640)	Sterne (2017)
Moderately disabling (N=58)	4,165 (7,668)						
Totally disabling (N=12)	6,324 (14,898)						

Abbreviations: ICH=intracranial haemorrhage, TIA=transient ischaemic attack, MI=myocardial infarction, SE=systemic embolism
*Scottish National Tariff (2013/2014)
**NHS reference cost (2011/2012)
***Double to include follow-up costs

Table 7.6 Cost of prescription drugs

Intervention	Size	Cost per pack (£)	Cost per day (£)	Cost per 3 months cycle (£)
Apixaban 2.5 mg	60	57.00	1.90	173.38
Apixaban 5 mg	56	53.20	1.90	173.38
Dabigatran	60	51.00	1.70	155.13
Rivaroxaban 15 mg	100	180.00	1.80	164.25
Rivaroxaban 20 mg	100	180.00	1.80	164.25
Warfarin (including monitoring)				77.25

Note: cost of warfarin inflated for 2017/2018
Source: National Institute for Health and Care Excellence. *BNF apixaban* (2018), National Institute for Health and Care Excellence. *BNF dabigatran* (2018), National Institute for Health and Care Excellence. *BNF rivaroxaban* (2018), National Institute for Health and Care Excellence. *Costing Report: Atrial Fibrillation* (2006).

Utility inputs

The utility parameters for the chronic health states were obtained from Sterne's model [180], where the parameter for the AF HRQoL was identified from a study that analysed factors such as age and comorbidities, determining the level of utilities in patients diagnosed with any type of AF [160]. The utilities for stroke and ICH were identified from a study looking at the long-term HRQoL in the 4 years following a stroke [195]; while the utility for MI was taken from a study exploring factors such as gender and social class believed to influence self-perceived health following an MI [196].

The utilities for acute health events were obtained by applying disutilities, assumed to last for 3 months, to the utilities of the chronic health states previously discussed. In particular, the disutilities for TIA, SE, stroke and major bleeding were obtained from a study assessing patient valuations of AF related health states [164]. The disutility for MI and ICH where obtained by subtracting the AF reference health state from the utility values for the acute phases identified respectively in Lacey (2003) [196] and Lenert (1997) [197].

The disutilities were capped at 0 and assumed to follow a Uniform distribution, ranging from 50-150% of the reported mean, whenever the uncertainty estimates and sample sizes were not available. Further, the utilities were multiplied by 0.25 to obtain a QALY value for a 3-month cycle [180]. Utility and disutility parameters with associated distributions are presented in Table 7.7.

Table 7.7 Utility and disutility

Component	Utility/disutility (SD)	Distribution
Event utility		
Stable AF	0.779	Normal (0.779, 0.0045)
Post Stroke	0.69 (0.18)	Normal (0.69, 0.0205)
Post MI	0.718 (0.243)	Normal (0.718, 0.0163)
Post ICH	0.74 (0.39)	Beta (3.941, 1.385)
Event disutility (first 3 months)		
Acute TIA	-0.131	Uniform (-0.197, - 0.066)
Acute SE	-0.131	Uniform (-0.197, - 0.066)
Acute Stroke	-0.590	Uniform (-0.885, - 0.295)
Acute MI*	0.683 (0.233) - Stable AF	Normal (0.683, 0.0156) - Stable AF
Acute Major bleeding	-0.03	Normal (-0.03, 0.001531)
Acute ICH*	0.60 - Stable AF	Normal (0.60, 0.064) - Stable AF

* distribution capped at 0
Abbreviations: MI=myocardial infarction, ICH=intracranial haemorrhage, TIA=transient ischaemic attack, SE=systemic embolism.
Source: Sterne et al. (2017).

Utility decrements were multiplied by the general population utility values used in Sterne's model [180] and estimated by Kind et all (1999) (Table 7.8) [76].

Table 7.8 General population utility values by age and gender

Age	Male			Female		
	mean (SD)	A	B	mean (SD)	α	β
45-54	0.84 (0.27)	314.4	65.0	0.85 (0.23)	544.1	96.0
55-64	0.78 (0.28)	330.4	93.2	0.81 (0.26)	526.6	123.5
65-74	0.78 (0.28)	388.5	109.6	0.78 (0.25)	551.7	155.6
75+	0.75 (0.28)	191.2	63.7	0.71 (0.28)	406.4	166.0

Source: Sterne et al. (2017)

7.3.5 Model assumptions

In this model, a series of assumptions were made. Firstly, ICH, major bleeding and stroke were assumed to have the same effect on future risk of death; nevertheless, the mortality rate in AF patients relative to the general population was not assumed to change according to age. Similarly, relative treatment effects and event rates were assumed to be constant over time.

Secondly, assumptions were made around the distribution of the cost of warfarin and the associated monitoring. The Uniform distribution ranging from 50% to 150% of the mean estimate reported in the NICE costing report on AF was assumed to reflect the uncertainty and the wide variation of warfarin and monitoring costs across clinical practice [194].

In addition, stroke and ICH long-term management costs were assumed to be the same, and the cost of MI was assumed to be double the amount of the reference cost to account for any potential follow-up cost [180]. A further assumption was made about no SE and ICH risk difference between dabigatran and warfarin. This is because it was not possible to establish, due to the limited number of SE events and no ICH events estimated in the dabigatran arm, the actual risk of SE and ICH for each treatment. For the same reason, patients on reduced dose rivaroxaban and warfarin were assumed to have the same risk of SE.

In this model, unlike other cost-effectiveness models on DOACs, no assumptions were made about treatment switching and discontinuation, as this was addressed in the time-to-event analysis where patients who switched or discontinued were censored. In addition, about a third of patients switched or discontinue anticoagulation within the first 2 years from therapy initiation.

Therefore, switching and discontinuation, in the long-term, may have limited effect on clinical outcomes. Full details on how treatment switching and discontinuation were included in the survival analysis were discussed in Chapters 5 and 6.

7.3.6 Model uncertainty

In addition to the base-case deterministic analyses, sensitivity analysis was carried out by means of PSA and one-way sensitivity analysis to assess the uncertainty around the findings and identify the key drivers. In particular, PSA, based on the probability distribution of each parameter, was used to assess parameter uncertainty [151, 168]. PSA represents parameters as distributions of possible mean values instead of single point estimates in a deterministic analysis.

Combination of cost-effectiveness pairs presented in the cost-effectiveness plane reflect the number of iterations run in the model; and the location of this across the four quadrants on the cost-effectiveness plane would indicate whether an intervention is less or more effective and less or more costly than the comparator. In particular, for an intervention to be cost-effective, the mean values would lie in the northeast quadrant indicating that the intervention is more costly but also more effective compared to other alternatives. On the other hand, if an intervention were less costly and less effective, mean values would mostly lie in the southwest quadrant. The dominance of the new intervention or the control would be shown on the northwest and southeast quadrant respectively [151].

The uncertainty around the cost-effectiveness results was also presented graphically by means of Cost-Effectiveness Acceptability Curves (CEAC), where the probability for each intervention of being cost-effective (on the y-axis) is plotted over a range of WTP thresholds (on the x-axis) [151, 168].

In addition, a series of one-way sensitivity analyses were carried out to assess the level of confidence and robustness of the findings. In the one-way sensitivity analyses, parameters thought to be key drivers are changed between their extreme values to evaluate the impact they have on the main conclusions of the cost-effectiveness analysis [151, 168]. The parameters selected were those that seemed to influence the most the conclusions of the assessed UK cost-effectiveness models. In particular, these were the risk of clinical events (tested at their lower and upper limits) and the cost of stroke and ICH incurred in the acute and chronic phase (increased and decreased by 20%). The risk of future clinical events, experienced as patients move across the model, may be determined by the previous event history; thus, the effect of previous events on future events was also tested.

Further, because stroke and ICH may have a substantial impact on quality of life of an individual, the long term disutilities associated with stroke and ICH were included in the one-way sensitivity analysis. Further, in the sensitivity analysis the NICE recommended threshold of £20,000 per QALY was changed to £12,936 to assess whether setting a threshold lower than the recommended may affect the probability of any DOAC or warfarin being cost-effective. The £12,936 threshold has been indicated by Claxton et al. (2015) as the most relevant for economic evaluation in a UK context [198]. The threshold was based on an analysis that evaluated the impact of the marginal increases and decreases in the NHS total expenditure for the year 2008 across 23 programme budget categories (such as infectious diseases, circulatory disease and cancers). The impact estimated was then linked to QALYs [198].

7.4 Results

7.4.1 Base-case analysis

Apixaban was found to be the most costly intervention over a patient's lifetime, but also yielded more QALYs than any other DOAC or warfarin (Table 7.9). When compared against warfarin, apixaban and dabigatran were cost-effective at the £20,000 threshold, with dabigatran yielding a greater INMB. Rivaroxaban was dominated by warfarin, being more effective and less costly. When all interventions were compared incrementally, ranked from least to most effective with associated total QALY gained, rivaroxaban was again dominated by warfarin, while dabigatran was shown to be more cost-effective than apixaban at the £20,000 threshold.

Table 7.9 Base-case deterministic analysis

Interventions	Total Cost (£)	Total QALY	Δ Cost (£)	Δ QALY	ICER	INMB
Compared to warfarin						
Warfarin	9,976	5.18				
Rivaroxaban	15,756	4.74	-5,780	-0.44	Dominated by warfarin	
Dabigatran	15,402	5.94	5,426	0.76	7,139	9,774
Apixaban	20,014	6.08	10,038	0.90	11,153	7,962
Incremental						
Rivaroxaban	15,756	4.74				
Warfarin	9,976	5.18	-5,780	0.44	Dominates rivaroxaban	
Dabigatran	15,402	5.94	5,426	0.76	7,139	9,774
Apixaban	20,014	6.08	4,612	0.14	32,943	-1,812

7.4.2 Base-case uncertainty

Probabilistic sensitivity analysis

The uncertainty in the differences in costs and health outcomes between treatment alternatives are presented graphically in the cost-effectiveness plane (Figure 7.3). For apixaban and dabigatran, the cost-effectiveness pairs reflecting the number of iterations run in the model mostly fall in the northeast quadrant. This suggests that apixaban and dabigatran may yield better health outcomes than warfarin; however, for rivaroxaban the iterations clearly fall into the northwest quadrant suggesting that rivaroxaban may be associated with poorer health outcomes than warfarin. Further, most of the uncertainty seems to come from uncertainty around QALYs.

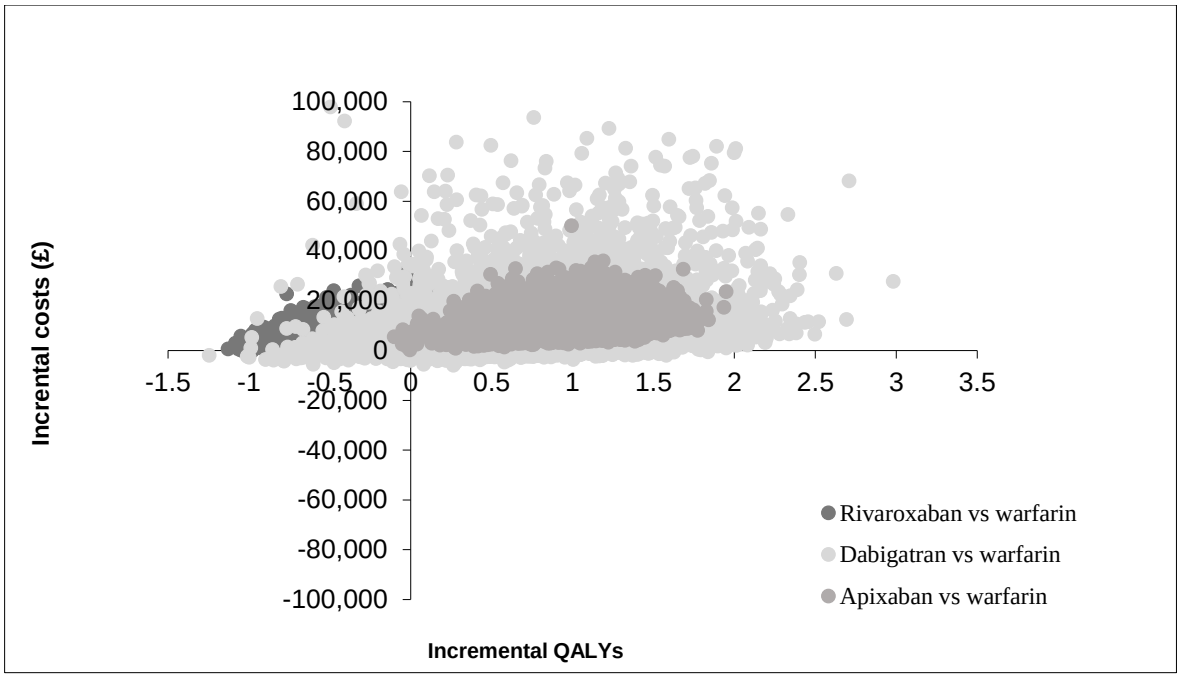


Figure 7.3 Cost-effectiveness plane base-case analysis

When assessing the uncertainty of DOACs of being cost-effective when compared against warfarin (Table 7.10), dabigatran showed a higher probability of being cost-effective (49.41 %) than apixaban (47.61%). Rivaroxaban being the least effective intervention had 0% probability of being cost-effective.

Table 7.10 Uncertainty of base-case analysis

Results (Probabilistic)	Intervention			
	Warfarin	Apixaban	Dabigatran	Rivaroxaban
£20,000 WTP				
Prob. cost-effect (%)	2.98	47.61	49.41	0.00
Prob. error (%)	97.02	53.29	50.59	100.00

The probabilities of DOACs being cost-effective compared to warfarin are presented graphically in Figure 7.4, where CEACs indicate how the probabilities of DOACs change according to a series of WTP. For instance, as shown in Figure 7.4, dabigatran and warfarin converge at £8,000 WTP, indicating that dabigatran is more cost-effective than warfarin only above a WTP of £8,000 per QALY. Similarly, apixaban has a higher probability of being cost-effective than warfarin only beyond a WTP of £11,000 per QALY.

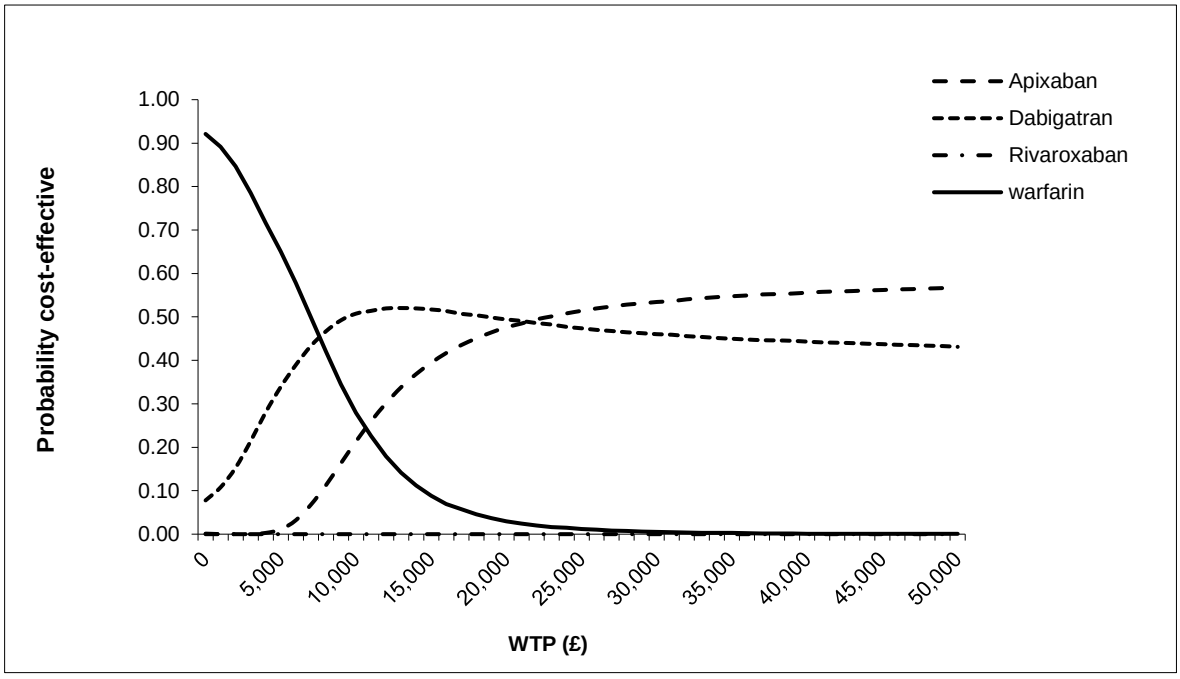


Figure 7.4 CEACs base-case analysis

One-way sensitivity analysis

As shown in Table 7.11, in the sensitivity analysis, assessing whether changing the threshold of £20,000 per QALY to £12,936 may affect the probability of any DOAC or warfarin being cost-effective, the probability of warfarin and dabigatran increased from 2.98% to 14.15% and from 49.41% to 52.04% respectively. By contrast, the probability of apixaban decreased from 47.61% to 33.81%. Rivaroxaban, dominated by warfarin, was still the least effective intervention with 0% probability of being cost-effective.

Table 7.11 Uncertainty sensitivity analysis

Results (Probabilistic)	Intervention			
	Warfarin	Apixaban	Dabigatran	Rivaroxaban
£12,936 WTP				
Prob. cost-effect (%)	14.15	33.81	52.04	0.00
Prob. error (%)	85.85	66.19	47.96	100.00

Apixaban

In the one-way sensitivity analysis for apixaban, the risk of ICH, major bleeding, and mortality were identified as the key clinical parameters influencing the ICER of £11,153 per QALY gained calculated in the base-case analysis. In particular, high uncertainty resulted by varying the risk of ICH (£6,123, £20,054), followed by major bleeding (£7,998, £17,044) and mortality (£9,619, £16,852). Although to a lesser extent, the risk of stroke (£10,532, £11,987) and MI (£10,338, £11,661) also had an impact on the base-case ICER. The risk of mortality due to previous events history was also found to influence the base-case findings: risk mortality due to previous bleeding (£9,901 and £12,298), risk of mortality due to ICH (£10,099, £12,119).

The cost of ICH had an impact on the base-case findings but only in the acute phase (£10,096, £12,105); indeed, the ICERs (£10,924, £11,227) resulting from the lifetime long-term cost of ICH in the chronic phase, diverged marginally from the ICER of the base-case analysis. All the other parameters tested had a negligible impact on the base-case ICER. The tornado diagram presenting the full one-way sensitivity analysis for apixaban is depicted in Figure 7.5.

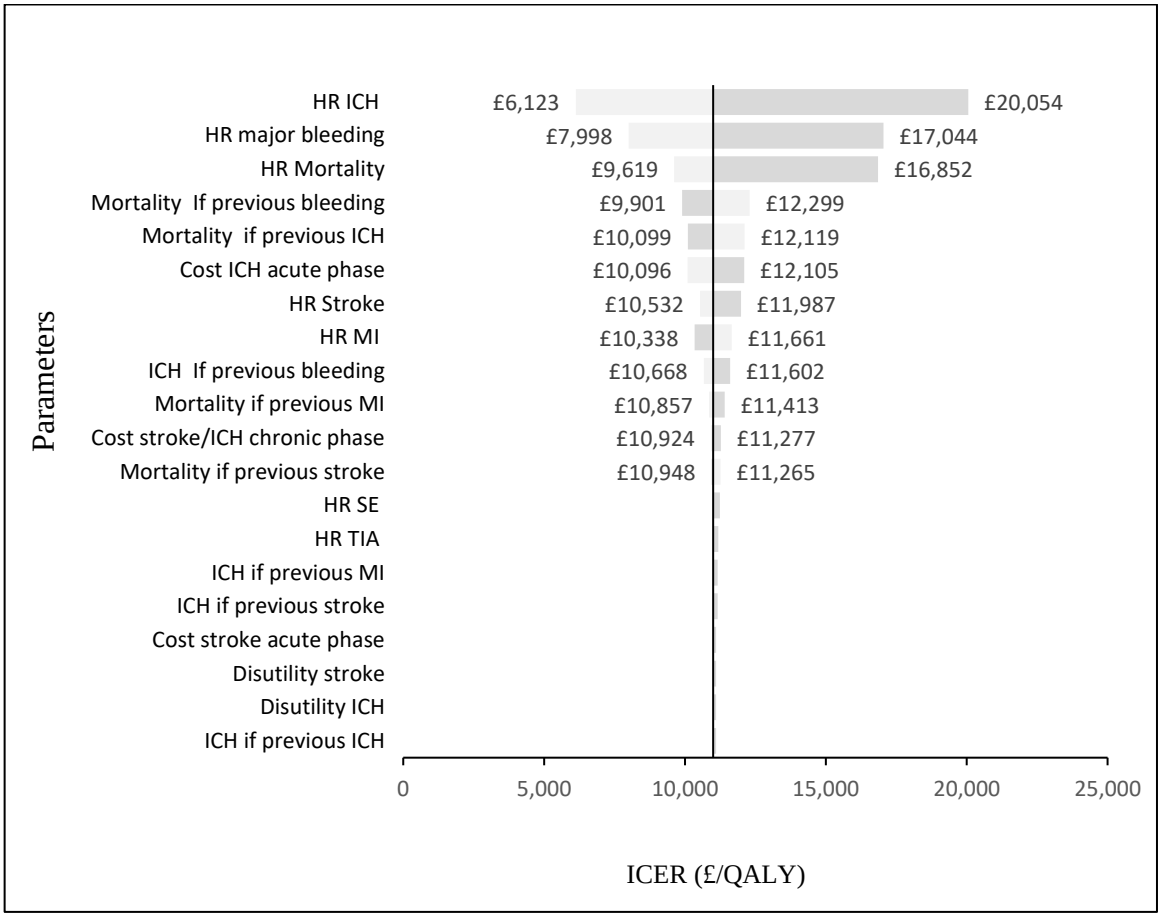


Figure 7.5 Apixaban vs. warfarin sensitivity analysis

Note: for HRs, the Lower and Upper 95% CI were tested. Cost and disutility parameters were tested with an increased and a decrease of 20% to their base-case values.
Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack,
ICH= Intracranial Haemorrhage, MI=myocardial.

Dabigatran

In the one-way sensitivity analysis for dabigatran, major bleeding, ICH and mortality a significant impact on the ICER of £7,139 per QALY gained calculated in the base-case analysis. In particular, high uncertainty resulted by varying the major bleeding (£3,398 £43,140), followed by ICH (£414, £39,463) and mortality (£-6,792, £6,333). Further, to a lesser extent, the risk of stroke (£6,224, £9,132) and MI (£5,424, £7,849) also had an impact on the base-case ICER. All the other parameters tested had a negligible impact on the base-case ICER. The tornado diagram presenting the full one-way sensitivity analysis for dabigatran is depicted in Figure 7.6.

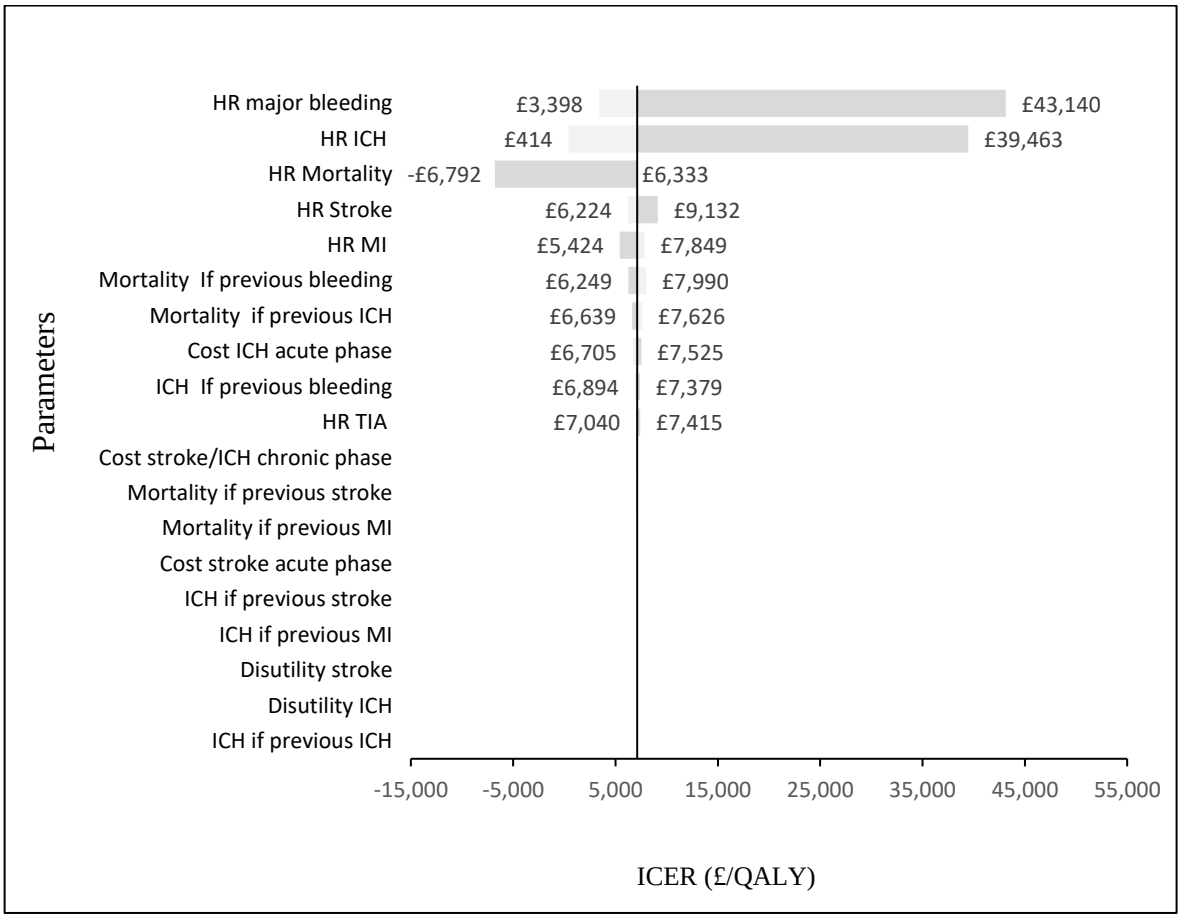


Figure 7.6 Dabigatran sensitivity analysis

Note: for HRs, the Lower and Upper 95% CI were tested. Cost and disutility parameters were tested with an increased and a decrease of 20% to their base-case values.
Abbreviations: TIA=transient ischaemic attack, ICH= Intracranial Haemorrhage, MI=myocardial.

7.4.3 Subgroup analysis

Deterministic analysis - standard dose

In the deterministic subgroup analysis assessing the cost-effectiveness of apixaban and rivaroxaban at their standard dose (Table 7.12), rivaroxaban was still dominated by warfarin being less costly and more effective. Standard dose apixaban was cost-effective when compared to warfarin (INMB £7,413) but not cost-effective in the incremental analysis (INMB £ -2,361) and at £20,000 threshold.

As pointed out, due to sample size, it was not possible to discern dabigatran into dose subgroups; therefore, in the subgroup analyses looking at different drug regimen the ICER and INMB for dabigatran were the same as the ones calculated in the base-case analysis indicating that dabigatran is cost-effective compared to warfarin.

Table 7.12 Deterministic subgroup analysis (standard dose)

Interventions	Total Cost (£)	Total QALY	Δ Cost (£)	Δ QALY	ICER	INMB
Compared to warfarin						
Warfarin	9,976	5.18				
Rivaroxaban	15,641	4.74	5,665	-0.51	Dominated by warfarin	
Dabigatran	15,402	5.94	5,426	0.76	7,139	9,774
Apixaban	20,563	6.08	10,587	0.90	11,763	7,413
Incremental						
Rivaroxaban	15,641	4.67				
warfarin	9,976	5.18	-5,665	0.51	Dominates rivaroxaban	
Dabigatran	15,402	5.94	5,426	0.76	7,139	9,774
Apixaban	20,563	6.08	5,161	0.14	36,864	-2,361

Probabilistic sensitivity analysis – standard dose

The uncertainty in the differences in costs and health outcomes between treatment alternatives at their standard dose are presented graphically in the cost-effectiveness plane (Figure 7.7). The iterations run in the model for the DOACs were distributed in a very similar fashion to the one observed in the base-case analysis; sitting on the northeast quadrant for apixaban and dabigatran, and sitting on the northwest quadrant for rivaroxaban. As mentioned, dabigatran could not be divided into subgroups, but it was included in Figure 7.7 for completeness.

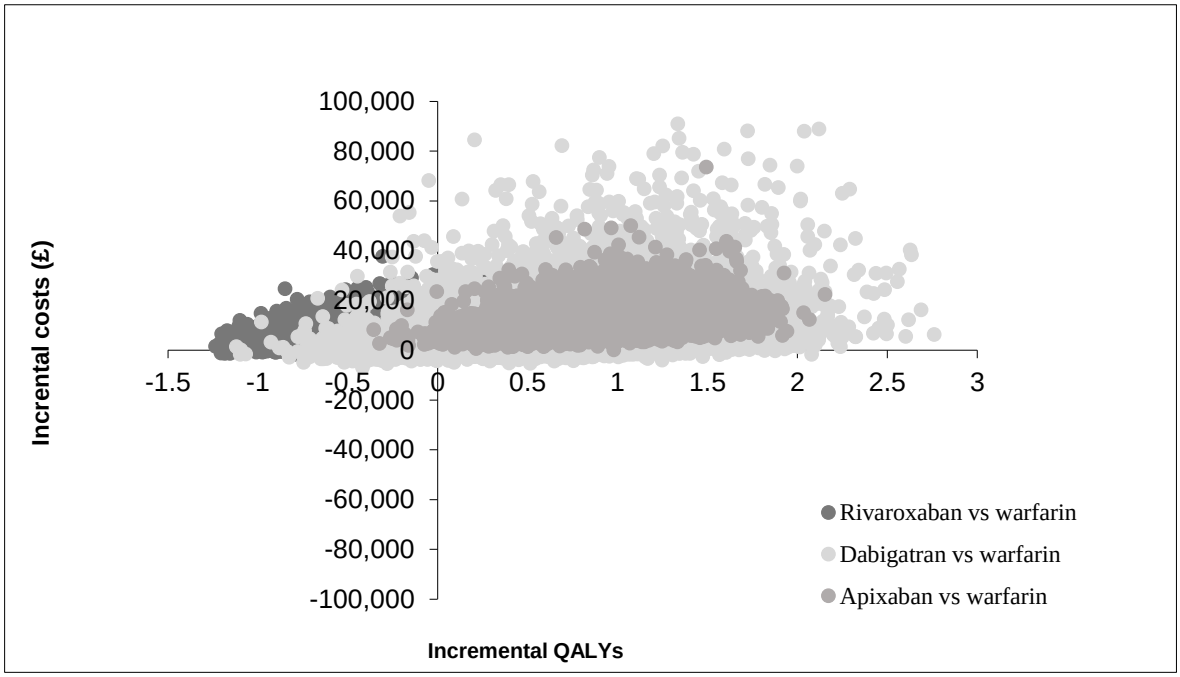


Figure 7.7 Cost-effectiveness plane subgroup analysis (standard dose)

When assessing the uncertainty of standard dose DOACs of being cost-effective against warfarin (Table 7.13), as in the base-case analysis, apixaban still showed a lower probability of being cost-effective (43.72%) than dabigatran (51.10%) at £20,000 threshold, and it is even less likely to be cost-effective at £12,936 threshold. Apixaban was found to be cost-effective at £12,936 threshold with a positive INMB of £1,113.

Rivaroxaban standard dose was found to be the least effective intervention with 0% probability of being cost-effective at £20,000 and £12,936 thresholds.

Table 7.13 Uncertainty subgroup analysis (standard dose)

Results (Probabilistic)	Intervention			
	Warfarin	Apixaban	Dabigatran	Rivaroxaban
£20,000 WTP				
Prob. cost-effect (%)	5.18	43.72	51.10	0.00
Prob. error (%)	94.82	56.28	48.90	100.00
£12,936 WTP				
Prob. cost-effect (%)	15.80	27.60	56.60	0.00
Prob. error (%)	84.20	72.40	43.40	100.00

The probability of DOACs (where apixaban and rivaroxaban are at their standard dose) being cost-effective compared to warfarin is presented graphically in Figure 7.8. The CEACs were very similar to the ones observed in the base-case analysis where apixaban was found to be cost-effective compared to warfarin only above a WTP of £11,000 per QALY. Similarly, dabigatran was shown to be cost-effective compared to warfarin only beyond a WTP of £7,000 per QALY.

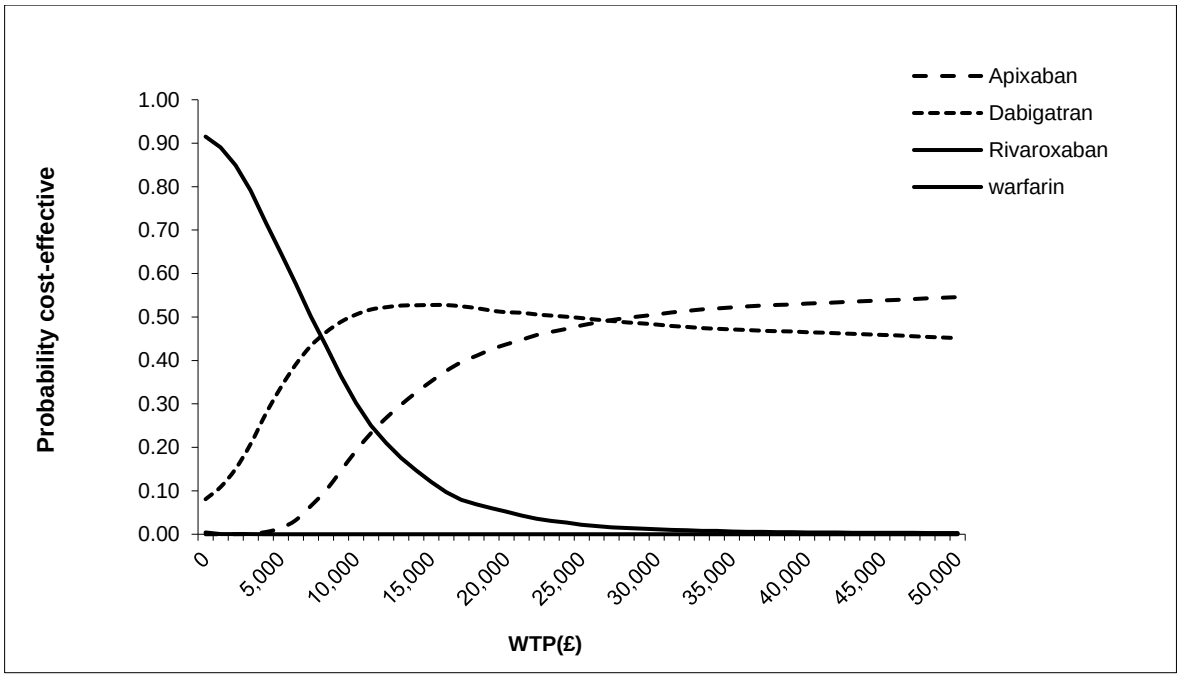


Figure 7.8 CEACs subgroup analysis (standard dose)

Deterministic analysis - reduced dose

In the deterministic subgroup analysis assessing the cost-effectiveness of reduced dose apixaban and rivaroxaban (Table 7.14), rivaroxaban was still dominated by warfarin being less costly and more effective. Standard dose apixaban was cost-effective when compared to warfarin (INMB £9,017) but not cost-effective in the incremental analysis (INMB £ -757) and at £20,000 threshold. As in standard dose analysis, dabigatran was included for completeness.

Table 7.14 Deterministic subgroup analysis (reduced dose)

Interventions	Total Cost (£)	Total QALY	Δ Cost (£)	Δ QALY	ICER	INMB
Compared to warfarin						
Warfarin	9,976	5.18				
Rivaroxaban	14,338	4.61	4,412	-0.57	Dominated by warfarin	
Dabigatran	15,402	5.94	5,426	0.76	7,139	9,774
Apixaban	20,159	6.14	10,183	0.96	10,607	9,017
Incremental						
Rivaroxaban	14,388	4.61				
warfarin	9,976	5.18	-4,412	0.57	Dominates rivaroxaban	
Dabigatran	15,402	5.94	5,426	0.76	7,139	9,774
Apixaban	20,159	6.14	4,757	0.20	23,785	-757

Probabilistic sensitivity analysis – reduced dose

The uncertainty in the differences in costs and health outcomes between treatment alternatives at their reduced dose are presented graphically in the cost-effectiveness plane (Figure 7.9). As in the base-case and subgroups analysis for standard dose apixaban and rivaroxaban, the scatter reflecting the number of iterations run in the model was clearly sitting in the northeast quadrant for apixaban and dabigatran (included for completeness). Comparatively, the iterations run for rivaroxaban are distributed more towards the northwest quadrant.

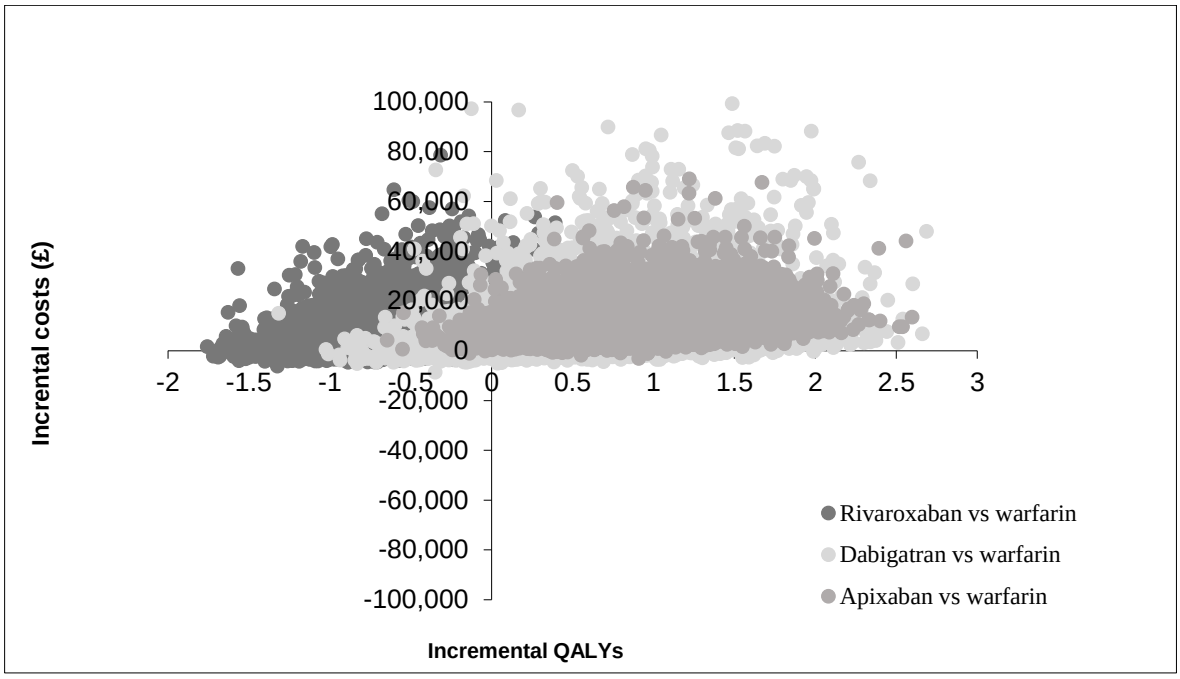


Figure 7.9 Cost-effectiveness plane subgroup analysis (reduced dose)

When assessing reduced dose apixaban and rivaroxaban uncertainty of being cost-effective against warfarin (Table 7.15), unlike in the base-case and the standard dose subgroup analysis, apixaban shows a higher probability of being cost-effective (47.46%) than dabigatran (47.00%) at £20,000 threshold. However, the probability of reduced dose apixaban of being cost-effective (37.50%) was even lower than the one observed for dabigatran at £12,936 WTP. Rivaroxaban reduced dose was found to be the least effective intervention with 0% probability of being cost-effective at £20,000 and £12,936 thresholds.

Table 7.15 Uncertainty subgroup analysis (reduced dose)

Results (Probabilistic)	Intervention			
	Warfarin	Apixaban	Dabigatran	Rivaroxaban
£20,000 WTP				
Prob. cost-effect (%)	5.47	47.46	47.00	0.07
Prob. error (%)	94.53	52.54	53.00	99.93
£12,936 WTP				
Prob. cost-effect (%)	16.00	37.50	46.10	0.40
Prob. error (%)	84.00	62.50	53.90	99.60

The probability of apixaban and rivaroxaban reduced dose being cost-effective compared to warfarin is presented graphically in Figure 7.10. As in previous analyses, apixaban and dabigatran were found to be cost-effective compared to warfarin only above a WTP of £11,000 and £7,000 per QALY respectively.

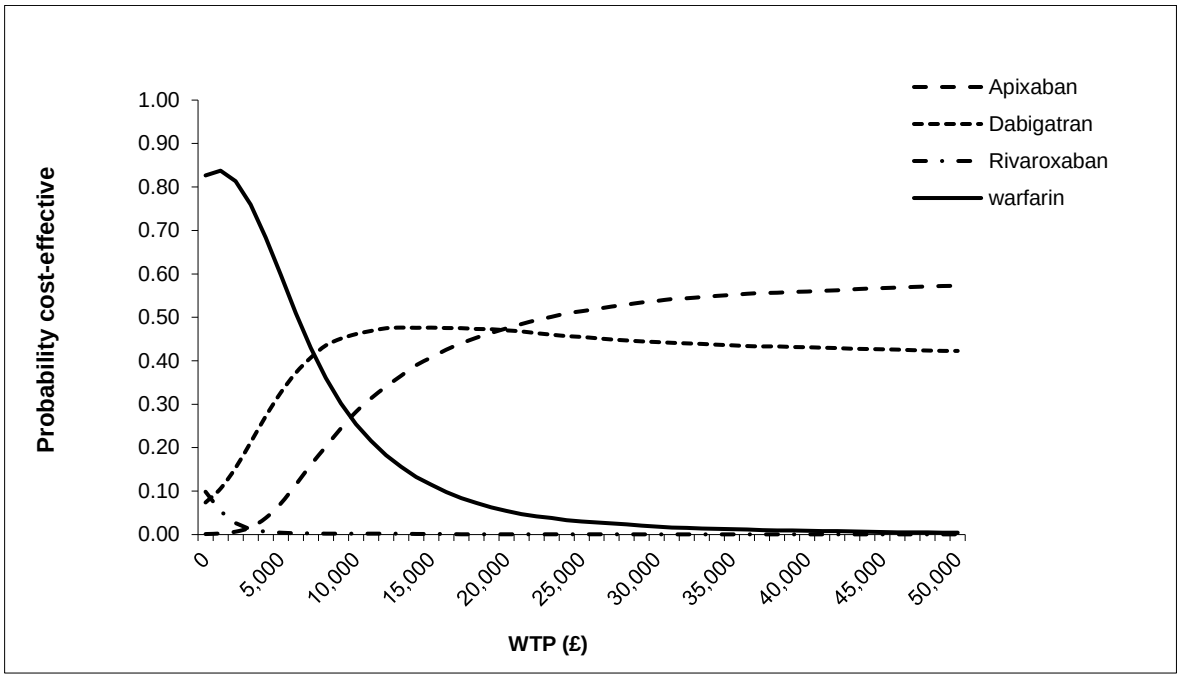


Figure 7.10 CEACs subgroup analysis (reduced dose)

Deterministic analysis - AF patients 70 years or older

In the deterministic subgroup analysis assessing the cost-effectiveness of DOACs in a cohort of AF patients 70 years or older (Table 7.16), apixaban was cost-effective compared to warfarin. Dabigatran and rivaroxaban were dominated by warfarin, being more effective and less costly. The same results were observed when all interventions were compared incrementally.

Table 7.16 Deterministic subgroup analysis (AF patients 70 years or older)

Interventions	Total Cost (£)	Total QALY	Δ Cost (£)	Δ QALY	ICER	INMB
Compared to warfarin						
Warfarin	8,760	4.18				
Rivaroxaban	10,739	3.45	1,979	-0.73	Dominated by warfarin	
Dabigatran	11,860	4.08	3,100	-0.10	Dominated by warfarin	
Apixaban	10,667	4.33	1,907	0.15	12,713	1,093
Incremental						
Rivaroxaban	10,739	3.45				
Dabigatran	11,860	4.08	1,121	0.63	1,779	
Warfarin	8,760	4.18	-3,100	0.10	Dominates dabigatran	
Apixaban	10,667	4.53	1,907	0.15	12,713	1,093

Probabilistic sensitivity analysis - AF patients 70 years or older

The uncertainty in the differences in costs and health outcomes between treatment alternatives in the AF patients 70 years or older are presented graphically in the cost-effectiveness plane (Figure 7.11). In this subgroup analysis, unlike base-case and previous subgroups analyses, the iterations run for apixaban and dabigatran did not fall into any specific quadrant, though the scatter for apixaban seems to be more distributed towards the northeast quadrant. Most of the uncertainty for apixaban and dabigatran seems to come from uncertainty around QALYs. The rivaroxaban scatter, falling in to the northwest quadrant, is hidden by the iterations run of dabigatran.

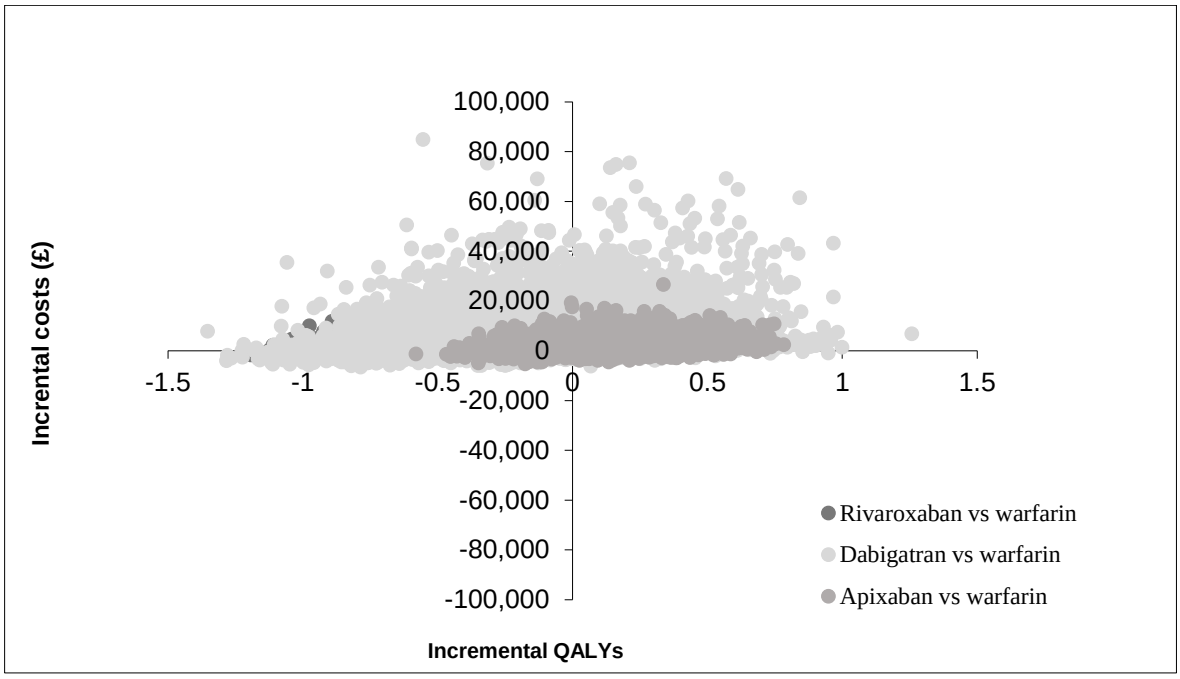


Figure 7.11 Cost-effectiveness plane subgroup analysis (AF patients 70 or older)

When assessing the uncertainty of DOACs being cost-effective against warfarin in the AF patients 70 years or older (Table 7.17), apixaban (55.30%) was shown to have a much higher probability of being cos-effective than dabigatran (12.04%) when the threshold was set at £20,000. At £12,936 WTP, apixaban was still more likely to have a higher probability of being cost-effective than any other DOAC or warfarin. As in the base-case and previous subgroup analyses, rivaroxaban was found to be the least effective intervention with 0% probability of being cost-effective at any threshold.

Table 7.17 Uncertainty subgroup analysis (AF patients 70 years or older)

Results (Probabilistic)	Intervention			
	Warfarin	Apixaban	Dabigatran	Rivaroxaban
£20,000 WTP				
Prob. cost-effect (%)	32.66	55.30	12.04	0.00
Prob. error (%)	67.34	44.70	87.96	100.00
£12,936 WTP				
Prob. cost-effect (%)	43.09	46.49	10.42	0.00
Prob. error (%)	56.91	53.51	89.58	100.00

The probabilities of DOACs being cost-effective against warfarin in the AF patients 70 years or older are presented graphically in Figure 7.12. Apixaban and warfarin CEACs converge at a WTP of about £12,000 indicating that apixaban is cost-effective at any WTP, with the WTP falling well below the £20,000 and £12,000 threshold.

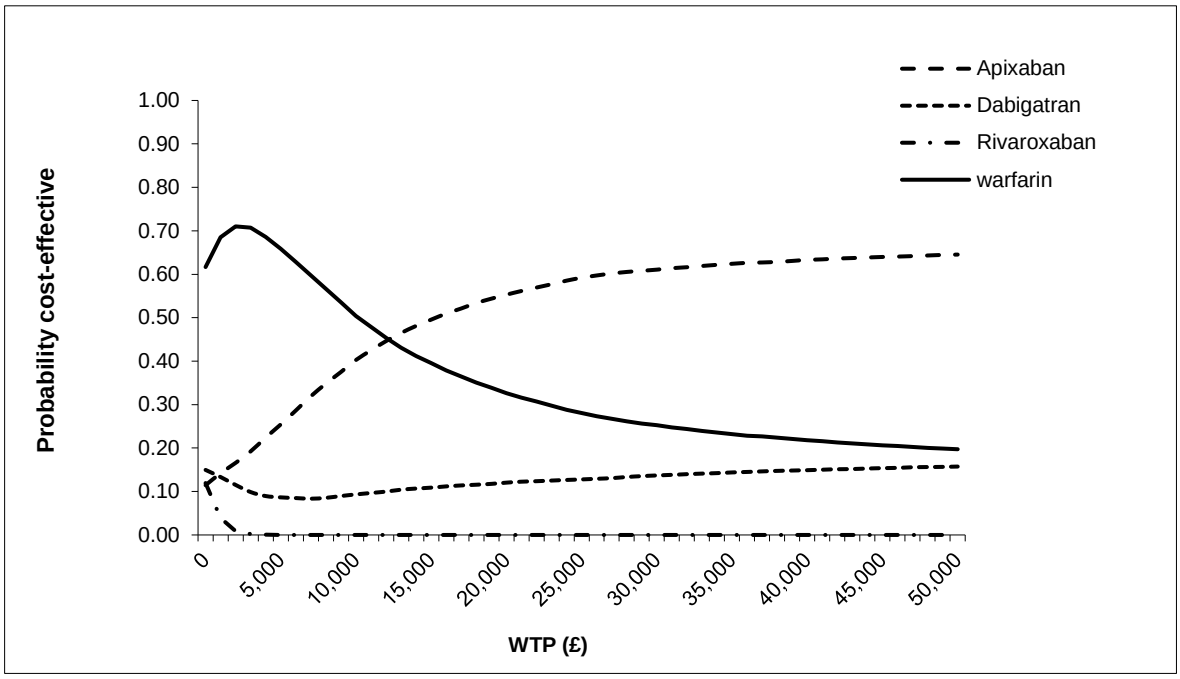


Figure 7.12 CEACs subgroup analysis (70 years or older AF patients)

7.5 Discussion

To date, economic evaluations, assessing the cost-effectiveness of DOACs in the AF population and adopting a UK perspective, were based on clinical efficacy taken from three pivotal RCTs: Aristotle (apixaban), RE-LY (dabigatran) and ROCKET-AF (rivaroxaban) [41-43]. RCTs provide robust evidence on efficacy measured under specific conditions, where randomisation is effective in minimising selection bias between two groups of individuals [59, 125]. Evidence from clinical practice offers the opportunity to investigate whether DOACs are cost-effective in real life under uncontrolled conditions. This type of evidence can complement cost-effectiveness conclusions generated with evidence coming from RCTs.

The findings from this analysis, to date the only UK economic evaluation using patient level data, are in line with existing evidence suggesting that apixaban and dabigatran are valid cost-effective options to warfarin for the prevention of stroke in the AF population [179-181]. However, the findings for the base-case analysis, based on the assumption discussed in Chapter 4 that prevalence and incidence of AF typically increase exponentially from 50 years onwards [89], were based on a cohort of patients starting anticoagulation at 50 years of age; thus including all age groups of patients potentially at risk of AF. This differs from the approach adopted in the other models assessed in this Chapter and the ones included in Limone's review (2013), where it was assumed that patients started anticoagulation at 70 years of age [178-182]. Nevertheless, this assumption was tested in the subgroup analysis and apixaban was still cost-effective compared to warfarin at the £20,000 threshold. However, for dabigatran, assumptions had to be made on the risk of SE, stroke and MI. In particular, given that only one event was observed for the mentioned outcomes, it was assumed that there was no risk difference between dabigatran and warfarin. Therefore, the results for the patients 70 years or older on dabigatran should be interpreted cautiously. While the same assumption was made for the risk of SE for the rivaroxaban group, the results obtained were comparable to those observed in other analyses.

According to the base-case findings, dabigatran may be indicated as the most favourable first line treatment having an ICER falling well below the £20,000 threshold, and a positive INMB greater than the one associated with apixaban compared to warfarin. Typically, the intervention associated with the greatest net benefit should be the preferred option even when less cost-effective than any other intervention [151, 156]. Similar findings were reported in Verhoef's study (2014), indicating dabigatran, still associated with substantial uncertainty, as the most cost-effective DOAC compared to warfarin in a UK setting [181].

Verhoef's study however, assumed the same level of disutilities between DOACs to account for switching [181]. This approach was criticised by the ERG considering the assumption of equivalent disutility not robust and likely to be against apixaban [37]. Nevertheless, in the base-case analysis, both ICERS for apixaban and dabigatran are well below the £20,000 threshold, and have nearly the same probability of being cost- effective compared to warfarin.

Unlike the existing evidence, rivaroxaban was not found to be cost-effective compared against warfarin. Nevertheless, in the studies including all DOACs and warfarin, rivaroxaban was dominated by either apixaban or dabigatran [179-182], thus, although cost-effective against warfarin, could not be recommended as first line treatment. The findings on rivaroxaban from this economic evaluation are mainly determined by the poor safety and effectiveness profile discussed in Chapter 6 assessing the comparative-effectiveness of DOACs against warfarin. While increased risk of stroke and ICH compared to warfarin certainly has a significant impact on long-term costs, the increased risk of mortality offset the total cost incurred in the long term as patients on rivaroxaban would die sooner than patients on other anticoagulation treatments. This would also cause rivaroxaban patients gaining fewer QALYs, thus resulting in not being cost-effective compared to warfarin or the other two DOACs.

Standard and reduced dose

In this subgroup analysis, the cost-effectiveness of DOACs was evaluated to establish in which subset of patients any of the treatments evaluated provides the greatest value for money. The difference in ICERs and uncertainty between DOAC standard and reduced doses suggests that apixaban is more cost-effective when administered in a reduced dose regimen, but high uncertainty was associated with the findings.

However, this still provides an indication on whether reduced dose apixaban is cost-effective in a real-world scenario. Previous studies have assessed the effectiveness of reduced dose apixaban in the UK [58, 62], but none evaluated the economic aspect.

As per clinical guidelines recommendation, reduced dose of DOACs should be offered to patients who are typically 80 years or older with one or more of the following: high risk of bleeding and moderate to severe renal or hepatic impairment [16-18, 35-37]. Therefore, this subgroup analysis may give an indication that apixaban is more cost-effective in the older rather than in the younger population with AF and potentially associated comorbidities. These findings are backed up by the comparative-effectiveness analysis in Chapter 6, where reduced dose apixaban and rivaroxaban show a better safety and effectiveness profile than their corresponding standard dose. While this is true for apixaban and rivaroxaban, it was not possible to discern, due to sample size, standard and reduced dose for dabigatran.

Sensitivity analysis

In the one-way sensitivity analyses, as in previous cost-effectiveness studies, the risk of ICH, major bleeding and mortality associated with DOACs was shown to influence the ICERs [178, 180-182]. This was particularly evident in the apixaban sensitivity analysis, where the ICERs estimated with lower and upper confidence interval of the risk of ICH (£6,123, £20,054), major bleeding (£7,998, £17,044) and mortality (£9,619, £16,852) diverged substantially from the ICER estimated in the base-case (£11,153). Similarly, in the dabigatran sensitivity analysis, the ICERs estimated with lower and upper confidence interval of the risk of major bleeding (£3,398, £43,140) ICH (£414, £39,463), and mortality (-£6,792, £6,333) diverged substantially from the ICER estimated in the base-case (£7,139).

Clearly, this difference is mirrored by the wide confidence intervals of fewer clinical events in a relatively small treatment arm, as in the case of dabigatran, and observed in the comparative-effectiveness analysis reported in Chapter 6.

For costs and disutilities, only costs were found to have a tangible impact on the base-case ICERs. In particular, the cost of ICH in the acute phase was shown to influence the base-case ICER of apixaban (£10,096, £12,119), and to a lesser extent the ICER of dabigatran (£6,705, £7,990). Existing evidence suggest that the INR monitoring costs has a significant impact on the cost-effectiveness of DOACs [173]. However, in this economic evaluation, the impact of INR monitoring cost was not tested as it was included in the cost of warfarin and assumed to vary significantly with a distribution ranging from -50% to +150% of the mean cost.

7.6 Limitations

This economic evaluation gives an indication on whether DOACs are valid cost-effective alternatives to warfarin for the prevention of stroke and associated comorbidities in real-world AF population. However, caution should be taken when interpreting the results as this study has some limitations and further research will be needed to reduce the level of uncertainty around the findings.

As previously pointed out, residual bias, sample size and the number of observed events may undermine the robustness of the findings, which in turn may translate into uncertain conclusions for the cost-effectiveness analysis. Specifically, uncertainty exists about the risk of mortality associated with rivaroxaban that may be caused by true association or bias from selective prescribing.

As elucidated in previous Chapters, methods to deal with selective prescribing and reduce confounding by indication exist, but as these issues can be mitigated and not entirely eliminated, there will always be a risk of residual selection bias. Thus, in the context of economic evaluation, observed and unobserved confounding represents a threat to the internal validity, as they might bias the incremental outcome, the incremental cost, or both, ultimately leading to incorrect decisions based on the estimated ICER. In addition, if dealing with unobserved confounding, it may be challenging to nest methods, such as instrumental variable and regression discontinuity, within a complex data structure of cost-effectiveness analysis [199].

Further, because of the relative small number of events observed in the dabigatran group, it was assumed that patients in the dabigatran and the warfarin group had the same risk of SE and ICH. Similarly, in the 70+ subgroup analysis, it was assumed that patients in the dabigatran and warfarin group had the same risk of SE and MI. These assumptions may underestimate or overestimate the effect of dabigatran. The assumption on the risk of SE is unlikely to bias the cost-effectiveness estimates as no difference in risk of SE between dabigatran and warfarin treatment group was observed in the RE-LY trial. The assumptions made on the risk of MI, however, would favour dabigatran over warfarin in the cost-effectiveness analysis; an increased risk of MI for dabigatran compared to the warfarin treatment group was reported in the RE-LY trial. On the other hand, the assumption made for the risk of ICH, would favour warfarin, as in the RE-LY trial a significant reduction of ICH was observed in the dabigatran treatment group [41]. Longer follow-up may reduce uncertainty, but as observed in the comparative-effectiveness analysis, much of the clinical events are experienced within the first two years from anticoagulation initiation.

Nevertheless, a longer follow-up would be useful for confirming the patterns of treatment switching having a substantial impact on final risk of experiencing clinical events. Another important limitation of the study is attributable to the nature of observational data and in particular to the associated bias and unmeasured residual confounding amply discussed in Chapter 5 [59, 118].

The lack of data on time spent in therapeutic INR range, not currently available in Scotland, is another potential limitation of this study, as existing evidence suggest that the INR range impact the cost-effectiveness of DOACs [181]. However, in RCTs, the effect of DOACs may have been overestimated, as the therapeutic INR between a range of 2.0 and 3.0, was not maintained for some patients randomised to warfarin [41-43]. The availability of data regarding the time spent within therapeutic range for patients on warfarin would allow for a more accurate comparative effectiveness analysis of DOACs versus warfarin in real life. A high INR range is associated with increased risk of thromboembolic and haemorrhagic events, thus INR, used in the HAS-BLED score, is important in determining the potential risk of bleeding. However, as described in Chapter 6, the HAS-BED score, a redefined version of HAS-BLED and not taking into account the labile INR item, can be used to estimate the risk of bleeding [84] and used to estimate the comparative effectiveness of DOACs versus warfarin. Therefore, the lack of data on time spent in therapeutic INR range is unlikely to have a meaningful impact on this economic evaluation. Similarly, data on serum creatinine level, to establish the level of severity for kidney impairment, are not currently available at the national level.

Arguably, economic evaluations should include societal perspective reflecting the impact that an intervention may have on the welfare of the whole society; thus, costs derived from productivity loss may be factored in to the cost-effectiveness equation. In the UK, however, economic evaluations assessing the relative efficiency of alternative interventions would adopt the NHS perspective [200].

7.7 Conclusions

This analysis, building on previous evidence from the comparative-effectiveness analysis, provides RWE that apixaban and dabigatran are at least as cost-effective as warfarin for the prevention of stroke and AF associated comorbidities. Apixaban seems to be more cost-effective than warfarin when administered as a reduced dose. Rivaroxaban, being the least effective intervention, is not cost-effective. However, while the findings for apixaban, coupled with those from the comparative-effectiveness analysis, may inform future research and an impact on future prescribing patterns, those concerning dabigatran and rivaroxaban should be interpreted cautiously.

Chapter 8 Summary of main findings, policy implications and future research

8.1 Main findings

The analyses undertaken in this thesis have explored the potentials and highlighted methodological issues associated with RWE in context of HTA. A major issue is that RWE may be inconsistently collected, and missing elements may lead to reduced statistically validity or may limit the potential for answering research questions. Sample size and the number of events are two other aspects that may challenge the use of RWE. In RCTs, the optimal number of events and participants needed to test hypothesis can be identified, whereas with real-world data it is not possible to determine the number of events need to obtain a robust statistical power. The biggest issue, however, is that RWE studies are subject to bias. While methods exist to control for bias, residual bias from unobserved factors may persist.

The initial AF cohort consisting of 278,286 patients with a known diagnosis of AF or atrial flutter was identified from hospitalisation records (SMR01) for the 1997 – 2015 study period, while the cohort of those on any OAC consisting of 160,405 patients was identified from the prescribing records (PIS) for the 2009 – 2015 study period. The use of patient level data including inpatient admissions, outpatient attendance, prescribing, care home and mortality records has allowed for the first time in Scotland to explore different methodologies concerning costing and comparative-effectiveness estimation to address a series of research questions. For instance, the level of granularity of the available patient level data has allowed for a more precise estimation of AF related health care costs compared to what has been previously done in Scotland.

The methodological aspects of using RWE were mostly explored in Chapter 5, where different methods for comparative-effectiveness analysis were tested and compared. The availability of patient level data has also allowed for conducting an economic evaluation that best reflects the epidemiology of AF and clinical practice in Scotland; clinical advice, prevalence and incidence suggest that AF increases exponentially from 50 years onwards [89]. Existing economic evaluations, using evidence from RCTs, have assumed that patients were 70 years of age when starting anticoagulation therapy [173, 178-182]. The economic evaluation carried out in this thesis was based on a cohort of patients 50 years and older; thus, including all age groups of patients potentially at risk of AF.
In the next paragraphs, the main findings followed by limitations and potential policy implications are discussed; in addition, areas of future research will be discussed.

8.1.1 The inpatient, outpatient, prescribing and care home costs associated with Atrial Fibrillation

Chapter 4 provided an overview of AF patient characteristics, including clinical and socio-economic factors, and aimed at meeting the following objective: estimate and examine the composition of direct and indirect medical costs in AF using Scottish linked health data.

In order to meet this objective, the existing evidence on cost of AF was reviewed to identify population, data, covariates, cost components and statistical methods used.
Being able to estimate costs using an incidence-based method using patient-level morbidity records, rather than using extrapolated rates with a prevalence-based approach, this thesis was able to improve on existing methods by offering more precise cost estimates.

This was also achieved by using a longer follow-up, a contemporary cohort of AF patients and the inclusion of a variety of healthcare and social care settings. The costs associated with hospital inpatient stay were identified as the cost component that contributed the most to the overall AF related healthcare costs. Most importantly, age was found to have a modest impact on overall healthcare costs, which were found to be fairly consistent across age groups; thus, reinforcing findings from previous studies suggesting that healthcare expenditure depends not only on patients' calendar age, but is also associated with remaining lifetime.

AF related costs incurred by females were higher than the costs observed in the male group. While there was little difference between the total costs and the distribution of costs for inpatient, outpatient and prescription costs, the difference seemed more pronounced when comparing the care home component of costs. AF patients from large urban area showed the highest estimated cost, while patients living in "very remote rural areas" incurred the lowest estimated cost. Further, patients from the most deprived areas seemed to incur a higher level of AF related costs compared to those living in the least deprived areas. A clear pattern of estimated costs decreasing from the most to the least deprived areas was observed.

This Chapter highlighted the importance of including all available cost components for establishing overall costs, as these often extend beyond hospitalisation.

8.1.2 Methods for comparative-effectiveness analysis

Chapter 5 provided an overview of different methods generally used in comparative-effectiveness analysis to estimate the ATE, and highlighted the typical challenges in addressing observed and unobserved confounding inherent to the nature of RWE. This chapter aimed at meeting the following objective: explore, with a focus on PS based methods and the DOACs case study, methodological challenges in using RWE to estimate comparative-effectiveness analysis.

In order to meet this objective, PS base methods including PS matching, IPW (IPTW, and IPTW combined with IPCW) and PS used as a covariate were tested. As in other RWE studies assessing the comparative effectiveness of DOACs the focus was on observed confounding, testing for unobserved confounding was beyond the scope of this Thesis, and therefore the assumption of 'no unobserved confounding' was made [114].

The PS estimation was carried out for each individual DOAC (apixaban, dabigatran and rivaroxaban), warfarin and the combined DOAC cohorts, resulting in five different PS models. For each model, CHA_2DS_2-VASc and HAS-BLED risk scores were calculated.

In this study, PS covariate adjustment was found to be the most robust method, PS matching and IPW methods also performed well, but were excluded to avoid further sample size reduction. While PS covariate adjustment is in theory more sensitive to distributional assumptions and PS specification [123, 128, 129], it was found to be the most robust method compared to PS matching and different variations of IPW (including IPTW, trimmed IPTW and IPTW combined with IPCW), as covariates showed the lowest standardised difference.

However, as results from this analysis suggest, the choice of the most adequate PS method depends on the characteristics of the available data; thus, the use of a single best method for reducing bias due to confounding should be avoided and several candidate models should be tested.

8.1.3 Comparative-effectiveness analysis

Chapter 6 assessed comparative-effectiveness and safety of DOACs and provided additional evidence to understand how effective DOACs are in Scottish clinical practice. Prior to conducting the comparative-effectiveness analysis, clinical outcomes were identified according to ICD-10 and OPCS-4 codes from SMR01 records; then CHA_2DS_2-VASc and HAS-BLED risk scores were calculated. In addition, an assessment of the proportionality assumption for the Cox proportional hazards model was carried out, and the identified violation of proportionality was addressed with the use of time varying covariates.

In the comparative-effectiveness analysis, the risk difference for stroke between apixaban and warfarin, although reduced, was not statistically significant; this differs from the risk reduction observed in the ARISTOLTE trial. In one observational study, Nielsen et al (2017) distinguished between doses [61] and found no risk differences for stroke between patients on apixaban 2.5 mg reduced dose and those on warfarin. Therefore, the risk of stroke reported in the analysis in Chapter 6 seems to be much closer to the one observed in clinical practice than the one reported in the ARISTOTLE study. In case of conflicting evidence, clinicians should balance that evidence and assess whether RCTs and RWE studies look at the same population, patients characteristics and outcomes.
In particular, clinicians may be interested in aspects not investigated in RCTs such as treatment interactions with concomitant treatments, and longer follow-up [2, 7].

For what concerns DOACs, clinicians may be interested in the effect of standard and reduced dose apixaban and rivaroxaban compared to warfarin not investigated in the pivotal RCTs [41-43]. Thus, although residual confounding will always be a threat to the validity of RWE, the finding on apixaban supported by existing studies [58, 61, 62] can be trusted.

Further, the comparative-effectiveness analysis confirmed the findings for the risk of GI bleeding reported in the ARISTOTLE clinical study [42] as no risk difference was found between apixaban and warfarin. However, the risk of GI bleeding differed from the one reported in another observational study (Ntaios (2017)) where a risk reduction was found for patients on apixaban [58]. When assessing the effectiveness of apixaban at standard and reduced dose, no risk differences were observed across clinical outcomes between apixaban and warfarin regardless of the dose administered.

No difference in risk was observed for most of the clinical outcomes when comparing efficacy and safety profiles of dabigatran versus warfarin. The estimated risk of stroke is in line with the risk observed in the RE-LY trial for patients receiving dabigatran at a reduced dose. Similarly, risk of GI bleeding observed in the RE-LY trial for patients treated with dabigatran 150 mg compared to those on warfarin, were confirmed in the comparative-effectiveness analysis and in a meta-analysis of observational studies [58]. Although the estimated difference in risk for most of the clinical outcomes for the dabigatran and warfarin comparisons was not statistically significant, the relatively small sample size and low number of events observed may undermine the robustness of the findings.

No difference in risk was observed for stroke and stroke composites between rivaroxaban and warfarin patients; those findings are comparable to those reported in the ROCKET-AF trial [43] and in an observational study Gorst-Rasmussen (2016) [60]. When assessing DOACs at different doses, an increased risk of GI bleeding was observed for patients treated with reduced dose rivaroxaban, but no differences were observed for patients receiving the standard dose compared to warfarin. While findings on the risk of mortality associated with reduced dose rivaroxaban are comparable to the risk reported in the ROCKET-AF trial [43], the increased risk of mortality observed for patients receiving the standard dose matches results reported by Nielsen (2017) [61] and Gorst-Rasmussen (2016) [60]. However, in these two observational studies, the increased risk of all-cause mortality was found for patients receiving rivaroxaban 15 mg reduced dose [60, 61].

This study provides real-world evidence that apixaban is as effective as warfarin for the prevention of stroke and AF associated comorbidities; the evidence concerning dabigatran and rivaroxaban should be interpreted cautiously. Although the cost-effectiveness estimates for dabigatran are not robust, due to limited number of events observed, they may still give an indication of whether dabigatran is cost-effective in preventing stroke and other AF related comorbidities in a real-world setting. For rivaroxaban, further research is needed to assess whether the association between the use of rivaroxaban and the risk of bleeding and mortality is a true association.

8.1.4 Economic evaluation

Chapter 7 focused on the economic evaluation of DOACs using RWE of effectiveness, and aimed at meeting the following objective: Update, with RWE data, existing cost-effectiveness analysis of DOACs for the prevention of stroke in patients with AF in Scotland.

In order to meet this objective, a review of existing economic models was undertaken to identify the most appropriate model in the UK context. Then, the Sterne model (2017), identified as the most suitable model and accounting for limitations identified in previous models, was updated with RWE. The findings from this analysis are in line with existing evidence suggesting that apixaban and dabigatran are cost-effective compared to warfarin for the prevention of stroke in the AF population [179-181]. However, apixaban and dabigatran were found to be cost-effective when patients entered the model at 50 years of age, thus including all patients potentially at risk of AF. Dabigatran, yielding a greater INMB than apixaban may be indicated as the most favourable first line treatment; similar findings were reported by Verhoef (2014), indicating dabigatran as the most cost-effective DOAC compared to warfarin in a UK setting [181]. Nevertheless, the probability of apixaban and dabigatran being cost-effective compared to warfarin was negligible. Further, this probability was more marked when assessing the cost-effectiveness of interventions at £12,936 threshold, where apixaban and dabigatran had a probability of about 30% and 50%, respectively, of being cost-effective. Rivaroxaban (at any dose) was not found to be cost-effective compared to warfarin; while existing evidence suggest the contrary [179-182], in the existing studies rivaroxaban is dominated by either apixaban or dabigatran.

8.2 Limitations

In Chapter 4, the cost analysis was to some extent limited by missing records or incomplete data, in addition to potential risk of clinical miscoding in the morbidity records. For instance, the poor level of completeness of CHI numbers or other patient identifiers still makes linking SMR00 with other datasets rather challenging. In addition, prescribing and care home data were only available from 2009 and 2012, so that their contribution to overall AF related costs might be underestimated.

Propensity score methods are powerful tools for addressing confounding by indication typical in newly marketed prescription drugs [145]. However, even after the application of these methods, residual bias due to observed confounders may persist. Propensity Score methods may reduce this bias due to observable confounders; however, other unmeasured or residual confounding, such as patients' tolerability, access to healthcare and selective prescribing may still bias the estimates. Therefore, the findings of the comparative-effectiveness-analysis (Chapter 6) and the economic evaluation (Chapter 7) should be interpreted in the light of this.

The relative small sample size of dabigatran, coupled with a few observed events of stroke, the composite of stroke including SE and TIA, and mortality due to stroke, was an important limitation of this Thesis. Although the findings concerning dabigatran in Chapter 6 are not robust, due to the limited number of events observed, they may still give an indication of the effect of dabigatran in preventing stroke and other AF related comorbidities in a real-world setting.

The relative small sample size and the relative small number of events observed in the dabigatran group, was also a major issue for the economic evaluation in Chapter 7, as several assumptions had to be made. In particular, it was assumed that patients in the dabigatran and the warfarin group had the same risk of SE and ICH. Similarly, in the 70+ subgroup analysis, it was assumed that patients in the dabigatran and warfarin group had the same risk of SE and MI. As explained in the limitation subsection of the economic evaluation in Chapter 7, these assumptions may underestimate or overestimate the effect of dabigatran.

The complexity of conducting analyses in the National-Safe-Haven platform was a further limitation of this study. While the National-Safe-Haven offers a high-powered computing service, secure analytic environment and secure file transfer [15], merging and manipulating datasets may be problematic due to data space restrictions. For this reason, it was not possible to create a single master file, where all the observations of the cohorts for each analysis were included. With more data space available, creating a single master file would be the first step towards a more efficient data manipulation.

8.3 Policy implications

Real-world evidence is increasingly important for informing healthcare policy issues; and the availability of government data, coupled with bespoke centres such as the Health Data Research UK, has the potential to answer various policy questions. This thesis has shown the potential of Scottish administrative data to generate RWE in the context of HTA, in general and with reference to AF and the DOACs case study.

Real-world evidence is increasingly used in the reimbursement process, as HTA agencies may request additional evidence to enhance the understanding of elements of uncertainty. In this context, RWE, providing a more comprehensive understanding of treatment safety and effectiveness, could be used for the reassessment of prescription drugs, for instance every 2 or 3 years, to warrant post-approval reimbursement. HTA decisions are often based on immature and limited RCT data, and therefore should be regularly revisited. Currently there is no formal process in place; however, this Thesis has demonstrated that Scottish administrative data can be used systematically to set up a framework where RWE can be used for the reassessment of prescription drugs by SMC.

Despite in present work sample size and the relative small number of events, particularly for dabigatran, were a major concern for the validity of the findings, more data will be available, and therefore more robust conclusions can be drawn. Thus, this Thesis can recommend the use of Scottish administrative data to generate RWE for the re-evaluation of HTA decisions.

Atrial fibrillation has significant implications on health care costs, putting additional financial burden on the NHS. Within a broader societal perspective, AF may also have a significant impact on productivity loss. Rohrbacker (2010) has demonstrated that benefit, costs, number of absence days due to sick leave and short-term disability were significant for individuals affected by AF or cardiac arrhythmias [84]. In particular, as shown in this thesis, different cost components contribute to overall AF related healthcare costs, with inpatient admission being the main driver. This is a pertinent finding that may well support future policies on opportunistic screening in the population at risk of AF, and especially in Scotland where 1 in 3 patients with AF are currently undiagnosed [25]. For instance, the European AF management guidelines and the Scottish Cross-Party Group ‘Heart Disease and Stroke’, has identified AF as a priority area and recommended that people who are 65 years or older and at risk of AF and associated comorbidities should be screened opportunistically in primary care, pharmacies or community settings. The group highlighted the need for rigorous screening in order to reduce inpatient admission and associated costs [25, 109].

This thesis provides additional evidence on the comparative and cost-effectiveness of DOACs, which would be of great value to the decision makers. In particular, the RWE from this study suggests that apixaban is at least as effective as warfarin for the prevention of stroke in the AF population; and cost-effective for patients who start anticoagulation treatment at the age of 50. Although in Scotland rivaroxaban is no longer recommended for the prevention of stroke in the AF population, the increased risk of bleeding and mortality associated with rivaroxaban, may raise concerns over its use where rivaroxaban is recommended for other indications or target populations. Therefore, the systematic used of administrative data can be used to identify issues that have the potential of changing the prescribing framework. For instance, based on the results from the Cardiovascular Outcomes for People Using Anti-coagulation Strategies (COMPASS) study [201], reduced dose rivaroxaban used in conjunction with aspirin is now recommended for restricted use for the prevention of atherothrombotic events in adults at high risk of ischaemic events [202].

8.4 Future research

Real-world evidence is increasingly used in HTA and is expected to have a growing importance in the context of the value-based pricing approach adopted by the UK government. The rationale for adopting a value-based pricing approach is that NICE appraisal may not adequately capture the full value of an intervention, where aspects such as innovation and social benefit are not factored in. For instance, RWE may allow for better estimates of healthcare costs and utilisation, thus increasing transparency in health outcomes. This, in turn, would let payers and manufacturers to assess performance on individual basis and pay only for good outcomes for drugs being used in clinical practice [203].

An important aspect of the value-based pricing is the pharmaceutical price regulation scheme, a voluntary agreement between the department of health of UK government and the pharmaceutical industry. The scheme allows companies to set the price for new branded drugs, thus incentivising innovation, but put a limit on the profit that companies can make from the NHS. Following expiry of the voluntary scheme in 2014, it has now been renegotiated coming into force on the 1th of January 2019 [203].

Future research, where the role of RWE in the context of HTA is assessed, should perhaps have a more defined value-based pricing approach. Using the DOACs case study, additional value of this class of drugs could be shown in a number of ways. Firstly, supplementary cost-effectiveness analysis should be carried out with additional evidence coming from the DOAC to DOAC non-randomised comparison, to establish which intervention is more cost-effective in clinical practice not just in relation to warfarin. This evidence is available for Scotland and reported in a recent study assessing the effectiveness of DOACs using the same patient level data utilised in this thesis [38]. Future research including this additional evidence would reduce the uncertainty regarding the cost-effectiveness of DOACs in the real-world AF population. Secondly, in order to include a wider societal perspective, future research may be able to include indirect costs associated with productivity-loss by, for example, linking morbidity and prescribing data to national data from the Department for Work and Pensions.

Further, given the advancement of pharmacogenomics in the medical field, up to date evidence may also include safety and efficacy data obtained with a pharmacogenomics approach as this would allow for a more defined value of an intervention.

Pharmacogenomics is a relatively new branch of pharmacology that studies drug response by correlating gene expression with the drug's efficacy and toxicity, in order to develop safer medicines "personalized" to a person genetic makeup [204]. While, to date, there is no strong evidence supporting the use of pharmacogenomics testing in improving safety and efficacy of DOACs, the genes of interest have been identified and future research will be needed to quantify the added value of testing. As recognised in a recent clinical review of the pharmacogenomics of DOACs, this will pose some fundamental policy questions on whether budgets should be allocated for the purchase on new oral anticoagulants, or improving current care with warfarin by implementing a pharmacogenetic-guided dosing system [205].

Additional comparative and cost-effectiveness analyses should also include evidence from Edoxaban. This DOAC was approved by SMC in 2015 for the prevention of stroke in the AF population [39], and in the near future RWE for Scotland should be available.

8.5 Conclusions

The body of work of this thesis contributes to defining the role of RWE in the context of HTA in a number of ways, but also acknowledges the challenges of using RWE and highlights its complementary role with RCTs in generating solid evidence for decision-making. This thesis has also shown how these challenges could be addressed, providing a case study for existing guidelines designed to guide on the use of observational data in comparative-effectiveness studies [114, 150]. Most importantly, this thesis has provided additional evidence concerning AF related health care costs, comparative-effectiveness and cost-effectiveness of DOACs that could potentially have relevant policy implications.

Appendices

Appendix I: Econometric modelling test to identify family and link function

Family distribution and link function	Log link		Identity link	
	Chi2	P-value	Chi2	P-value
Family				
Gamma	5.12	0.024	5.80	0.016
Poisson	45.44	0.000	45.77	0.000
Gaussian NLLS	126.91	0.000	134.17	0.000
Inverse Gaussian or Wald	247.88	0.000	254.09	0.000
Link				
Pearson Correlation Test		0.445		0.419
Pregibon Link Test		0.384		0.392
Modified Hosmer and Lemeshow		0.032		0.015

Appendix II: Regression interactions and Odd Ratios conversion

Table II-1 Regression interactions

Covariates	Probability (1st modelling part)		Probability (2nd modelling part)	
	Coefficient (95%CI)	Std. Err	Coefficient (95%CI)	Std. Err
Interaction: age (year) - Charlson score (1 comorbidity)				
50-54	Reference			
55-59	-0.260 (-0.408, -0.111)	0.076	-0.073 (-0.173, 0.028)	0.051
60-64	-0.365 (-0.511, -0.219)	0.074	-0.108 (-0.204, -0.013)	0.049
65-69	-0.400 (-0.541, -0.260)	0.072	-0.131 (-0.224, -0.038)	0.048
70-74	-0.425 (-0.563, -0.287)	0.070	-0.184 (-0.276, -0.093)	0.047
75-79	-0.500 (-0.637, -0.363)	0.070	-0.199 (-0.288, -0.109)	0.046
80-84	-0.550 (-0.687, -0.412)	0.070	-0.203 (-0.295, -0.110)	0.047
85-89	-0.595 (-0.738, -0.453)	0.073	-0.243 (-0.341, -0.145)	0.050
90-max	-0.697 (-0.852, -0.543)	0.079	-0.268 (-0.383, -0.153)	0.059
Interaction: age (year) - Charlson score (>1 comorbidities)				
50-54	Reference			
55-59	-0.176 (-0.385, 0.034)	0.107	-0.016 (-0.120, 0.088)	0.053
60-64	-0.188 (-0.390, 0.015)	0.103	-0.121 (-0.218, -0.024)	0.050
65-69	-0.257 (-0.453, -0.061)	0.100	-0.230 (-0.325, -0.135)	0.048
70-74	-0.385 (-0.578, -0.193)	0.098	-0.345 (-0.436, -0.254)	0.047
75-79	-0.495 (-0.686, -0.304)	0.097	-0.432 (-0.522, -0.342)	0.046
80-84	-0.622 (-0.813, -0.430)	0.098	-0.483 (-0.576, -0.390)	0.047
85-89	-0.758 (-0.953, -0.563)	0.100	-0.550 (-0.647, -0.452)	0.050
90-max	-0.890 (-1.096, -0.685)	0.105	-0.712 (-0.825, -0.599)	0.058

Table II-2 Regression coefficients conversion to Odd Ratios of healthcare resources utilisation

Covariates	OR (95%CI)
Age group (years)	
50-54	Reference
55-59	1.35 (1.25 -1.45)
60-64	1.53 (1.42 -1.65)
65-69	1.65 (1.53 -1.77)
70-74	1.72 (1.60 -1.85)
75-79	1.92 (1.78 -2.06)
80-84	1.76 (1.64 -1.90)
85-89	1.67 (1.54 -1.80)
90-max	1.33 (1.22 -1.45)
Sex	
Male	Reference
Female	1.02 (1.00 -1.04)
Date of admission	1.23 (1.17 -1.28)
Follow-up years	
1	Reference
2	0.20 (0.19 -0.20)
3	0.16 (0.15 -0.16)
4	0.14 (0.14 -0.14)
5	0.13 (0.12 -0.13)
SIMD	
1	Reference
2	1.02 (0.99 -1.05)
3	0.97 (0.94 -1.01)
4	0.94 (0.91 -0.98)
5	0.94 (0.91 -0.97)

Note: the probability of using healthcare resources increases up to 80 years. No differences in resource utilisation is observed between males and females. In the 5 years post AF incident event, a small decrement in resource utilisation is observed over time. A marginal increase is observed for patients living in the most deprived areas compared with patients living in areas with the lowest level of deprivation.

Table II-2 Regression coefficients conversion to Odd Ratios of healthcare resources utilisation (continued)

Covariates	OR (95%CI)
Geography	
Large urban	Reference
Other urban	0.84 (0.82 -0.87)
Accessible small towns	0.82 (0.79 -0.85)
Accessible rural	0.78 (0.75 -0.81)
Remote small towns	0.85 (0.79 -0.90)
Remote rural	0.70 (0.66 -0.74)
Very remote small towns	0.65 (0.59 -0.71)
Very remote rural	0.66 (0.61 -0.71)
Health boards	
Great Glasgow and Clyde	Reference
Lothian	0.95 (0.91 -0.99)
Lanarkshire	1.01 (0.97 -1.05)
Ayrshire and Arran	0.65 (0.62 -0.68)
Grampian	1.02 (0.98 -1.07)
Tayside	0.61 (0.58 -0.63)
Fife	0.91 (0.86 -0.96)
Highland	0.82 (0.77 -0.87)
Forth Valley	0.58 (0.55 -0.61)
Dumfries and Galloway	0.71 (0.67 -0.76)
Borders	0.56 (0.53 -0.60)
Western Isles	0.27 (0.24 -0.31)
Orkney	0.61 (0.53 -0.72)
Shetland	0.51 (0.44 -0.59)
Comorbidity	
no comorbidities	Reference
1 comorbidity	1.74 (1.53 -1.97)
>1 comorbidities	2.20 (1.83 -2.64)

Note: the use of health or social care services decreases significantly in areas other than large urban. Aside from Lanarkshire and Grampian, patients are less likely to utilise health or social care services than patients in Great Glasgow and Clyde. Patients with comorbidities are more likely to make use of health or social care services than those with no comorbidities.

Appendix III: Conditional standardised difference

For a continuous covariate, the conditional standardised difference is obtained in the following steps:

Step 1 - the continuous baseline covariate, denoted by X, is calculated according to Equation III-1[127]$X = \alpha_0 + \alpha_1 T + \alpha_2 Z + \alpha_3 T \times Z + \varepsilon, \varepsilon \sim N(0, \sigma^2$

Equation (III-1)

Where Z indicates the estimated propensity score and T indicates treatment status (T=1 indicates treated, while T=0 indicates untreated)

Step 2 – for each subject in the sample, as shown in Equation III-2 [127], the following quantity is estimated

$$\frac{\hat{\alpha}_1 + \hat{\alpha}_3 Z}{\hat{\sigma}}$$

Equation (III-2)

Step 3 – The estimated X (Equation III-1) is then used to estimate the standardised difference as shown in Equation III-3 [127]

$$d = \frac{(\bar{x}_{treatment} - \bar{x}_{control})}{\sqrt{((s^2_{treatment} + s^2_{control})/2)}}$$

Equation (III-3)

Where $\bar{x}$ *treatment* and $\bar{x}$ *control* are the sample mean of the treated and untreated group respectively; S^2*treatment* and S^2*control* are the sample variance of the treated and untreated group respectively.

For a continuous covariate, the conditional standardised difference is obtained in the following steps:

Step 1 - the continuous baseline covariate, denoted by X, is calculated according to Equation III-1

$$logit\ (\Pr(x = 1)) = \alpha_0 + \alpha_1 T + \alpha_2 Z + \alpha_3 T \times Z$$

Equation (III-4)

Where Z indicates the estimated propensity score and T indicates treatment status (T=1 indicates treated, while T=0 indicates untreated)

Step 2 – for each subject in the sample, as shown in Equation III-2 [127], the following quantity is estimated

$$\frac{\hat{\alpha}_1 + \hat{\alpha}_3 Z}{\hat{\sigma}}$$

Equation (III-5)

Step 3 - The estimated predicted probability (Equation III-4) is then used to estimate the standardised difference as shown in Equation III-6 [127]

$$d = \frac{(\hat{p}_{treatment} - \hat{p}_{control})}{\sqrt{(((\hat{p}_{treatment}(1 - \hat{p}_{treatment})) + (\hat{p}_{control}(1 - \hat{p}_{control}))/2)}}$$

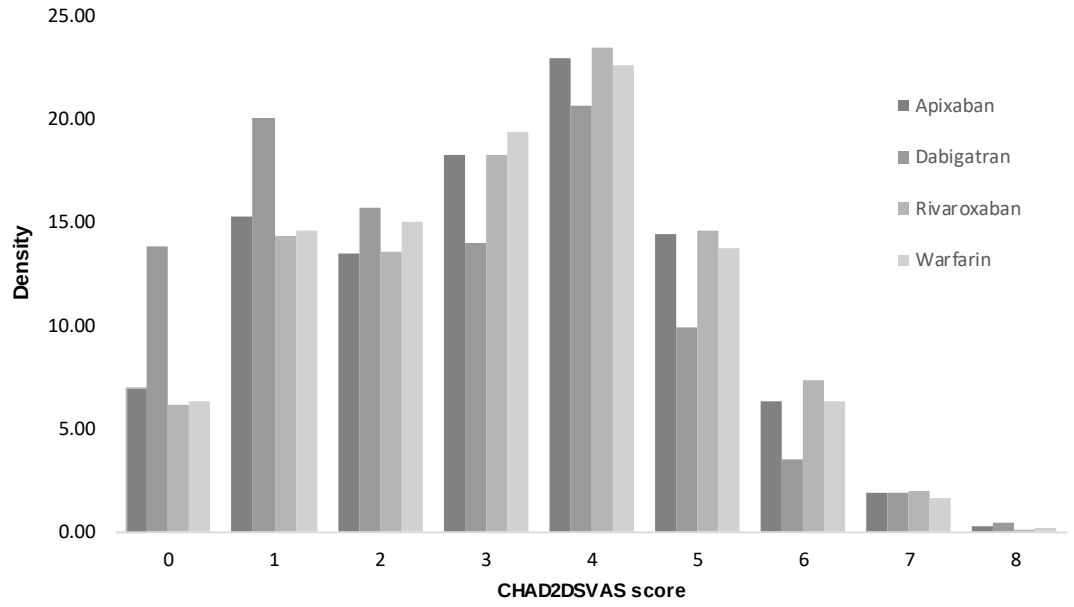
Equation (III-6)

Where $\hat{p}$ *treatment* and $\hat{p}$ *control* are the prevalence or mean of the dichotomous variable of the treated and untreated group respectively.

Appendix IV: Collinearity test (Table I)

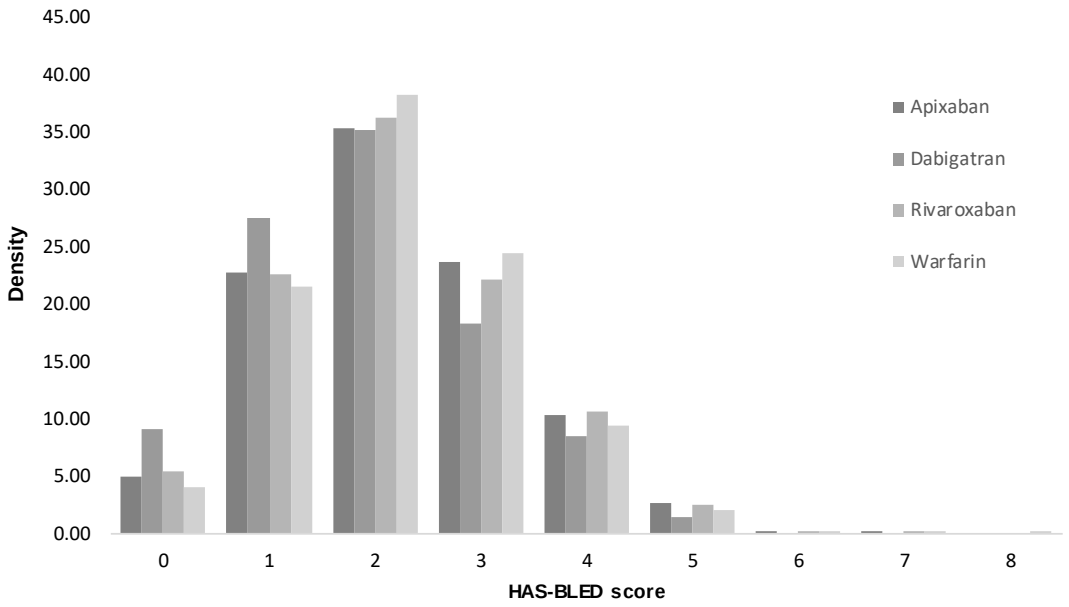
Covariates				Standard dose		Reduced dose	
	Apixaban	Dabigatran	Rivaroxaban	Apixaban 5 mg	Rivaroxaban 20 mg	Apixaban 2.5 mg	Rivaroxaban 15 mg
Age	2.55	2.54	2.54	2.55	2.54	2.52	2.53
Sex							
Women	1.44	1.44	1.44	1.44	1.44	1.44	1.44
SIMD (Scottish index of multiple deprivation)							
2	1.65	1.68	1.69	1.65	1.68	1.67	1.68
3	1.64	1.69	1.7	1.65	1.69	1.67	1.68
4	1.63	1.67	1.69	1.64	1.69	1.65	1.66
5	1.63	1.64	1.65	1.63	1.65	1.63	1.63
CHA_2DS_2-VASc score							
0-2	2.79	2.79	2.79	2.79	2.79	2.81	2.81
>=4	6.41	6.4	6.41	6.38	6.39	6.43	6.42
HAS-BLED							
>=3	2.78	2.78	2.79	2.76	2.78	2.80	2.80
Comorbidity							
1 comorbidity	1.23	1.22	1.23	1.23	1.23	1.23	1.23
>1 comorbidity	1.67	1.66	1.67	1.67	1.67	1.67	1.67
Stroke or TIA	1.33	1.32	1.33	1.32	1.33	1.32	1.32
Vascular disease	1.25	1.25	1.25	1.25	1.25	1.25	1.25
Hypertension	1.98	2.01	1.99	2.00	1.99	2.00	2.00
Diabetes mellitus	1.27	1.26	1.26	1.27	1.27	1.26	1.26
Cancer	1.11	1.11	1.11	1.11	1.11	1.11	1.10
Drug causing bleeding	1.26	1.26	1.26	1.26	1.26	1.26	1.26
Mean	1.98	1.98	1.99	1.97	1.98	1.98	1.98

Appendix V: Baseline characteristics for patients on anticoagulants



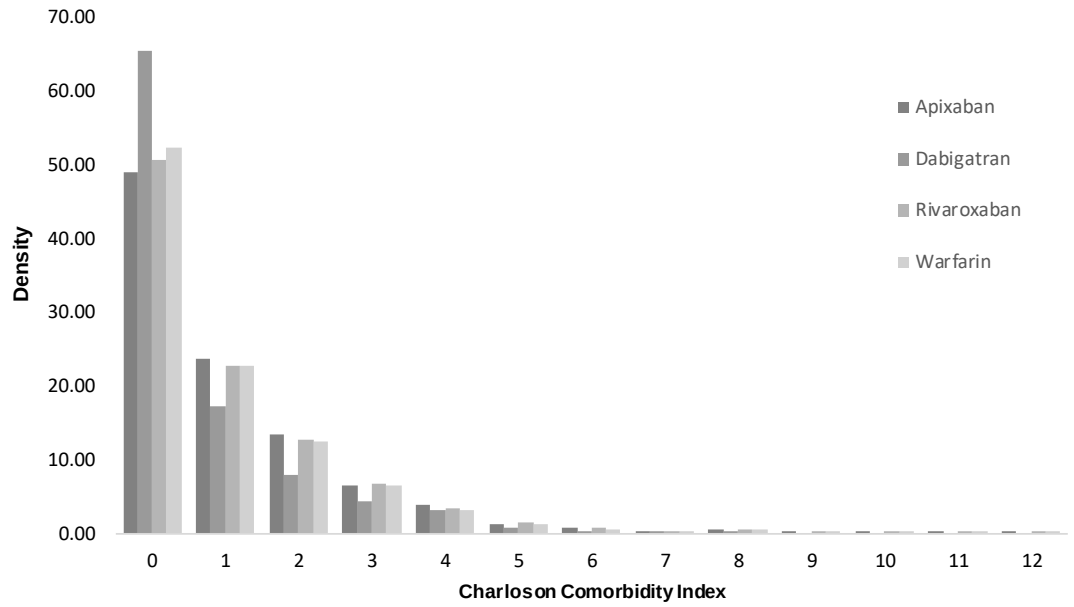
CHA2DS2-VAS score	Apixaban N(%)	Dabigatran N(%)	RivaroxabanN(%)	Warfarin N(%)
0	260 (7.02)	67 (13.84)	210 (6.20)	1,683 (6.38)
1	566 (15.27)	97 (20.04)	487 (14.37)	3,864 (14.64)
2	501 (13.52)	76 (15.70)	459 (13.55)	3,961 (15.01)
3	678 (18.29)	68 (14.05)	619 (18.27)	5,123 (19.41)
4	851 (22.96)	100 (20.66)	795 (23.47)	5,971 (22.63)
5	535 (14.44)	48 (9.92)	496 (14.64)	3,622 (13.73)
6	234 (6.31)	17 (3.51)	249 (7.35)	1,672 (6.34)
7	72 (1.94)	9 (1.86)	68 (2.01)	439 (1.66)
8	9 (0.24)	<5 (1.03)	5 (0.15)	52 (0.20)

Figure V-1 CHA_2DS_2VAS distribution by DOACs and warfarin



HAS-BLED score	Apixaban N(%)	Dabigatran N(%)	RivaroxabanN(%)	Warfarin N(%)
0	184 (4.96)	44 (9.09)	181 (5.34)	1,065 (4.04)
1	842 (22.72)	133 (27.48)	768 (22.67)	5,696 (21.59)
2	1,312 (35.40)	170 (35.12)	1,229 (36.28)	10,117 (38.34)
3	877 (23.66)	89 (18.39)	752 (22.20)	6,448 (24.44)
4	385 (10.39)	41 (8.47)	361 (10.66)	2,467 (9.35)
5	96 (2.59)	7 (1.45)	87 (2.57)	533 (2.02)
6	9 (0.24)	0 (0.00)	9 (0.27)	57 (0.22)
7	<5 (0.13)	0 (0.00)	<5 (0.15)	<5 (0.02)
8	0 (0.00)	0 (0.00)	0 (0.00)	<5 (0.02)

Figure V-2 HAS-BLED distribution by DOACs and warfarin



CCI	Apixaban N(%)	Dabigatran N(%)	RivaroxabanN(%)	Warfarin N(%)
0	1,814 (48.25)	317 (65.50)	1,714 (50.59)	13,831 (52.42)
1	883 (23.83)	83 (17.15)	769 (22.70)	5,974 (22.64)
2	496 (13.38)	39 (8.06)	432 (12.75)	3,282 (12.44)
3	241 (6.50)	21 (4.34)	231 (6.82)	1,698 (6.43)
4	140 (3.78)	15 (3.10)	113 (3.34)	823 (3.12)
5	51 (1.38)	<5 (1.03)	50 (1.48)	339 (1.28)
6	26 (0.70)	<5 (1.03)	23 (0.68)	170 (0.64)
7	9 (0.24)	<5 (1.03)	13 (0.38)	49 (0.19)
8	20 (0.54)	<5 (1.03)	20 (0.59)	116 (0.44)
9	14 (0.38)	0 (0.00)	10 (0.30)	64 (0.24)
10	6 (0.16)	0 (0.00)	7 (0.21)	25 (0.09)
11	<5 (0.13)	0 (0.00)	<5 (0.15)	9 (0.03)
12	<5 (0.13)	0 (0.00)	<5 (0.15)	7 (0.03)

Figure V-3 Charlson Comorbidity Index distribution by DOACs and warfarin

Appendix VI: Proportionality hazard assumption (number of events, event rates and HRs estimates with time intervals)

Table VI-1 Number of events and event rates estimated with time intervals

Outcome	time intervals (2 years)	Apixaban Events (event rates)	Dabigatran Events (event rates)	Rivaroxaban Events (event rates)	Warfarin Events (event rates)
Stroke all	0-730	61 (2.27)	5 (0.56)	85 (2.33)	704 (1.12)
	731-1,460	no event	5 (0.56)	< 5 (0.03)	286 (0.45)
	1,461-2,186	no event	no event	< 5 (0.03)	68 (0.11)
	0-2,186	61 (2.27)	10 (1.12)	87 (2.39)	1,058 (1.68)
Stroke or SE	0-730	65 (2.42)	5 (0.56)	87 (2.39)	754 (1.20)
	731-1,460	no event	5 (0.56)	< 5 (0.05)	307 (0.49)
	1,461-2,186	no event	no event	< 5 (0.03)	69 (0.11)
	0-2,186	65 (2.42)	10 (1.12)	90 (2.47)	1,130 (1.80)
Stroke or SE or TIA	0-730	81 (3.02)	6 (0.67)	100 (2.47)	941 (1.50)
	731-1,460	no event	6 (0.67)	< 5 (0.08)	375 (0.60)
	1,461-2,186	no event	no event	< 5 (0.03)	83 (0.13)
	0-2,186	81 (3.02)	12 (1.35)	104 (2.85)	1,399 (2.22)
Stroke or SE or death	0-730	240 (8.95)	29 (3.25)	378 (10.38)	2,720 (4.32)
	731-1,460	< 5 (0.04)	18 (2.02)	26 (0.71)	1,280 (2.03)
	1,461-2,186	no event	no event	< 5 (0.05)	313 (0.50)
	0-2,186	241 (8.98)	47 (5.27)	406 (11.14)	4,313 (6.85)
MI	0-730	46 (1.71)	< 5 (0.45)	30 (0.82)	423 (0.67)
	731-1,460	< 5 (0.04)	< 5 (0.11)	< 5 (0.08)	158 (0.25)
	1,461-2,186	no event	no event	< 5 (0.00)	30 (0.05)
	0-2,186	47 (1.75)	5 (0.56)	33 (0.91)	611 (0.97)
Major bleed	0-730	127 (4.73)	26 (2.92)	223 (6.21)	1,718 (2.73)
	731-1,460	no event	6 (0.67)	8 (0.22)	645 (1.02)
	1,461-2,186	no event	no event	< 5 (0.00)	111 (0.18)
	0-2,186	127 (4.73)	32 (3.59)	231 (6.34)	2,474 (3.93)
ICH	0-730	16 (0.60)	no event	18 (0.49)	117 (0.19)
	731-1,460	no event	< 5 (0.11)	no event	53 (0.08)
	1,461-2,186	no event	no event	no event	6 (0.01)
	0-2,186	16 (0.60)	< 5 (0.11)	18 (0.49)	176 (0.28)
GI bleed	0-730	43 (1.60)	14 (1.57)	74 (2.03)	565 (0.90)
	731-1,460	no event	no event	< 5 (0.08)	200 (0.32)
	1,461-2,186	no event	no event	no event	33 (0.05)
	0-2,186	43 (1.60)	14 (1.57)	77 (2.11)	798 (1.27)
Mortality (all-cause)	0-730	176 (6.56)	24 (2.69)	292 (8.02)	1,980 (3.15)
	731-1,460	< 5 (0.04)	13 (1.46)	24 (0.66)	978 (1.55)
	1,461-2,186	no event	no event	< 5 (0.03)	244 (0.39)
	0-2,186	177 (6.60)	37 (4.15)	317 (8.70)	3,202 (5.09)
Mortality (stroke)	0-730	10 (0.37)	< 5 (0.22)	19 (0.52)	106 (0.17)
	731-1,460	no event	< 5 (0.11)	< 5 (0.03)	46 (0.07)
	1,461-2,186	no event	no event	no event	6 (0.01)
	0-2,186	10 (0.37)	3 (0.34)	20 (0.55)	158 (0.25)
Mortality (cardiovascular)	0-730	100 (3.73)	13 (1.46)	146 (4.01)	1,050 (1.67)
	731-1,460	no event	< 5 (0.45)	12 (0.33)	467 (0.74)
	1,461-2,186	no event	no event	< 5 (0.03)	99 (0.16)
	0-2,186	100 (3.73)	17 (1.91)	159 (4.36)	1,616 (2.57)

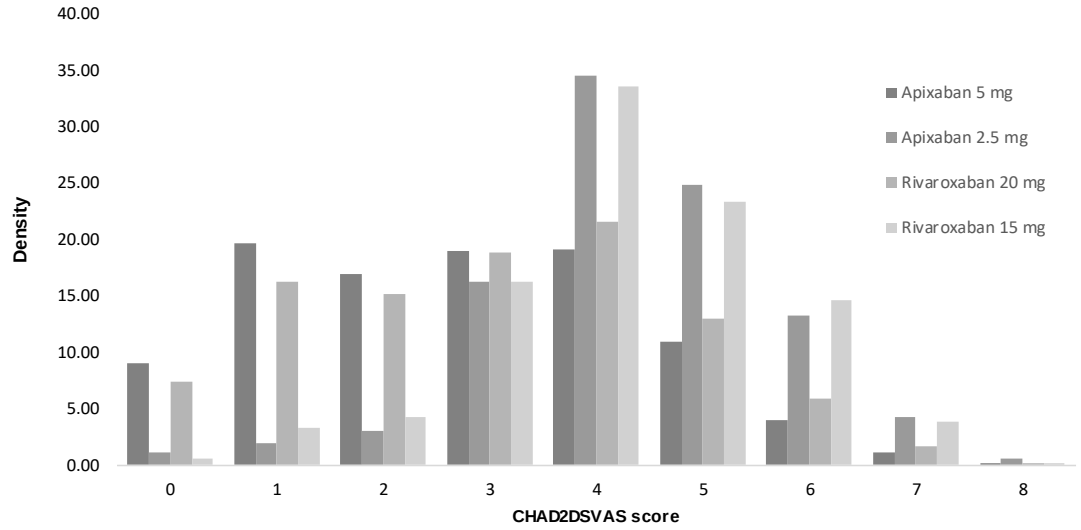
Note: due to disclosure restrictions, in the case of fewer than five events, the sign < 5 was used instead. Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack, ICH=intracranial haemorrhage, GI=gastrointestinal

Table VI-2 HRs estimated with time intervals

Outcome	time intervals (2 years)	Apixaban vs. warfarin HR 95% CI	Dabigatran vs. warfarin HR 95% CI	Rivaroxaban vs. warfarin HR 95% CI
Stroke all	0-730	0.96 (0.70, 1.30)	0.52 (0.21, 1.25)	1.16 (0.91, 1.48)
	731-1,460	not estimated	1.71 (0.70, 4.18)	0.19 (0.03, 1.39)
	1461-2,186	not estimated	not estimated	35.85 (4.64, 262.03)
	0-2,186	0.94 (0.69, 1.27)	0.79 (0.42, 1.47)	1.10 (0.87, 1.40)
Stroke SE	0-730	0.95 (0.70, 1.27)	0.48 (0.20, 1.16)	1.10 (0.87, 1.40)
	731-1,460	not estimated	1.61 (0.66, 3.93)	0.36 (0.09, 1.47)
	1,461-2,186	not estimated	not estimated	35.54 (4.73, 267.10)
	0-2,186	0.93 (0.69, 1.25)	0.73 (0.39, 1.37)	1.06 (0.84, 1.34)
Stroke or SE Or TIA	0-730	0.94 (0.73, 1.23)	0.46 (0.21, 1.03)	1.00 (0.80, 1.25)
	731-1,460	not estimated	1.49 (0.66, 3.37)	0.41 (0.13, 1.30)
	1,461-2,186	not estimated	not estimated	23.79 (3.18, 178.02)
	0-2,186	0.93 (0.72, 1.21)	0.70 (0.39, 1.23)	0.97 (0.78, 1.20)
Stroke or SE or TIA	0-730	1.03 (0.89, 1.21)	0.73 (0.50, 1.05)	1.33 (1.18, 1.50)
	731-1,460	0.63 (0.09, 4.49)	1.43 (0.89, 2.29)	1.19 (0.80, 1.78)
	1,461-2,186	not estimated	not estimated	16.18 (3.98, 65.81)
	0-2,186	1.02 (0.88, 1.19)	0.89 (0.66, 1.18)	1.32 (1.18, 1.47)
MI	0-730	1.37 (0.95, 1.98)	0.67 (0.25, 1.79)	0.76 (0.52, 1.13)
	731-1,460	4.83 (0.64, 36.27)	0.53 (0.07, 3.78)	0.96 (0.30, 3.09)
	1,461-2,186	not estimated	not estimated	not estimated
	0-2,186	1.41 (0.98, 2.03)	0.63 (0.26, 1.52)	0.77 (0.53, 1.12)
Major bleed	0-730	0.89 (0.73, 1.09)	0.94 (0.64, 1.39)	1.30 (1.12, 1.52)
	731-1,460	not estimated	0.85 (0.38, 1.92)	0.73 (0.36, 1.48)
	1,461-2,186	not estimated	not estimated	not estimated
	0-2,186	0.88 (0.72, 1.08)	0.92 (0.65, 1.30)	1.26 (1.09, 1.46)
ICH	0-730	1.39 (0.74, 2.61)	not estimated	1.24 (0.72, 2.12)
	731-1,460	not estimated	not estimated	not estimated
	1,461-2,186	not estimated	not estimated	not estimated
	0-2,186	1.36 (0.73, 2.53)	0.38 (0.05, 2.69)	1.11 (0.66, 1.88)
GI bleed	0-730	0.88 (0.62, 1.26)	1.76 (1.03, 3.01)	1.37 (1.05, 1.78)
	731-1,460	not estimated	not estimated	0.87 (0.27, 2.78)
	1,461-2,186	not estimated	not estimated	not estimated
	0-2,186	0.87 (0.61, 1.25)	1.38 (0.81, 2.35)	1.32 (1.02, 1.71)
Mortality (all-cause)	0-730	1.06 (0.88, 1.27)	0.80 (0.54, 1.20)	1.41 (1.23, 1.62)
	731-1,460	0.87 (0.12, 6.23)	1.35 (0.78, 2.34)	1.47 (0.96, 2.23)
	1,461-2,186	not estimated	not estimated	10.61 (1.47, 76.52)
	0-2,186	1.05 (0.87, 1.25)	0.93 (0.67, 1.29)	1.40 (1.24, 1.60)
Mortality (stroke)	0-730	0.79 (0.37, 1.73)	1.39 (0.34, 5.69)	1.60 (0.93, 2.78)
	731-1,460	not estimated	2.83 (0.37, 21.61)	2.00 (0.25, 16.11)
	1,461-2,186	not estimated	not estimated	not estimated
	0-2,186	0.77 (0.35, 1.67)	1.68 (0.53, 5.33)	1.57 (0.92, 2.68)
Mortality (cardiovascular)	0-730	1.17 (0.91, 1.49)	0.84 (0.49, 1.46)	1.37 (1.13, 1.65)
	731-1,460	not estimated	0.92 (0.34, 2.47)	1.59 (0.87, 2.88)
	1,461-2,186	not estimated	not estimated	22.63 (3.07, 167.05)
	0-2,186	1.15 (0.90, 1.47)	0.86 (0.53, 1.38)	1.39 (1.16, 1.66)

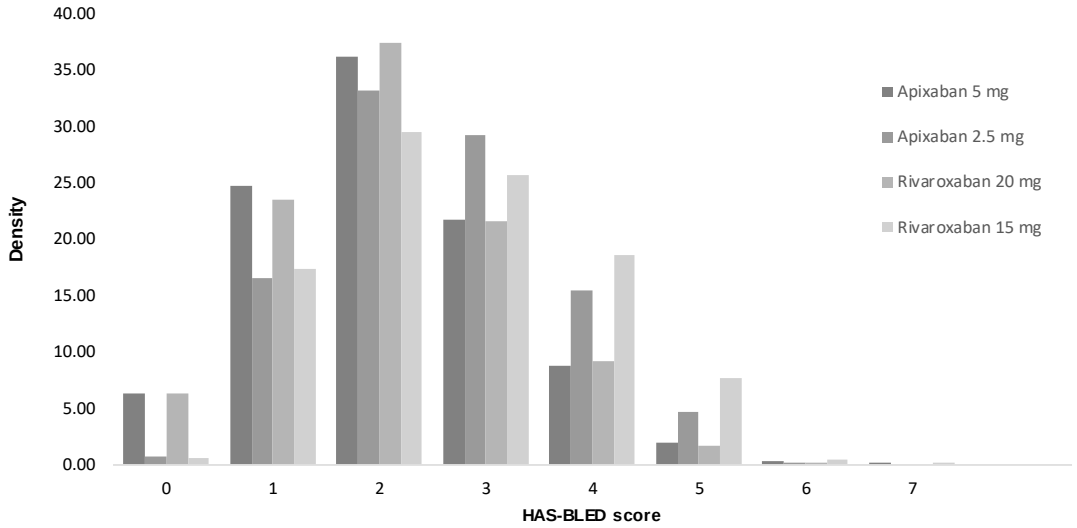
Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack, ICH=intracranial haemorrhage, GI=gastrointestinal

Appendix VII: Baseline characteristics for patients on anticoagulants (standard and reduced dose)



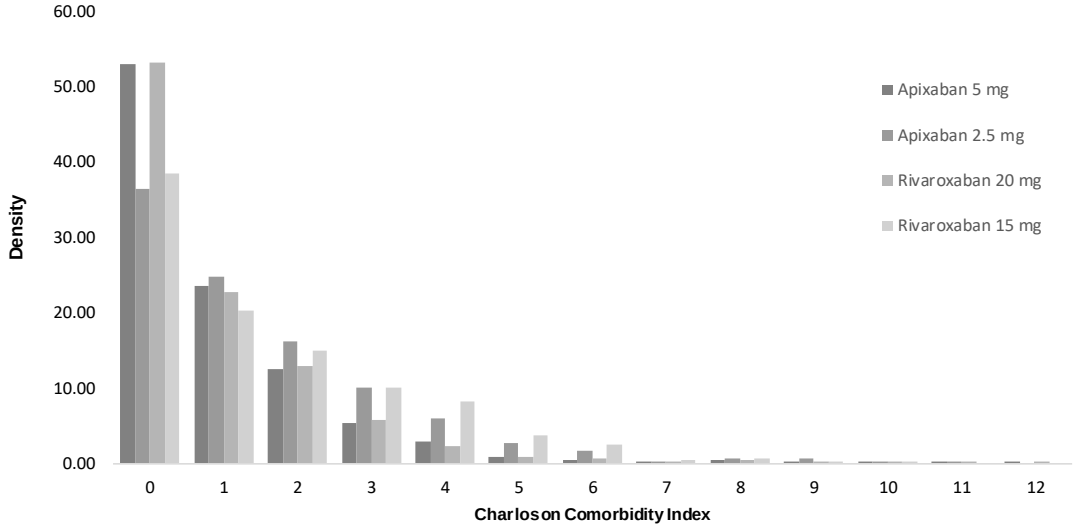
CHA2DS2-VAS score	Apixaban 5 mg N(%)	Apixaban 2.5 mg N(%)	Rivaroxaban 20 mg N(%)	Rivaroxaban 15 N(%)
0	250 (8.98)	10 (1.08)	203 (7.42)	3 (0.58)
1	548 (19.68)	18 (1.95)	445 (16.27)	17 (3.28)
2	473 (16.99)	28 (3.04)	417 (15.25)	22 (4.25)
3	528 (18.97)	150 (16.27)	515 (18.83)	84 (16.22)
4	532 (19.11)	319 (34.60)	589 (21.54)	174 (33.59)
5	306 (10.99)	229 (24.84)	354 (12.94)	121 (23.36)
6	112 (4.02)	122 (13.23)	163 (5.96)	76 (14.67)
7	32 (1.15)	40 (4.34)	45 (1.65)	20 (3.86)
8	<5 (0.18)	6 (0.65)	<5 (0.18)	<5 (0.97)

Figure VII-1 CHA_2DS_2VAS distribution by DOACs (standard and reduced dose)



HASBLED score	Apixaban 5 mg N(%)	Apixaban 2.5 mg N(%)	Rivaroxaban 20 mg N(%)	Rivaroxaban 15 N(%)
0	177 (6.36)	7 (0.76)	174 (6.36)	3 (0.58)
1	689 (24.75)	153 (16.59)	643 (23.51)	90 (17.37)
2	1,006 (36.14)	306 (33.19)	1,022 (37.37)	153 (29.54)
3	607 (21.80)	270 (29.28)	592 (21.65)	133 (25.68)
4	243 (8.73)	142 (15.40)	251 (9.18)	96 (18.53)
5	53 (1.90)	43 (4.66)	47 (1.72)	40 (7.72)
6	8 (0.29)	<5 (0.54)	6 (0.22)	<5 (0.97)
7	<5 (0.18)	0 (0.00)	0 (0.00)	<5 (0.97)
8	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)

Figure VII-2 HAS-BLED distribution by DOACs (standard and reduced dose)



CCI	Apixaban 5 mg N(%)	Apixaban 2.5 mg N(%)	Rivaroxaban 20 mg N(%)	Rivaroxaban 15 N(%)
0	1,477 (53.05)	337 (36.55)	1,458 (53.31)	200 (38.61)
1	655 (23.53)	228 (24.73)	624 (22.82)	105 (20.27)
2	347 (12.46)	149 (16.16)	354 (12.94)	78 (15.06)
3	148 (5.32)	93 (10.09)	158 (5.78)	52 (10.04)
4	84 (3.02)	56 (6.07)	62 (2.27)	43 (8.30)
5	26 (0.93)	25 (2.71)	24 (0.88)	19 (3.67)
6	11 (0.40)	15 (1.63)	16 (0.59)	13 (2.51)
7	6 (0.22)	3 (0.33)	8 (0.29)	<5 (0.97)
8	13 (0.47)	7 (0.76)	15 (0.55)	<5 (0.97)
9	8 (0.29)	6 (0.65)	10 (0.37)	<5 (0.97)
10	<5 (0.18)	<5 (0.54)	<5 (0.18)	<5 (0.97)
11	<5 (0.18)	<5 (0.54)	<5 (0.18)	0 (0.00)
12	<5 (0.18)	0 (0.00)	<5 (0.18)	0 (0.00)

Figure VII-3 Charlson Comorbidity Index distribution by DOACs (standard and reduced dose)

Appendix VIII: Proportionality hazard assumption (plots of goodness of fit)

Apixaban versus warfarin

Seemingly, for the majority of the clinical outcomes assessed, plots were straight lines but not necessarily parallel, and were crossing at different time points. Figure VIII-1 assessing the proportional hazards assumption for stroke, TIA and SE in the apixaban versus warfarin comparison, clearly indicates proportionality violation, where log-cumulative hazard lines are crossing several times. Comparable trends of proportionality violation were observed in Figure VIII-2 looking at MI, ICH, major and GI bleeding. The log-cumulative hazard plots for mortality due to stroke and, arguably mortality due to cardiovascular conditions, presented in Figures VIII-3 b and VIII-3 c, are the only ones that show proportionality.

Dabigatran versus warfarin

Overall proportionality violation is observed for most of the clinical outcomes. Unlike to what is observed in the apixaban plots, the log-cumulative hazard lines for stroke, TIA and SE are not crossing (Figure VIII-4 a–VIII-4 c). However, these plots are not straight lines either; therefore, it is difficult to establish whether the proportionality assumption is met. Due to a small dabigatran sample size and a relatively small number of events, the plot for ICH only included the log-cumulative hazard lines for warfarin (Figure VIII-5c). Other differences with apixaban plots are seen for mortality due stroke and cardiovascular conditions (Figure VIII-6 b, VIII-6 c), where crossing lines clearly show proportionality violation.

Rivaroxaban versus warfarin

As for apixaban and dabigatran, proportionality violation can be observed throughout most of the clinical outcomes. Violation is particularly evident in the plots showing MI (Figure VIII-8 a) and mortality due to stroke (Figure VIII-9 b) where log-cumulative hazard lines are crossing several times.

Table VIII-1 shows that overall, age and CCI where the variables violating the proportionality assumption across all treatments and clinical outcomes.

Figure VIII-1 a (stroke-all)

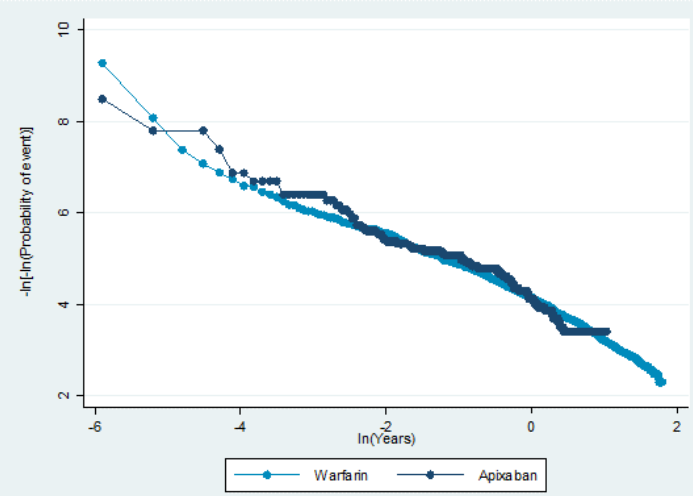


Figure VIII-1 b (stroke or SE)

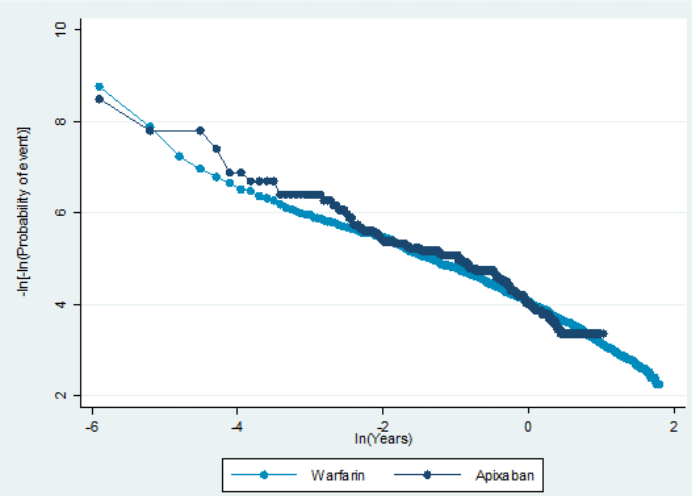


Figure VIII-1 c (stroke or SE or TIA)

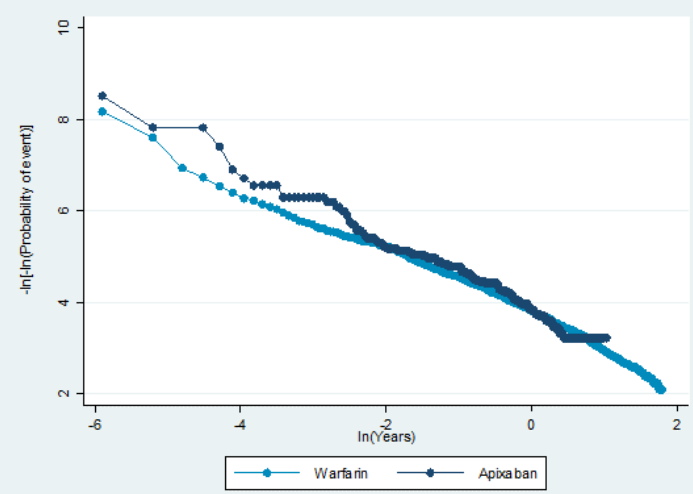


Figure VIII-1 d (stroke or mortality-all-cause)

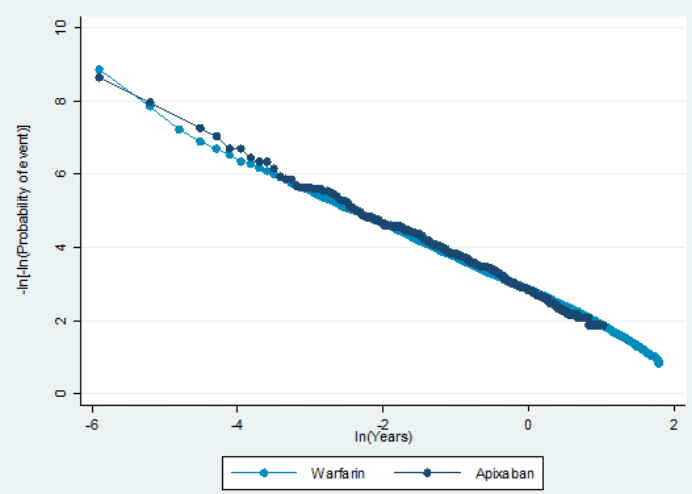


Figure VIII-1: log-cumulative hazard plot for stroke, SE and TIA (apixaban vs. warfarin)
Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack

Figure VIII-2 a (MI)

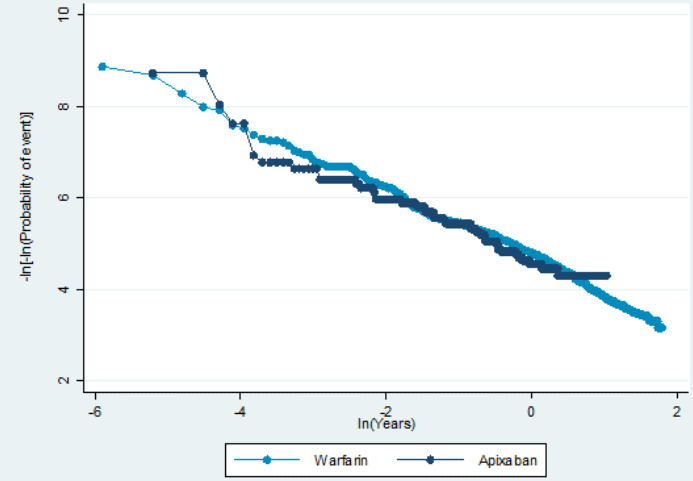


Figure VIII-2 b (major bleeding)

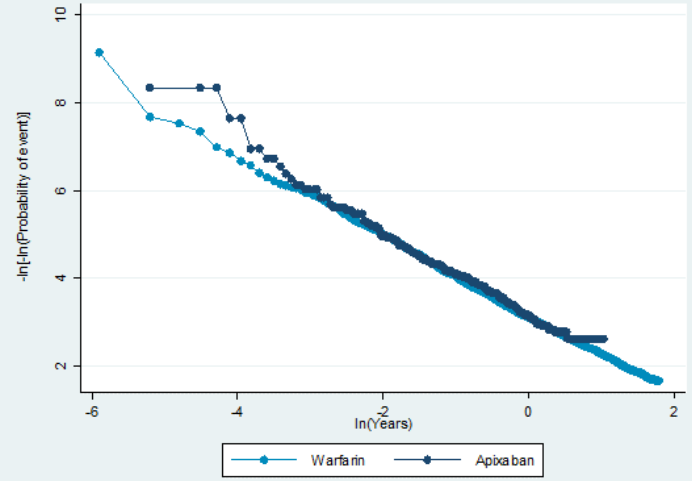


Figure VIII-2 c (ICH)

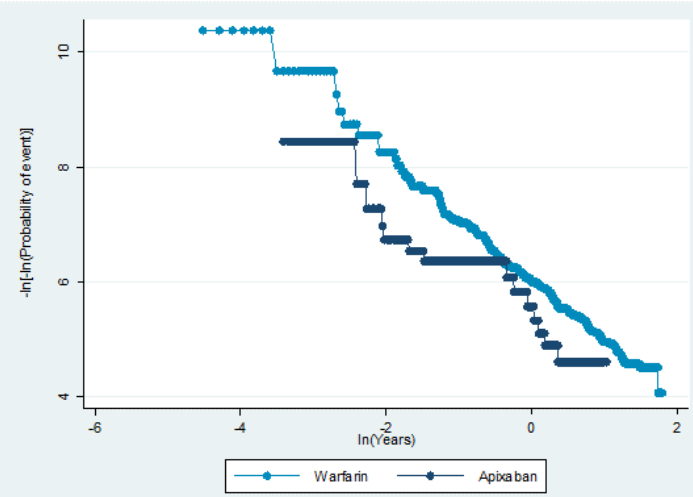


Figure VIII-2 d (GI bleeding)

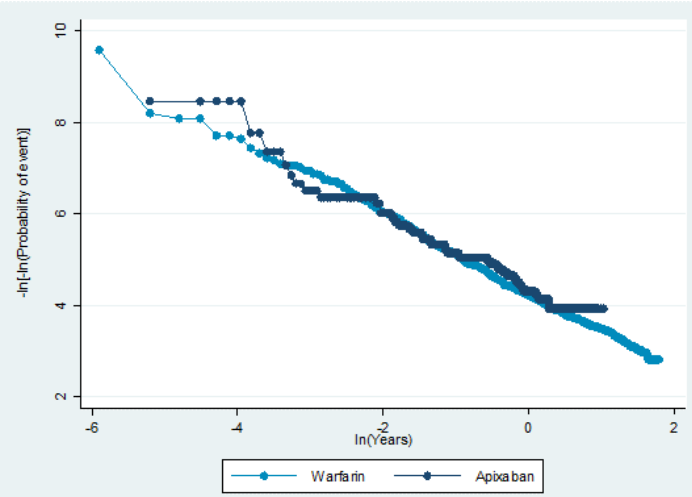


Figure VIII-2: log-cumulative hazard plot for MI, ICH and bleeding (apixaban vs. warfarin)
Abbreviations: MI=myocardial, ICH= intracranial haemorrhage, GI=gastrointestinal

Figure VIII-3 a (mortality-all-cause)

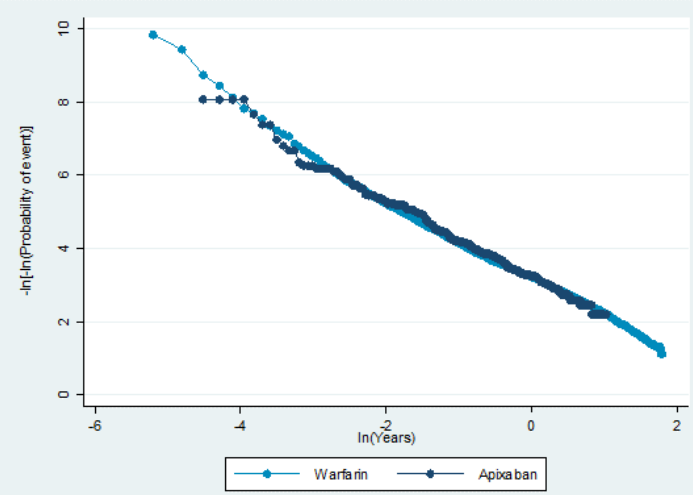


Figure VIII-3 b (mortality-stroke)

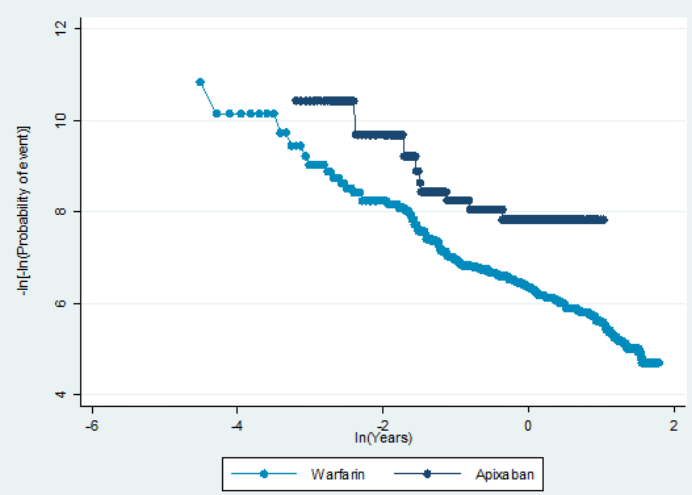


Figure VIII-3 c (mortality-cardiovascular)

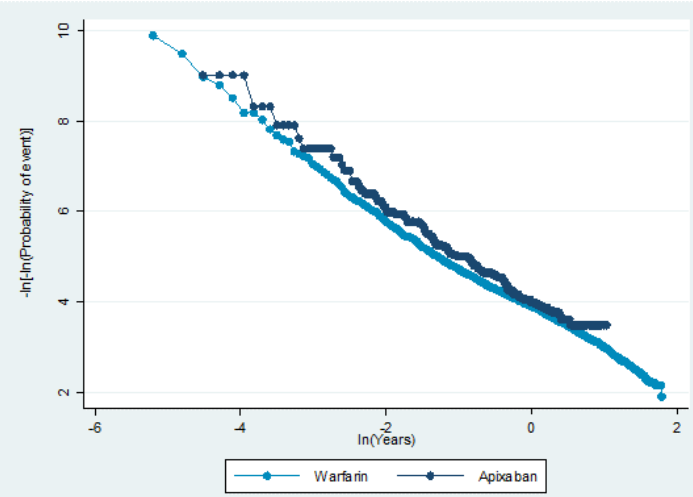


Figure VIII-3: log-cumulative hazard plot for mortality (apixaban vs. warfarin)

Figure VIII-4 a (stroke-all)

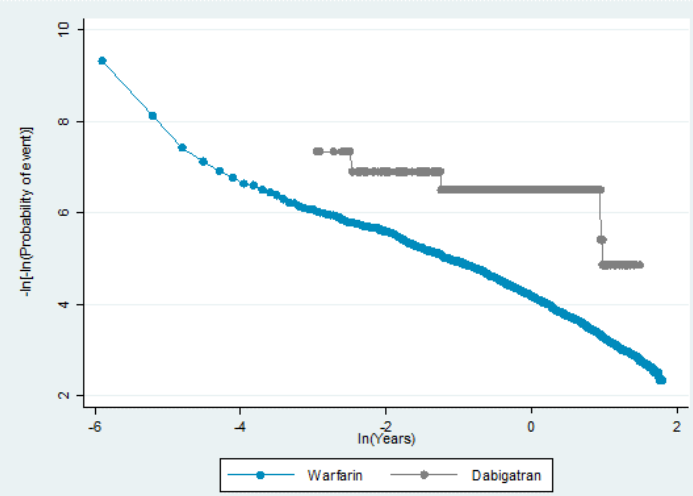


Figure VIII-4 b (stroke or SE)

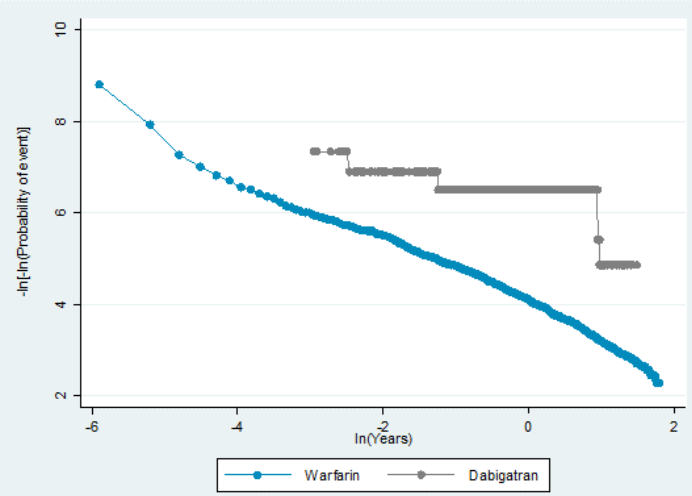


Figure VIII-4 c (stroke or SE or TIA)

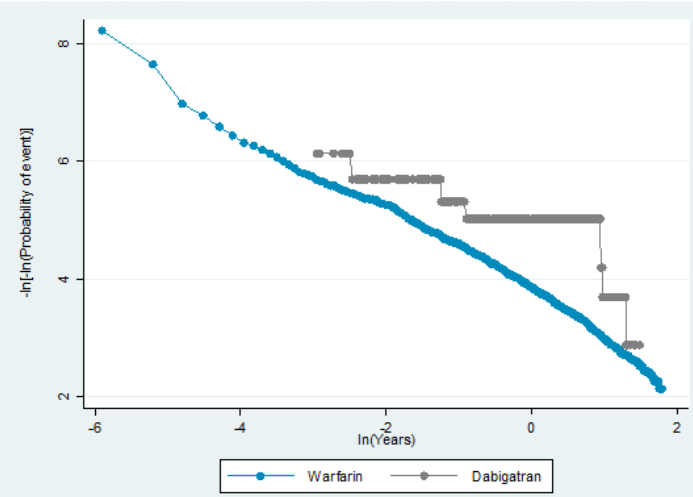


Figure VIII-4 d (stroke or mortality-all-cause)

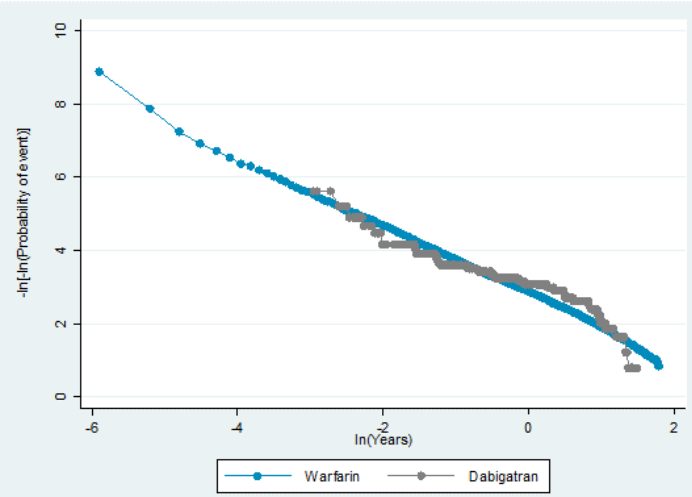


Figure VIII-4: log-cumulative hazard plot for stroke, SE and TIA (dabigatran vs. warfarin)
Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack

Figure VIII-5 a (MI)

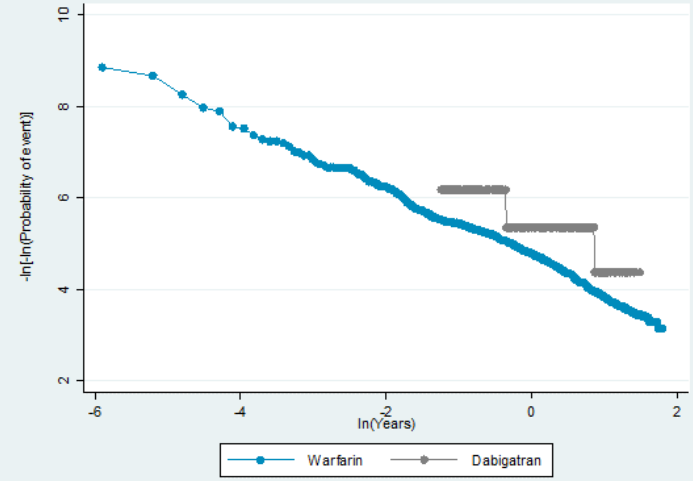


Figure VIII-5 b (major bleeding)

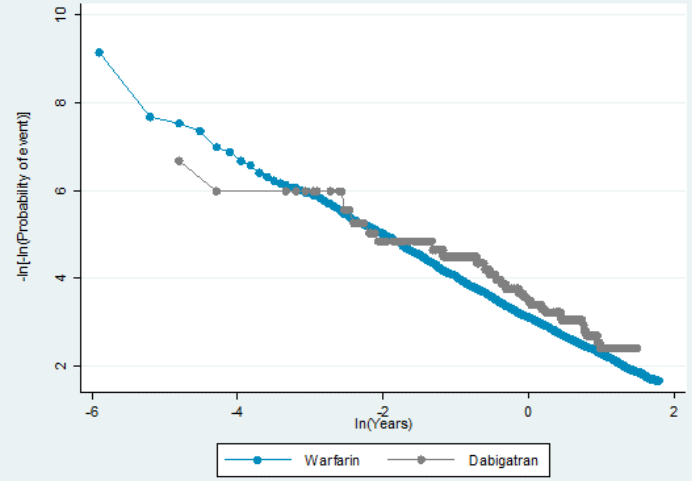


Figure VIII-5 c (ICH)

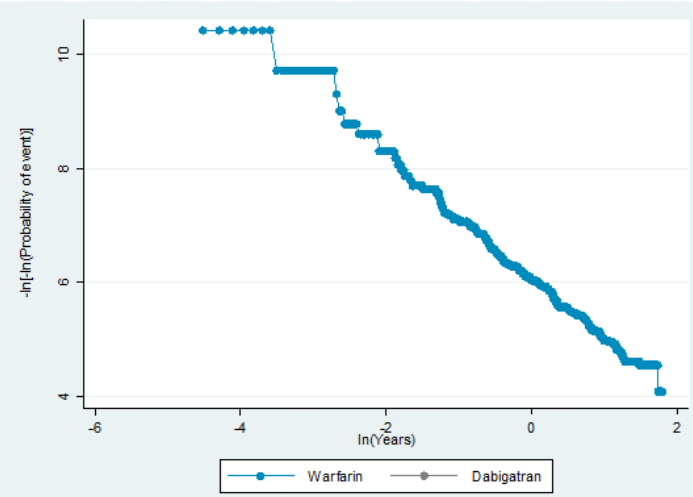


Figure VIII-5 d (GI bleeding)

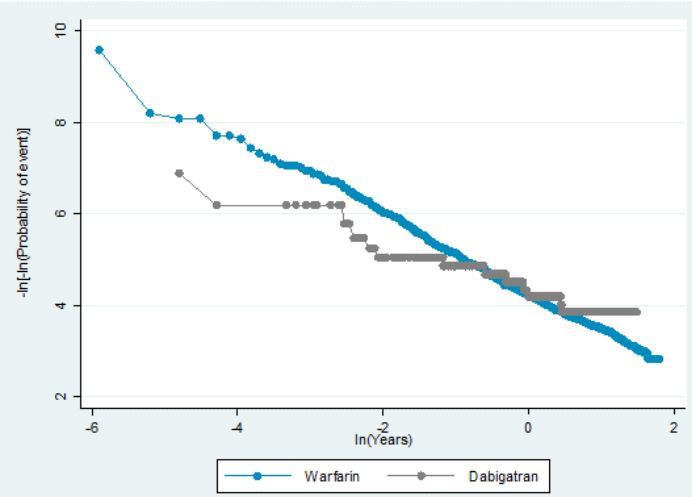


Figure VIII-5: log-cumulative hazard plot for MI, ICH and bleeding (dabigatran vs. warfarin)
Abbreviations: MI=myocardial, ICH= intracranial haemorrhage, GI=gastrointestinal
Due to disclosure restrictions, in the case of fewer than five events, "<5" was reported.

Figure VIII-6 a (mortality-all-cause)

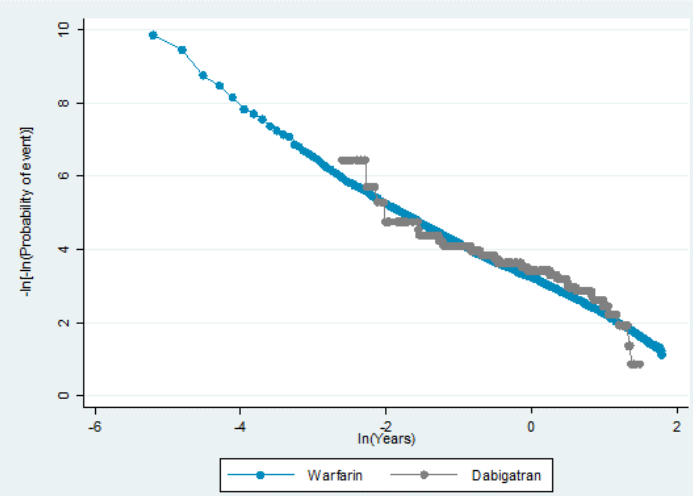


Figure VIII-6 b (mortality-stroke)

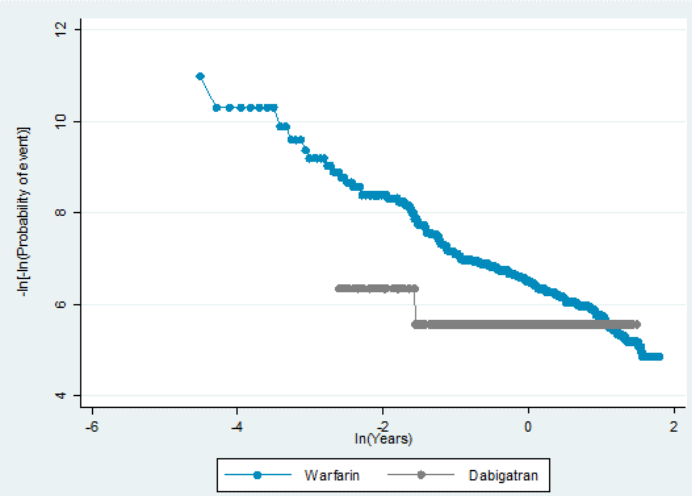


Figure VIII-6 c (mortality-cardiovascular)

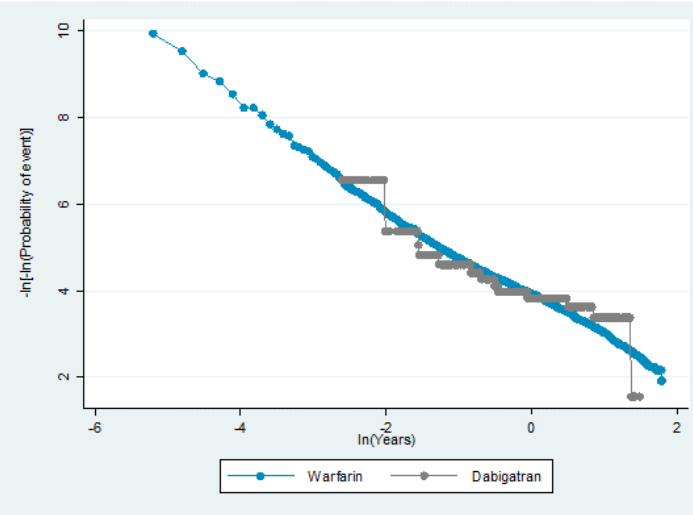


Figure VIII-6: log-cumulative hazard plot for mortality (dabigatran vs. warfarin)

Figure VIII-7 a (stroke-all)

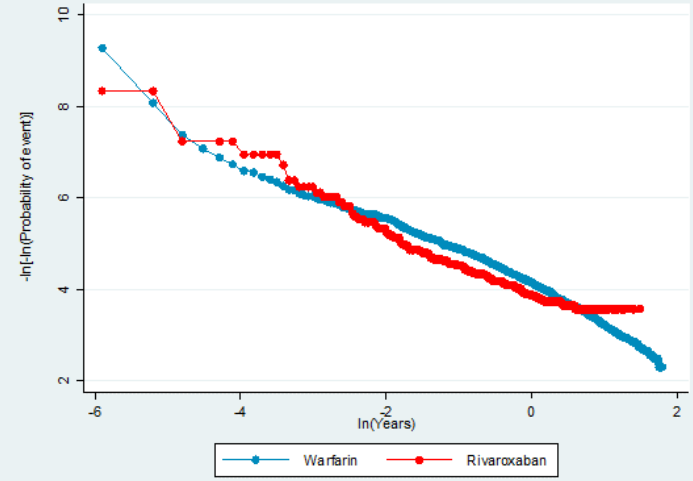


Figure VIII-7 b (stroke or SE)

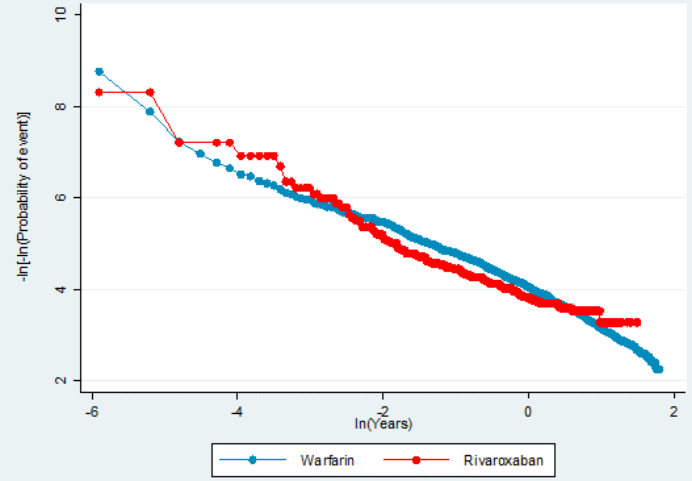


Figure VIII-7 c (stroke or SE or TIA)

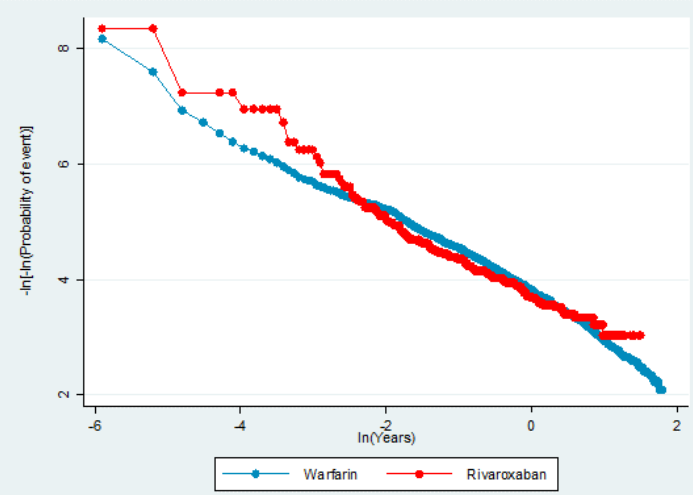


Figure VIII-7 d (stroke or mortality-all-cause)

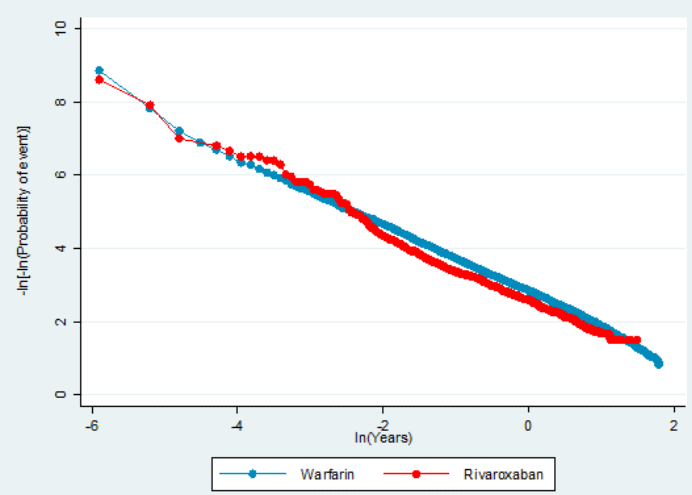


Figure VIII-7: log-cumulative hazard plot for stroke, SE and TIA (rivaroxaban vs. warfarin)
Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack

Figure VIII-8 a (MI)

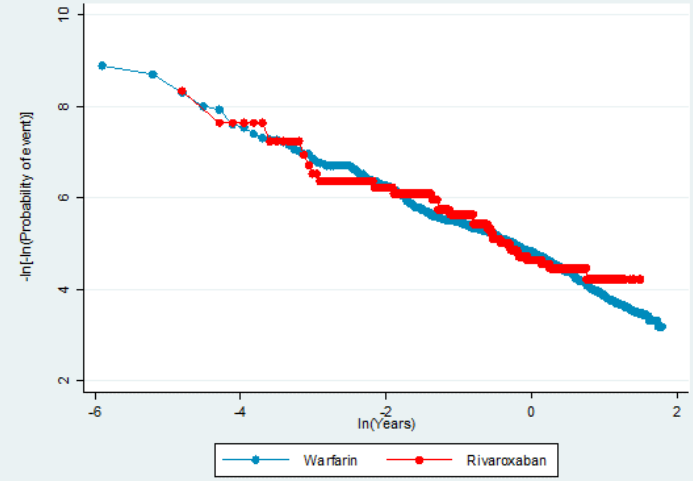


Figure VIII-8 b (major bleeding)

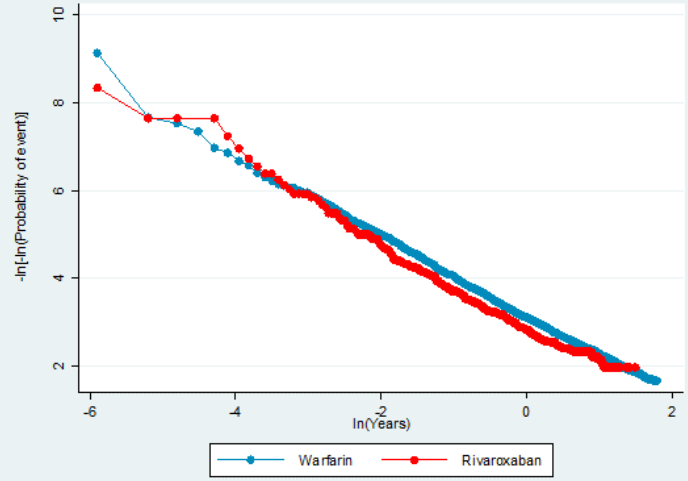


Figure VIII-8 c (ICH)

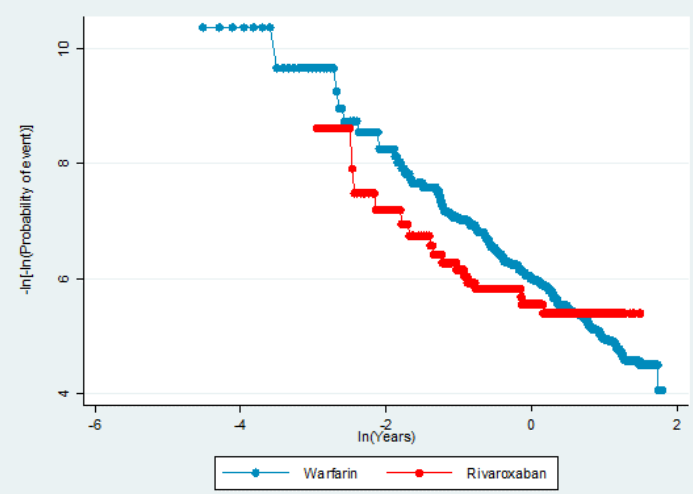


Figure VIII-8 d (GI bleeding)

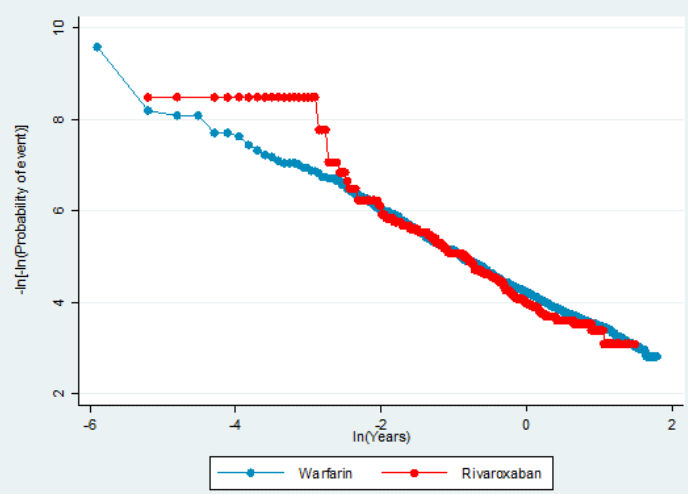


Figure VIII-8: log-cumulative hazard plot for MI, ICH and bleeding (rivaroxaban vs. warfarin)
Abbreviations: MI=myocardial, ICH= intracranial haemorrhage, GI=gastrointestinal

Figure VIII-9 a (mortality-all-cause)

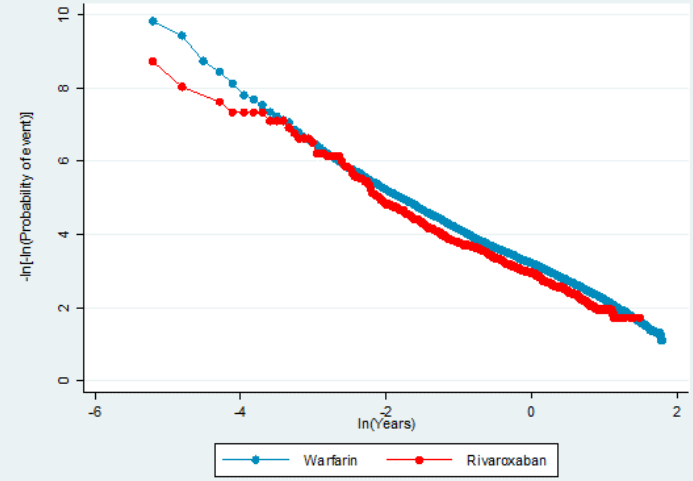


Figure VIII-9 b (mortality-stroke)

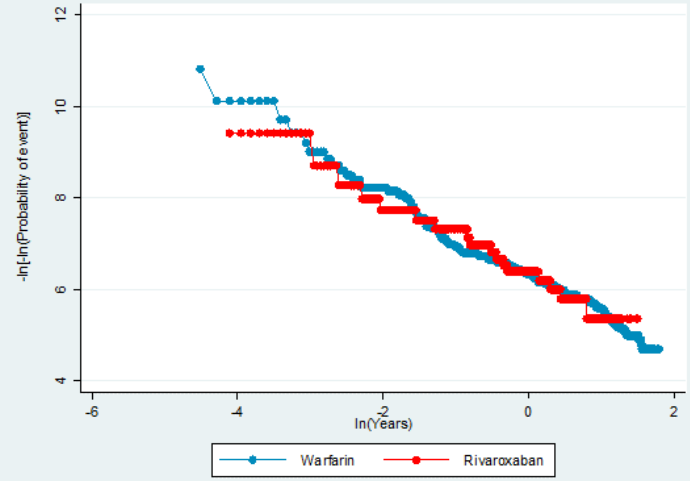


Figure VIII-9 c (mortality-cardiovascular)

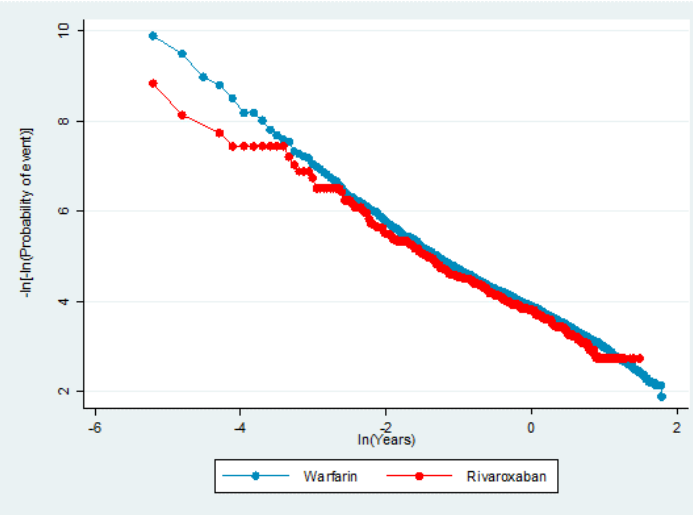


Figure VIII-9: log-cumulative hazard plot for mortality (rivaroxaban vs. warfarin)

Table VIII-1 Proportionality hazard assumption test of significance

Outcome	p-value						
	treatment	age	sex	CCI	SIMD	year	PS
Apixaban							
Stroke all	0.150	0.001	0.297	0.001	0.336	0.904	0.130
Stroke or SE	0.088	0.001	0.182	0.001	0.370	0.910	0.189
Stroke or SE or TIA	0.133	0.001	0.283	0.001	0.124	0.905	0.095
Stroke or SE or mortality (all-cause)	0.164	0.000	0.015	0.001	0.927	0.501	0.948
MI	0.979	0.872	0.495	0.001	0.460	0.120	0.298
Major bleeding	0.961	0.560	0.948	0.001	0.451	0.886	0.086
ICH	0.335	0.039	0.345	0.001	0.445	0.385	0.523
GI bleeding	0.839	0.503	0.416	0.018	0.724	0.183	0.625
Mortality (all-cause)	0.102	0.007	0.072	0.000	0.371	0.263	0.388
Mortality (stroke)	0.573	0.984	0.861	0.002	0.944	0.573	0.899
Mortality (cardiovascular)	0,080	0.907	0.777	0.001	0.955	0.171	0.124
Dabigatran							
Stroke all	0.134	0.001	0.311	0.001	0.065	0.799	0.069
Stroke or SE	0.136	0.001	0.232	0.001	0.054	0.837	0.091
Stroke or SE or TIA	0.111	0.001	0.321	0.001	0.063	0.800	0.078
Stroke or SE or mortality (all-cause)	0.184	0.001	0.573	0.001	0.218	0.582	0.069
MI	0.691	0.941	0.653	0.001	0.566	0.173	0.646
Major bleeding	0.633	0.903	0.668	0.025	0.936	0.936	0.072
ICH	1.000	0.283	0.319	0.002	0.496	0.522	0.531
GI bleeding	0.520	0.713	0.486	0.658	0.474	0.179	0.825
Mortality (all-cause)	0.517	0.005	0.053	0.001	0.763	0.314	0.774
Mortality (stroke)	0.272	0.670	0.945	0.003	0.791	0.714	0.871
Mortality (cardiovascular)	0.197	0.969	0.772	0.001	0.829	0.240	0.888
Rivaroxaban							
Stroke all	0.082	0.006	0.200	0.001	0.228	0.969	0.828
Stroke or SE	0.075	0.004	0.132	0.001	0.299	0.945	0.879
Stroke or SE or TIA	0.182	0.011	0.183	0.001	0.060	0.994	0.380
Stroke or SE or mortality (all-cause)	0.791	0.001	0.173	0.001	0.368	0.439	0.595
MI	0.704	0.998	0.442	0.001	0.229	0.291	0.527
Major bleeding	0.365	0.409	0.945	0.005	0.366	0.817	0.701
ICH	0.074	0.215	0.496	0.006	0.694	0.307	0.578
GI bleeding	0.309	0.852	0.309	0.077	0.229	0.168	0.744
Mortality (all-cause)	0.657	0.010	0.154	0.001	0.902	0.254	0.691
Mortality (stroke)	0.585	0.436	0.786	0.038	0.575	0.416	0.312
Mortality (cardiovascular)	0.415	0.681	0.304	0.001	0.678	0.191	0.835

Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack, ICH=intracranial haemorrhage, GI=gastrointestinal

Appendix IX: Kaplan Meier curves

Apixaban versus warfarin

The crude estimates over time indicate that overall the probability of surviving is higher for patients on warfarin than those on apixaban. While for apixaban the survival curves are estimated within the first two years from anticoagulation initiation, for warfarin patients the estimates expand beyond the 2 years across all clinical outcomes. This is mainly due to patients on apixaban being exposed to treatment for a much shorter period of time; consequently, patients on apixaban overall experienced far fewer clinical events than patients on warfarin. The Kaplan Meier curves indicate that overall at 2-year follow-up, the proportion of patients free of event is higher in the warfarin rather than in the apixaban group (Figure IX-1–IX-3). This suggests that, over time, patients on apixaban are more likely to have an event, such as stroke or bleeding, than patients who are on warfarin; thus, the probability of surviving is higher for patients on warfarin then those on apixaban. While this seems to be true for the composite of stroke and mortality (Figure IX-1 d), MI, ICH (Figure IX-2 a, IX-2 c) and mortality (Figure IX-3 a, IX-3 c), no difference in the probability of surviving is observed for the various composites of stroke including SE and TIA(Figure IX-1 a – IX-2 c) and bleeding (Figure IX-2 a, IX-2 c) in the first year from anticoagulation initiation.

Dabigatran versus warfarin

The Kaplan Meier curves indicate that overall at 2-year follow-up, the proportion of patients free of event is higher in the dabigatran rather than in the warfarin group (Figure IX-4–IX-6). However, the probability of survival from MI for patients on dabigatran is based on less than 10 events. While this seems to be true for most of the outcomes, the probability of survivals seems to be inverted for GI bleeding (Figure IX-5 d) and mortality due to stroke (Figure IX-6 a).

Rivaroxaban versus warfarin

The Kaplan Meier curves indicate that overall at 2-year follow-up, the proportion of patients free of event is higher in the warfarin rather than in the apixaban group (Figure IX-7–IX-9). While this is true for the composite of stroke and mortality (Figure IX-7 d), ICH (Figure IX-8 c), bleeding (Figure IX-8 b, Figure IX-8 d) and mortality (Figure IX-9 a – IX-9 c), the probability of survivals seems to be inverted for MI (Figure IX-8 a). No difference in the probability of surviving is observed for the various composites of stroke including SE and TIA (Figure IX-7 a – IX-7 c).

Figure IX-1 a (stroke-all)

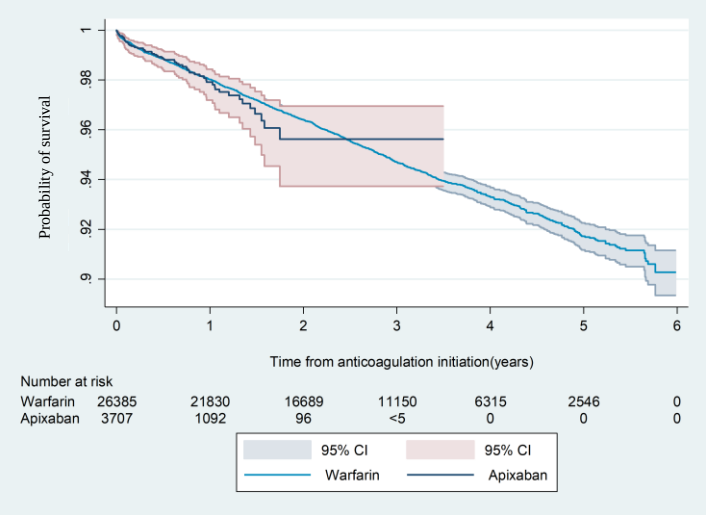


Figure IX-1 b (stroke or SE)

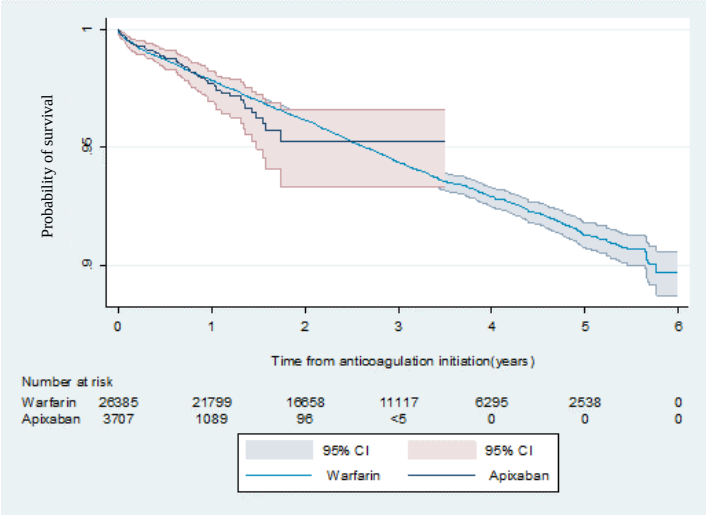


Figure IX-1 c (stroke or SE or TIA)

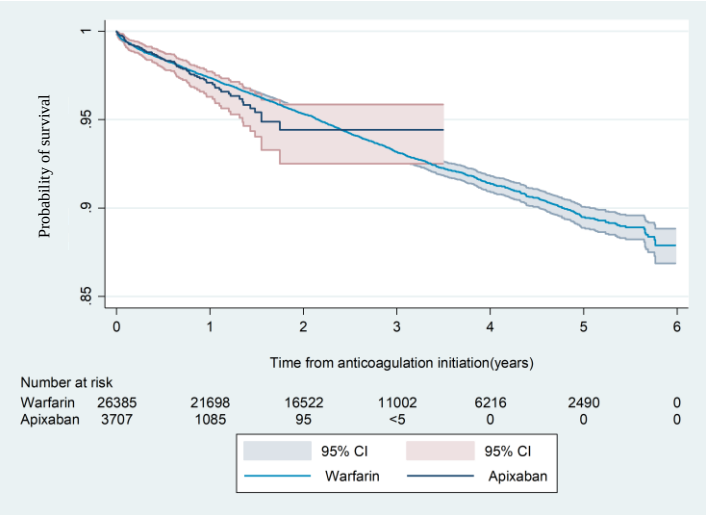


Figure IX-1 d (stroke or mortality-all-cause)

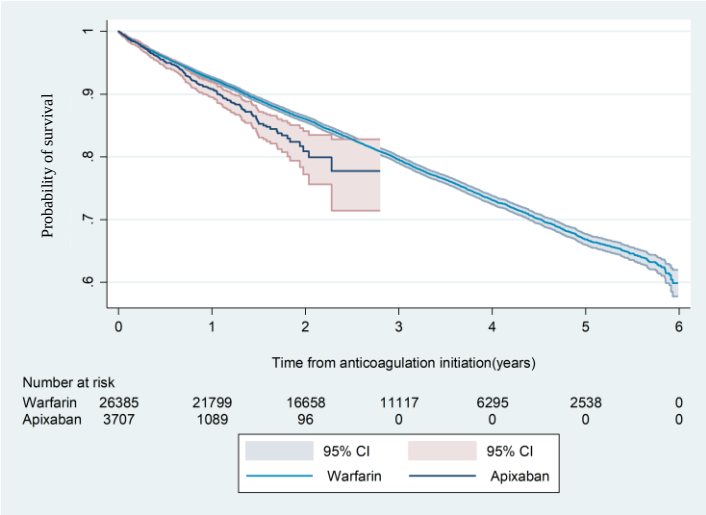


Figure IX-1: Stroke, SE and TIA (apixaban vs. warfarin)
Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack

Figure IX-2 a (MI)

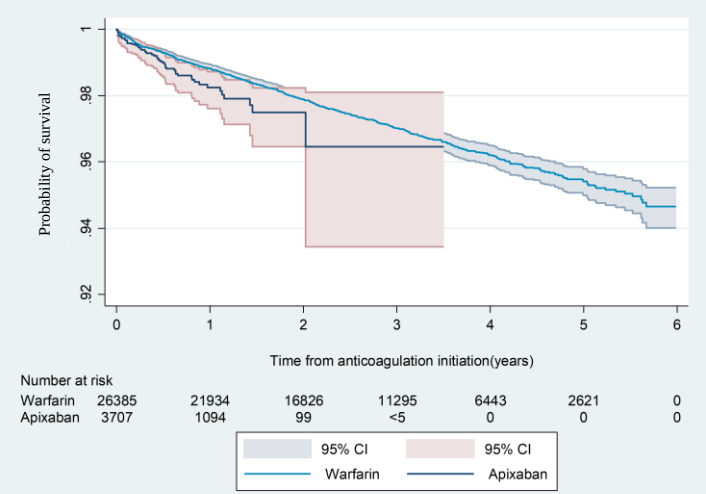


Figure IX-2 b (major bleeding)

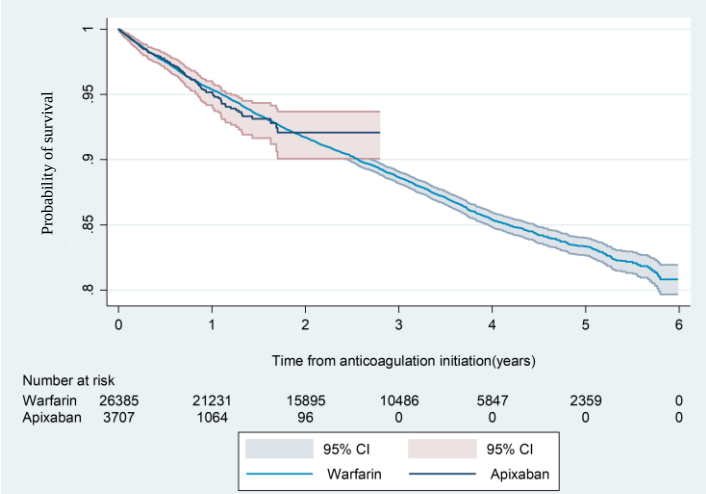


Figure IX-2 c (ICH)

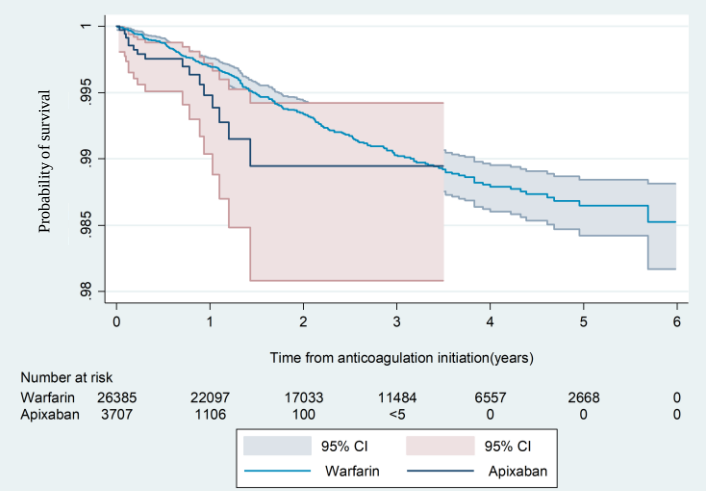


Figure IX-2 d (GI bleeding)

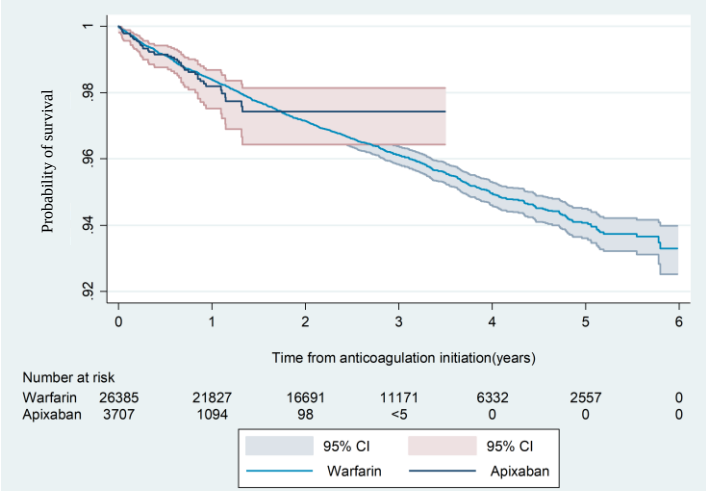


Figure IX-2: MI, ICH and bleeding (apixaban vs. warfarin)

Abbreviations: MI=myocardial, ICH= intracranial haemorrhage, GI=gastrointestinal

Figure IX-3 a (mortality-all-cause)

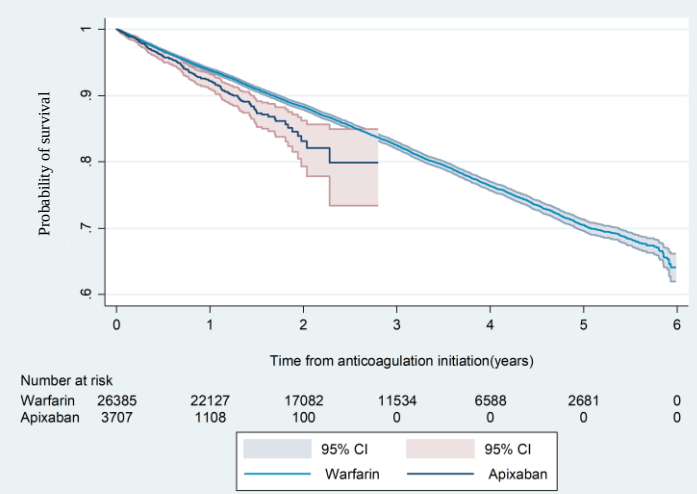


Figure IX-3 b (mortality-stroke)

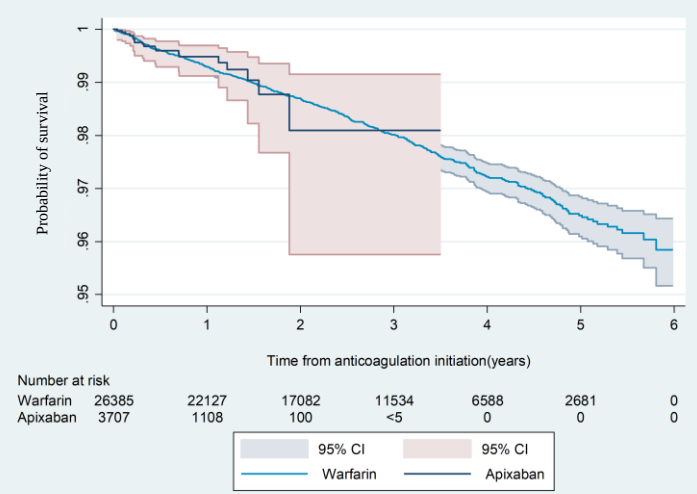


Figure IX-3 c (mortality-cardiovascular)

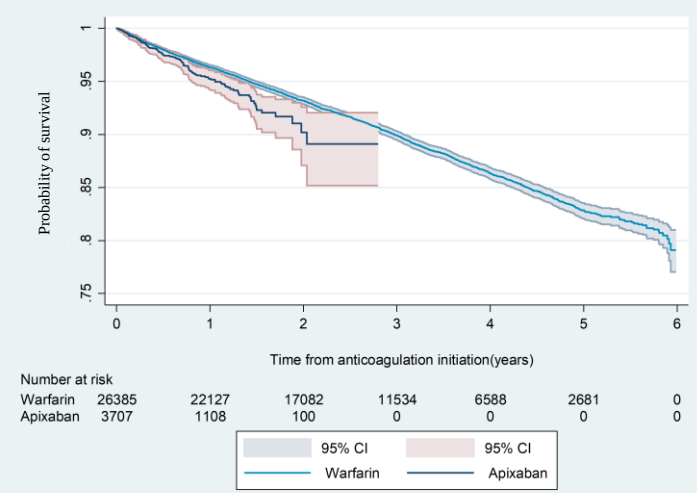


Figure IX-3: Mortality (apixaban vs. warfarin)

Figure IX-4 a (stroke-all)

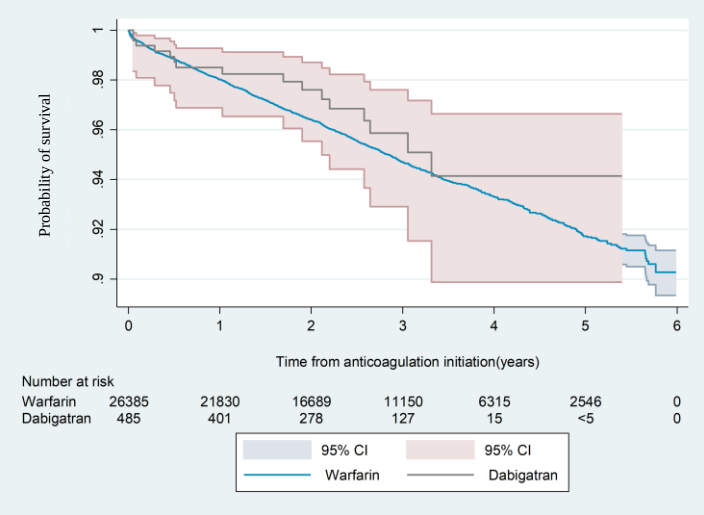


Figure IX-4 b (stroke or SE)

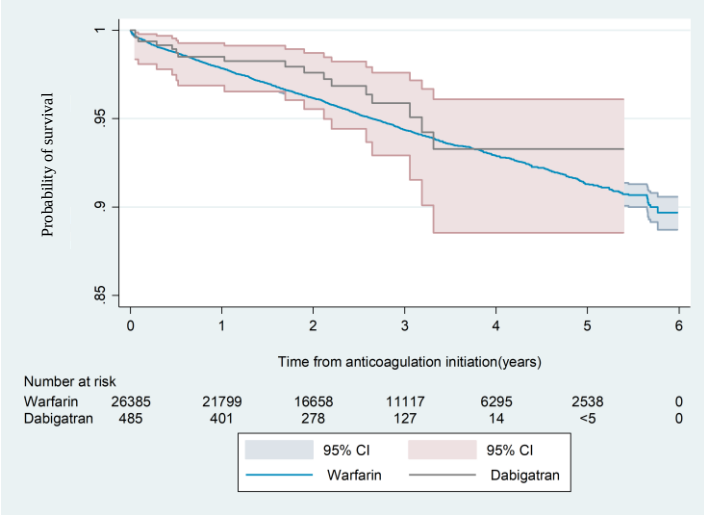


Figure IX-4 c (stroke or SE or TIA)

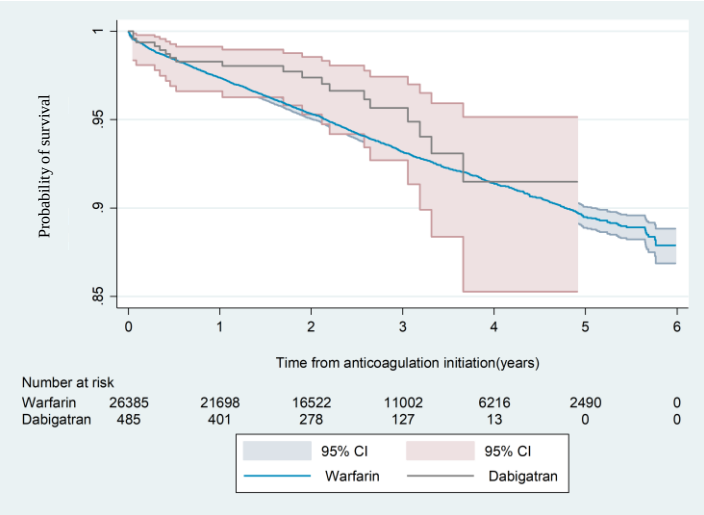


Figure IX-4 d (stroke or mortality-all-cause)

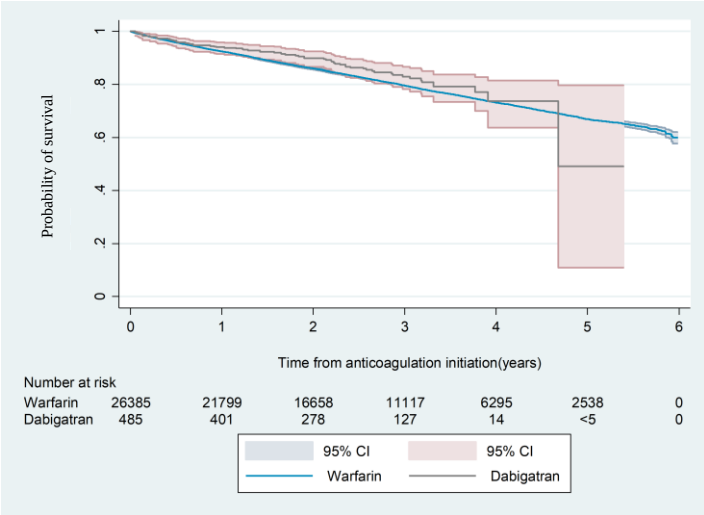


Figure IX-4: Stroke, SE and TIA (dabigatran vs. warfarin)

Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack

Figure IX-5 a (MI)

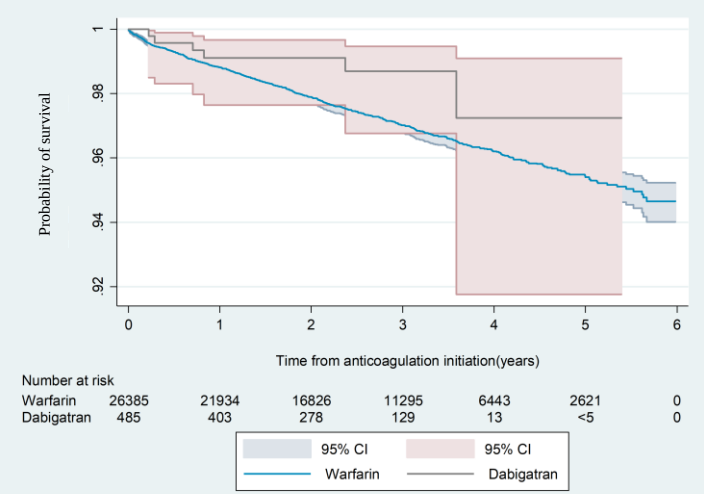


Figure IX-5 b (major bleeding)

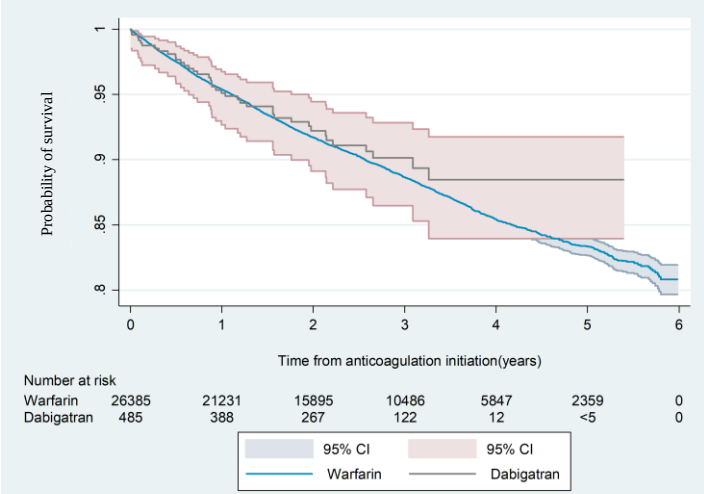


Figure IX-5 c (ICH)

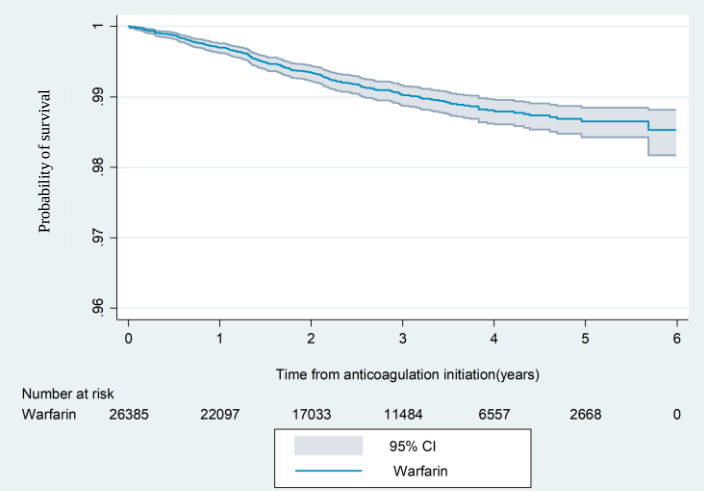


Figure IX-5 d (GI bleeding)

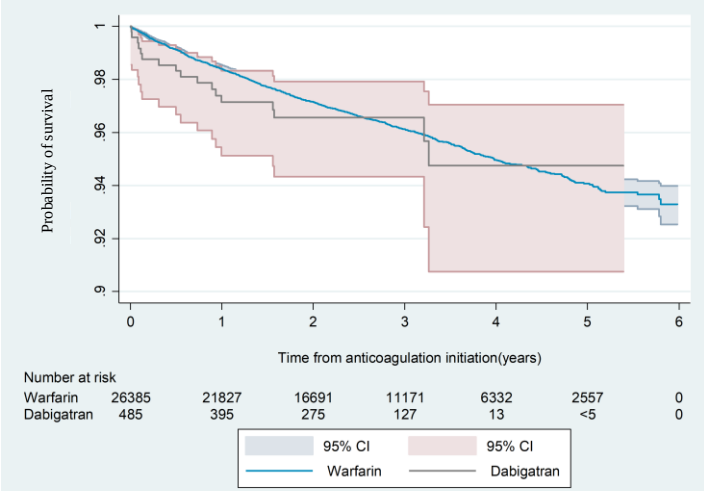


Figure IX-5: MI, ICH and bleeding (dabigatran vs. warfarin)

Note: due to disclosure restrictions, in the case of fewer than five events, the Kaplan Meier curve was not reported

Abbreviations: MI=myocardial, ICH= intracranial haemorrhage, GI=gastrointestinal

Figure IX-6 a (mortality-all-cause)

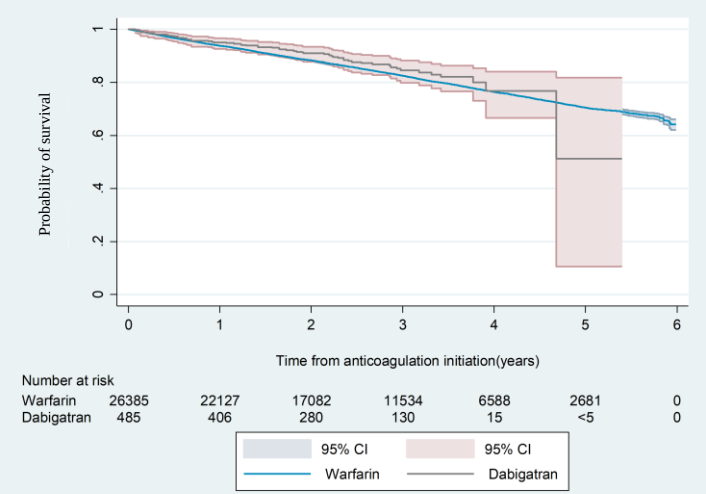


Figure IX-6 b (mortality-stroke)

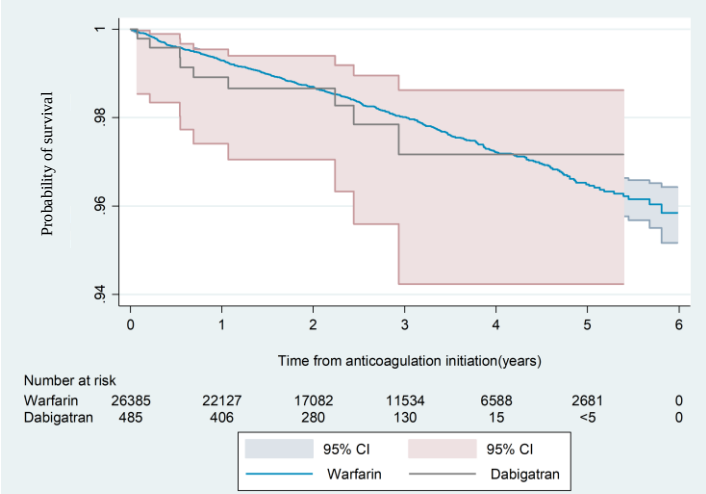


Figure IX-6 c (mortality-cardiovascular)

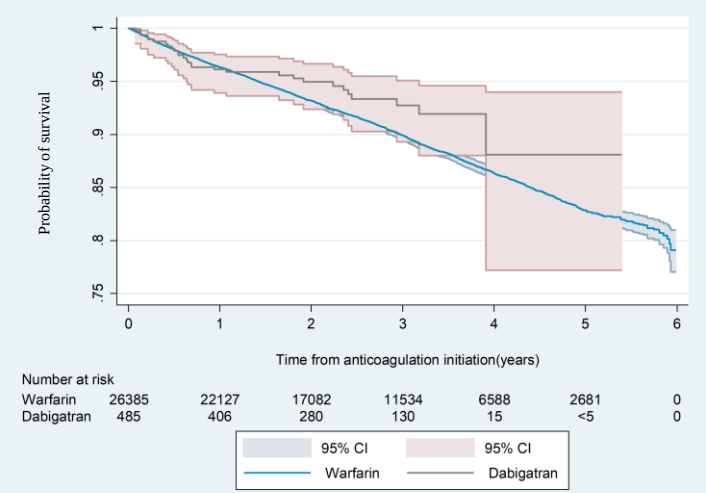


Figure IX-6: Mortality (dabigatran vs. warfarin)

Figure IX-7 a (stroke-all)

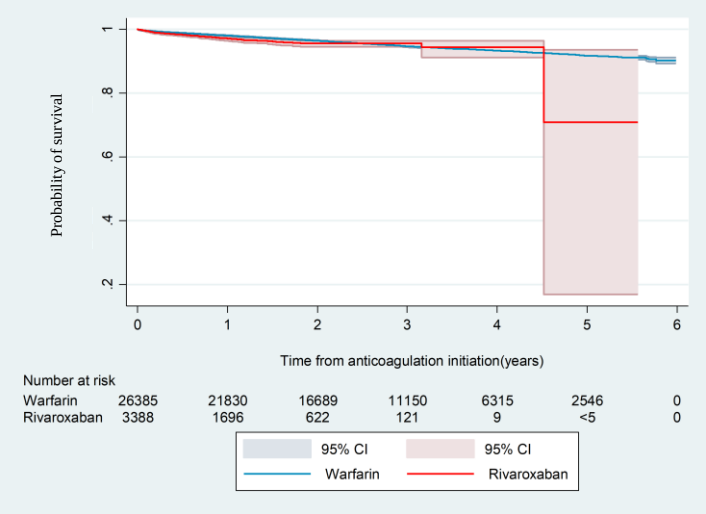


Figure IX-7 b (stroke or SE)

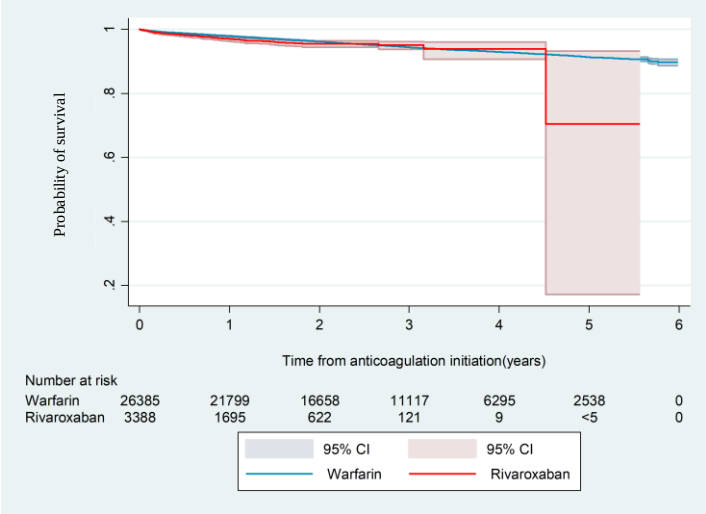


Figure IX-7 c (stroke or SE or TIA)

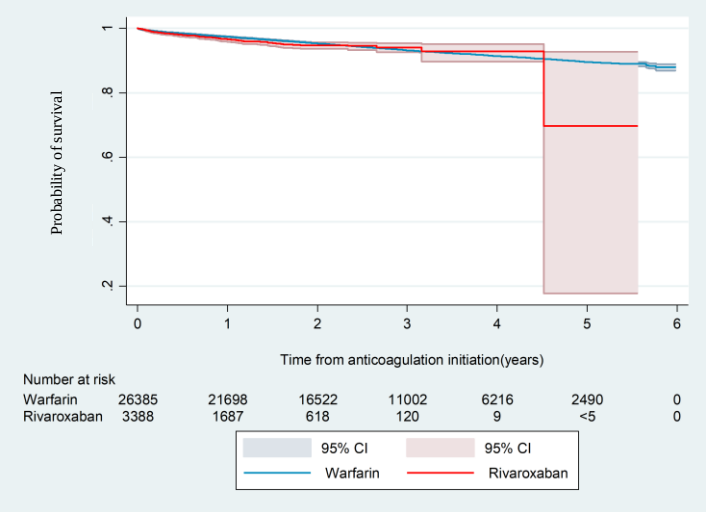


Figure IX-7 d (stroke or mortality-all-cause)

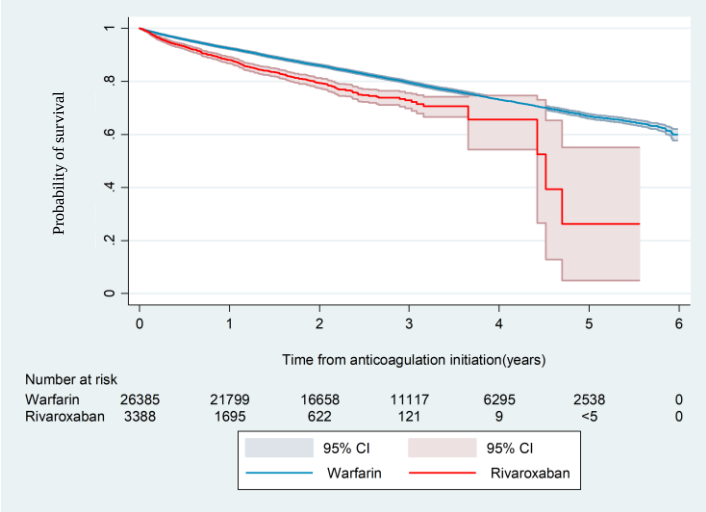


Figure IX-7: Stroke, SE and TIA (rivaroxaban vs. warfarin)

Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack

Figure IX-8 a (MI)

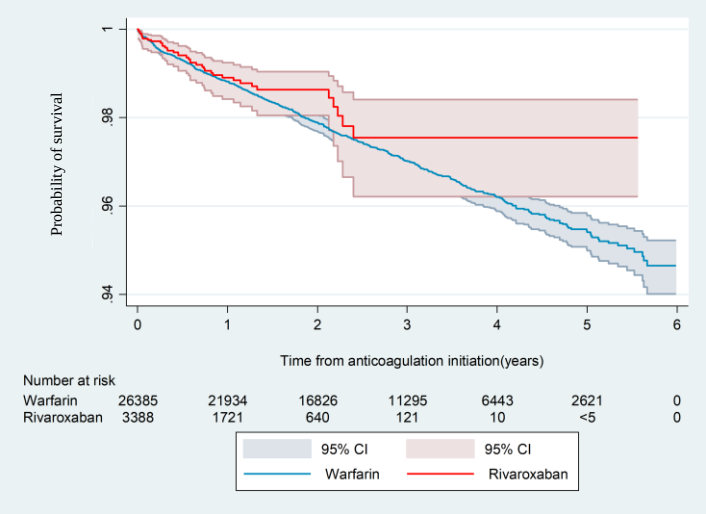


Figure IX-8 b (major bleeding)

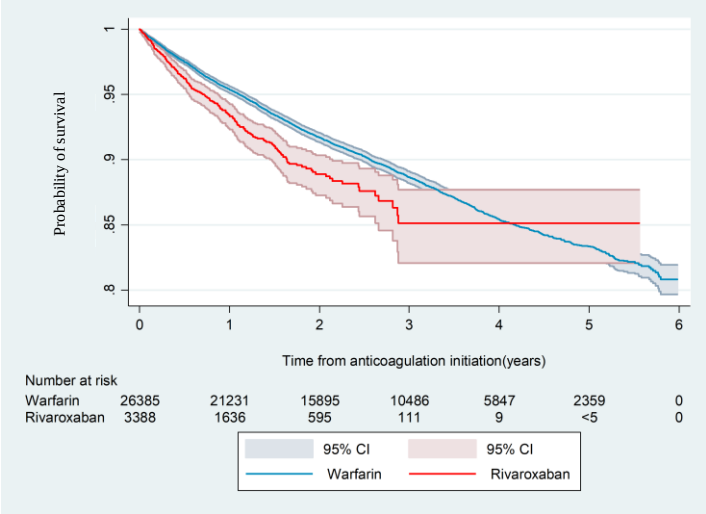


Figure IX-8 c (ICH)

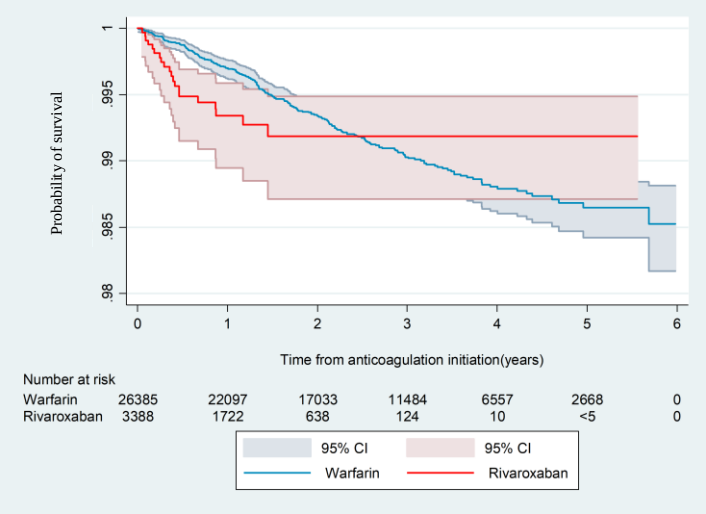


Figure IX-8 d (GI bleeding)

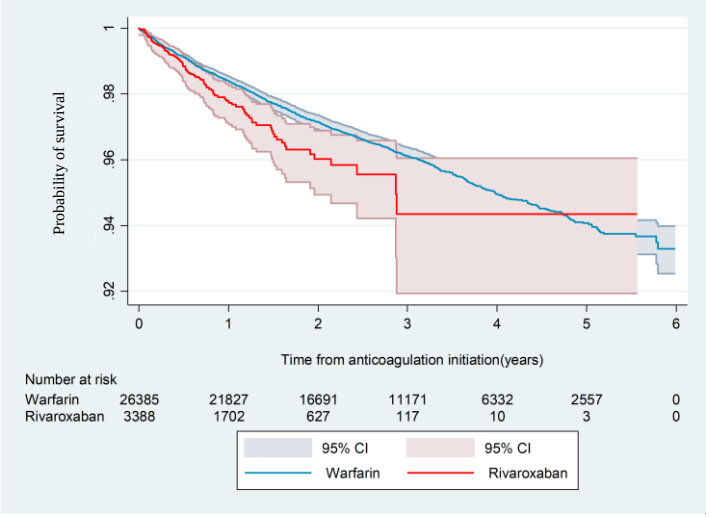


Figure IX-8: MI, ICH and bleeding (rivaroxaban vs. warfarin)

Abbreviations: MI=myocardial, ICH= intracranial haemorrhage, GI=gastrointestinal

Figure IX-9 a (mortality-all-cause)

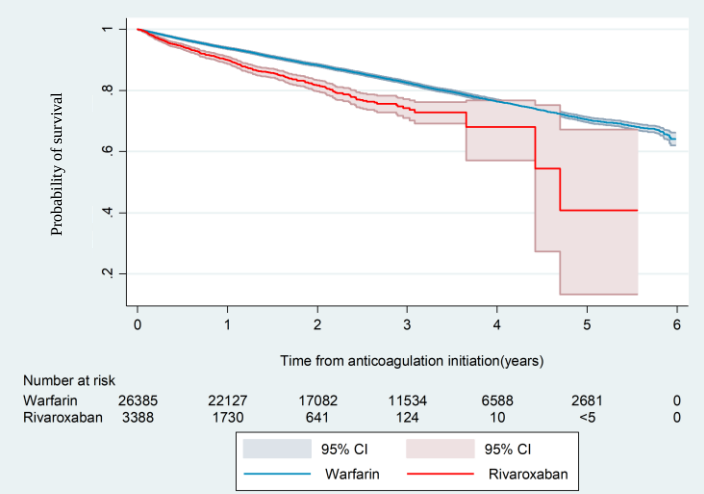


Figure IX-9 b (mortality-stroke)

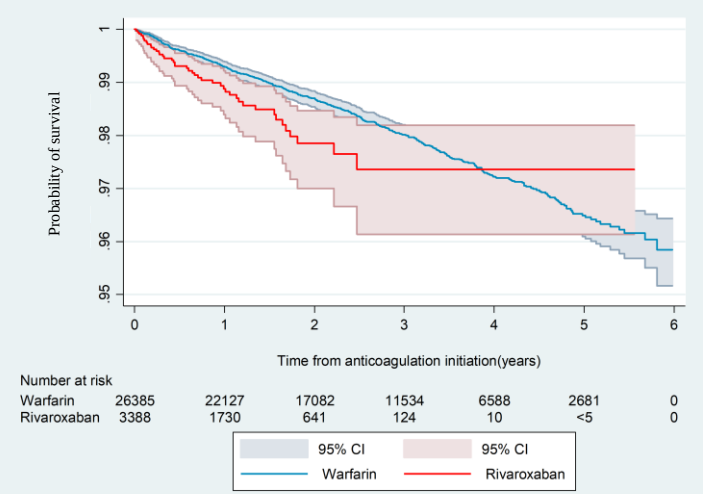


Figure IX-9 c (mortality-cardiovascular)

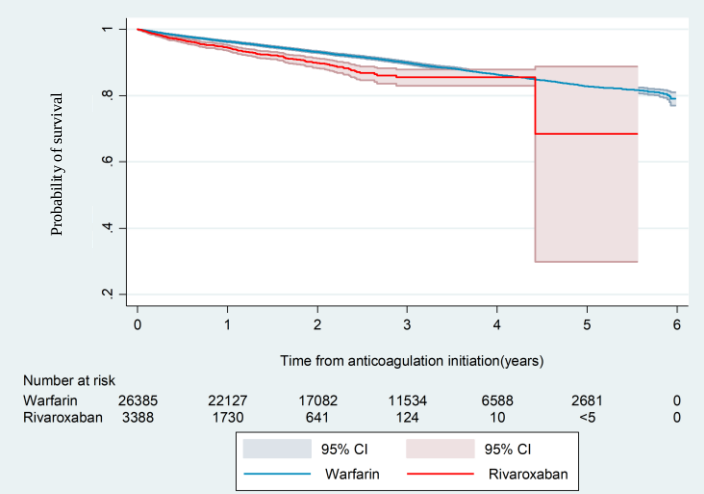


Figure IX-9: Mortality (rivaroxaban vs. warfarin)

Appendix X: Cumulative incidence curves

Apixaban versus warfarin

While for apixaban, the cumulative incidence curves were estimated within the first two years from anticoagulation initiation, for warfarin patients the estimates expanded beyond the 2 years across all clinical outcomes. As pointed out this is mainly due to patients on apixaban being exposed to the treatment for a much shorter period of time and experiencing most of the events within the first 2 years from anticoagulation initiation. In general, the probability of an event to occur, as depicted in the cumulative incidence curves (Appendix X, Figure X-1–Figure X-3), indicated that overall within the first 2 years from therapy initiation, patients on apixaban were more likely to have an event than patients who were on warfarin; nevertheless, in the first year from starting therapy, the difference in risk seemed to be negligible.

Dabigatran versus warfarin

Overall, the probability of experiencing a clinical event in the dabigatran group is lower than the probability observed in the warfarin treatment arm (Appendix X, Figure X-4–Figure X-6). The cumulative incidence curves reflecting the probability of experiencing an event were not reported if associated with less than five events (Figure X-5 a, Figure X-5 c and Figure X-6 b). As shown in Figure X-5 b, the direction of the probability of experiencing a major bleeding event changes over time. Further, Figure X-5 d seems to suggest that at 2 years from anticoagulation initiation, patients on dabigatran are more likely to experience a GI bleeding event.

Rivaroxaban versus warfarin

For most of the clinical outcomes, the probability for an event to occur seems to be greater in rivaroxaban than in the warfarin treatment group (Appendix X, Figure XI-7–X-9); however, the cumulative incidence curve for MI does not seem to follow a clear pattern. Following the first 2 years from anticoagulation initiation, the probability of SE or TIA occurring appear to increase for warfarin patients; but only a few events are observed (Figure X-7 c).

Figure X-1 a (stroke-all)

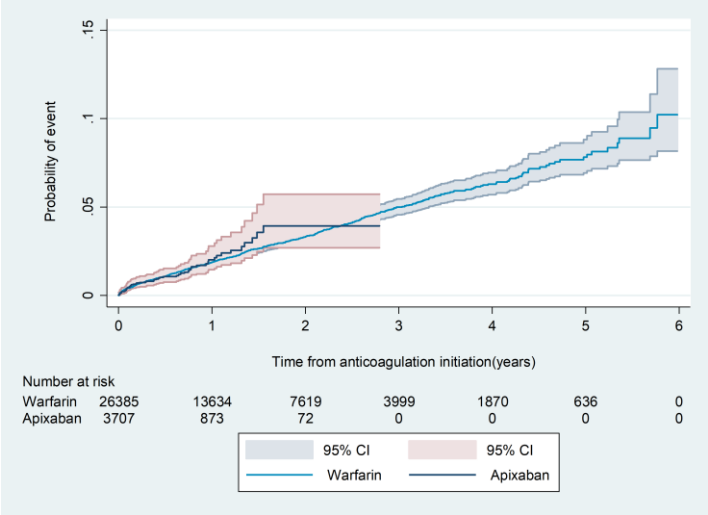


Figure X-1 b (stroke or SE)

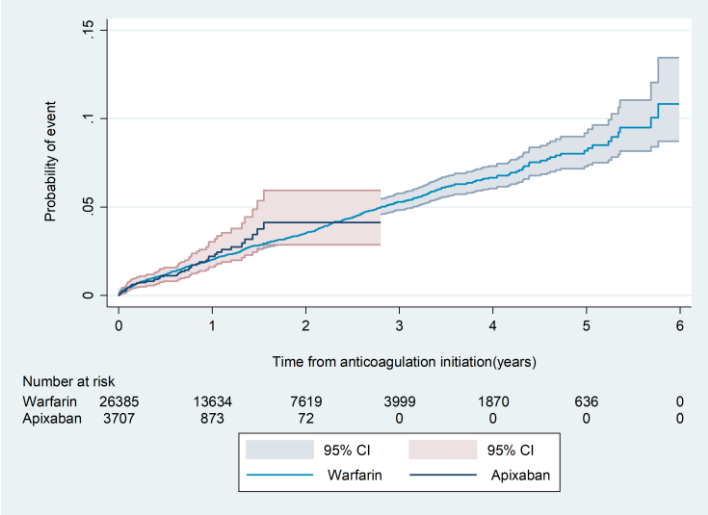


Figure X-1 c (stroke or SE or TIA)

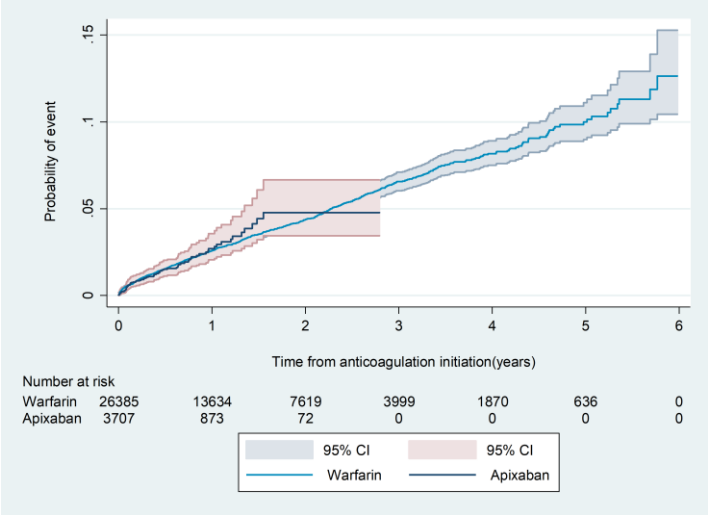


Figure X-1 d (stroke or SE or mortality-all-cause)

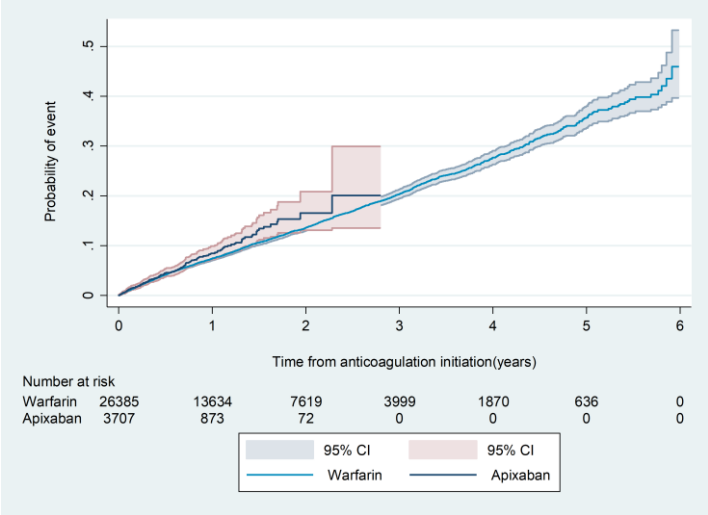


Figure X-1: Stroke, SE and TIA (apixaban vs. warfarin)

Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack

Figure X-2 a (MI)

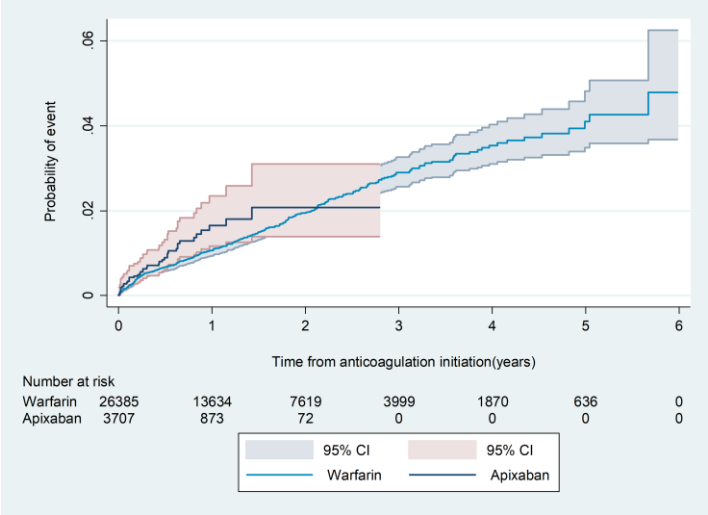


Figure X-2 b (major bleeding)

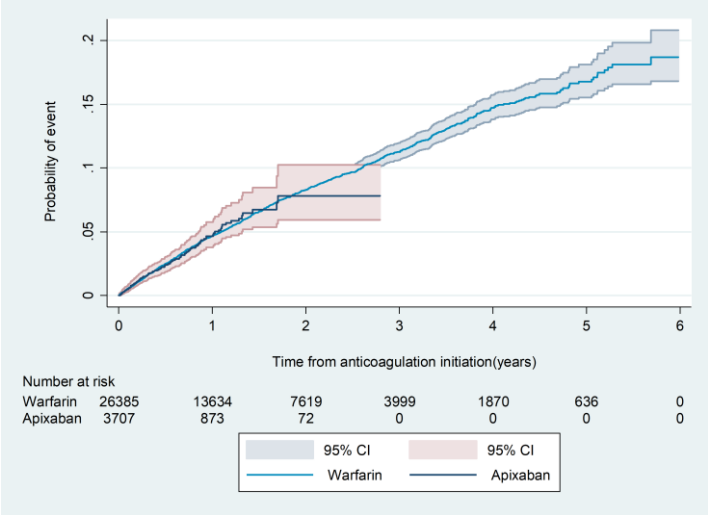


Figure X-2 c (ICH)

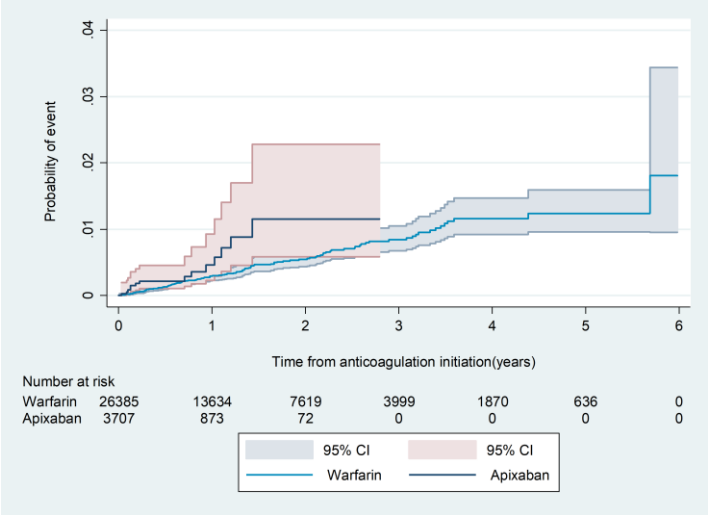


Figure X-2 d (GI bleeding)

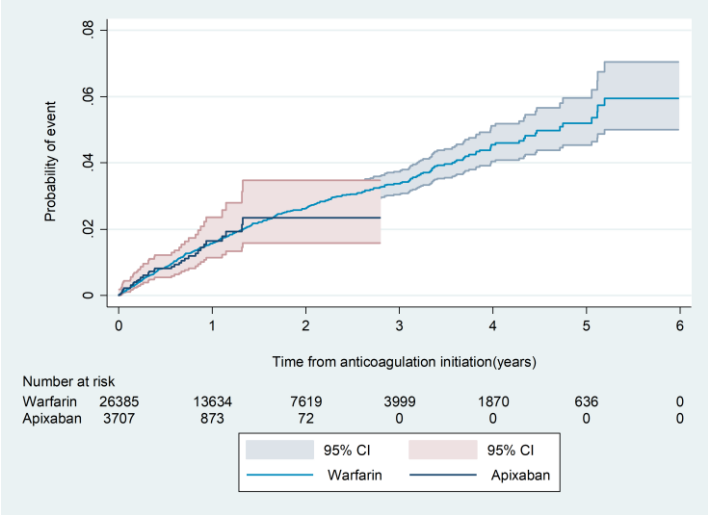


Figure X-2: MI, ICH and bleeding (apixaban vs. warfarin)

Abbreviations: MI=myocardial, ICH= intracranial haemorrhage, GI=gastrointestinal

Figure X-3 a (mortality-all-cause)

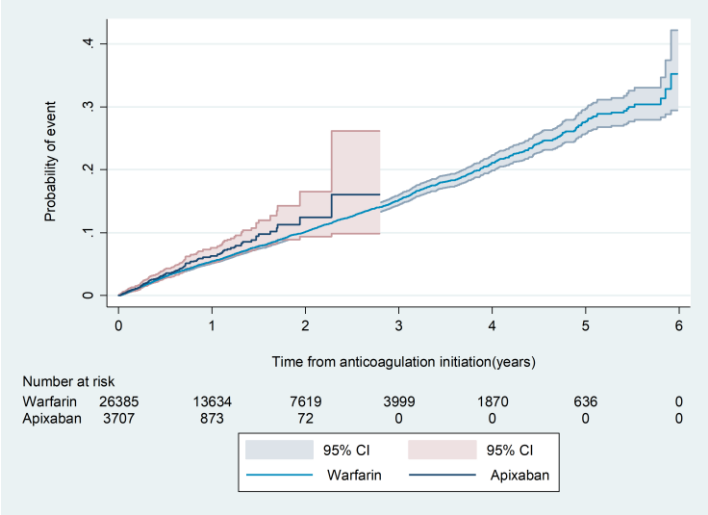


Figure X-3 b (mortality-stroke)

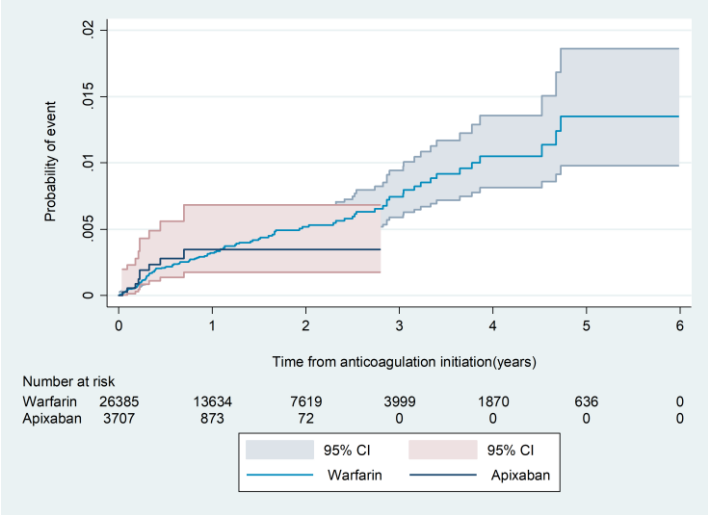


Figure X-3 c (mortality-cardiovascular)

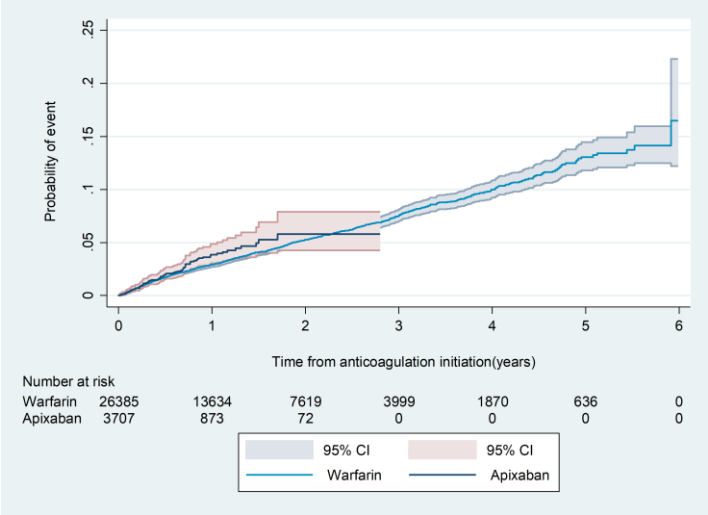


Figure X-3: Mortality (apixaban vs. warfarin)

Figure X-4 a (stroke-all)

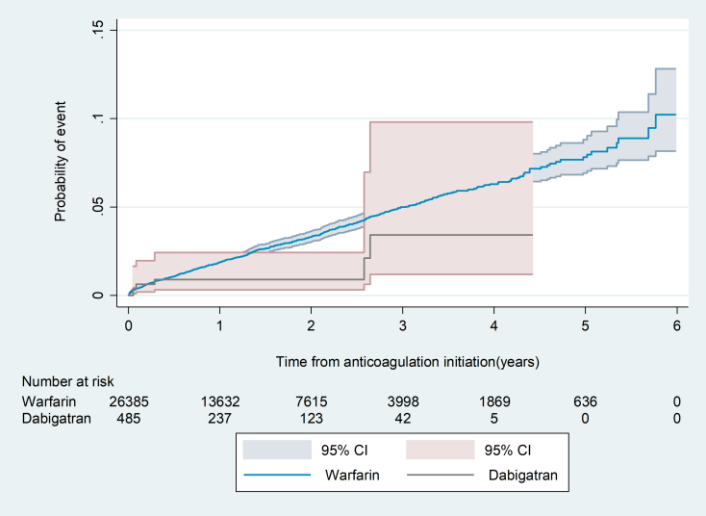


Figure X-4 b (stroke or SE)

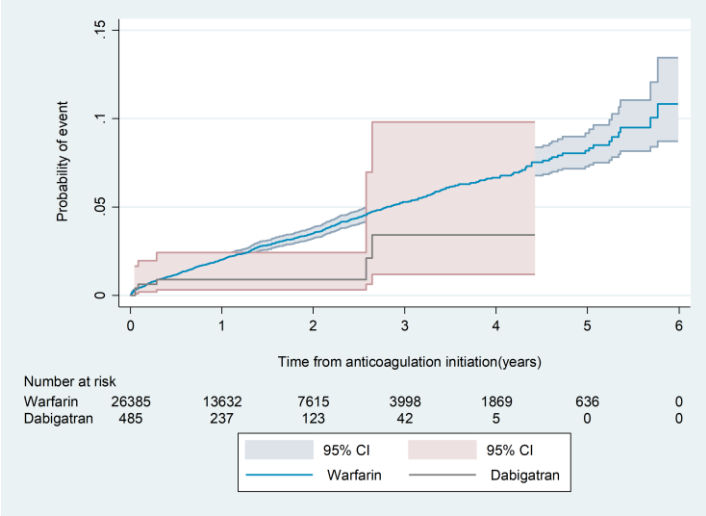


Figure X-4 c (stroke or SE or TIA)

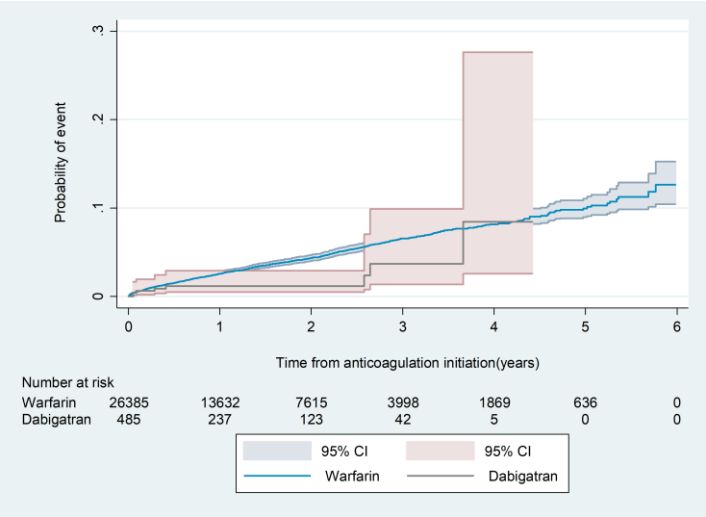


Figure X-4 d (stroke or mortality-all-cause)

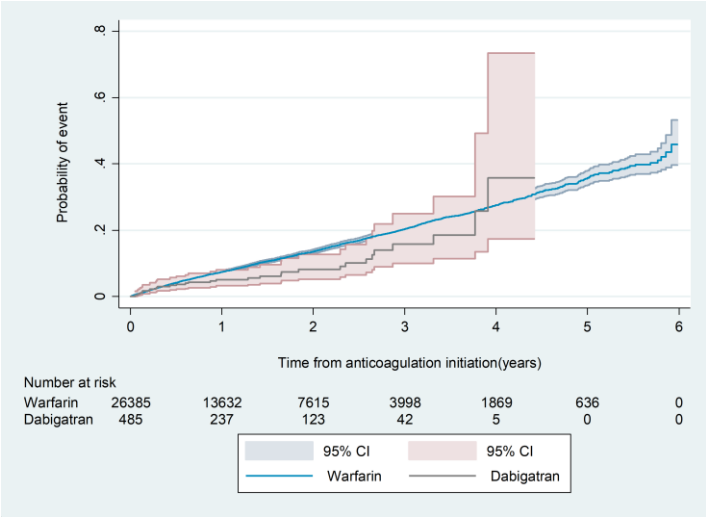
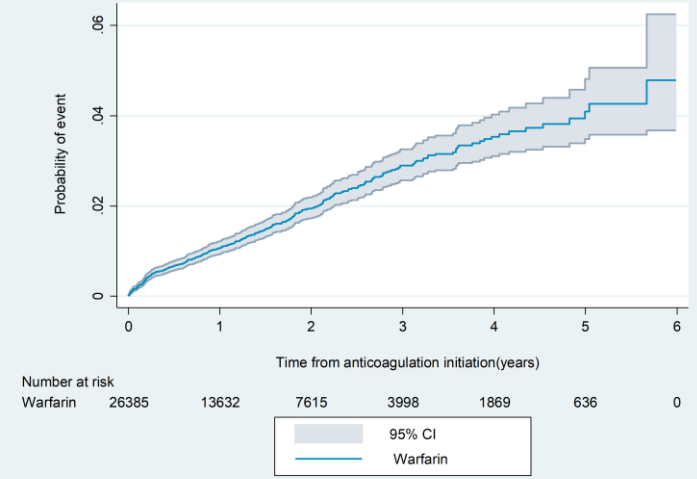


Figure X-4: Stroke, SE and TIA (dabigatran vs. warfarin)

Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack

Figure X-5 a (MI)



Figurer X-5 b (major bleeding)

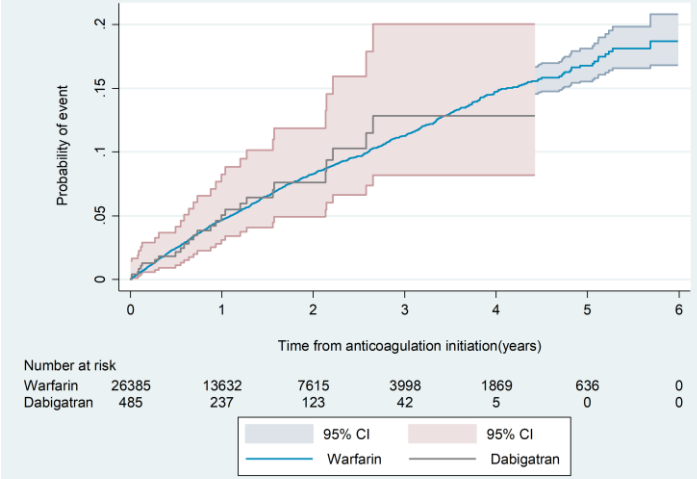


Figure X-5 c (ICH)

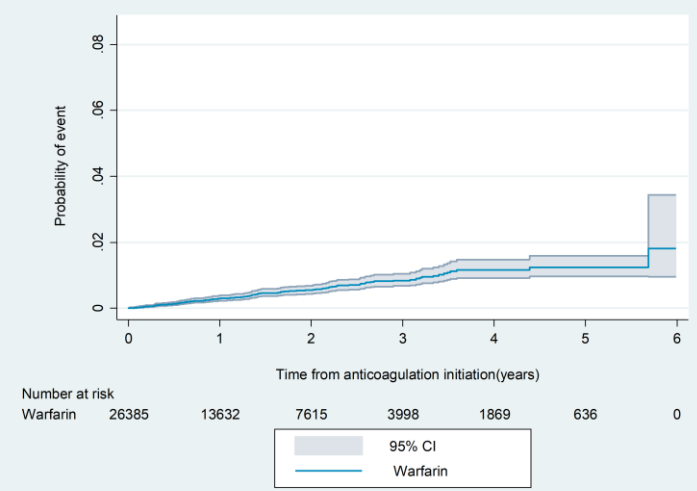
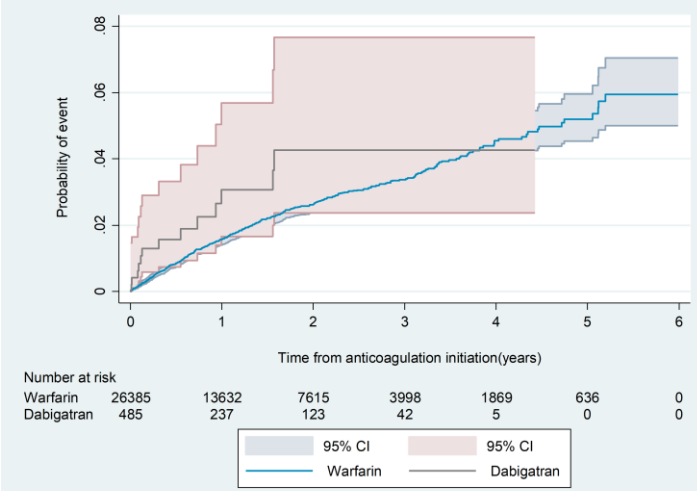


Figure X-5 d (GI bleeding)



Figurer X-5: MI, ICH and bleeding (dabigatran mg vs. warfarin)

Note: due to disclosure restrictions, in the case of fewer than five events, the cumulative incidence curve was not reported

Abbreviations: MI=myocardial, ICH= intracranial haemorrhage, GI=gastrointestinal

Figure X-6 a (mortality-all-cause)

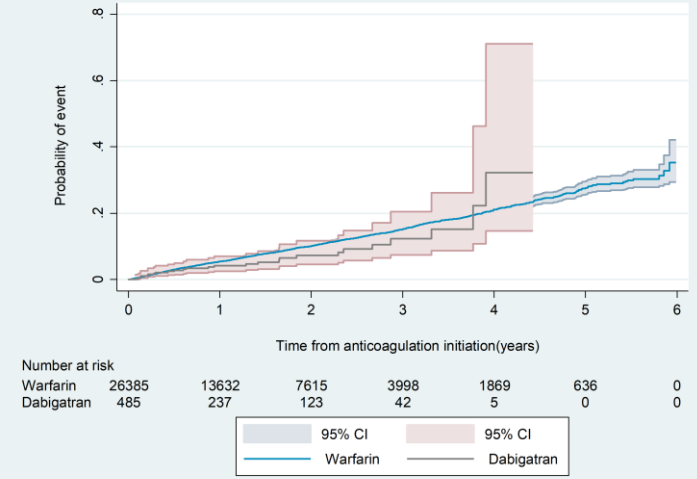


Figure X-6 b (mortality-stroke)

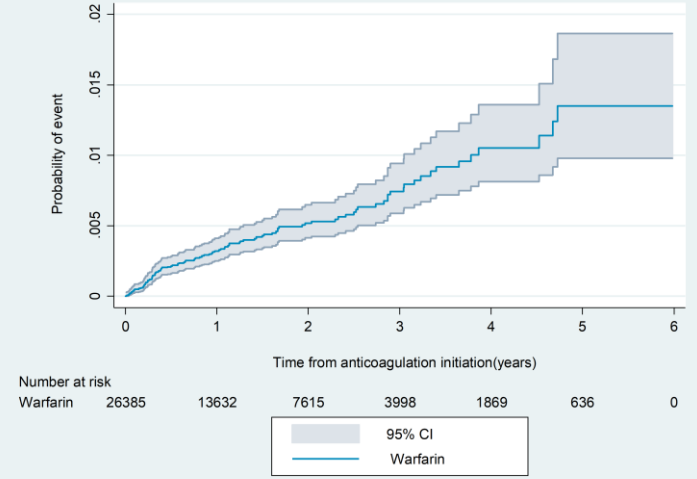


Figure X-6 c (mortality-cardiovascular)

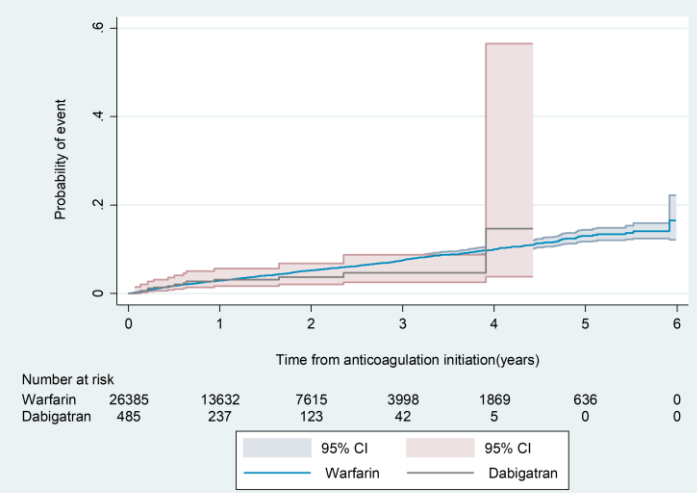


Figure X-6: Mortality (dabigatran vs. warfarin)

Note: due to disclosure restrictions, in the case of fewer than five events, the cumulative incidence curve was not reported

Figure X-7 a (stroke-all)

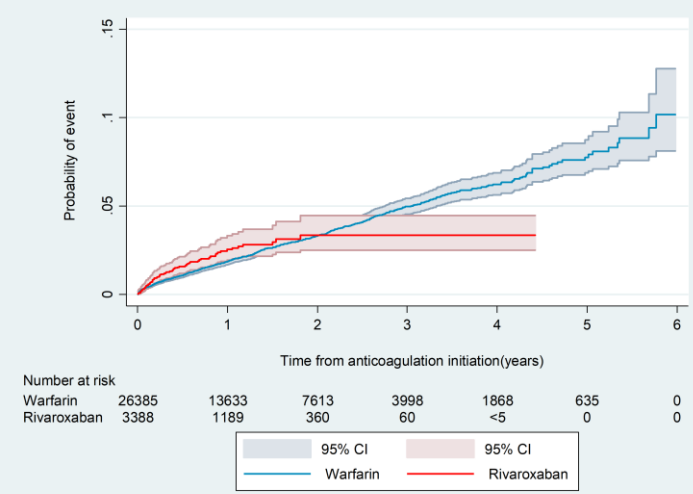


Figure X-7 b (stroke or SE)

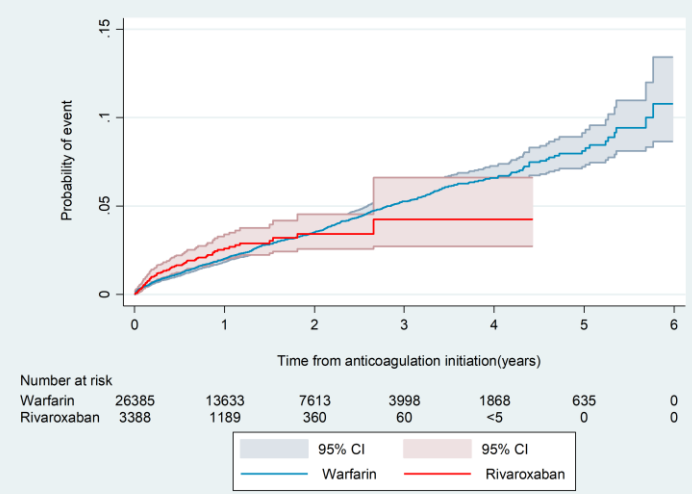


Figure X-7 c (stroke or SE or TIA)

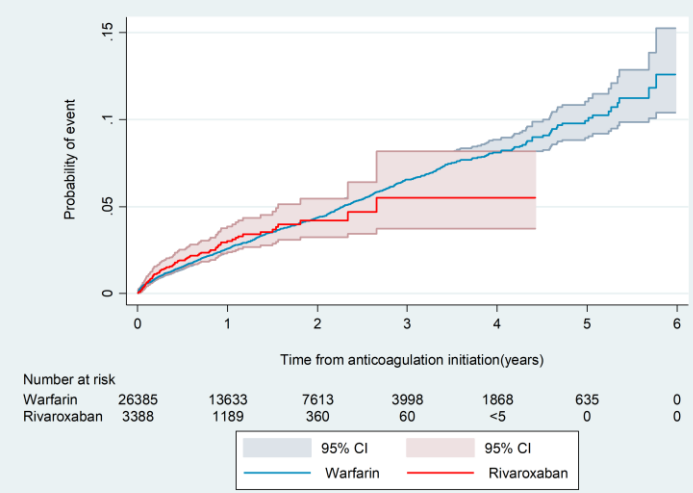


Figure X-7 d (stroke or mortality-all-cause)

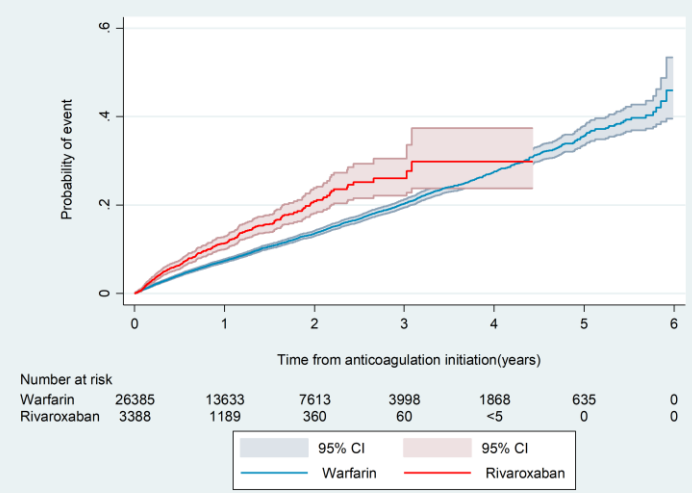


Figure X-7: Stroke, SE and TIA (rivaroxaban vs. warfarin)

Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack

Figure X-8 a (MI)

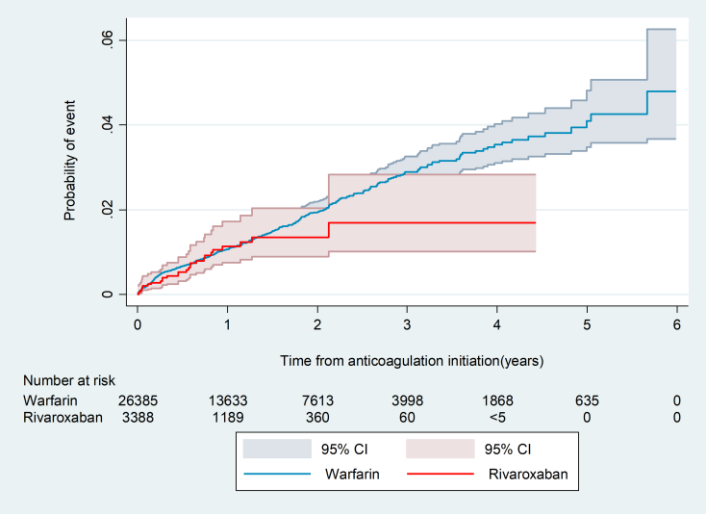


Figure X-8 b (major bleeding)

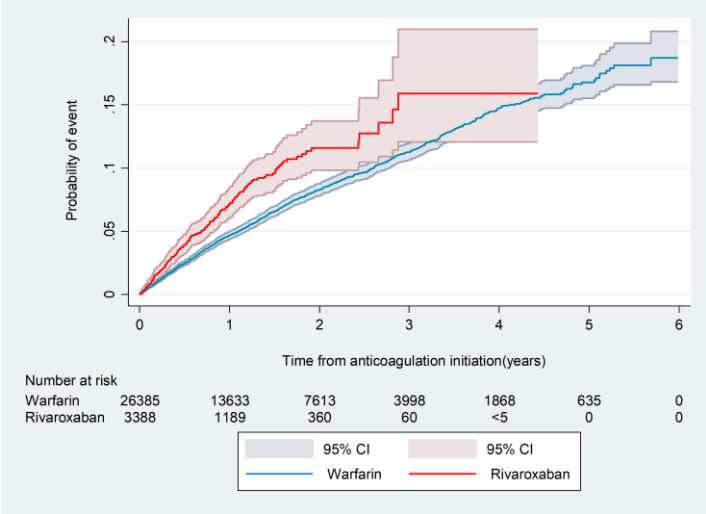


Figure X-8 c (ICH)

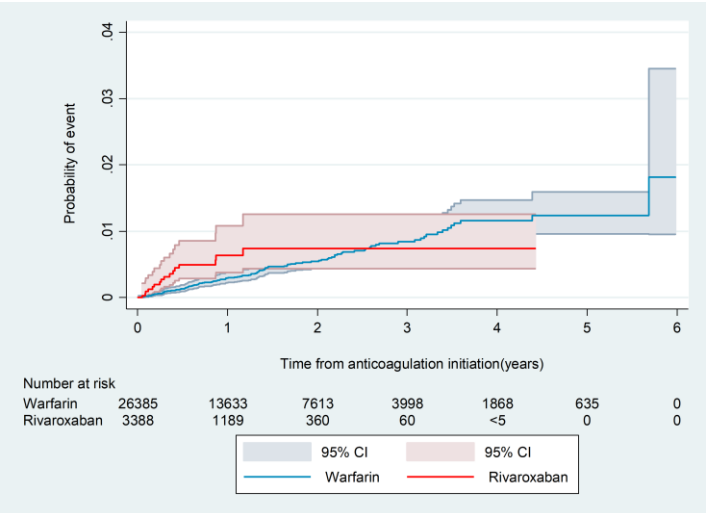


Figure X-8 d (GI bleeding)

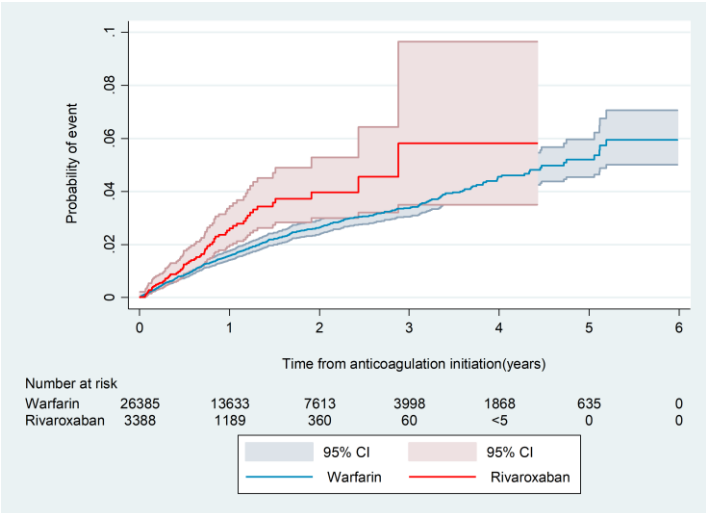


Figure X-8: MI, ICH and bleeding (rivaroxaban mg vs. warfarin)

Abbreviations: MI=myocardial, ICH= intracranial haemorrhage, GI=gastrointestinal

Figure X-9 a (mortality-all-cause)

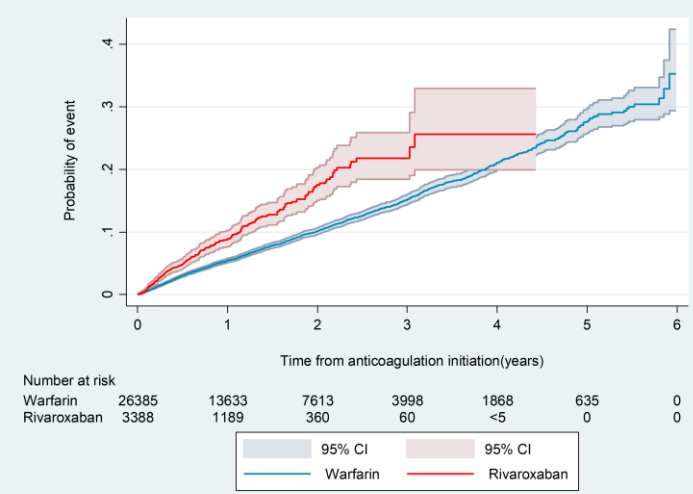


Figure X-9 b (mortality-stroke)

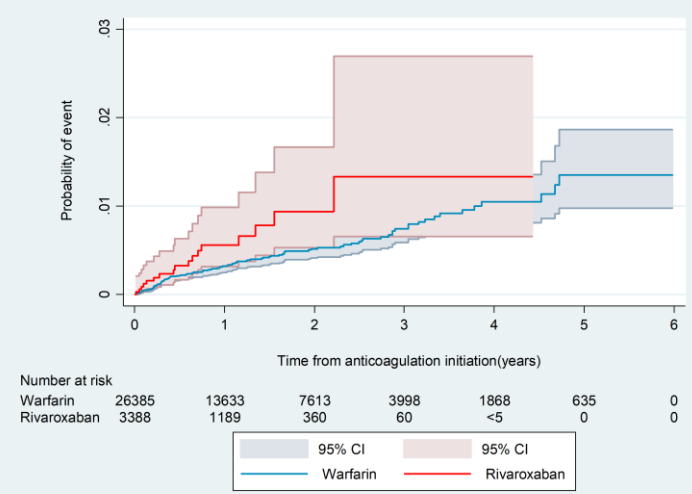


Figure X-9 c (mortality-cardiovascular)

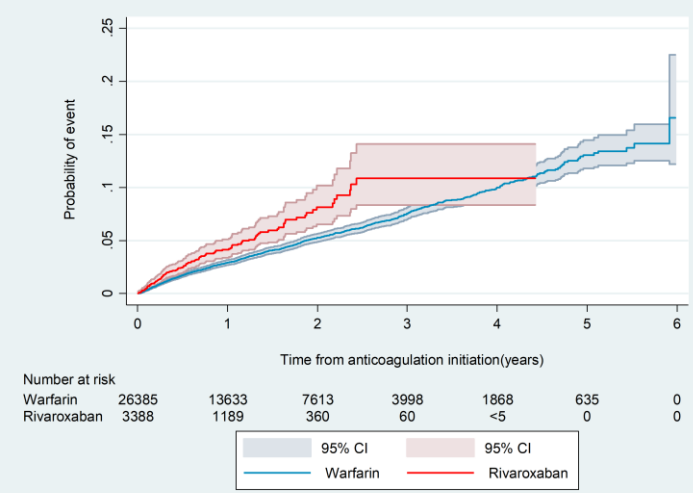


Figure X-9: Mortality (rivaroxaban vs. warfarin)

Appendix XI: Hazard ratios for 6 years follow-up

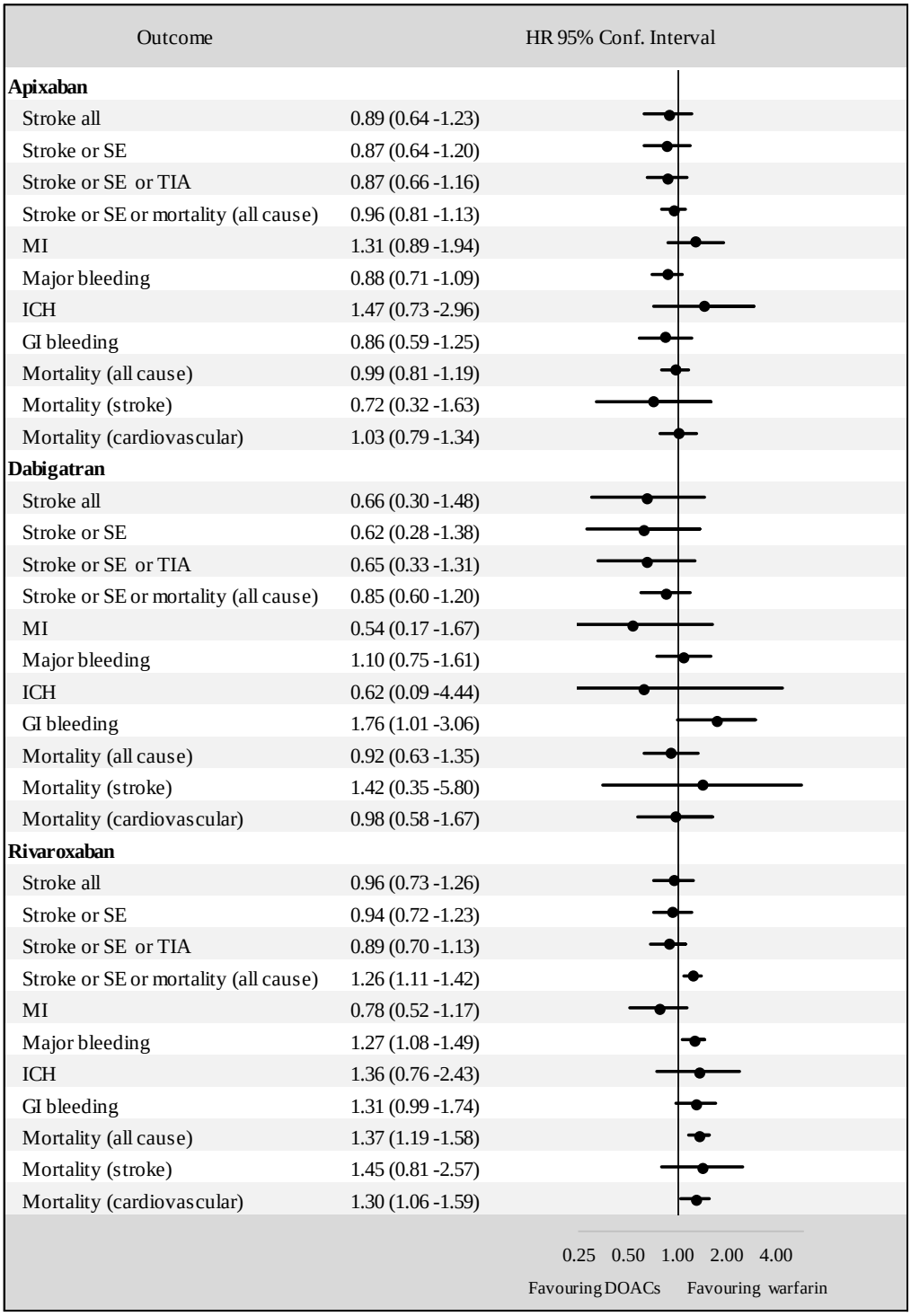


Figure XI-1: Hazard ratios for 6 years since first prescription, DOACs vs. warfarin
Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack, MI=myocardial infarction, ICH=intracranial haemorrhage, GI=gastrointestinal.

Appendix XII: Proportionality hazard assumption (plots of goodness of fit for standard and reduced dose)

Apixaban versus warfarin

Very similar trends can be observed when assessing proportionality assumptions violation plots for apixaban reduced and standard dose (Figure XII-1–XII-6). However, violation for stroke and stroke composites in apixaban standard dose (Figure XII-1 a–XII-1c) seems to be less evident.

Rivaroxaban versus warfarin

The plots assessing proportionality for the standard dose of 20 mg rivaroxaban seem to follow the same patterns observed in the proportionality assessment of any dose rivaroxaban. Comparatively, When assessing proportionality violation of reduced rivaroxaban, the assumption of proportionality seem to be satisfied for some clinical outcomes such stroke (Figure XII-10 a), MI (Figure XII-11 a), ICH (Figure XII-11 c), GI bleeding (Figure XII-11 d),all-cause mortality (Figure IXI-12 a), and mortality due to cardiovascular conditions (Figure XII-12 c).

Table XII-1 (standard dose) and XII-2 (reduced dose) shows that overall, age and CCI where the variables violating the proportionality assumption across all treatments and clinical outcomes.

Figure XII-1 a (stroke-all)

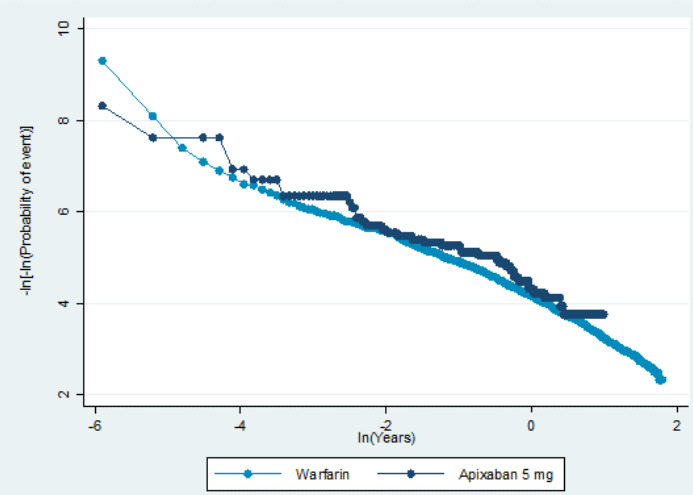


Figure XII-1 b (stroke or SE)

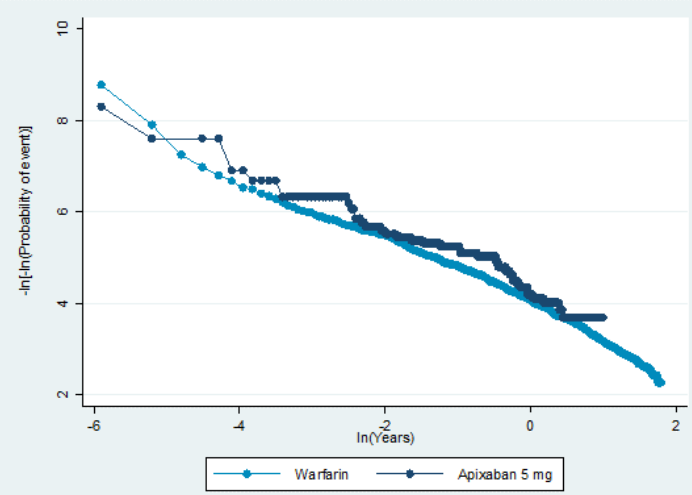


Figure XII-1 c (stroke or SE or TIA)

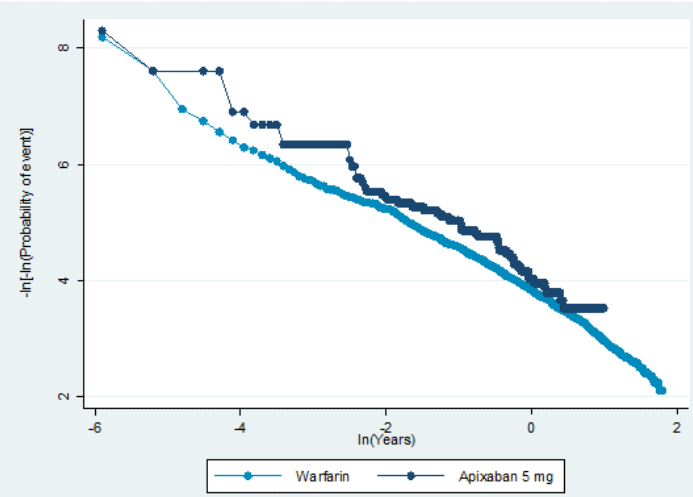


Figure XII-1 d (stroke or mortality-all-cause)

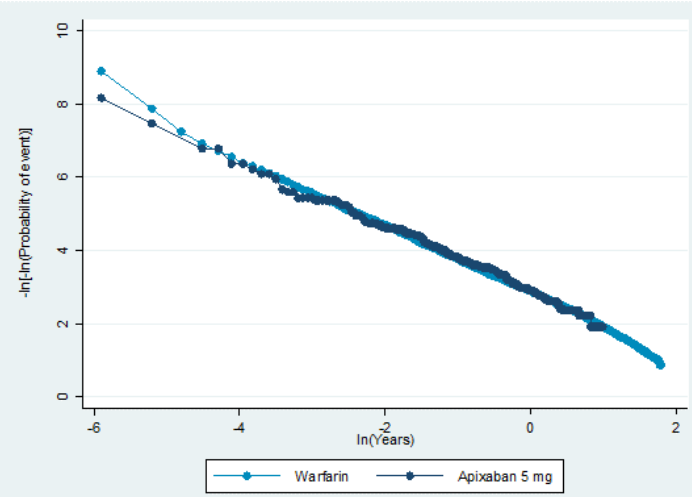


Figure XII-1 log-cumulative hazard plot for stroke, SE and TIA (apixaban 5 mg vs. warfarin)

Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack

Figure XII-2 a (MI)

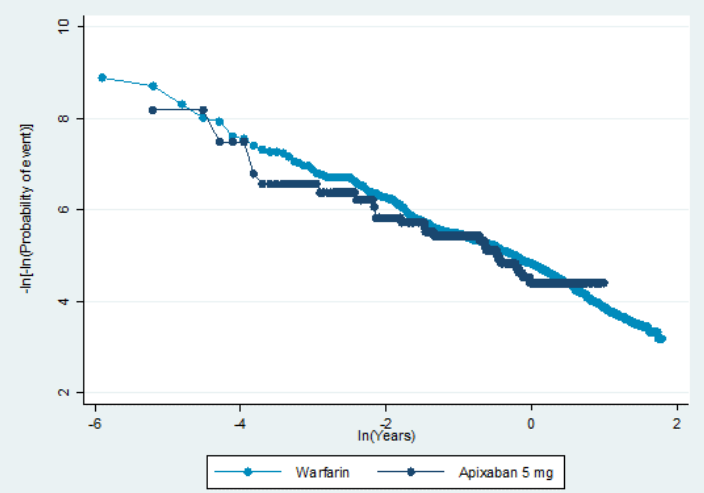


Figure XII-2 b (major bleeding)

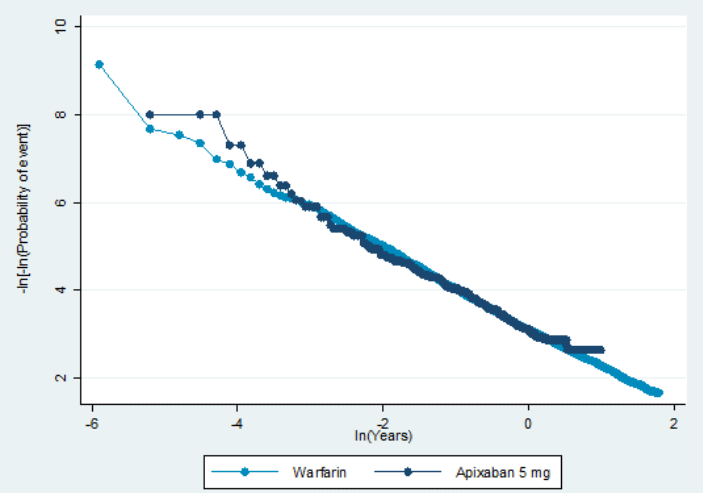


Figure XII-2 c (ICH)

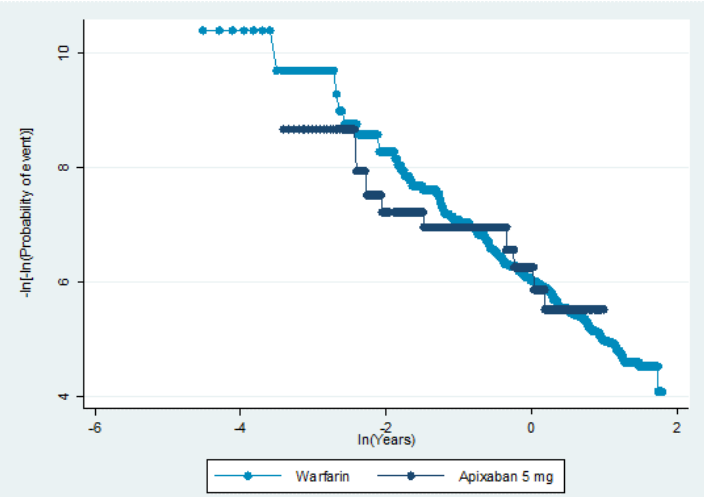


Figure XII-2 d (GI bleeding)

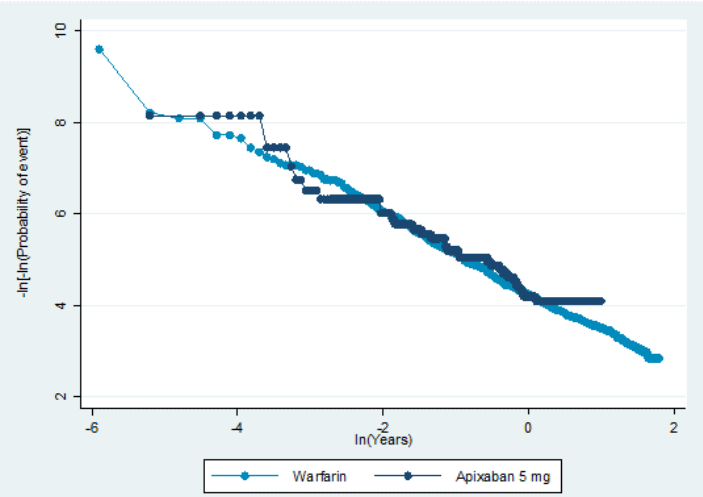


Figure XII-2: log-cumulative hazard plot for MI, ICH and bleeding (apixaban 5 mg vs. warfarin)

Abbreviations: MI=myocardial, ICH= intracranial haemorrhage, GI=gastrointestinal

Figure XII-3 a (mortality-all-cause)

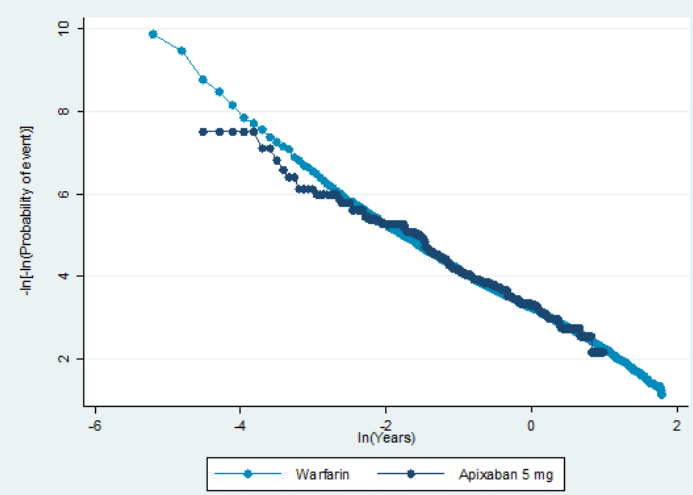


Figure XII-3 b (mortality-stroke)

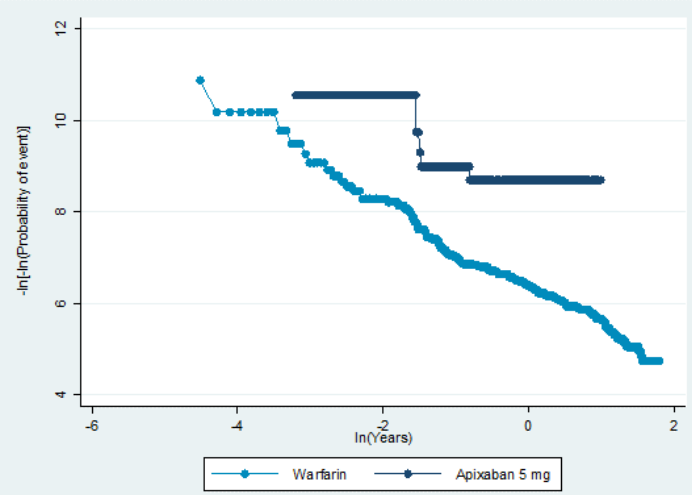


Figure XII-3 c (mortality-cardiovascular)

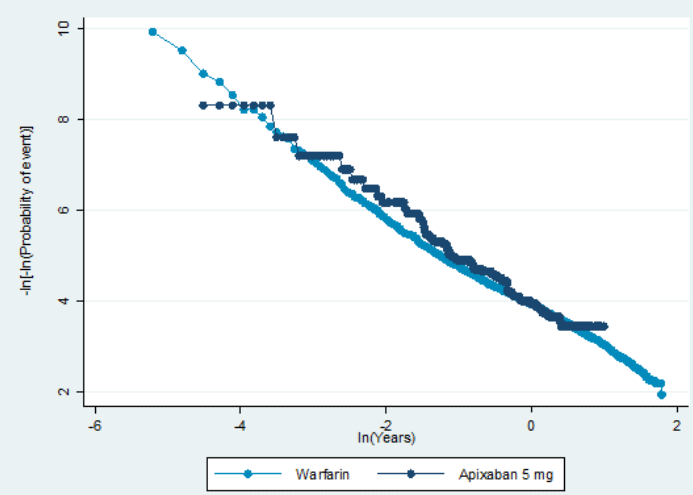


Figure XII-3: log-cumulative hazard plot for mortality (apixaban 5 mg vs. warfarin)

Figure XII-4 a (stroke-all)

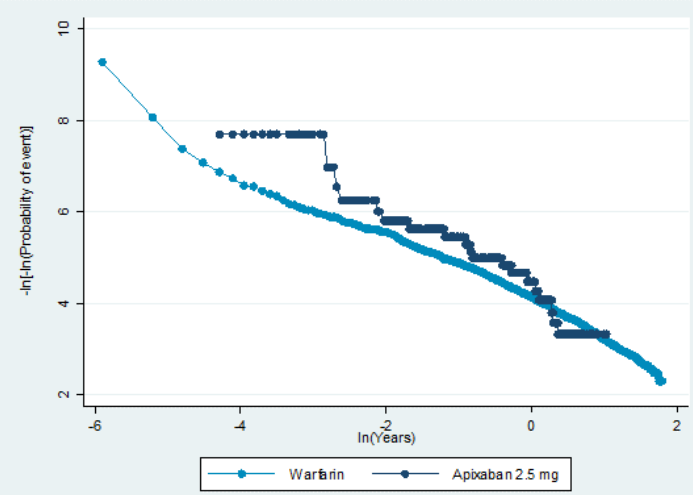


Figure XII-4 b (stroke or SE)

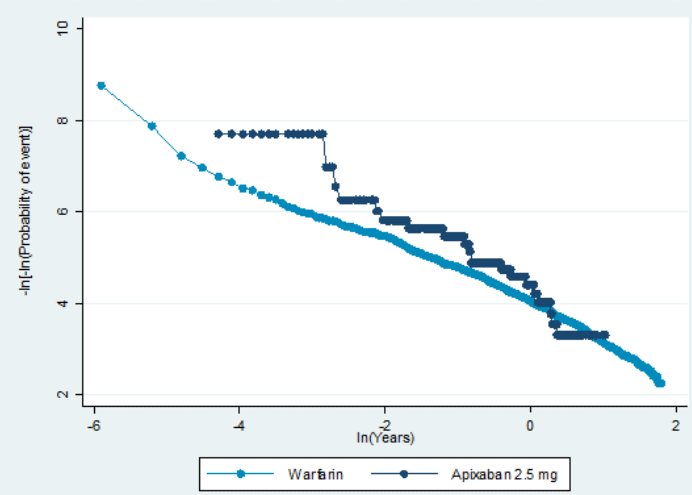


Figure XII-4 c (stroke or SE or TIA)

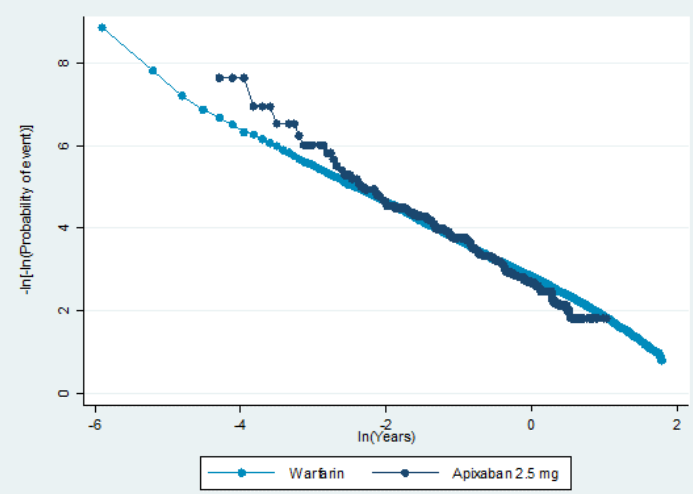


Figure XII-4 d (stroke or mortality-all-cause)

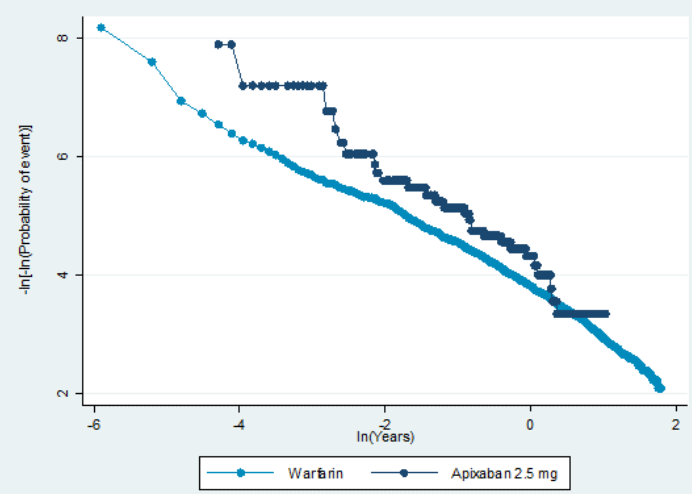


Figure XII-4 log-cumulative hazard plot for stroke, SE and TIA (apixaban 2.5 mg vs. warfarin)
Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack

Figure XII-5 a (MI)

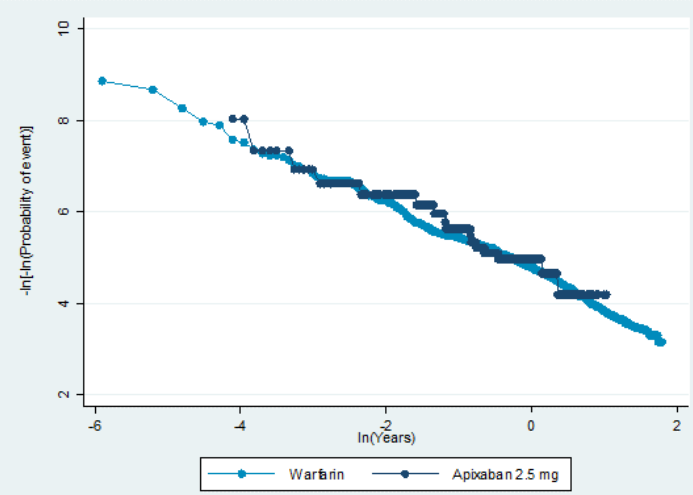


Figure XII-5 b (major bleeding)

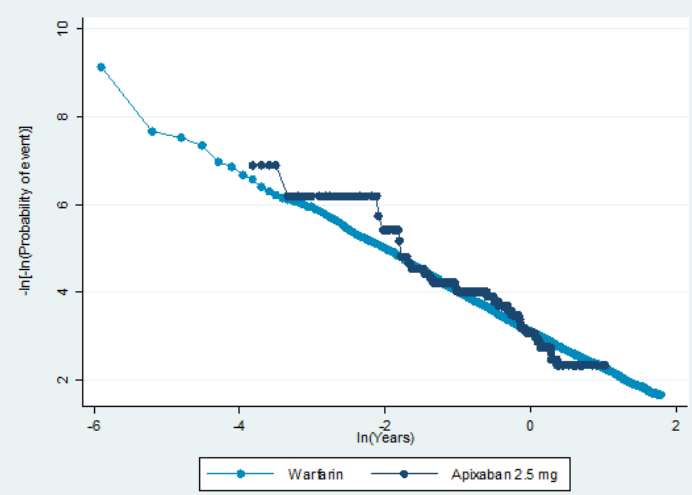


Figure XII-5 c (ICH)

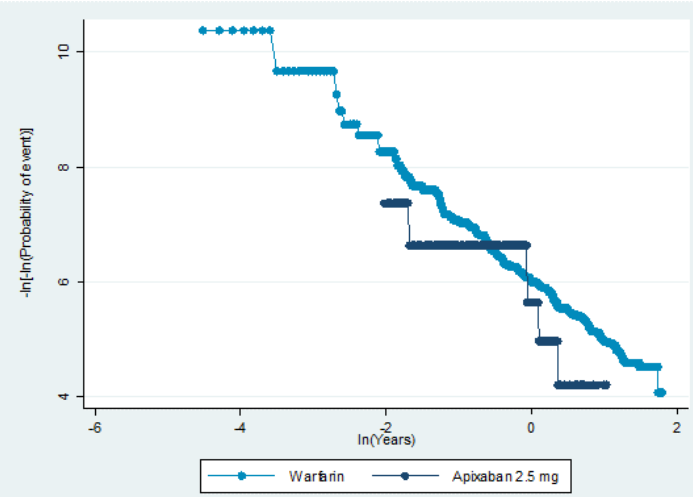


Figure XII-5 d (GI bleeding)

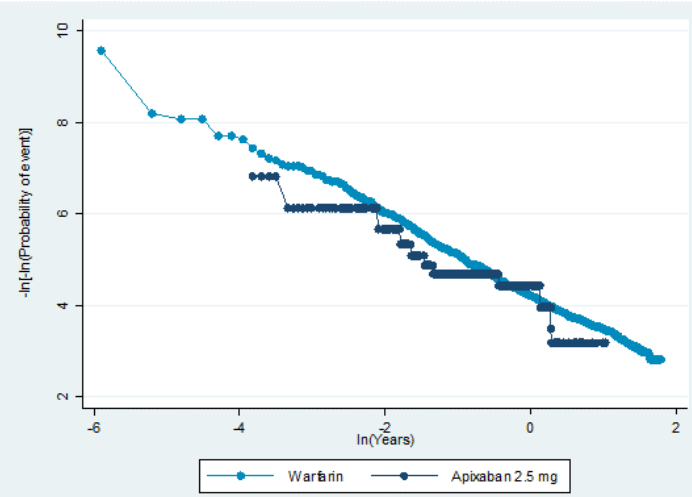


Figure XII-5: log-cumulative hazard plot for MI, ICH and bleeding (apixaban 2.5 mg vs. warfarin)
Abbreviations: MI=myocardial, ICH= intracranial haemorrhage, GI=gastrointestinal

Figure XII-6 a (mortality-all-cause)

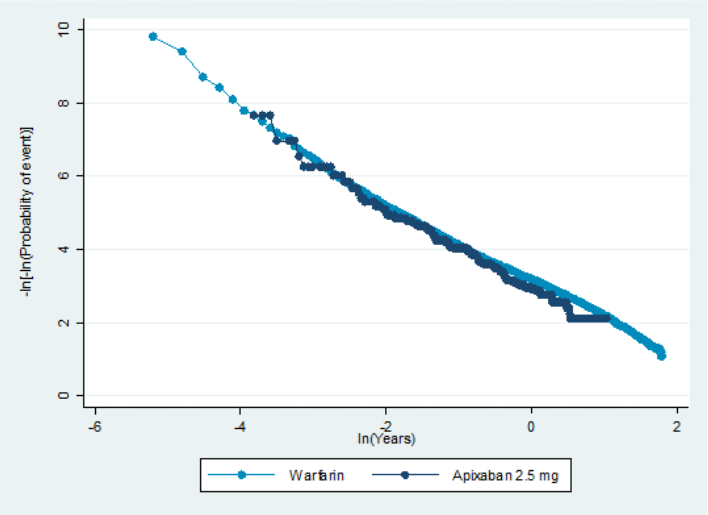


Figure XII-6 b (mortality-stroke)

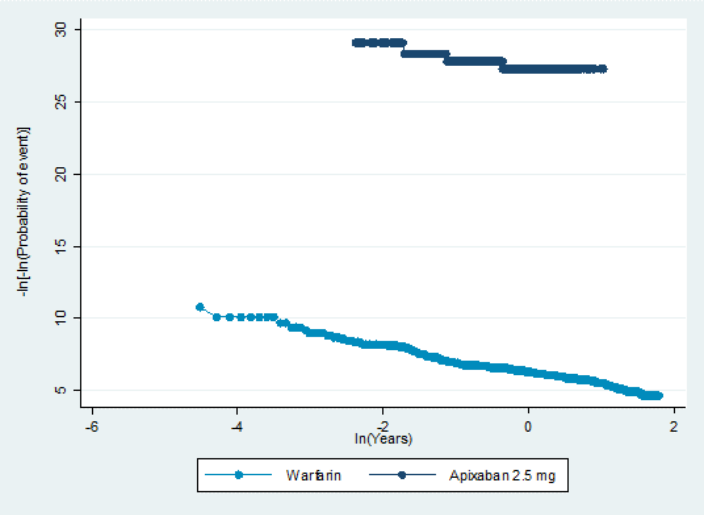


Figure XII-6 c (mortality-cardiovascular)

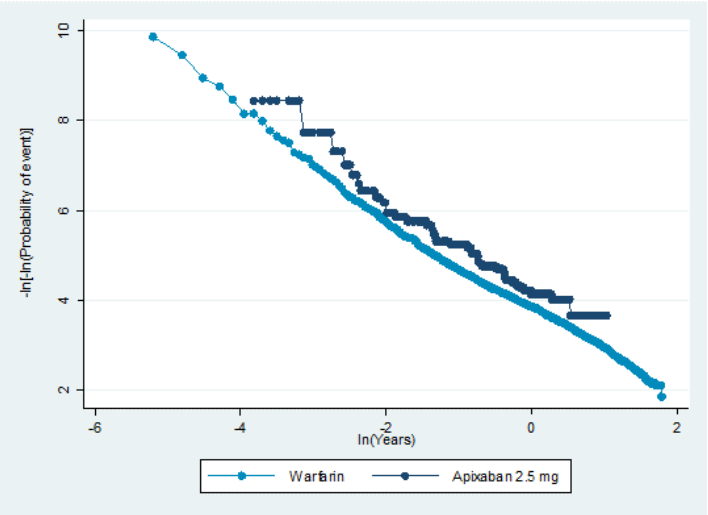


Figure XII-6: log-cumulative hazard plot for mortality (apixaban 2.5 mg vs. warfarin

Figure XII-7 a (stroke-all)

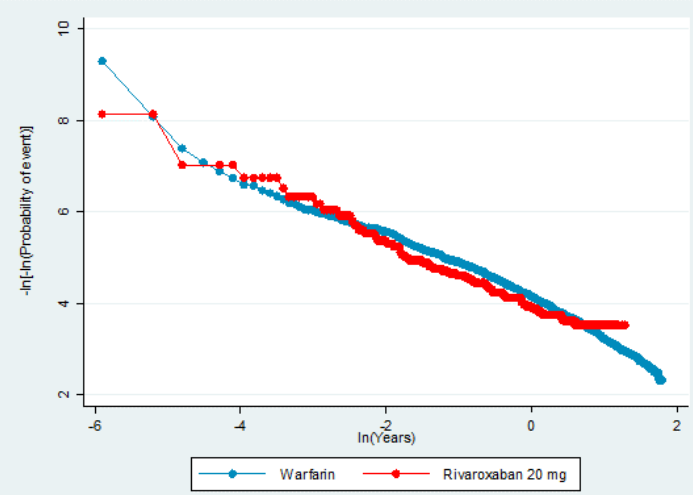


Figure XII-7 b (stroke or SE)

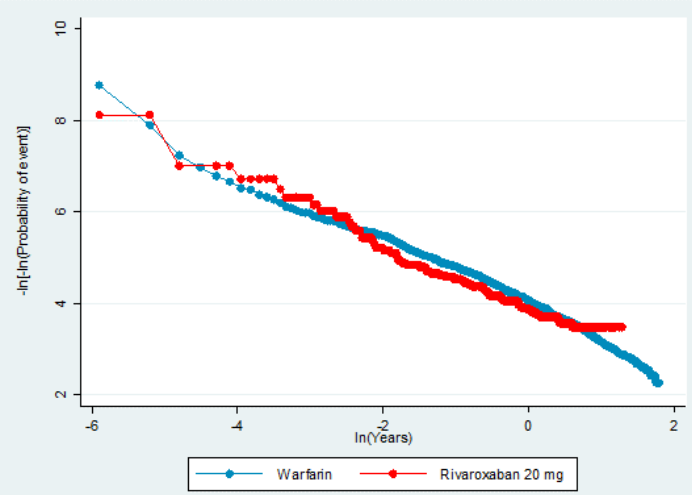


Figure XII-7 c (stroke or SE or TIA)

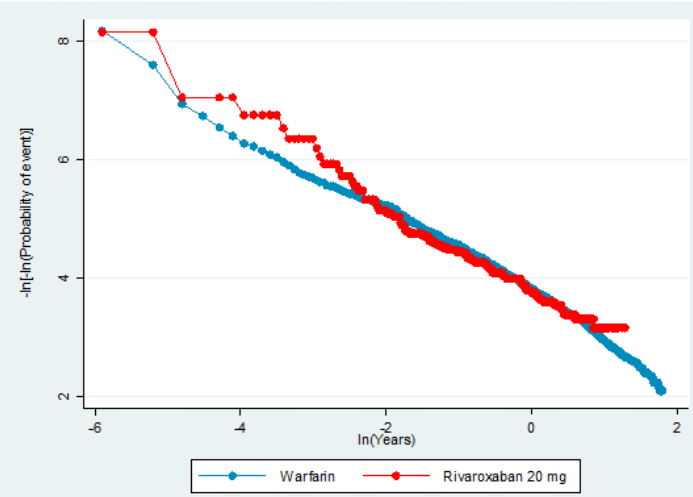


Figure XII-7 d (stroke or mortality-all-cause)

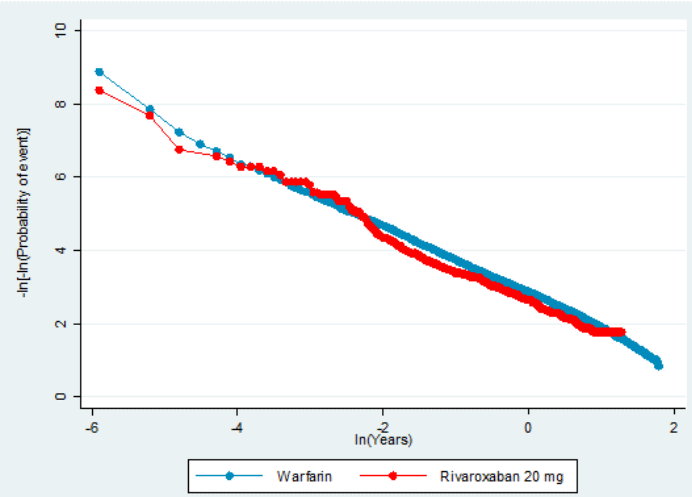


Figure XII-7: log-cumulative hazard plot for stroke, SE and TIA (rivaroxaban 20 mg vs. warfarin)
Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack

Figure XII-8 a (MI)

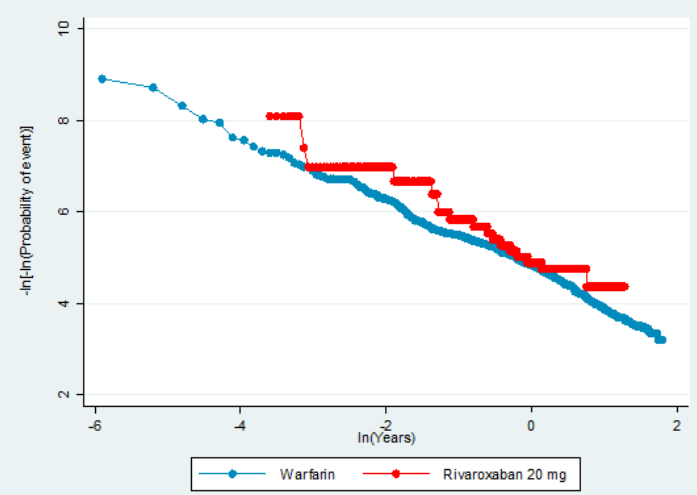


Figure XII-8 b (major bleeding)

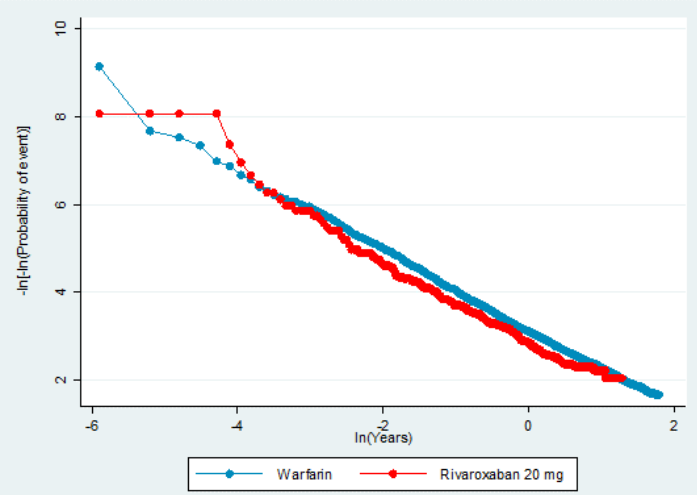


Figure XII-8 c (ICH)

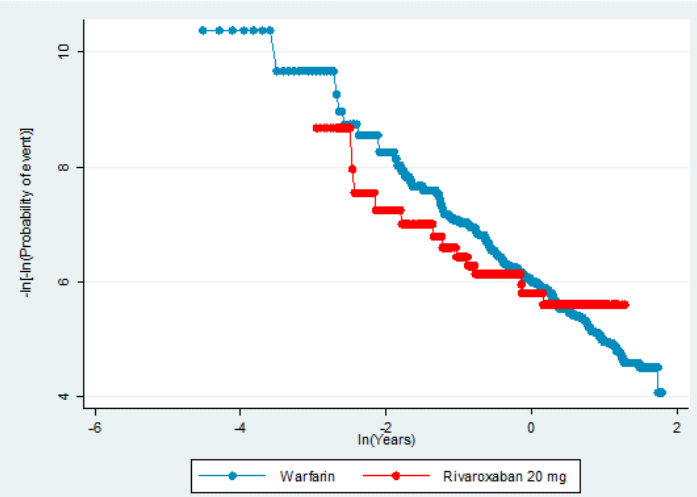


Figure XII-8 d (GI bleeding)

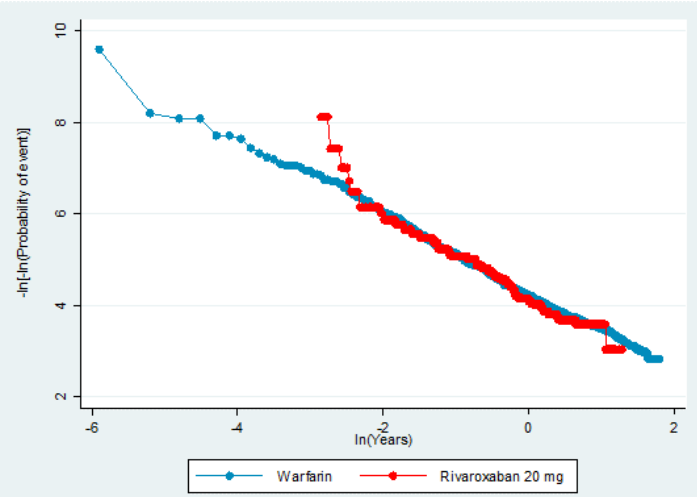


Figure XII-8: log-cumulative hazard plot for MI, ICH and bleeding (rivaroxaban 20 mg vs. warfarin)
Abbreviations: MI=myocardial, ICH= intracranial haemorrhage, GI=gastrointestinal

Figure XII-9 a (mortality-all-cause)

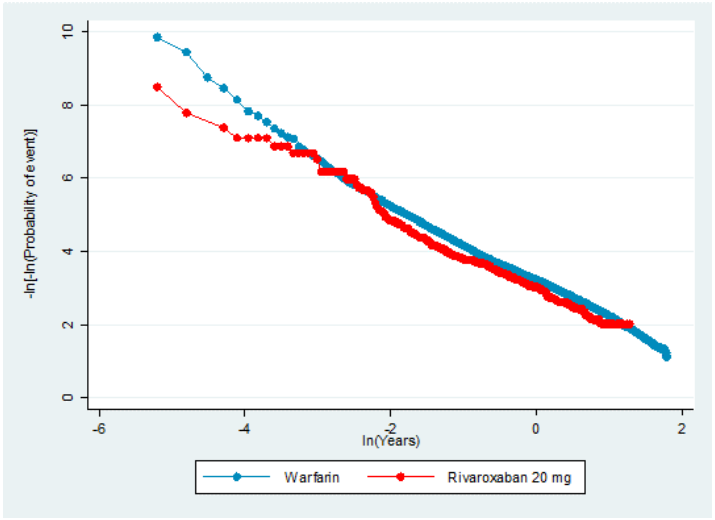


Figure XII-9 b (mortality-stroke)

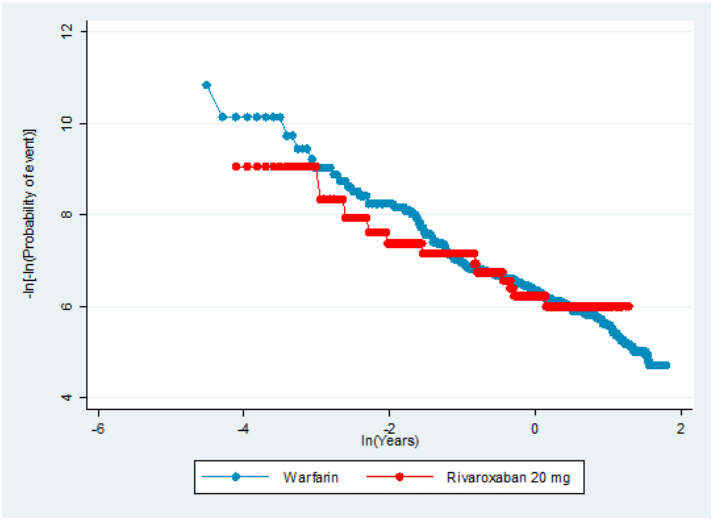


Figure XII-9 c (mortality-cardiovascular)

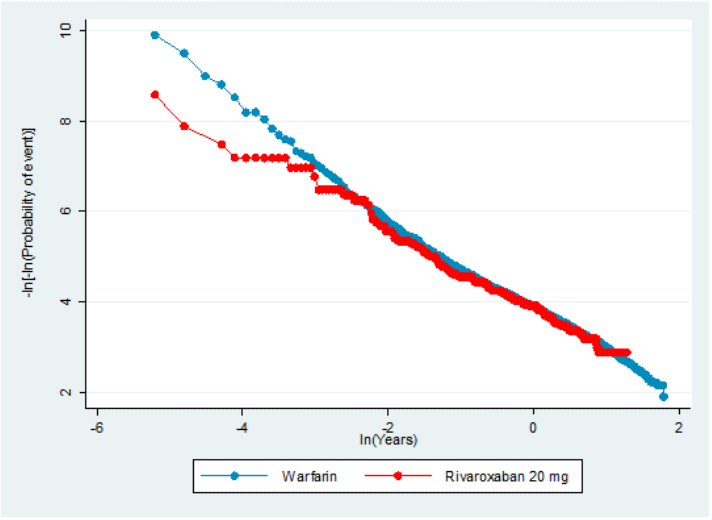


Figure XII-9: log-cumulative hazard plot for mortality (rivaroxaban 20 mg vs. warfarin)

Figure XII-10 a (stroke-all)

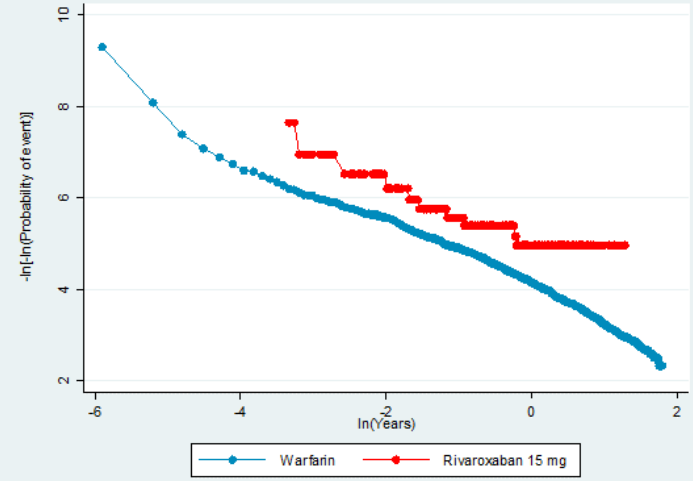


Figure XII-10 b (stroke or SE)

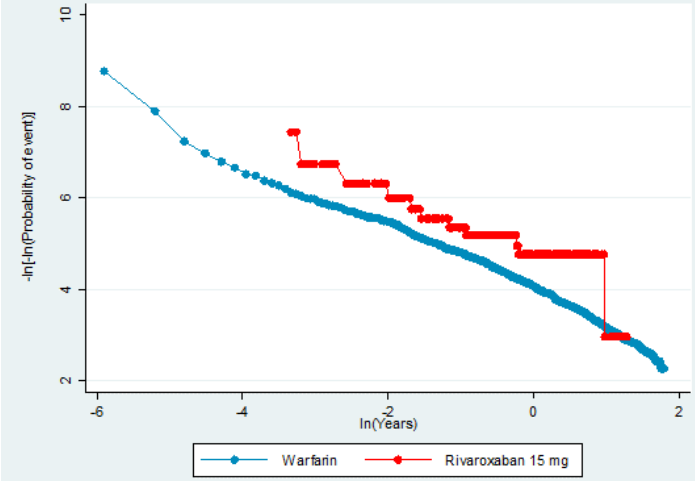


Figure XII-10 c (stroke or SE or TIA)

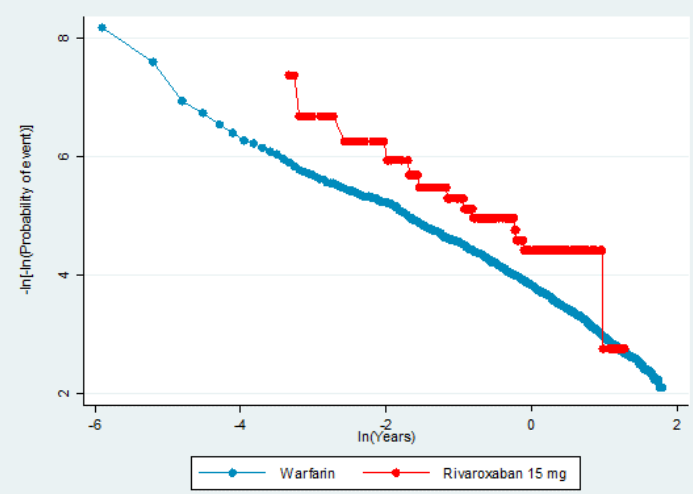


Figure XII-10 d (stroke or mortality-all-cause)

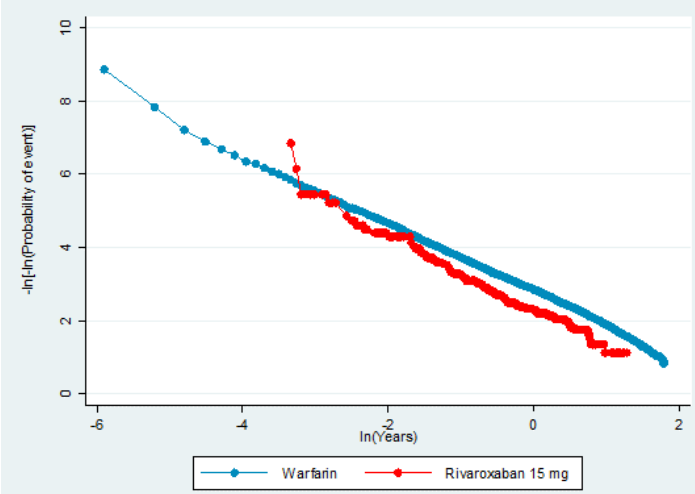


Figure XII-10: log-cumulative hazard plot for stroke, SE and TIA (rivaroxaban 15 mg vs. warfarin)
Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack

Figure XII-11 a (MI)

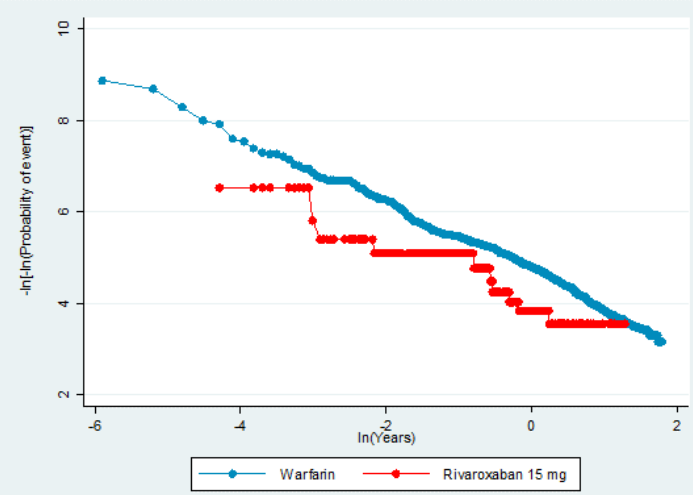


Figure XII-11 b (major bleeding)

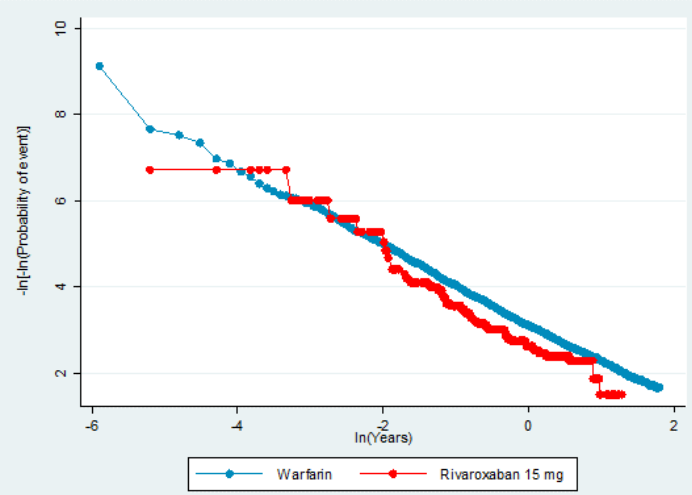


Figure XII-11 c (ICH)

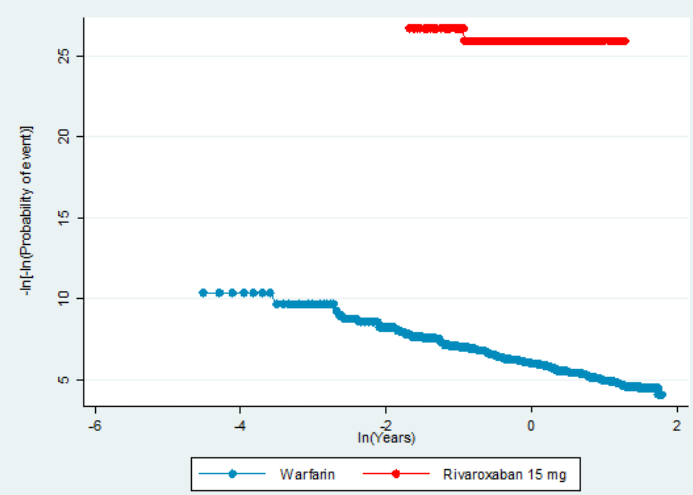


Figure XII-11 d (GI bleeding)

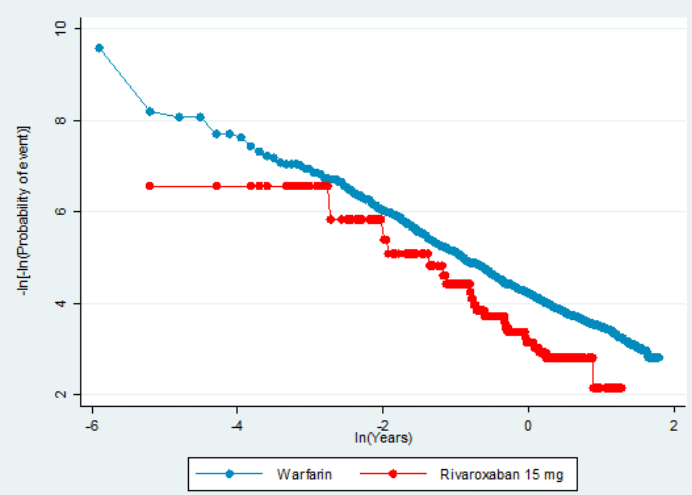


Figure XII-11: log-cumulative hazard plot for MI, ICH and bleeding (rivaroxaban 15 mg vs. warfarin)
Abbreviations: MI=myocardial, ICH= intracranial haemorrhage, GI=gastrointestinal

Figure XII-12 a (mortality-all-cause)

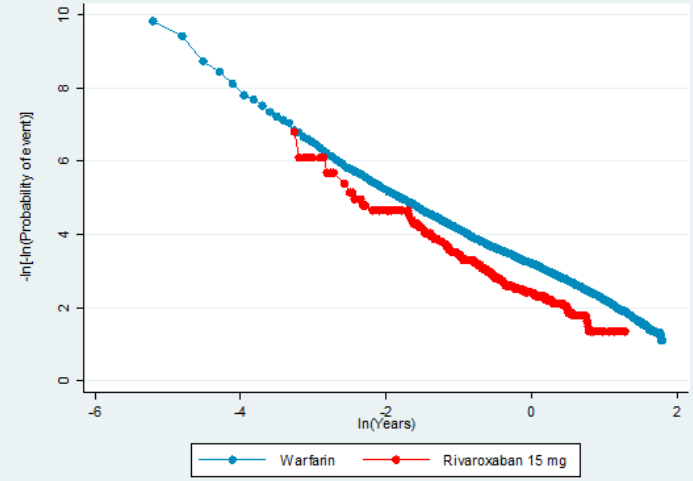


Figure XII-12 b (mortality-stroke)

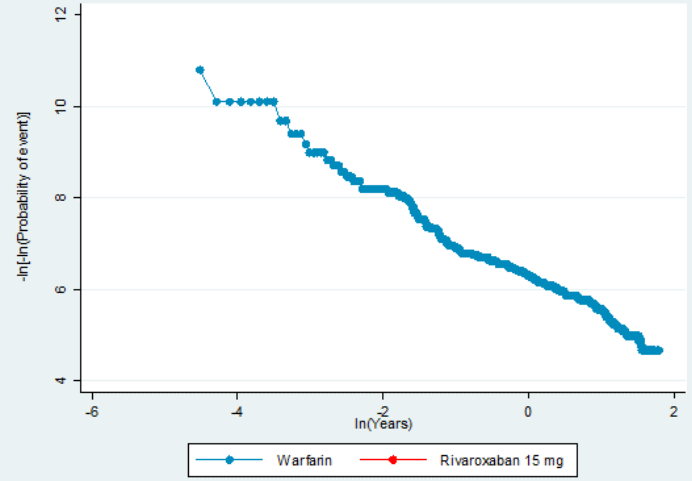


Figure XII-12 c (mortality-cardiovascular)

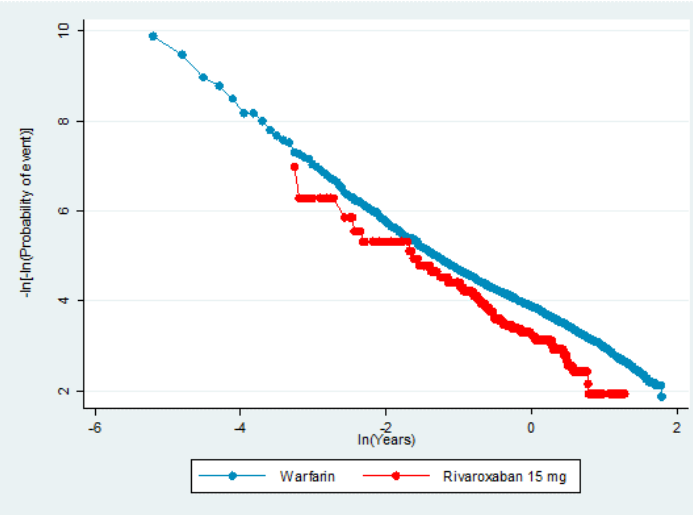


Figure XII-12: log-cumulative hazard plot for mortality (rivaroxaban 15 mg vs. warfarin)

Table XII-1 Proportionality hazard assumption test of significance (standard dose)

Outcome	p-value						
	treatment	age	sex	CCI	SIMD	year	PS
Apixaban 5 mg							
Stroke all	0.154	0.065	0.125	0.001	0.293	0.878	0.289
Stroke or SE	0.168	0.055	0.074	0.001	0.344	0.902	0.339
Stroke or SE or TIA	0.287	0.069	0.107	0.001	0.124	0.901	0.257
Stroke or SE or mortality (all-cause)	0.930	0.004	0.114	0.001	0.997	0.493	0.464
MI	0.787	0.826	0.461	0.001	0.449	0.212	0.969
Major bleeding	0.214	0.946	0.931	0.002	0.152	0.695	0.155
ICH	0.395	0.170	0.367	0.010	0.730	0.450	0.506
GI bleeding	0.756	0.632	0.596	0.017	0.284	0.127	0.105
Mortality (all-cause)	0.648	0.038	0.130	0.001	0.391	0.250	0.854
Mortality (stroke)	0.115	0.727	0.877	0.020	0.926	0.387	0.403
Mortality (cardiovascular)	0.084	0.964	0.863	0.004	0.946	0.148	0.758
Rivaroxaban 20 mg							
Stroke all	0.359	0.001	0.259	0.001	0.417	0.940	0.709
Stroke or SE	0.318	0.001	0.173	0.001	0.467	0.913	0.774
Stroke or SE or TIA	0.589	0.001	0.247	0.001	0.086	0.936	0.491
Stroke or SE or mortality (all-cause)	0.881	0.001	0.241	0.001	0.499	0.471	0.885
MI	0.834	0.973	0.514	0.001	0.098	0.277	0.166
Major bleeding	0.741	0.328	0.941	0.005	0.861	0.893	0.188
ICH	0.162	0.261	0.562	0.005	0.456	0.358	0.278
GI bleeding	0.246	0.921	0.302	0.105	0.527	0.173	0.772
Mortality (all-cause)	0.593	0.002	0.243	0.001	0.877	0.291	0.960
Mortality (stroke)	0.621	0.835	0.886	0.008	0.374	0.637	0.066
Mortality (cardiovascular)	0.696	0.820	0.200	0.001	0.632	0.258	0.698

Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack, ICH=intracranial haemorrhage, GI=gastrointestinal

Table XII-2 Proportionality hazard assumption test of significance (reduced dose)

Outcome	p-value						
	treatment	**age**	**sex**	**CCI**	**SIMD**	**year**	**PS**
Apixaban 2.5 mg							
Stroke all	0.108	0.245	0.105	0.001	0.124	0.946	0.077
Stroke or SE	0.095	0.218	0.060	0.001	0.145	0.982	0.091
Stroke or SE or TIA	0.273	0.268	0.110	0.001	0.134	0.932	0.068
Stroke or SE or mortality (all-cause)	0.061	0.001	0.072	0.001	0.888	0.536	0.779
MI	0.464	0.965	0.557	0.001	0.607	0.189	0.880
Major bleeding	0.189	0.943	0.842	0.001	0.278	0.707	0.180
ICH	0.109	0.358	0.532	0.009	0.949	0.395	0.883
GI bleeding	0.508	0.646	0.566	0.042	0.673	0.134	0.889
Mortality (all-cause)	0.087	0.006	0.107	0.001	0.323	0.309	0.216
Mortality (stroke)	0.982	0.691	0.913	0.001	0.942	0.610	0.939
Mortality (cardiovascular)	0.333	0.502	0.857	0.001	0.977	0.187	0.528
Rivaroxaban 15 mg							
Stroke all	0.503	0.000	0.413	0.001	0.499	0.794	0.209
Stroke or SE	0.335	0.000	0.278	0.001	0.533	0.796	0.200
Stroke or SE or TIA	0.273	0.000	0.455	0.001	0.219	0.825	0.054
Stroke or SE or mortality (all-cause)	0.261	0.001	0.017	0.001	0.989	0.505	0.772
MI	0.707	0.950	0.650	0.001	0.342	0.121	0.896
Major bleeding	0.399	0.087	0.629	0.003	0.350	0.970	0.165
ICH	0.824	0.083	0.330	0.009	0.196	0.486	0.791
GI bleeding	0.840	0.735	0.371	0.037	0.422	0.210	0.946
Mortality (all-cause)	0.462	0.132	0.058	0.001	0.526	0.233	0.682
Mortality (stroke)	0.448	0.516	0.769	0.001	0.767	0.624	0.417
Mortality (cardiovascular)	0.112	0.398	0.581	0.001	0.855	0.170	0.356

Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack, ICH=intracranial haemorrhage, GI=gastrointestinal

Appendix XIII: Kaplan Meier curves (standard and reduced dose)

Apixaban versus warfarin

The crude estimates over time indicate the same probability of surviving for patients on either warfarin or standard dose apixaban (Figure XIII-3 – XIII-6). While this is true for most of the outcomes, the probability of surviving from MI and ICH (Figure XIII-3 a, XIII-3 c) is greater for warfarin group than in the in the standard dose apixaban group; however, the probability of survival is inverted for the mortality due to stroke outcome (Figure XIII-3 B). The crude estimates for the reduced dose apixaban versus warfarin comparison indicate that the probability of surviving is higher for patients on warfarin than those on apixaban for each of the outcome assessed (Figure XIII-4 – XIII-6).

Rivaroxaban versus warfarin

The crude estimates over time indicate that the probability of surviving is higher for patients on warfarin than those on rivaroxaban standard dose (Figure XIII-7 – XIII-9). While this is true for most of the outcomes, the probability of surviving is inverted for patients experiencing an MI (Figure XIII-8 a). As for patients on standard dose rivaroxaban, over time, the probability of surviving is lower for patients on reduced dose rivaroxaban than for those on warfarin (Figure XIII-10–XIII-12).

Figure XIII-1 a (stroke-all)

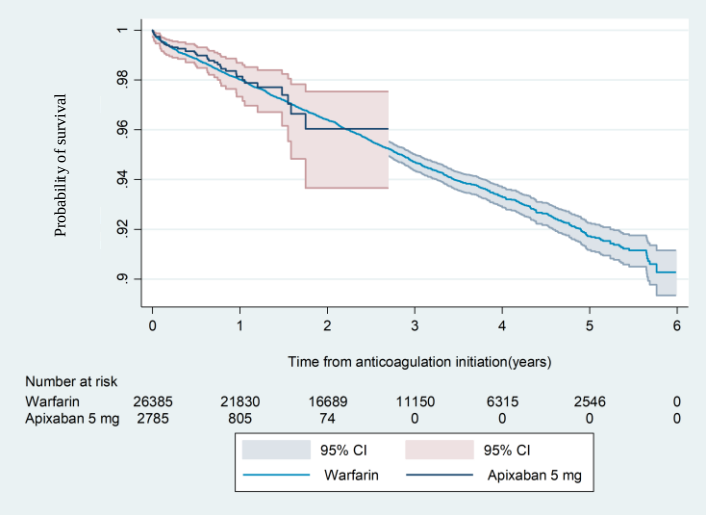


Figure XIII-1 b (stroke or SE)

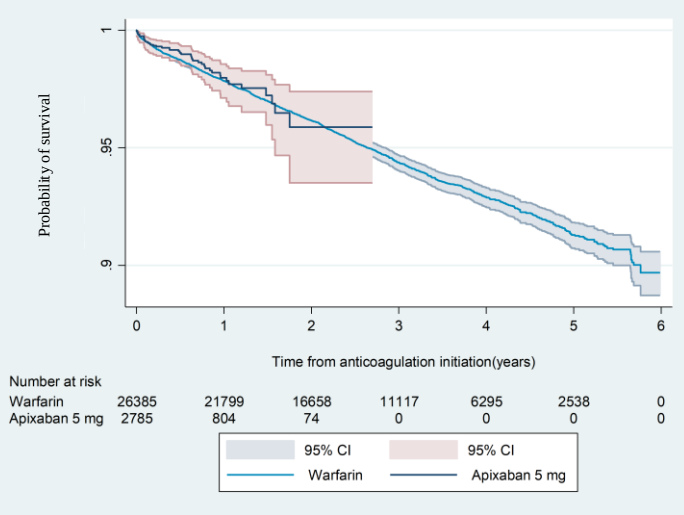


Figure XIII-1 c (stroke or SE or TIA)

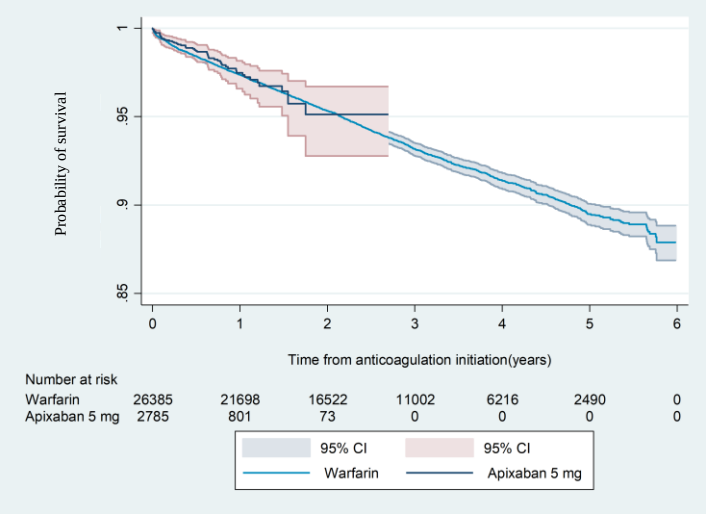


Figure XIII-1 d (stroke or mortality-all-cause)

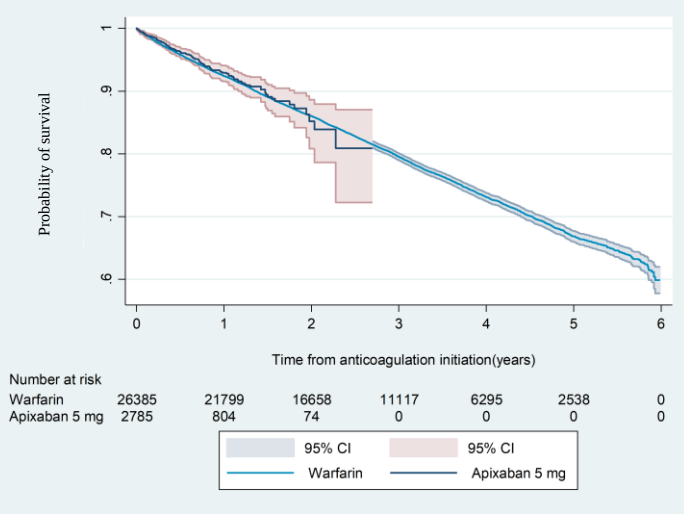


Figure XIII-1 Stroke, SE and TIA (apixaban 5 mg vs. warfarin)

Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack

Figure XIII-2 a (MI)

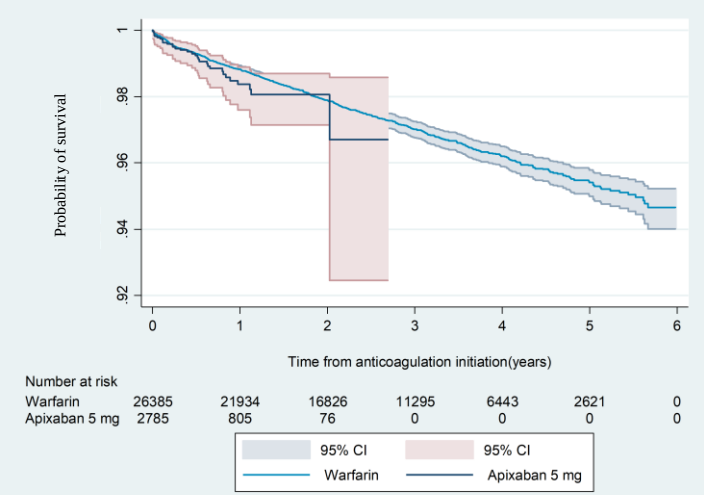


Figure XIII-2 b (major bleeding)

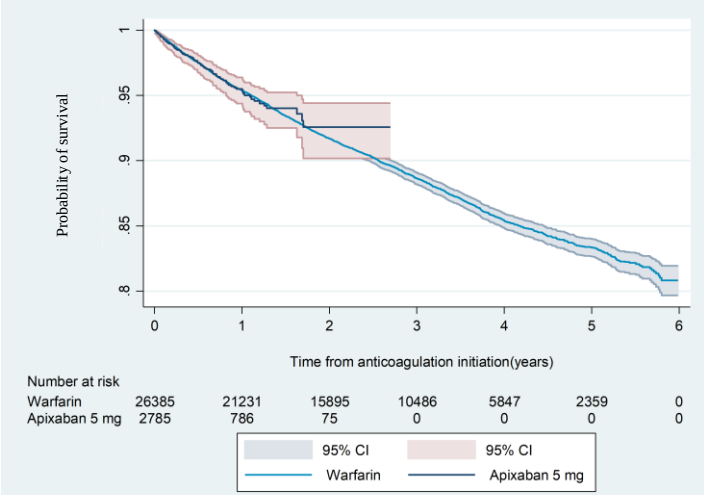


Figure XIII-2 c (ICH)

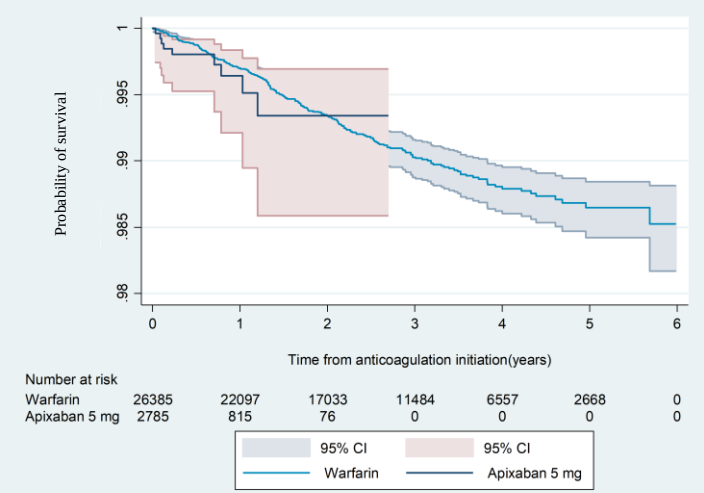


Figure XIII-2 d (GI bleeding)

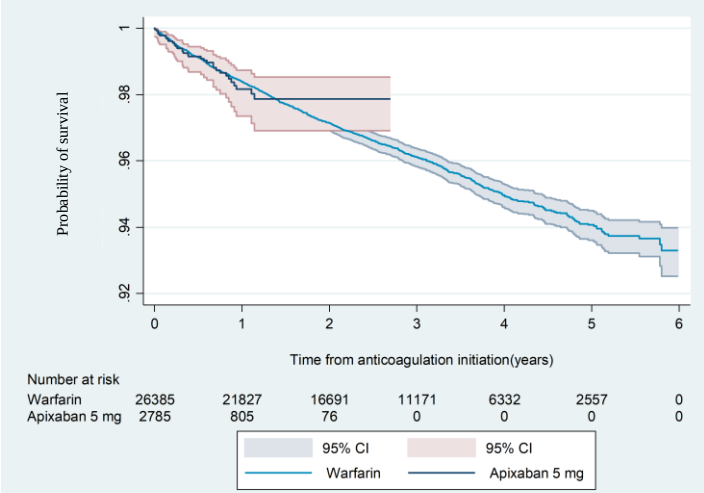


Figure XIII-2: MI, ICH and bleeding (apixaban 5 mg vs. warfarin)

Abbreviations: MI=myocardial, ICH= intracranial haemorrhage, GI=gastrointestinal

Figure XIII-3 a (mortality-all-cause)

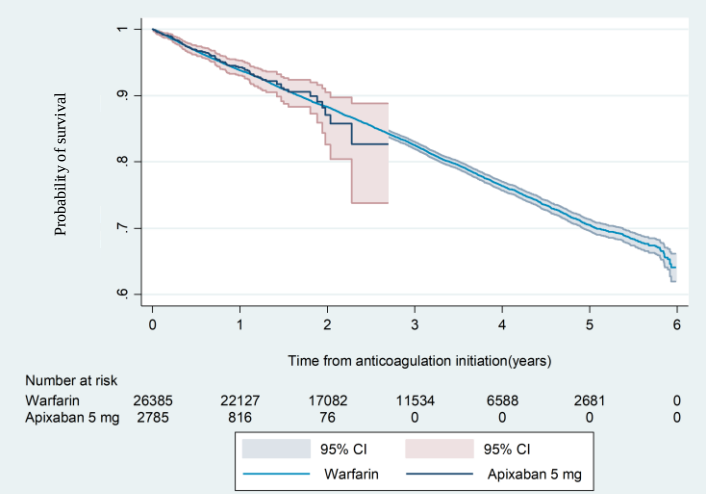


Figure XIII-3 b (mortality-stroke)

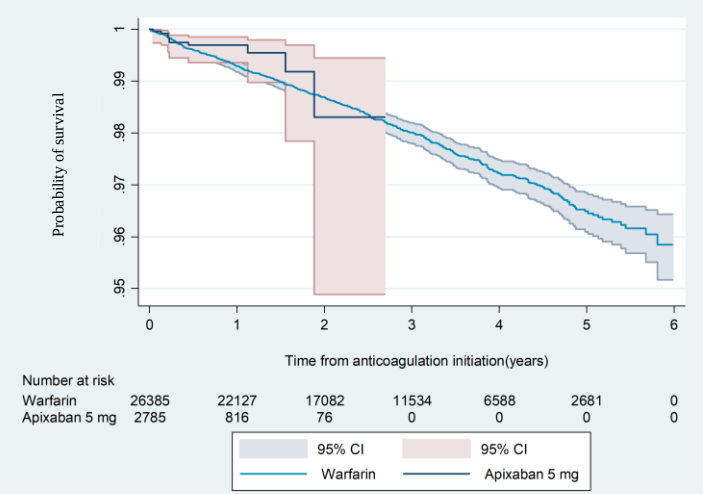


Figure XIII-3 c (mortality-cardiovascular)

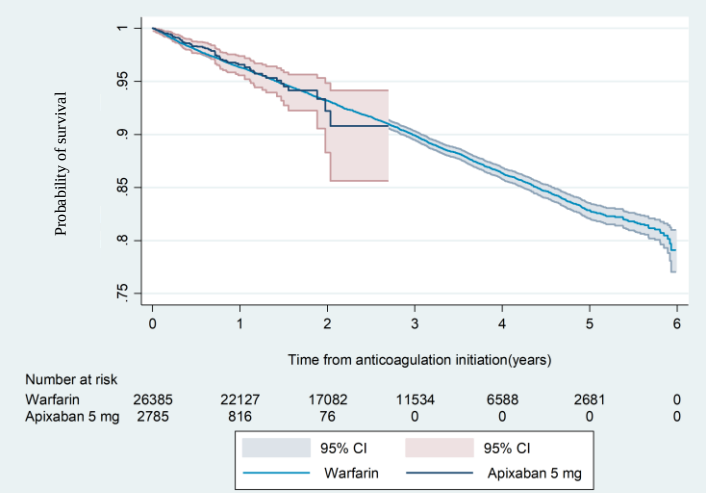


Figure XIII-3: Mortality (apixaban 5 mg vs. warfarin)

Figure XIII-4 a (stroke-all)

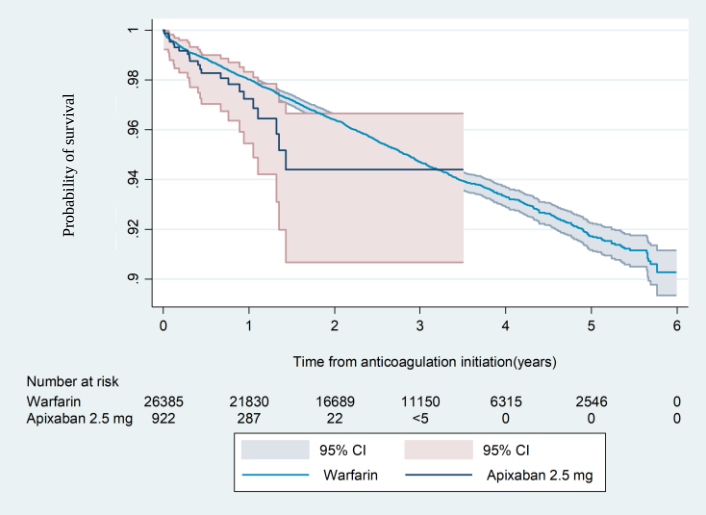


Figure XIII-4 b (stroke or SE)

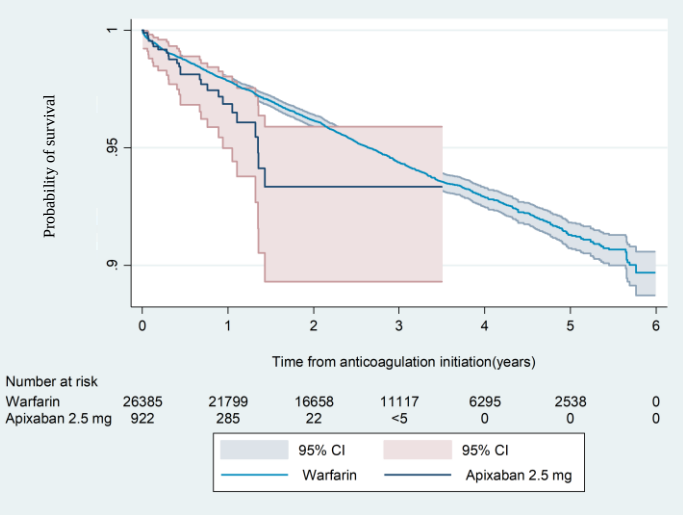


Figure XIII-4 c (stroke or SE or TIA)

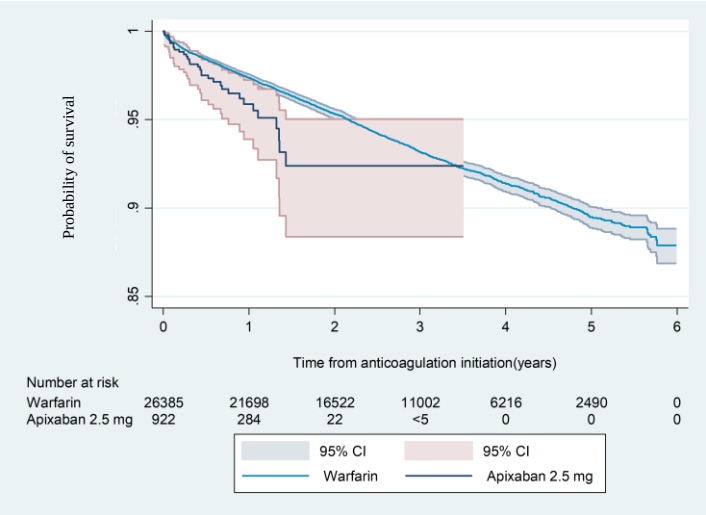


Figure XIII-4 d (stroke or mortality-all-cause)

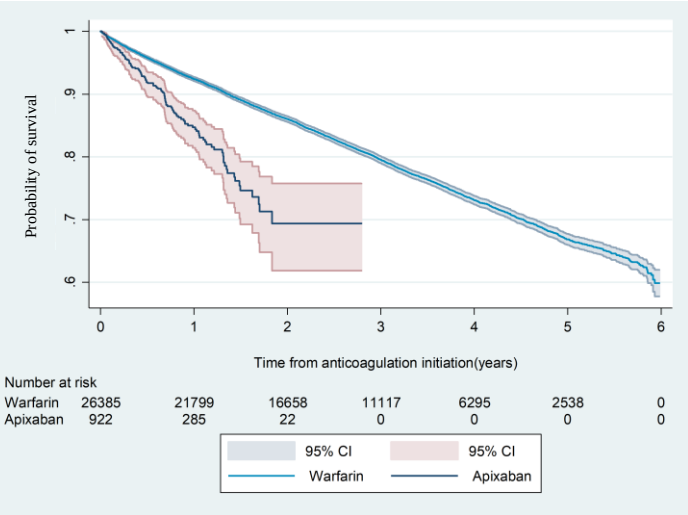


Figure XIII-4 Stroke, SE and TIA (apixaban 2.5 mg vs. warfarin)

Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack

Figure XIII-5 a (MI)

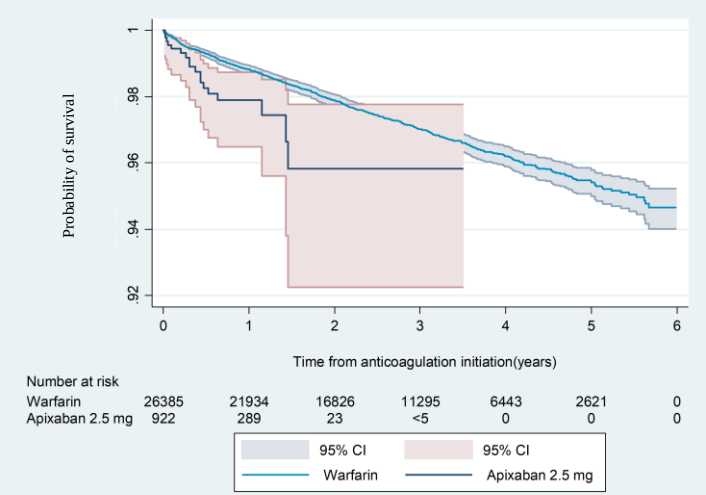


Figure XIII-5 b (major bleeding)

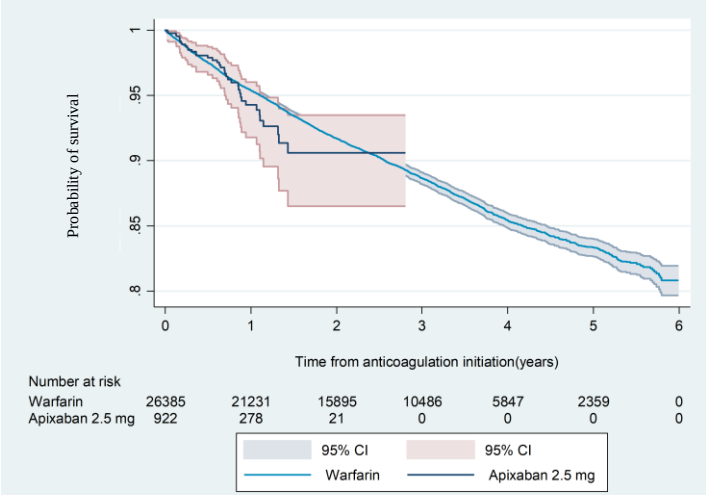


Figure XIII-5 c (ICH)

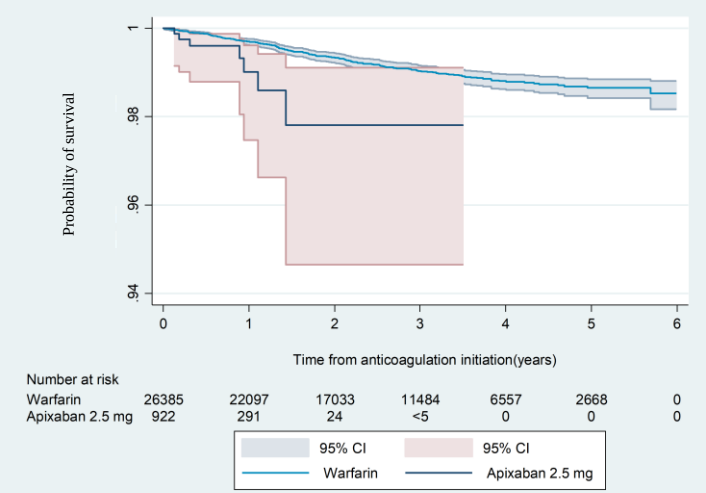


Figure XIII-5 d (GI bleeding)

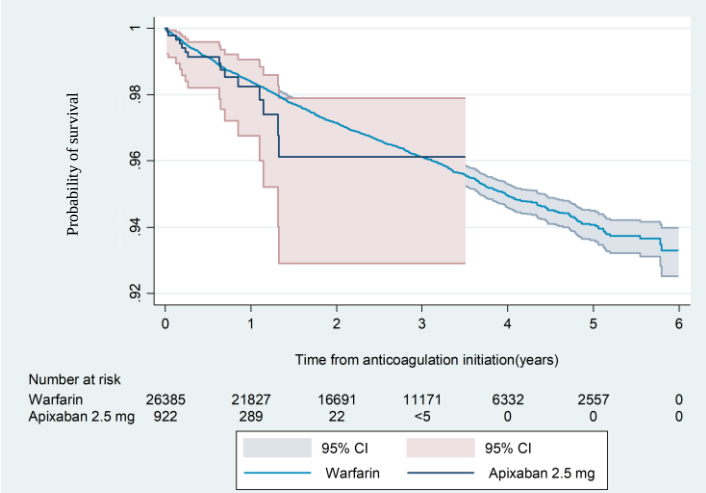


Figure XIII-5: MI, ICH and bleeding (apixaban 2.5 mg vs. warfarin)

Abbreviations: MI=myocardial, ICH= intracranial haemorrhage, GI=gastrointestinal

Figure XIII-6 a (mortality-all-cause)

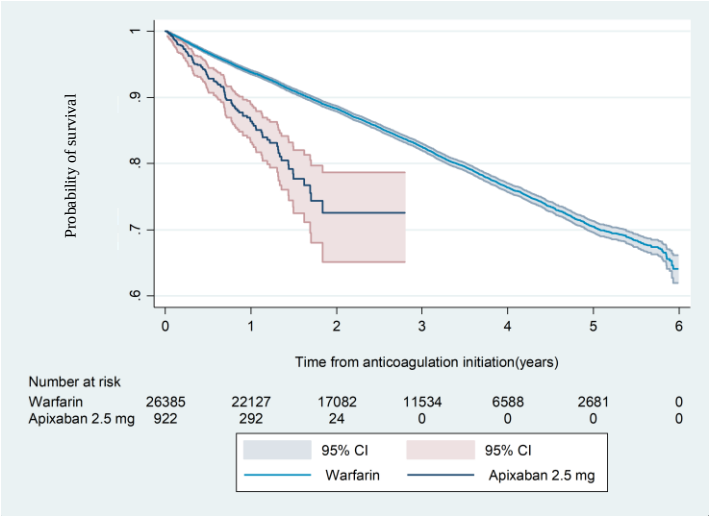


Figure XIII-6 b (mortality-stroke)

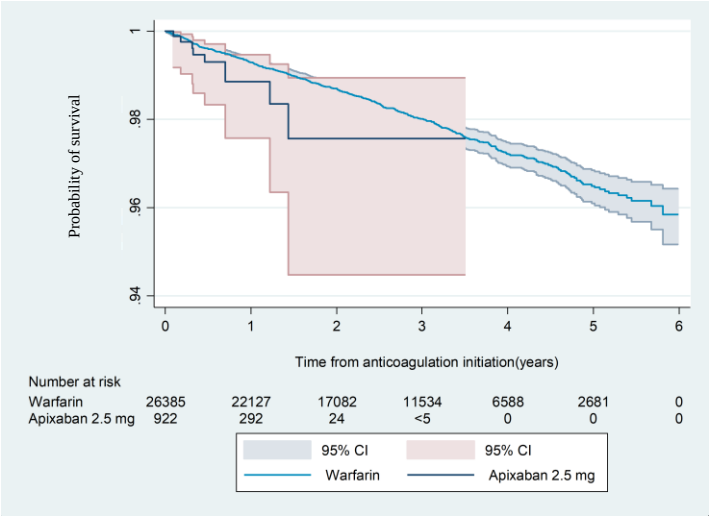


Figure XIII-6 c (mortality-cardiovascular)

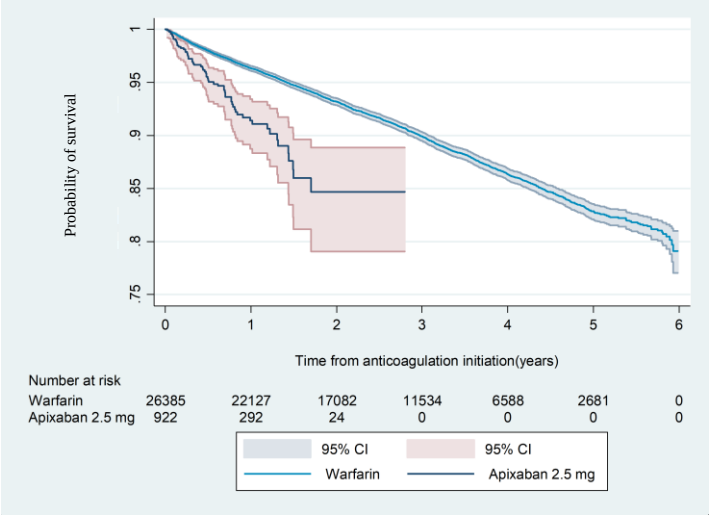


Figure XIII-6: Mortality (apixaban 2.5 mg vs. warfarin)

Figure XIII-7 a (stroke-all)

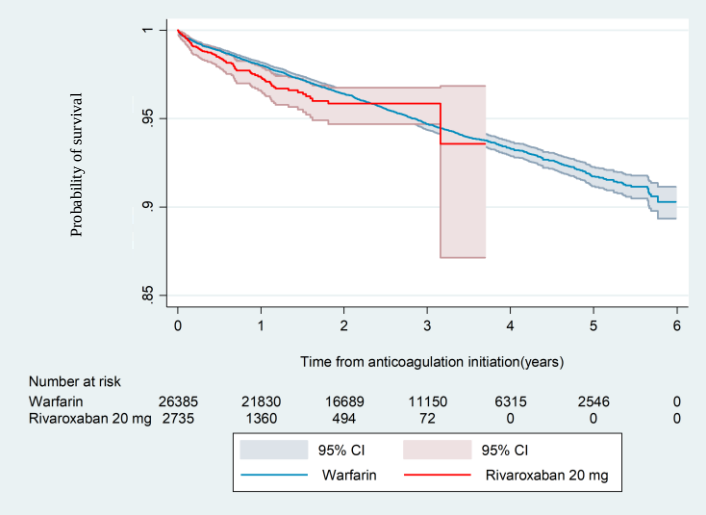


Figure XIII-7 b (stroke or SE)

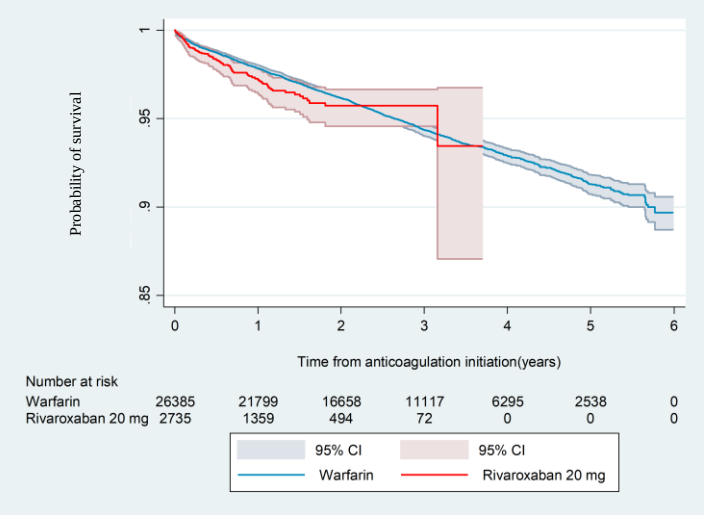


Figure XIII-7 c (stroke or SE or TIA)

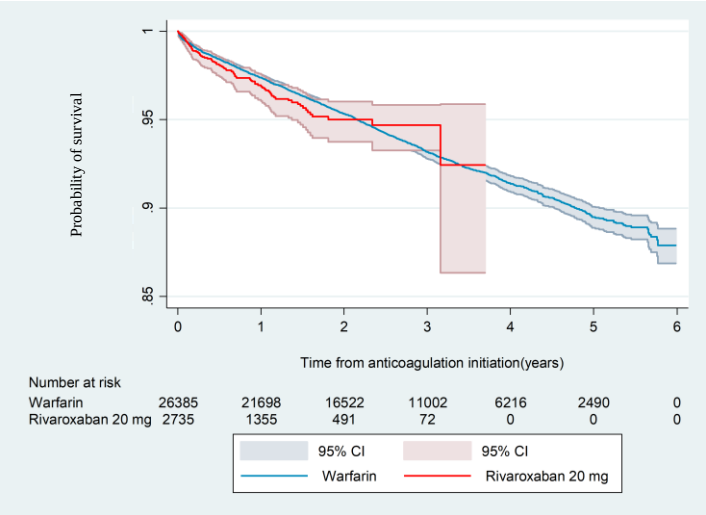


Figure XIII-7 d (stroke or mortality-all-cause)

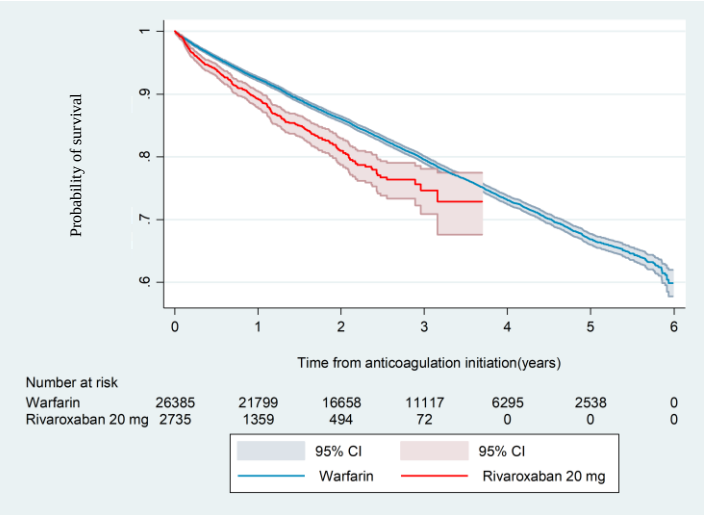


Figure XIII-7: Stroke, SE and TIA (rivaroxaban 20 mg vs. warfarin)
Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack

Figure XIII-8 a (MI)

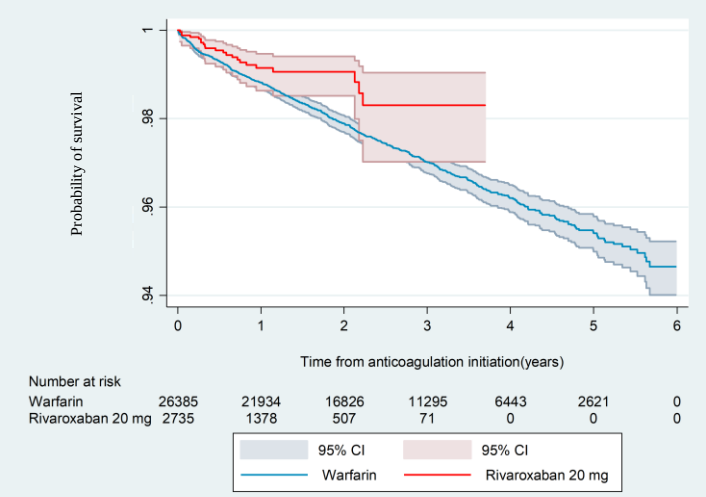


Figure XIII-8 b (major bleeding)

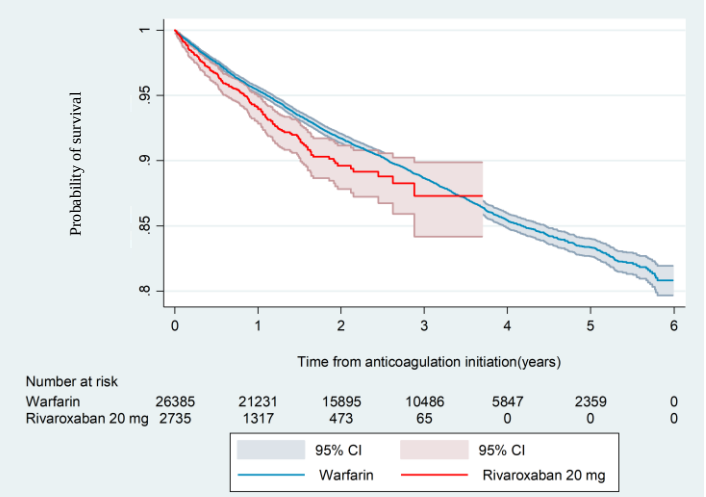


Figure XIII-8 c (ICH)

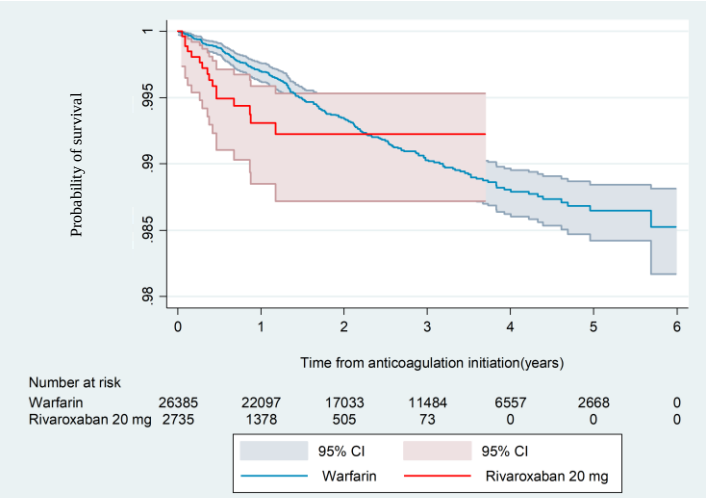


Figure XIII-8 d (GI bleeding)

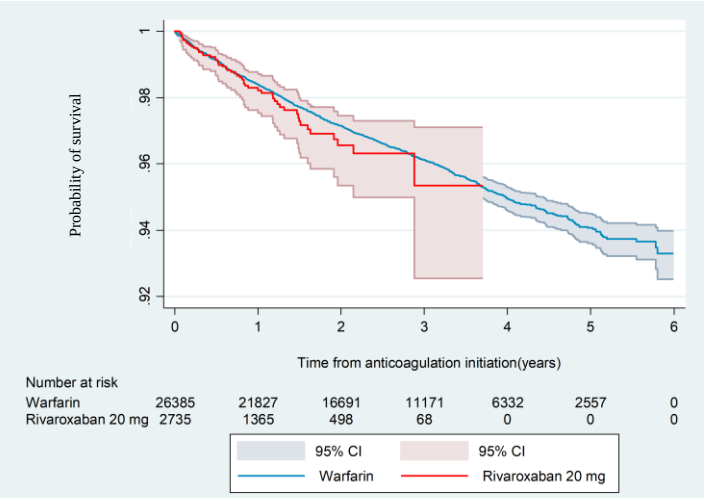


Figure XIII-8: MI, ICH and bleeding (rivaroxaban 20 mg vs. warfarin)
Abbreviations: MI=myocardial, ICH= intracranial haemorrhage, GI=gastrointestinal

Figure XIII-9 a (mortality-all-cause)

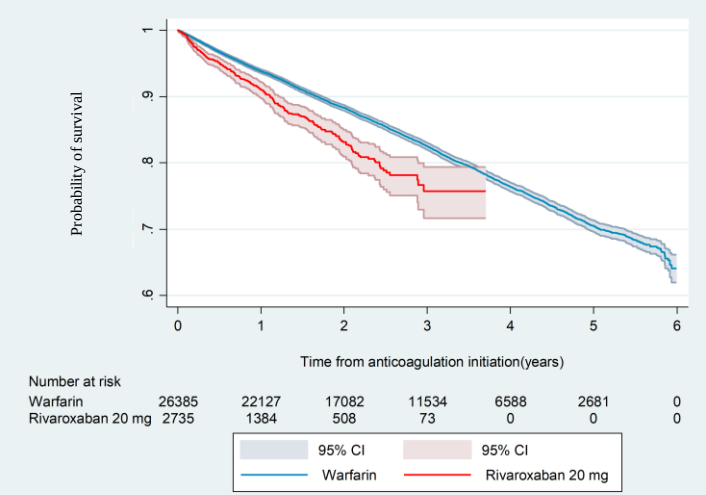


Figure XIII-9 b (mortality-stroke)

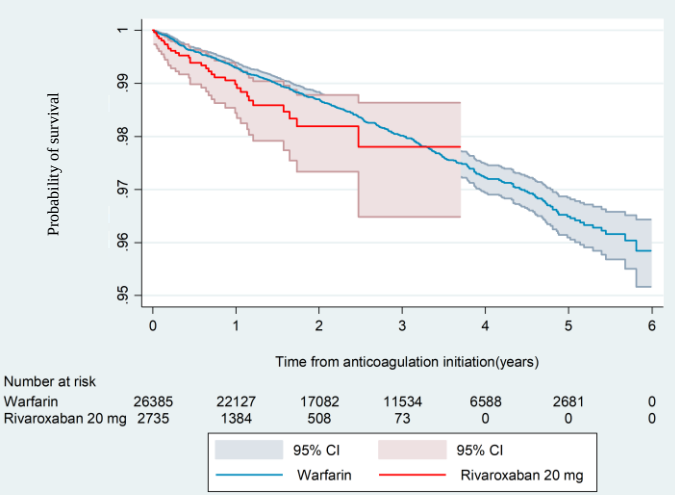


Figure XIII-9 c (mortality-cardiovascular)

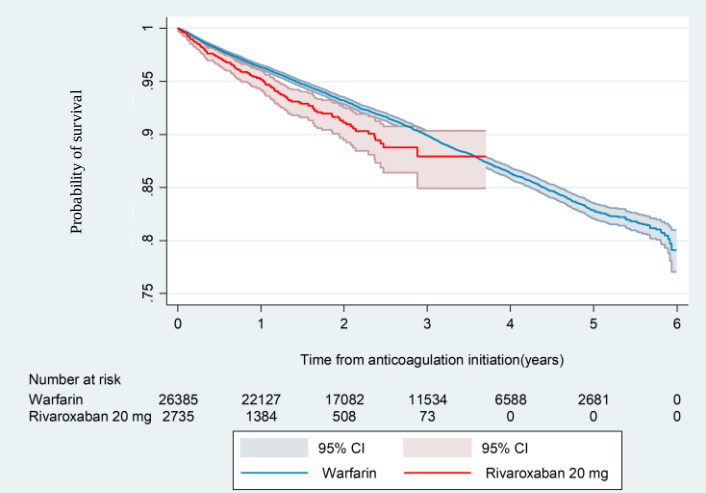


Figure XIII-9: Mortality (rivaroxaban 20 mg vs. warfarin)

Figure XIII-10 a (stroke-all)

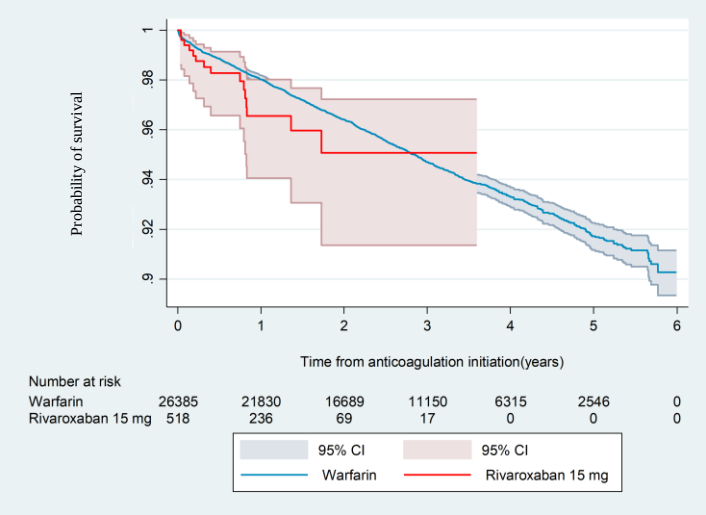


Figure XIII-10 b (stroke or SE)

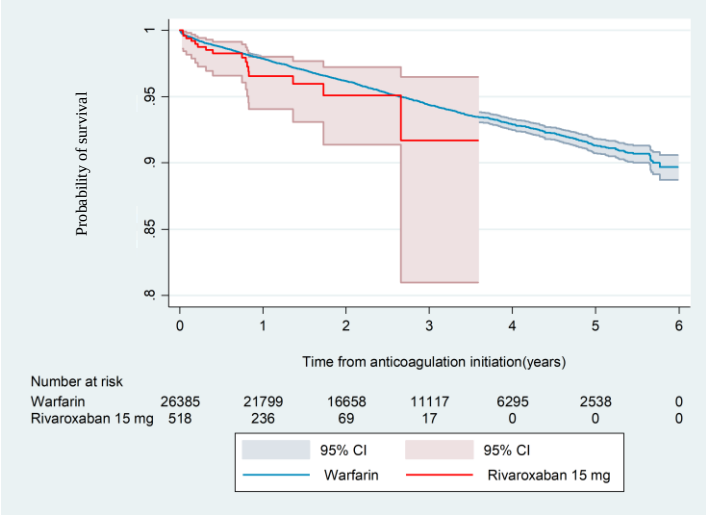


Figure XIII-10 c (stroke or SE or TIA)

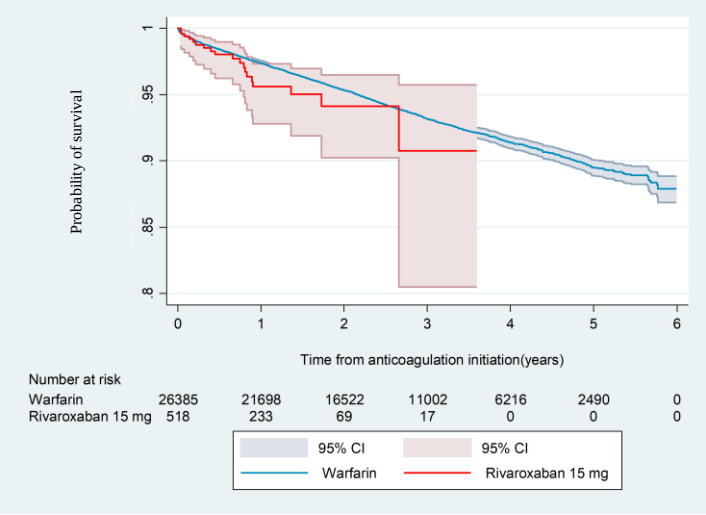


Figure XIII-10 d (stroke or mortality-all-cause)

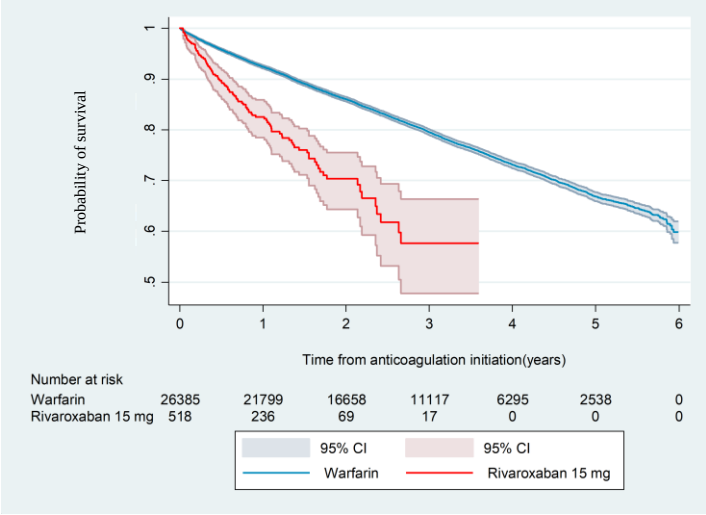


Figure XIII-10: Stroke, SE and TIA (rivaroxaban 15 mg vs. warfarin)
Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack

Figure XIII-11 a (MI)

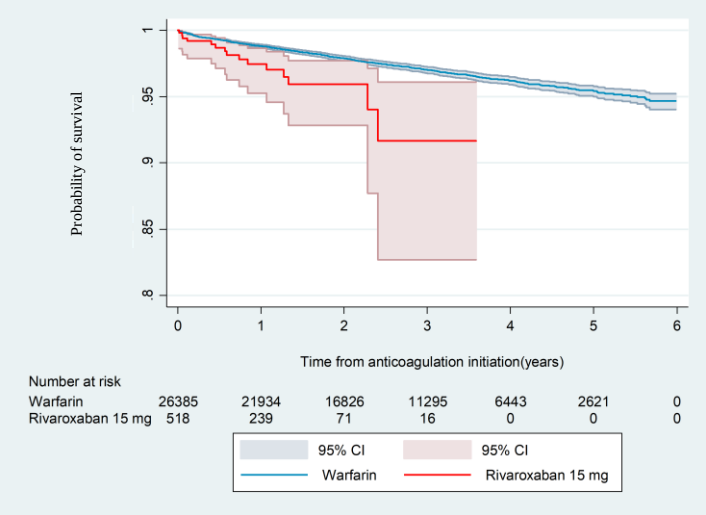


Figure XIII-11 b (major bleeding)

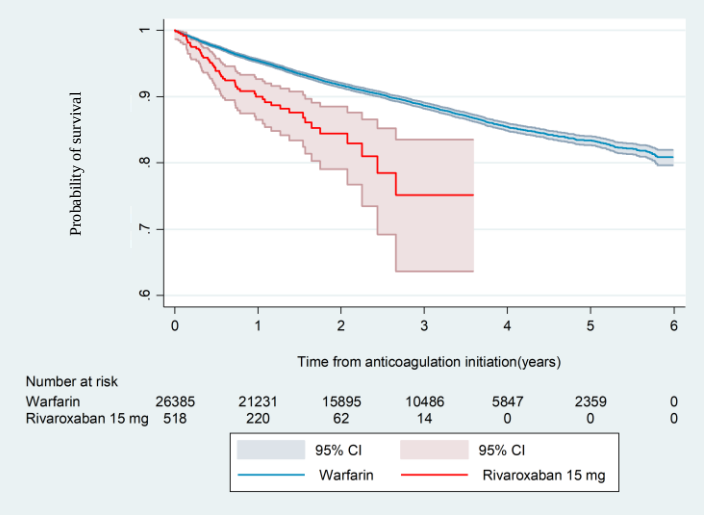


Figure XIII-11 c (ICH)

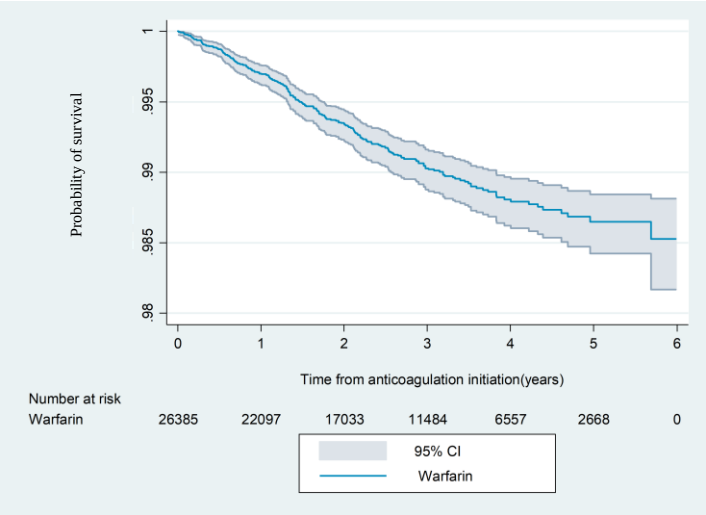


Figure XIII-11 d (GI bleeding)

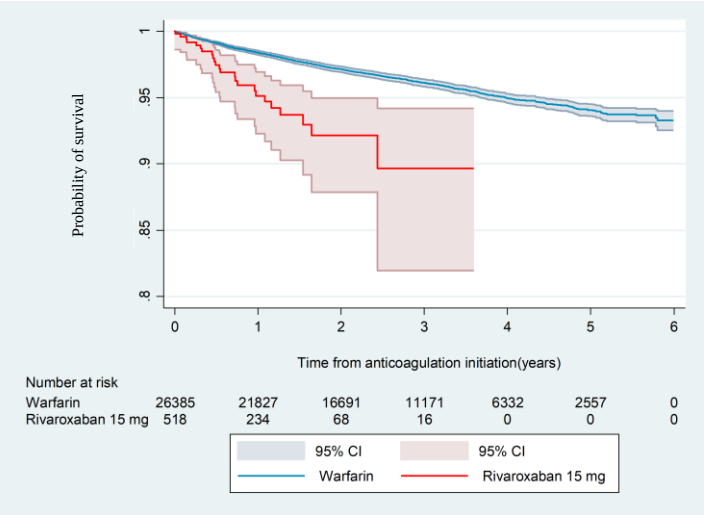


Figure XIII-11: MI, ICH and bleeding (rivaroxaban 15 mg vs. warfarin)

Note: due to disclosure restrictions, in the case of fewer than five events, the Kaplan Meier curve was not reported

Abbreviations: MI=myocardial, ICH= intracranial haemorrhage, GI=gastrointestinal

Figure XIII-12 a (mortality-all-cause)

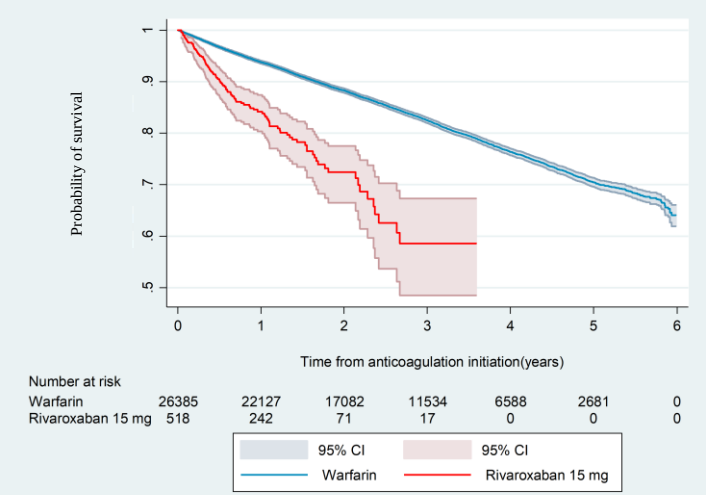


Figure XIII-12 b (mortality-stroke)

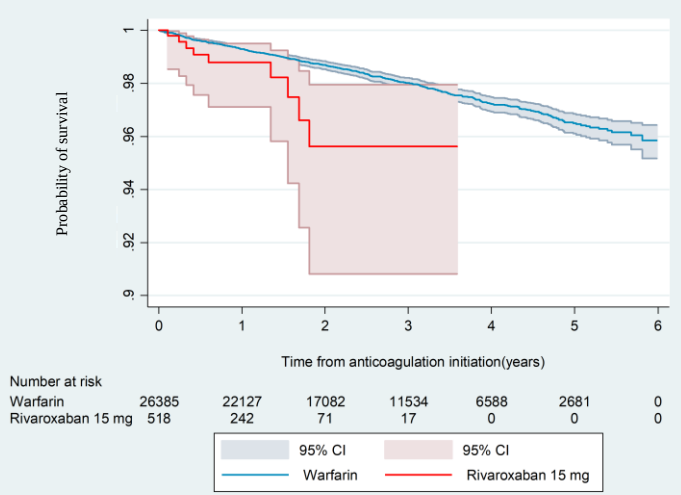


Figure XIII-12 c (mortality-cardiovascular)

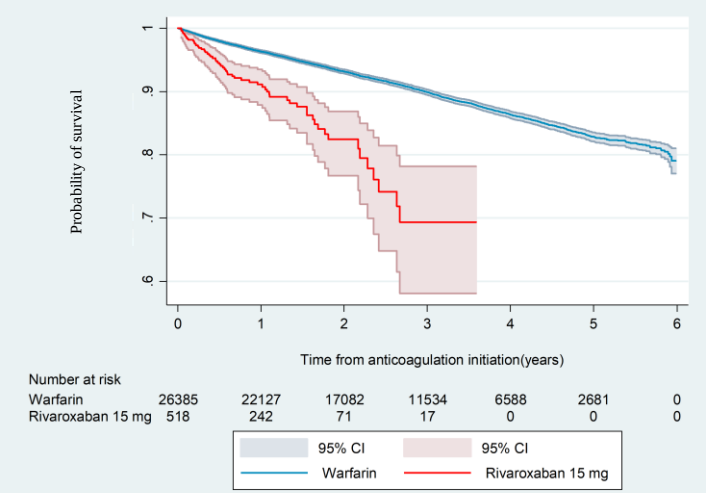


Figure XIII-12: Mortality (rivaroxaban 15 mg vs. warfarin)

Appendix XIV: Cumulative incidence curves (standard and reduced dose)

Apixaban versus warfarin

The cumulative incidence curves for apixaban standard and reduced dose are presented in Appendix XIV (Figure XIV-1–XIV-6). Overall, the cumulative incidence curves for standard dose apixaban (Figure XIV-1–XIV-3) did not show any difference in the probability of events occurring at 2 years since first prescription for most of the clinical outcomes. Nevertheless, patients on standard dose apixaban appeared to be at a greater risk of experiencing MI or ICH than those on warfarin. The probability of experiencing an MI event, however, seemed to be greater only after about 6 months from anticoagulation initiation (Figure XIV-2 a). On the other hand, for most of the clinical outcomes at 2 years from anticoagulation initiation, the probability of an event to occur seemed to be greater in the reduced dose apixaban than in the warfarin treatment group (Figure XIV-4–XIV-6). However, the cumulative incidence curves, indicating the occurrence of major and GI bleeding in the reduced dose apixaban group, did not seem to follow a clear pattern (Figure XIV-5 b, XIV-5 d).

Rivaroxaban versus warfarin

The cumulative incidence curves for rivaroxaban standard and reduced dose are presented in Appendix XIV (Figure XIV-7–XIV-12). Overall, for most of the clinical outcomes at 2 years from anticoagulation initiation, the probability of an event to occur seemed to be greater in the standard dose rivaroxaban than in the warfarin treatment group (Figure XIV-7–XIV-9). However, patients on standard dose rivaroxaban seemed to be at a greater risk of experiencing MI, following the 2 years since treatment initiation, than those on warfarin. Aside from risk of MI, similar patterns were observed in the cumulative incidence curves of reduced dose rivaroxaban (Figure XIV-10–XIV-12).

Figure XIV-1 a (stroke-all)

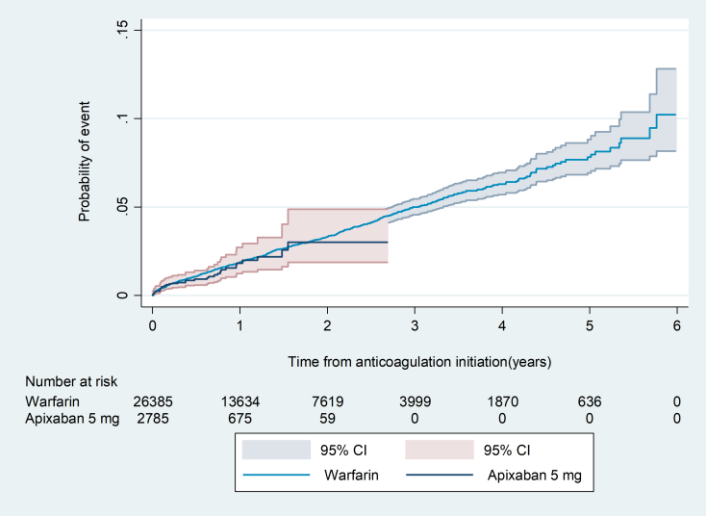


Figure XIV-1 b (stroke or SE)

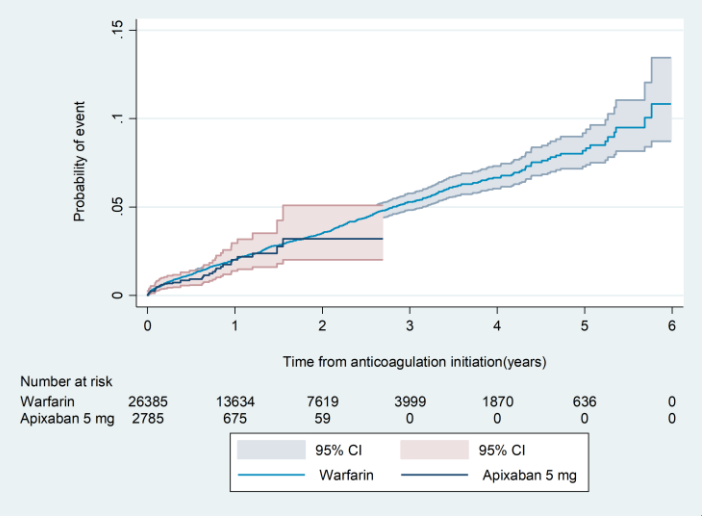


Figure XIV-1 c (stroke or SE or TIA)

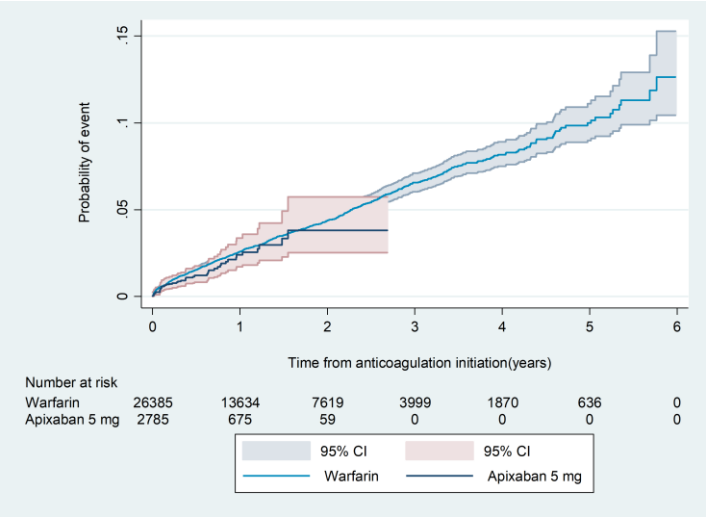


Figure XIV-1 d (stroke or SE or mortality-all-cause)

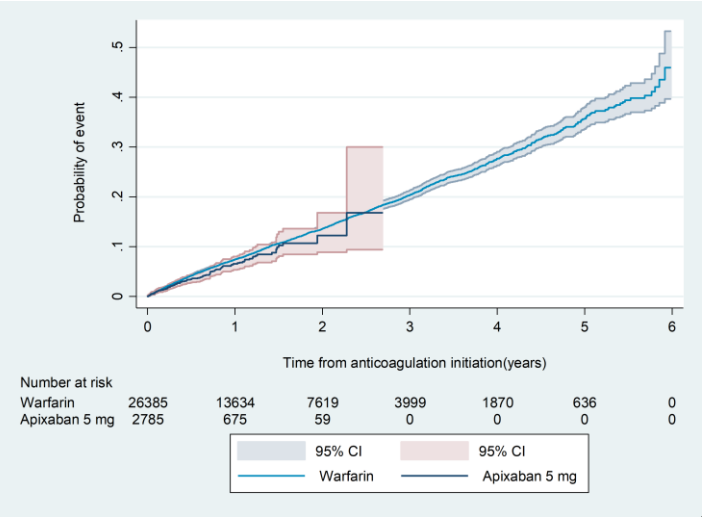


Figure XIV-1: Stroke, SE and TIA (apixaban 5 mg vs. warfarin)
Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack

Figure XIV-2 a (MI)

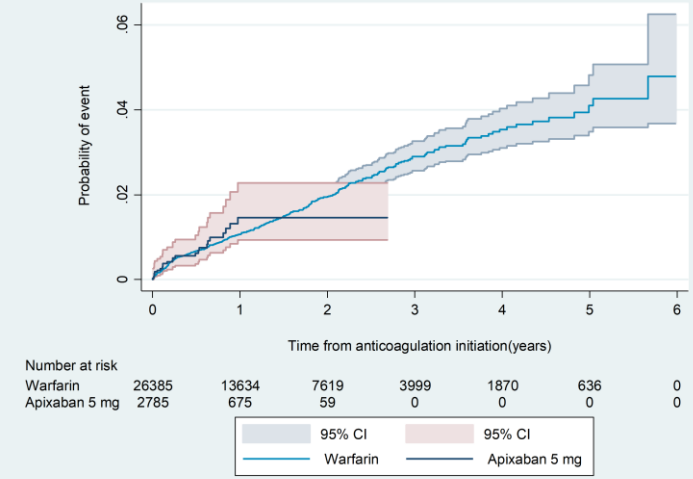


Figure XIV-2 b (major bleeding)

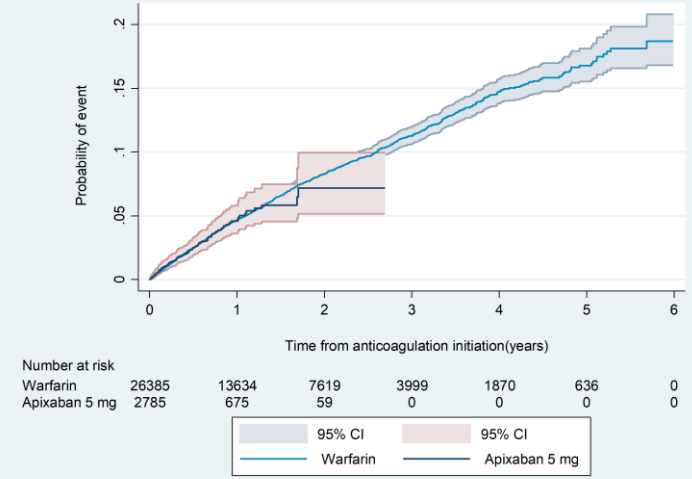


Figure XIV-2 c (ICH)

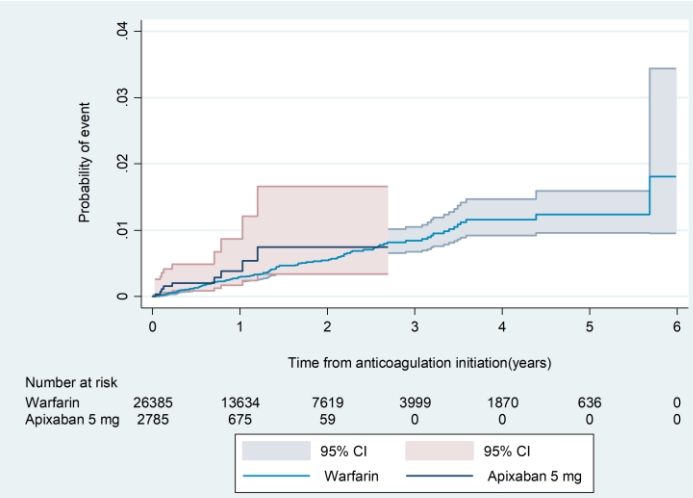


Figure XIV-2 d (GI bleeding)

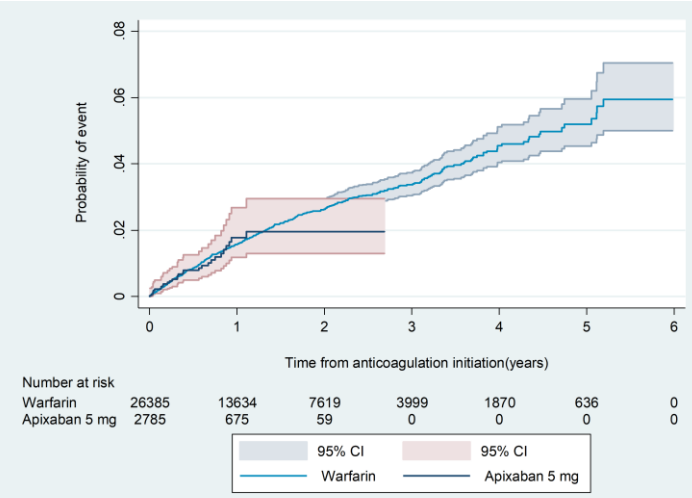


Figure XIV-2: MI, ICH and bleeding (apixaban 5 mg vs. warfarin)

Abbreviations: MI=myocardial, ICH= intracranial haemorrhage, GI=gastrointestinal

Figure XIV-3 a (mortality-all-cause)

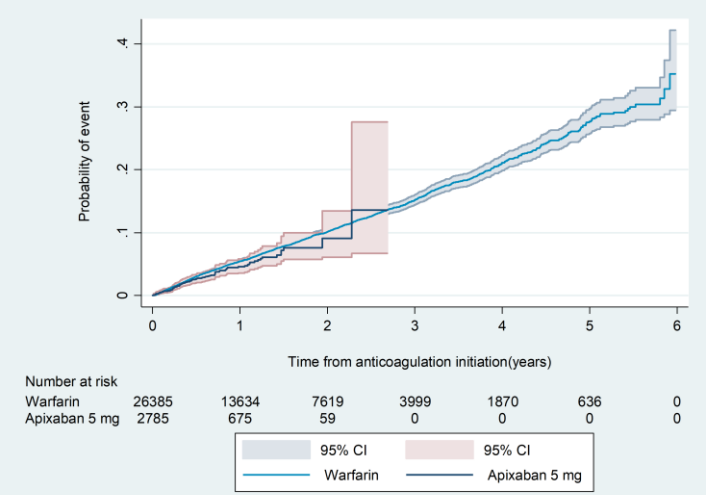


Figure XIV-3 b (mortality-stroke)

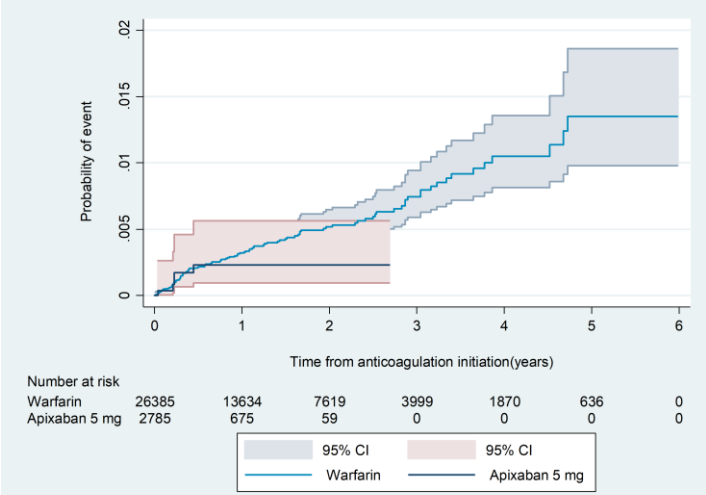


Figure XIV-3 c (mortality-cardiovascular)

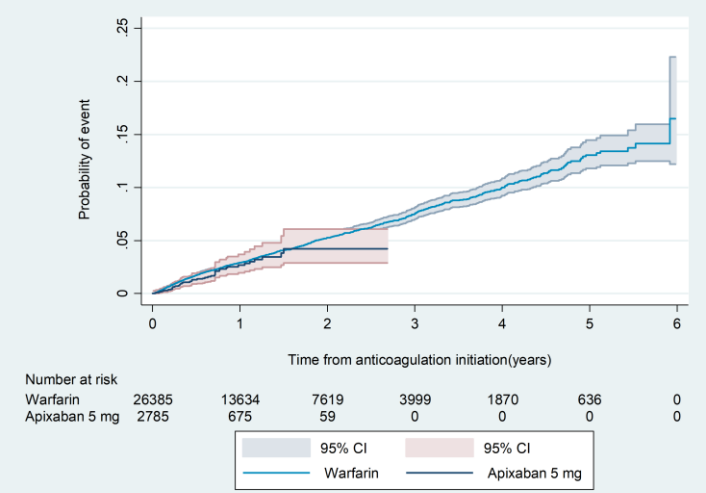


Figure XIV-3: Mortality (apixaban 5 mg vs. warfarin)

Figure XIV-4 a (stroke-all)

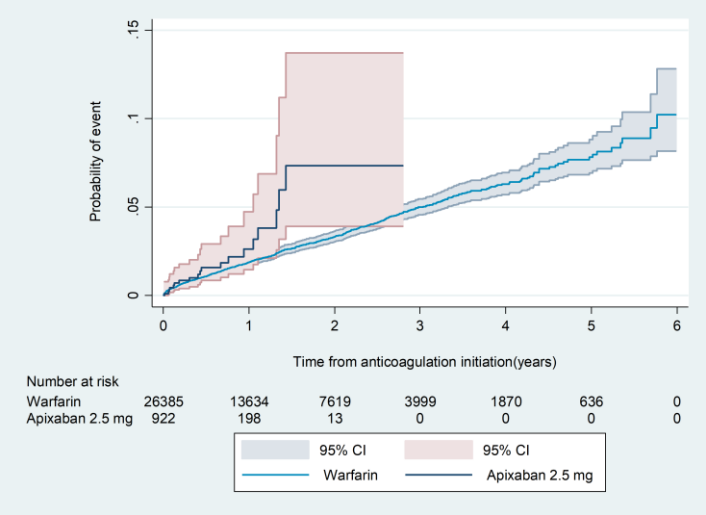


Figure XIV-4 b (stroke or SE)

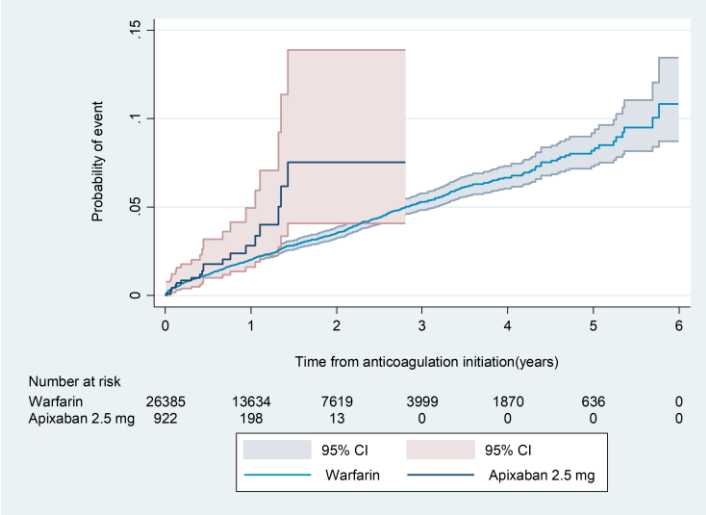


Figure XIV-4 c (stroke or SE or TIA)

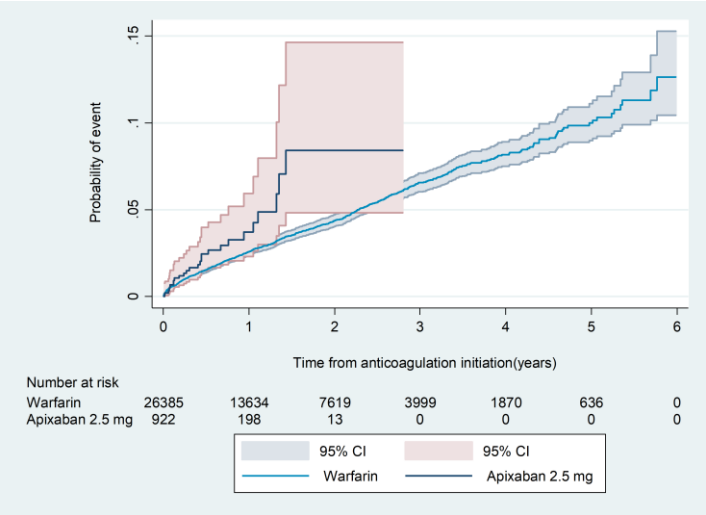


Figure XIV-4 d (stroke or SE or mortality-all-cause)

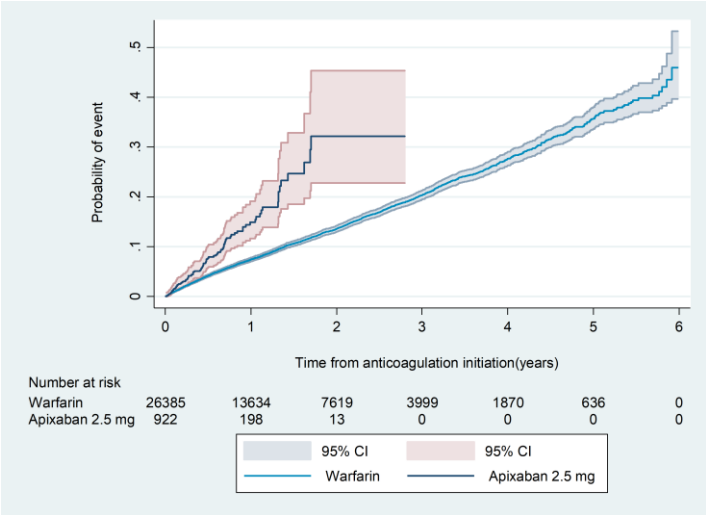


Figure XIV-4: Stroke, SE and TIA (apixaban 2.5 mg vs. warfarin)

Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack

Figure XIV-5 a (MI)

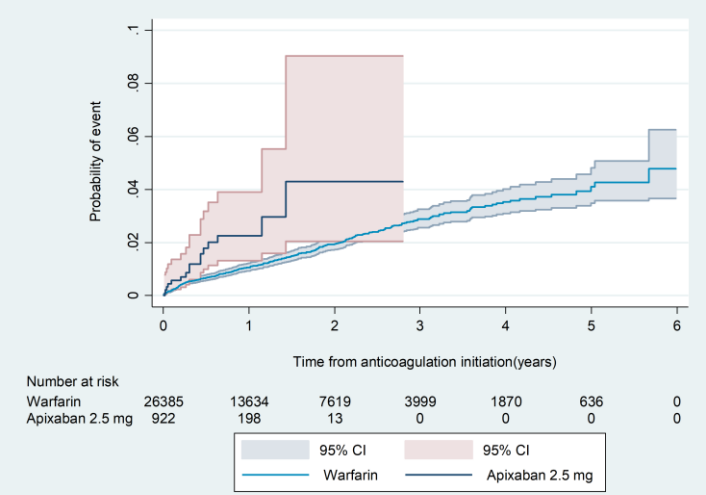


Figure XIV-5 b (major bleeding)

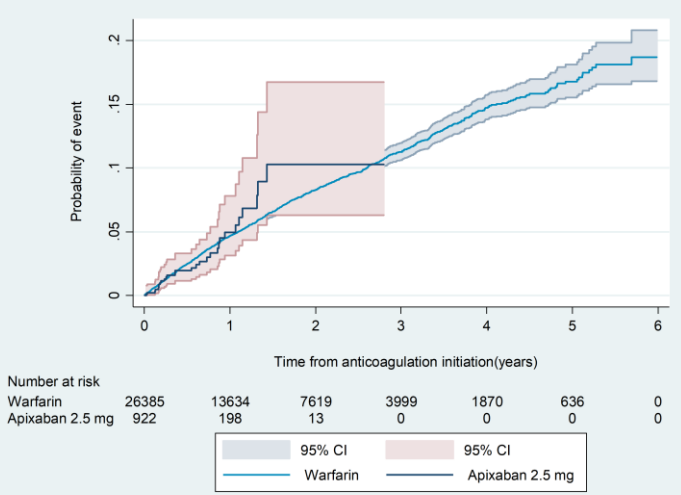


Figure XIV-5 c (ICH)

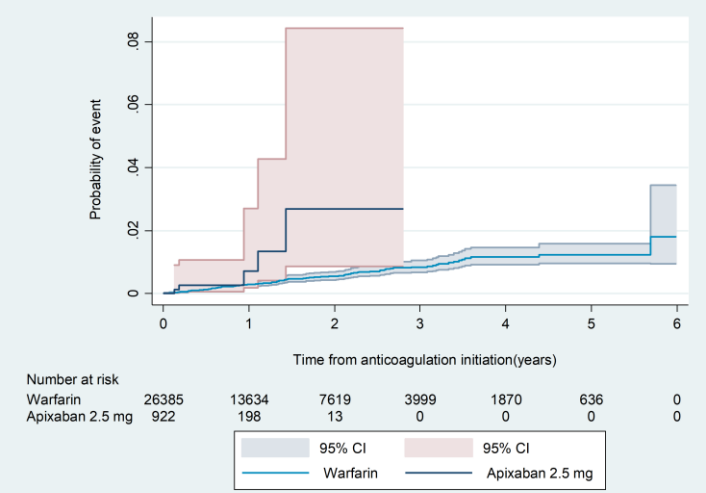


Figure XIV-5 d (GI bleeding)

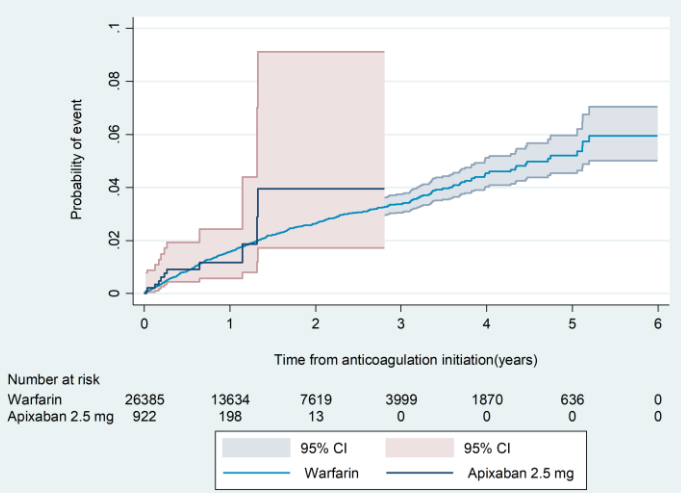


Figure XIV-5: MI, ICH and bleeding (apixaban 2.5 mg vs. warfarin)

Abbreviations: MI=myocardial, ICH= intracranial haemorrhage, GI=gastrointestinal

Figure XIV-6 a (mortality-all-cause)

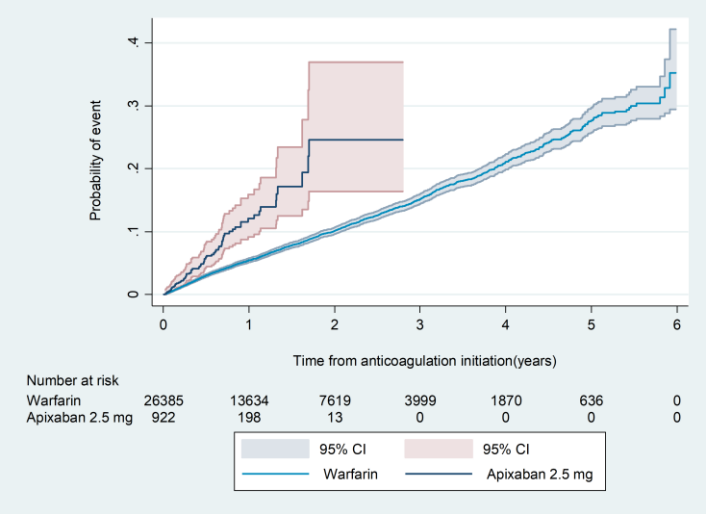


Figure XIV-6 b (mortality-stroke)

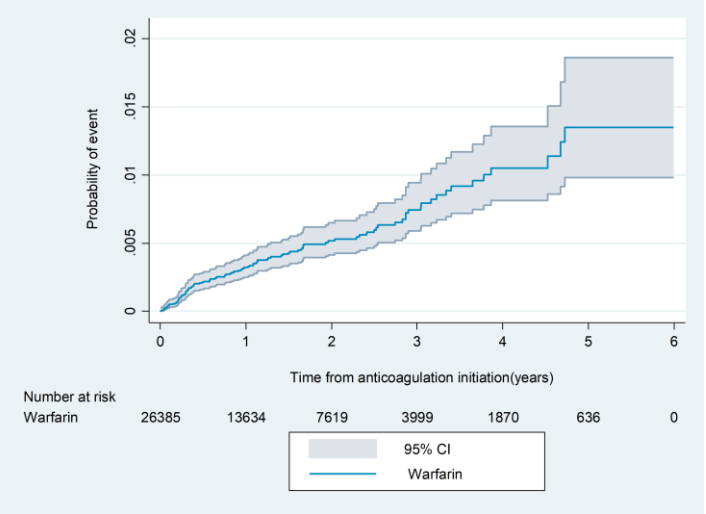


Figure XIV-6 c (mortality-cardiovascular)

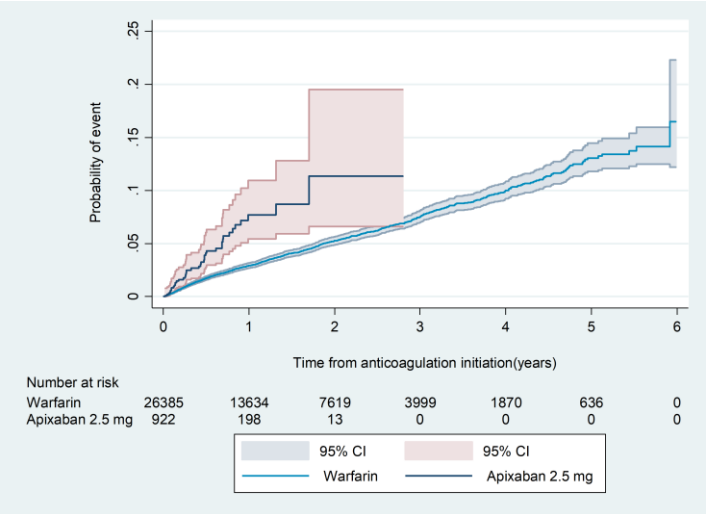


Figure XIV-6: Mortality (apixaban 2.5 mg vs. warfarin)

Note: due to disclosure restrictions, in the case of fewer than five events, the cumulative incidence curve was not reported

Figure XIV-7 a (stroke-all)

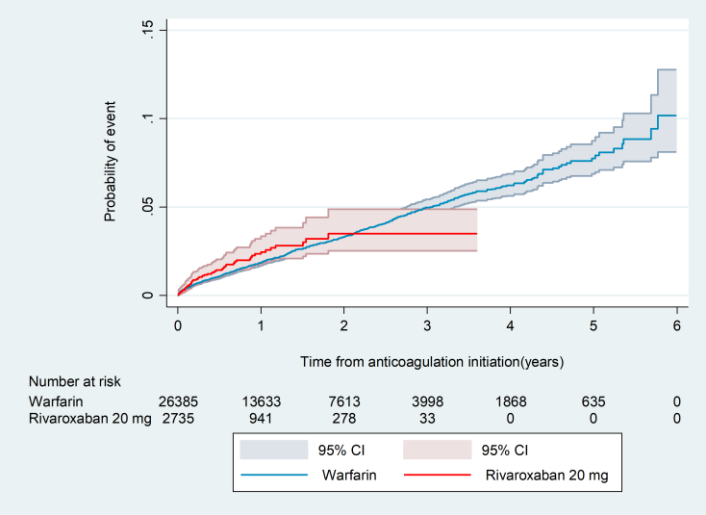


Figure XIV-7 b (stroke or SE)

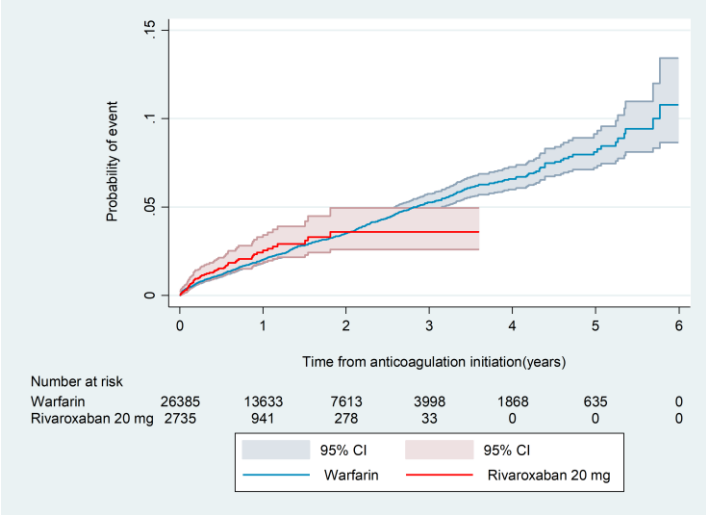


Figure XIV-7 c (stroke or SE or TIA)

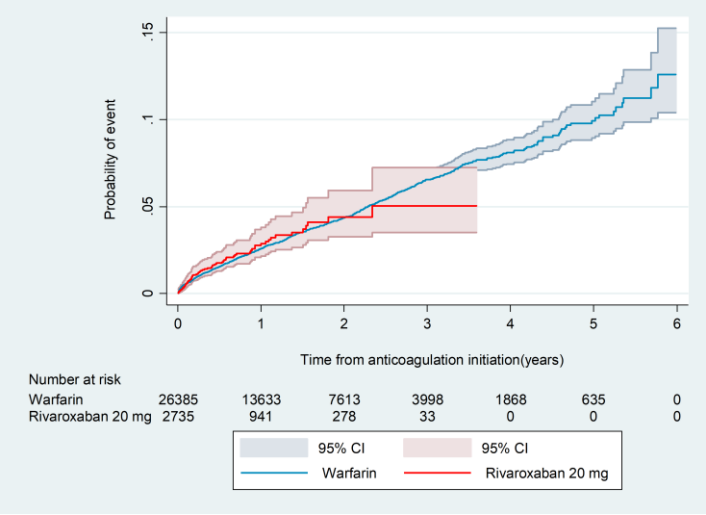


Figure XIV-7 d (stroke or mortality-all-cause)

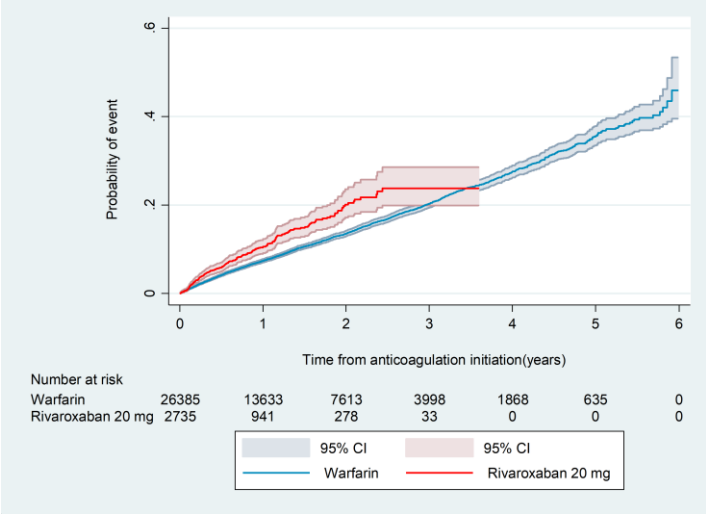


Figure XIV-7: Stroke, SE and TIA (rivaroxaban 20 mg vs. warfarin)

Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack

Figure XIV-8 a (MI)

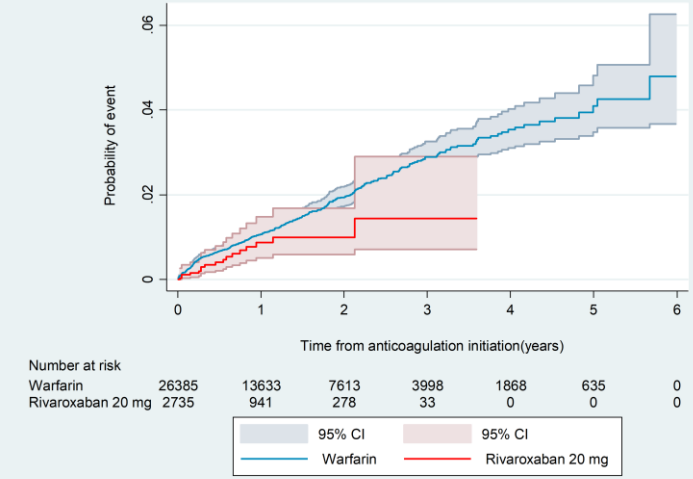


Figure XIV-8 b (major bleeding)

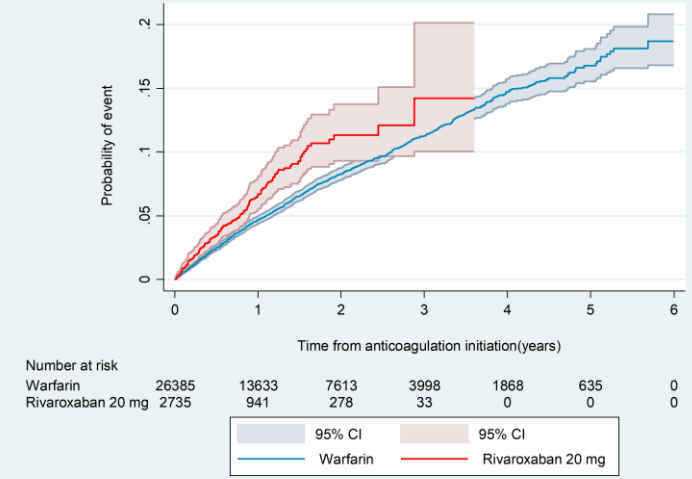


Figure XIV-8 c (ICH)

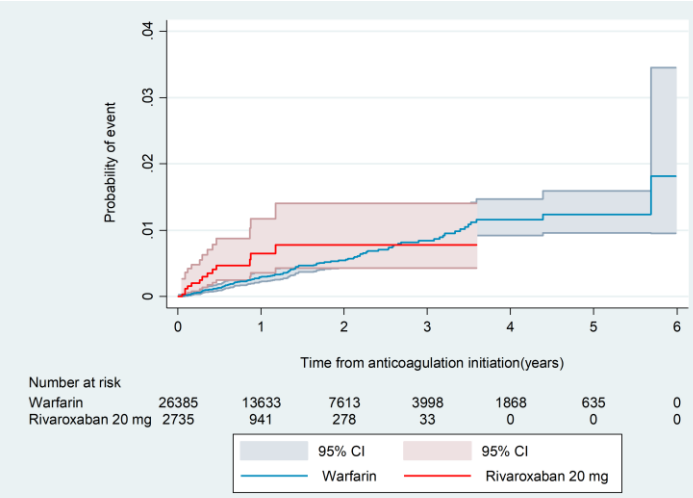


Figure XIV-8 d (GI bleeding)

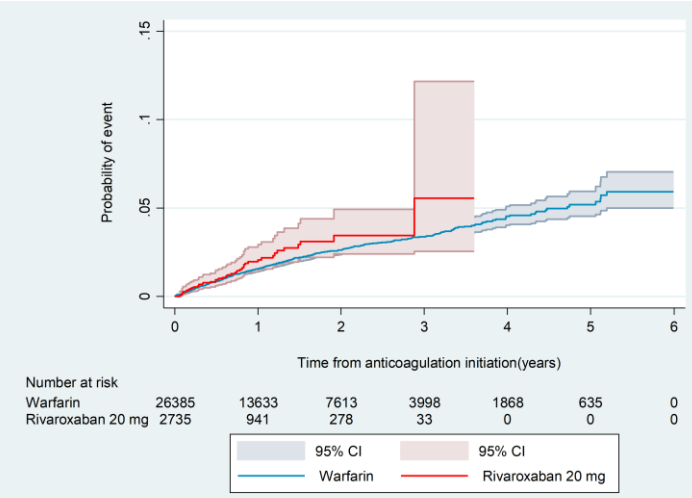


Figure XIV-8: MI, ICH and bleeding (rivaroxaban 20 mg vs. warfarin)

Abbreviations: MI=myocardial, ICH= intracranial haemorrhage, GI=gastrointestinal

Figure XIV-9 a (mortality-all-cause)

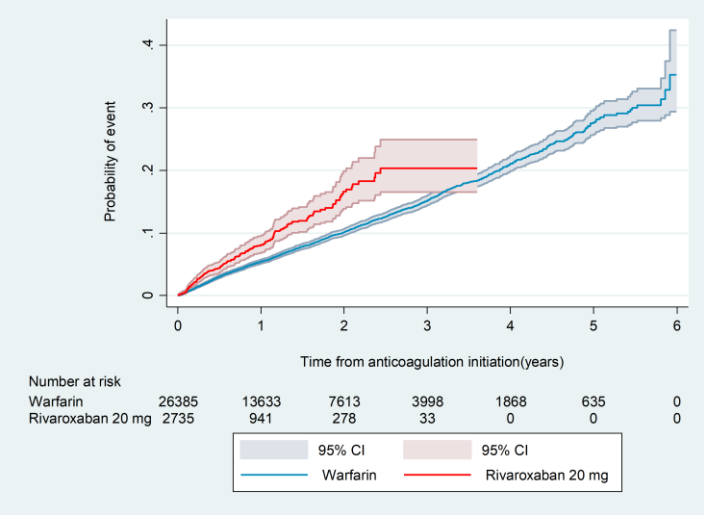


Figure XIV-9 b (mortality-stroke)

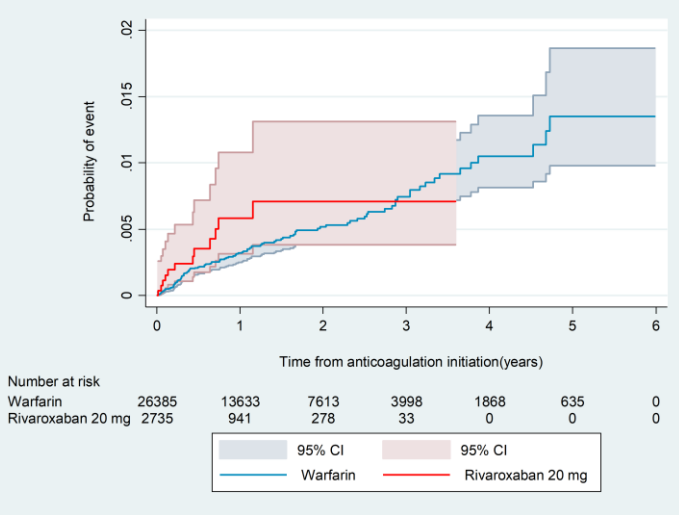


Figure XIV-9 c (mortality-cardiovascular)

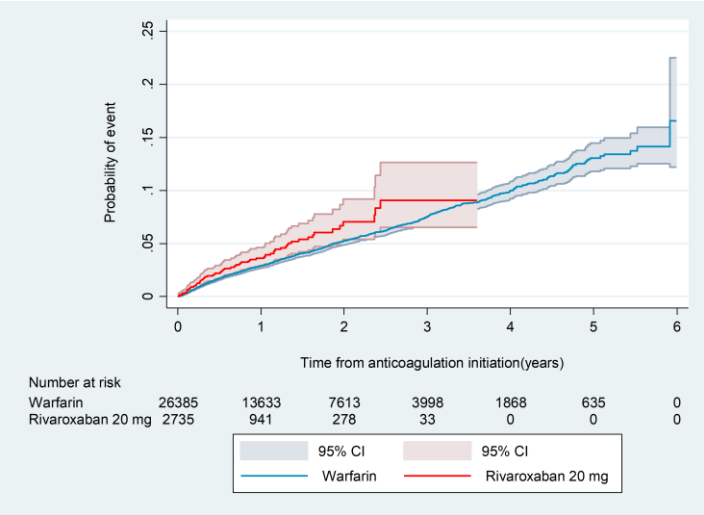


Figure XIV-9: Mortality (rivaroxaban 20 mg vs. warfarin)

Figure XIV-10 a (stroke-all)

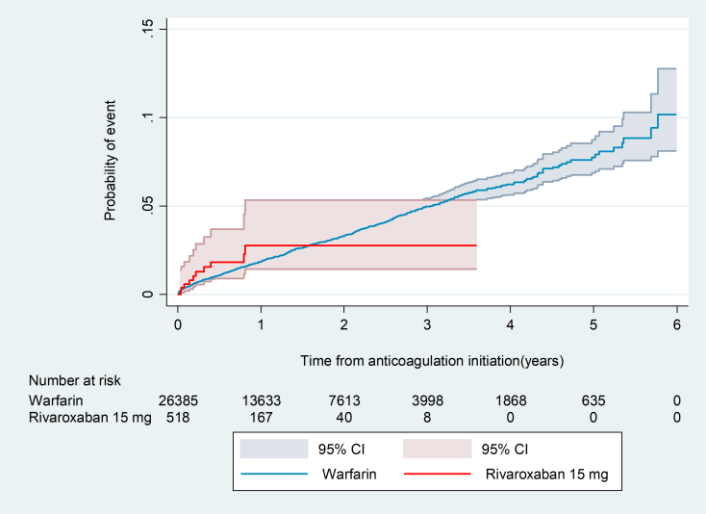


Figure XIV-10 b (stroke or SE)

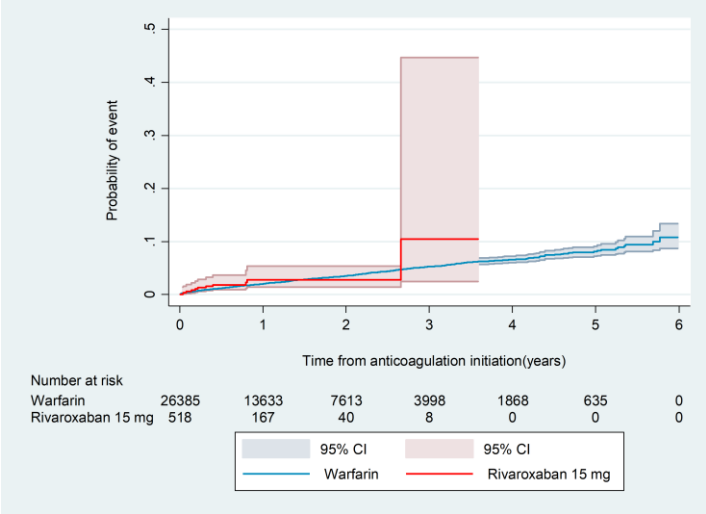


Figure XIV-10 c (stroke or SE or TIA)

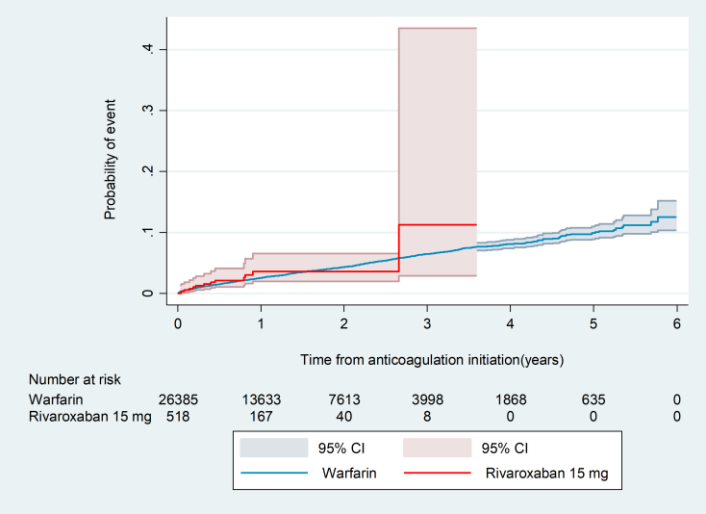


Figure XIV-10 d (stroke or mortality-all-cause)

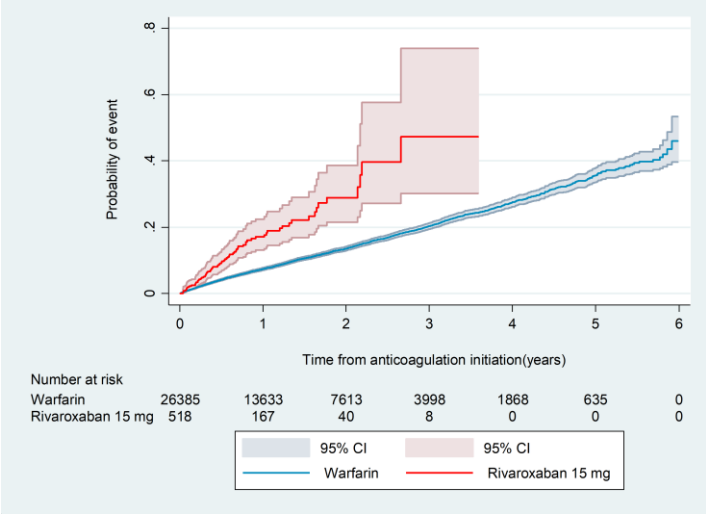


Figure XIV-10: Stroke, SE and TIA (rivaroxaban 15 mg vs. warfarin)
Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack

Figure XIV-11 a (MI)

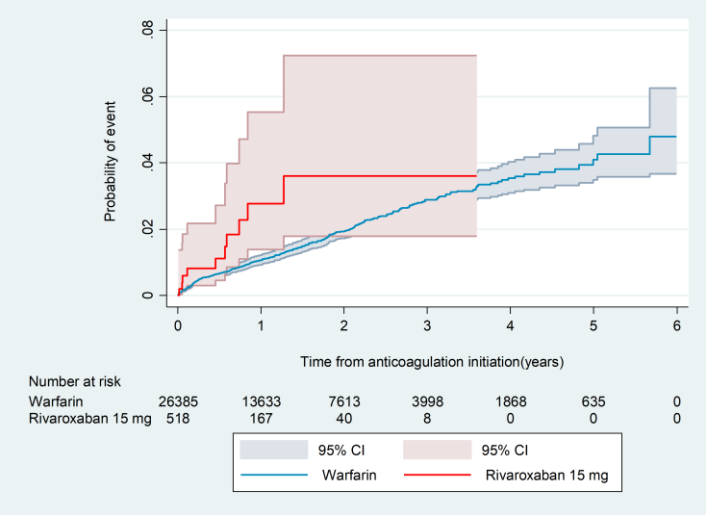


Figure XIV-11 b (major bleeding)

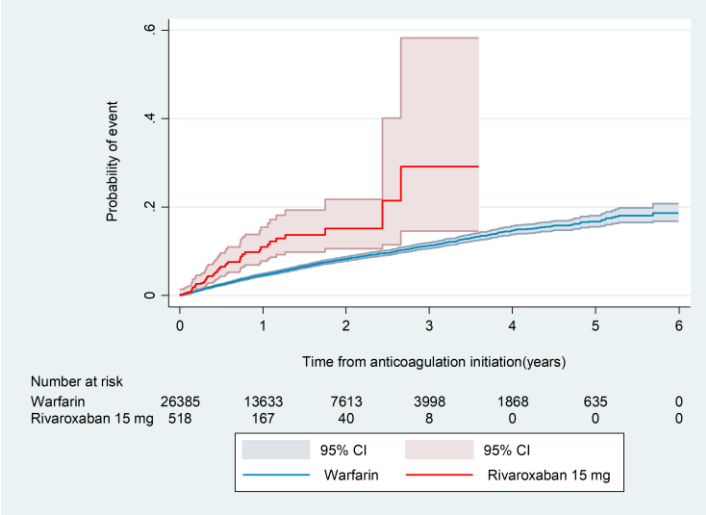


Figure XIV-11 c (ICH)

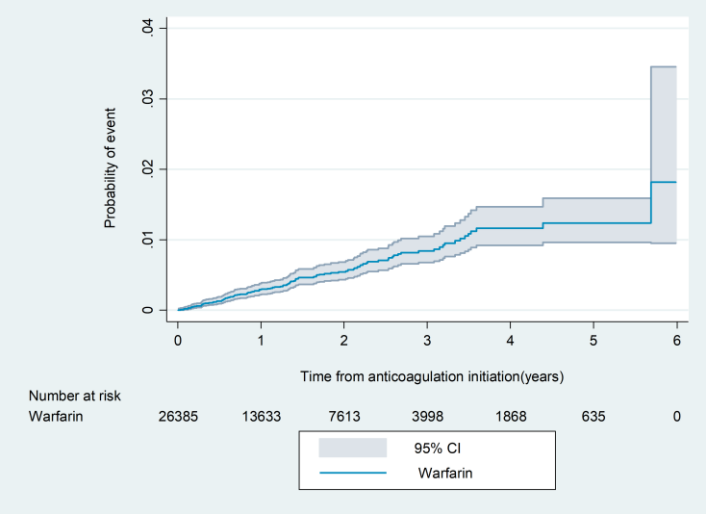


Figure XIV-11 d (GI bleeding)

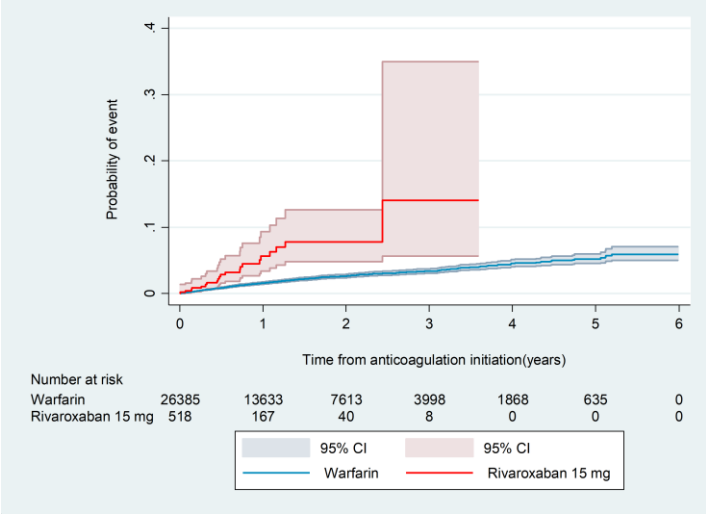


Figure XIV-11: MI, ICH and bleeding (rivaroxaban 15 mg vs. warfarin)

Note: due to disclosure restrictions, in the case of fewer than five events, the cumulative incidence curve was not reported

Abbreviations: MI=myocardial, ICH= intracranial haemorrhage, GI=gastrointestinal

Figure XIV-12 a (mortality-all-cause)

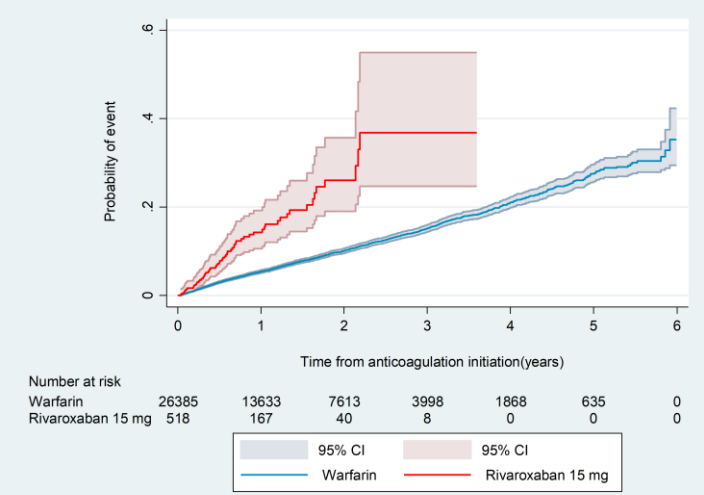


Figure XIV-12 b (mortality-stroke)

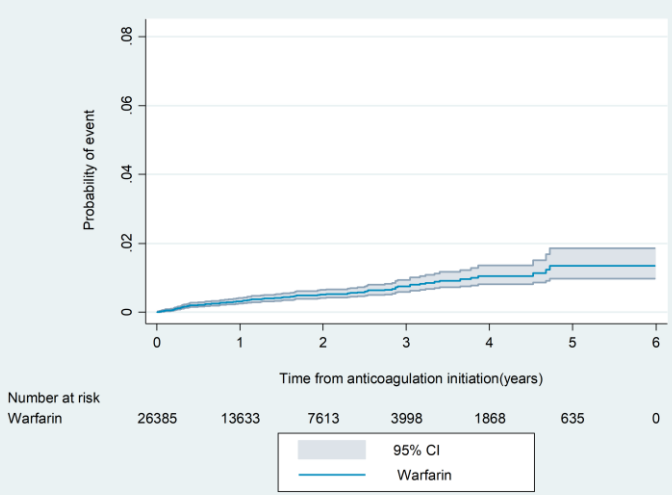


Figure XIV-12 c (mortality-cardiovascular)

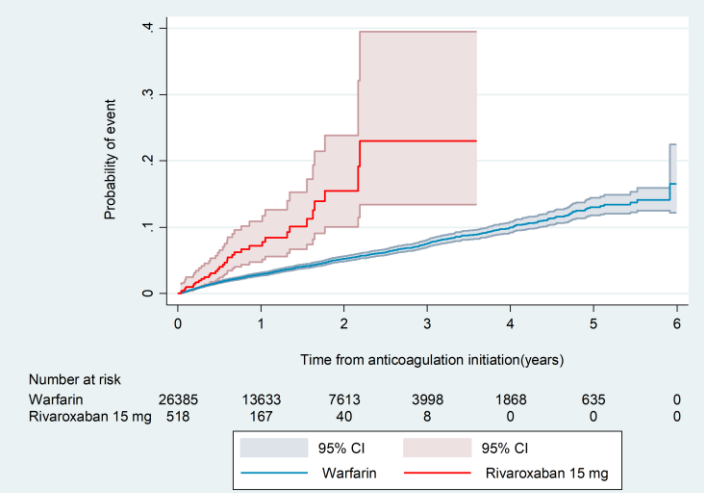


Figure XIV-12: Mortality (rivaroxaban 15 mg vs. warfarin)

Note: due to disclosure restrictions, in the case of fewer than five events, the cumulative incidence curve was not reported

Appendix XV: Hazard ratios for 6 years follow-up (standard and reduced dose)

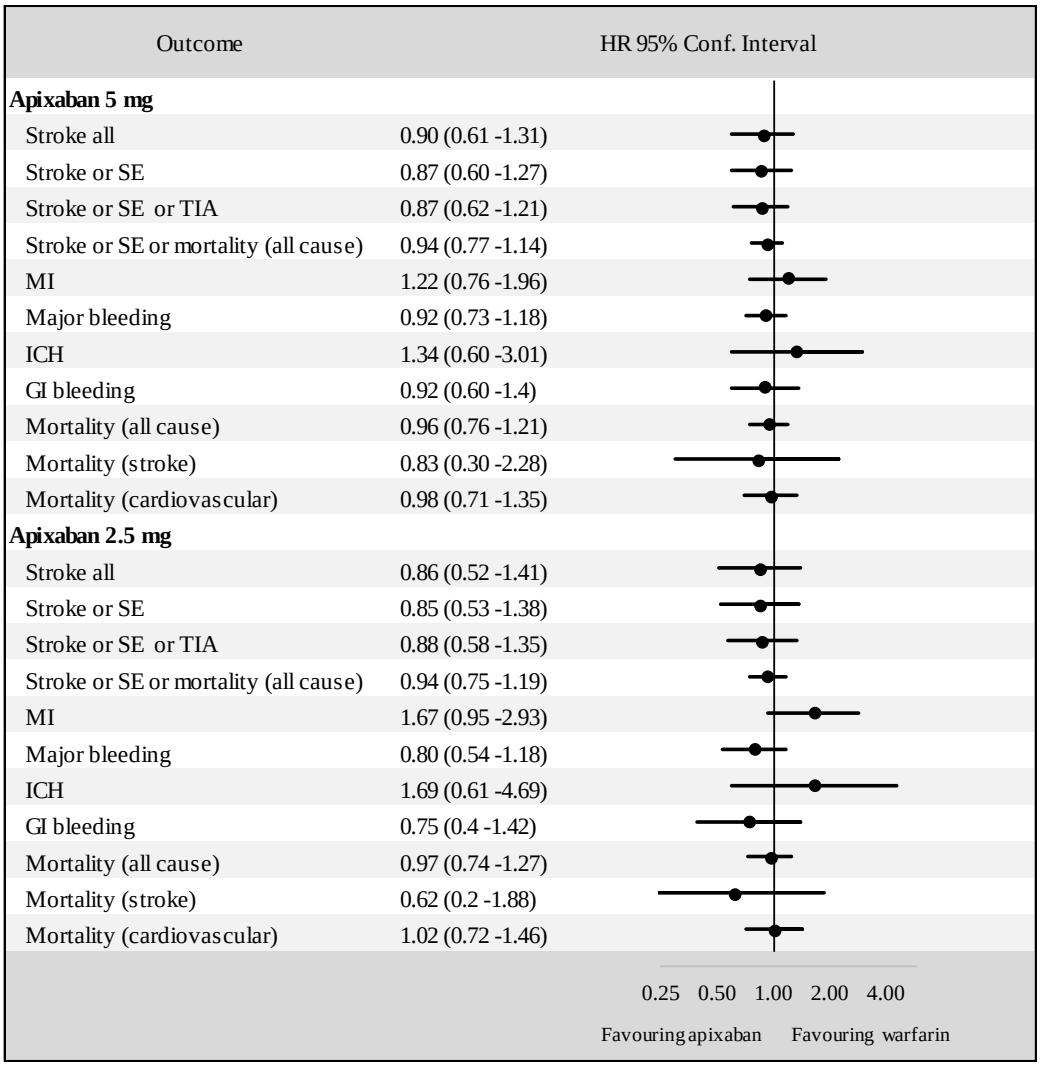


Figure XV-1: Hazard ratios 6 years since first prescription, apixaban (standard and reduced dose) vs. warfarin

Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack, MI=myocardial infarction, ICH=intracranial haemorrhage, GI=gastrointestinal.

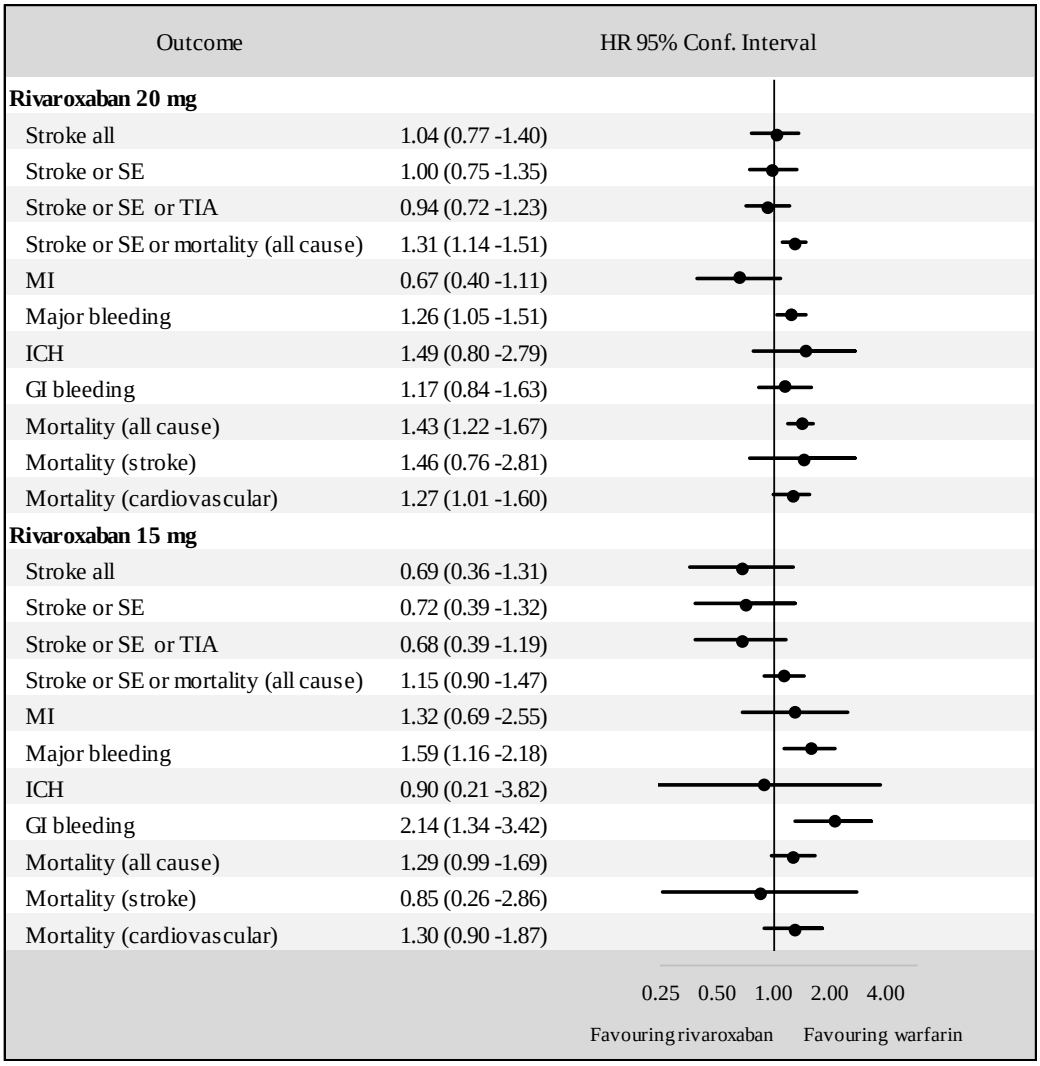


Figure XIV-2: Hazard ratios 6 years since first prescription, rivaroxaban (standard and reduced dose) vs. warfarin

Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack, MI=myocardial infarction, ICH=intracranial haemorrhage, GI=gastrointestinal.

Appendix XVI: Comparison of results with RCTs and observational studies

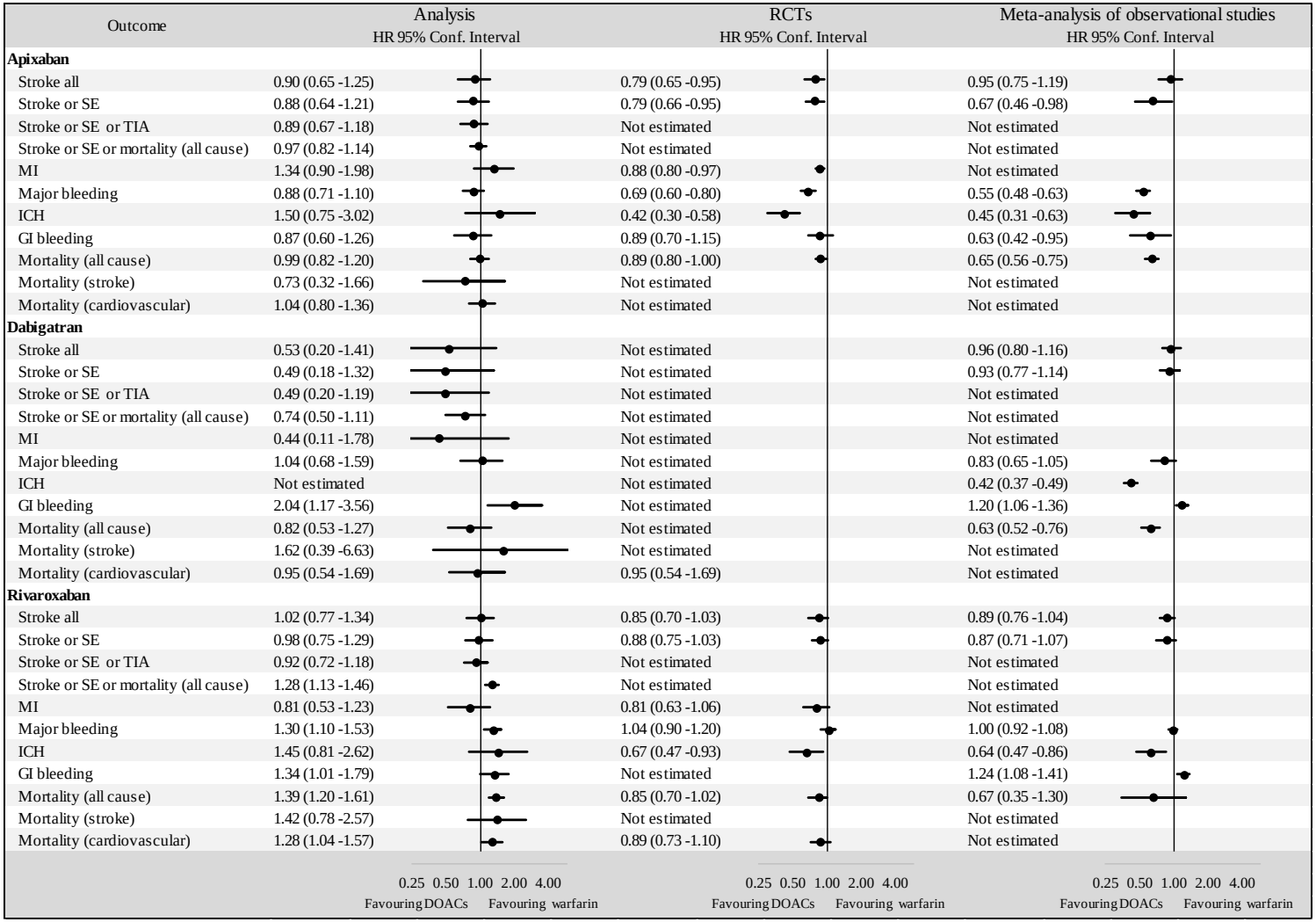


Figure XVI-1 Comparison of results with RCTs and observational studies

Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack, MI=myocardial infarction, ICH=intracranial haemorrhage, GI=gastrointestinal

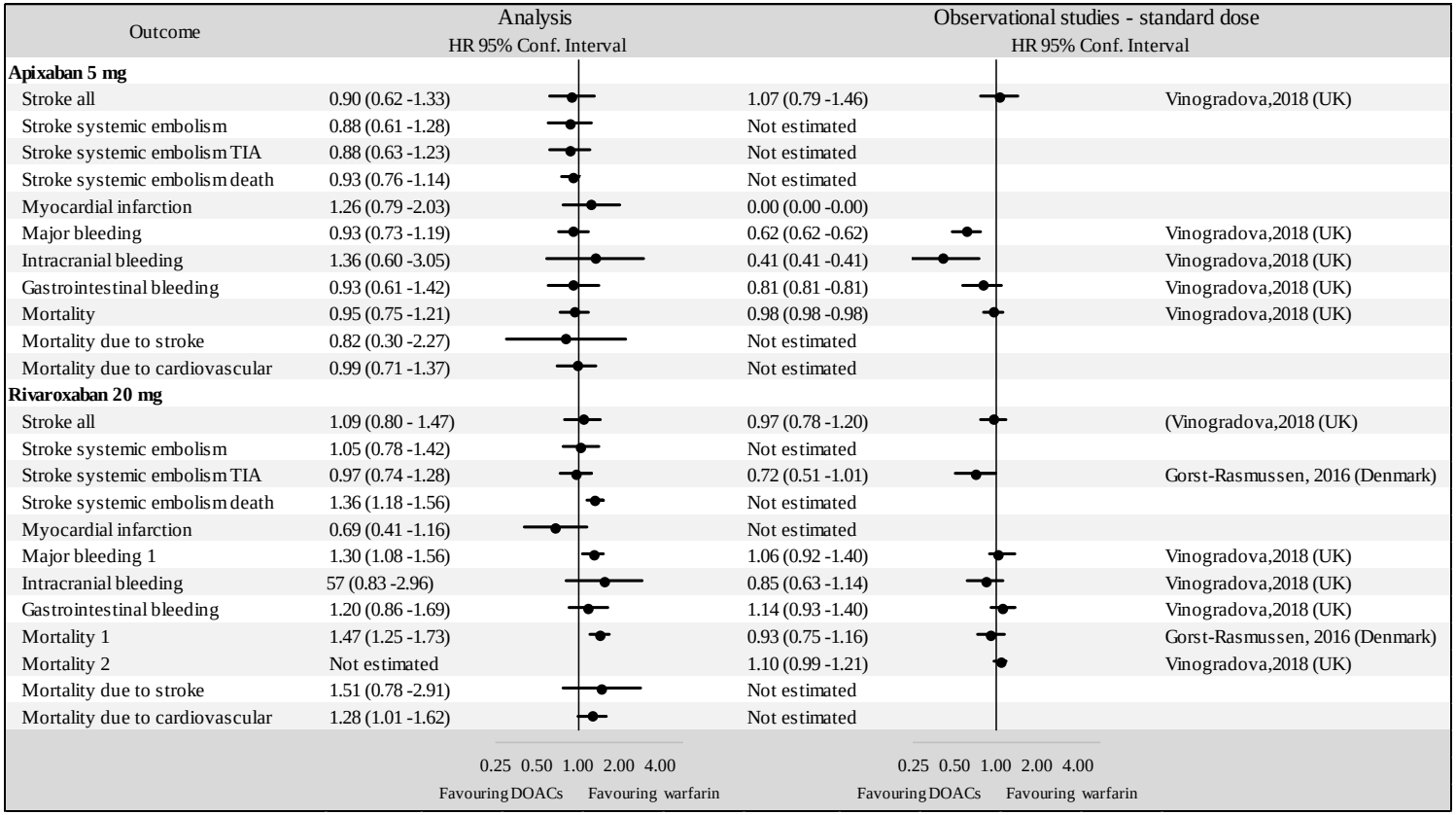


Figure XVI- 2 Comparison of results with RCTs and observational studies (standard dose)
Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack, MI=myocardial infarction, ICH=intracranial haemorrhage, GI=gastrointestinal

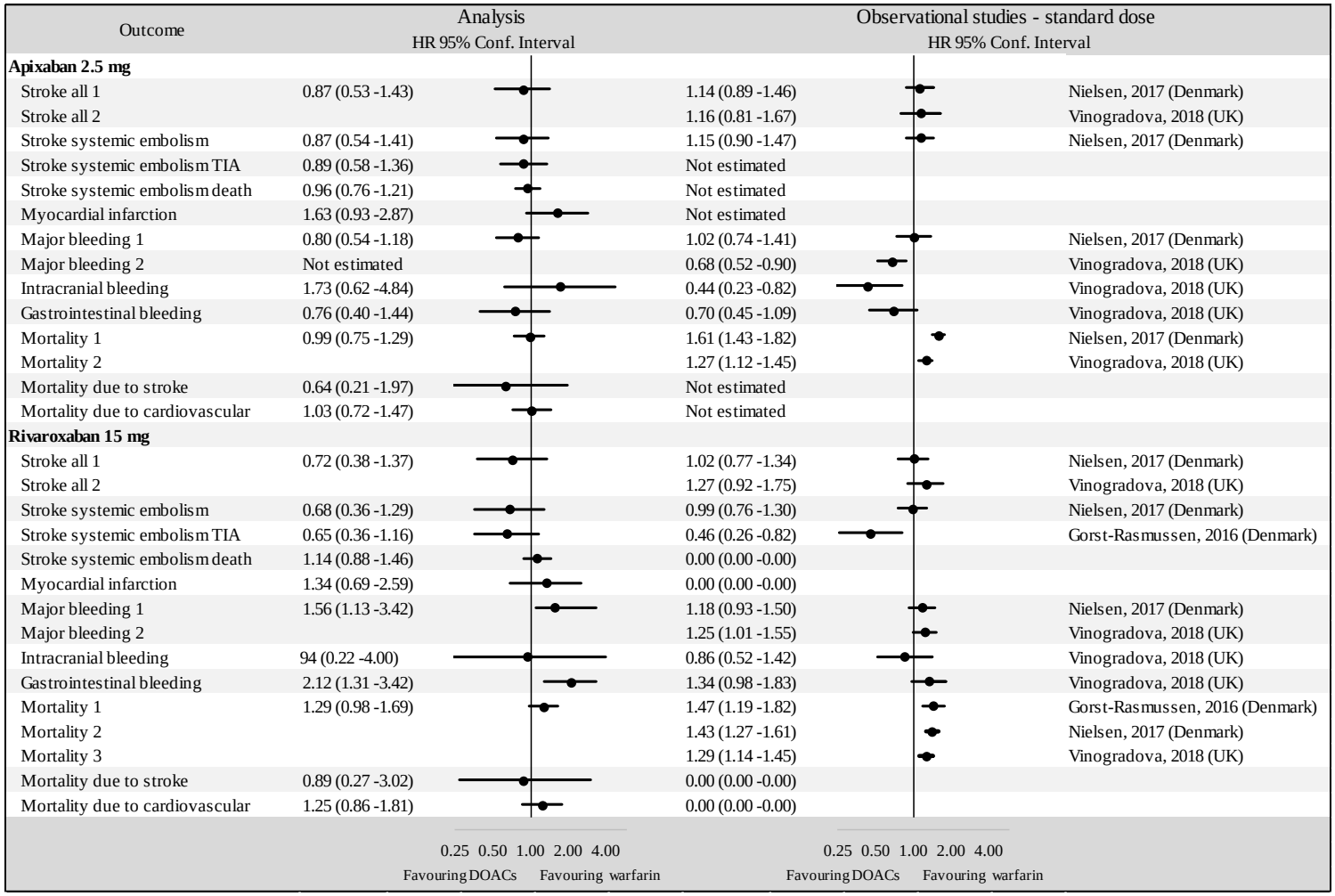


Figure XVI-3 Comparison of results with RCTs and observational studies (reduced dose)
Abbreviations: SE=systemic embolism, TIA=transient ischaemic attack, MI=myocardial infarction, ICH=intracranial haemorrhage, GI=gastrointestinal

Appendix XVII: Literature search strategy

1. Atrial fibrillation/
2. Markov/
3. Markov state transition model/
4. Markov simulation/
5. Markov chain/
6. Markov processes/
7. semi-Markov/
8. 2 or 3 or 4 or 5 or 6 or 7
9. decision analys*.mp.
10. decision analytic/
11. decision tree/
12. decision model/
13. 9 or 10 or 11 or 12
14. warfarin/
15. coumarins/
16. vitamin K antagonists/
17. Apixaban/
18. Dabigatran/
19. Rivaroxaban/
20. Anticoagulants/
21. Aspirin/
22. OAC/
23. NOAC/
24. DOAC/
25. 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24
26. cost-effectiveness
27. cost effectiveness
28. 26 or 27
29. 8 and 13 and 25 and 28

Appendix XVIII: CHEERS Checklist of items to include when reporting economic evaluations of health interventions

Table XVIII-1 CHEERS Checklist (Dorian,2014)

Section/item	Item No	Recommendation	Reported on page No /line No
Title and abstract			
Title	1	Identify the study as an economic evaluation or use more specific terms such as "cost-effectiveness analysis", and describe the interventions compared.	Page 1897.
Abstract	2	Provide a structured summary of objectives, perspective, methods (including study design and inputs), results (including base-case and uncertainty analyses), and conclusions.	Page 1897.
Introduction			
Background and objectives	3	Provide an explicit statement of the broader context for the study. Present the study question and its relevance for health policy or practice decisions.	Page 1898.
Methods			
Target population and subgroups	4	Describe characteristics of the base-case population and subgroups analysed, including why they were chosen.	Descriptive characteristics presented in page 1898-1900. Descriptive of the subgroup analysed in presented in Appendix SB.
Setting and location	5	State relevant aspects of the system(s) in which the decision(s) need(s) to be made.	Page 1898.
Study perspective	6	Describe the perspective of the study and relate this to the costs being evaluated.	Page 1898.
Comparators	7	Describe the interventions or strategies being compared and state why they were chosen.	Page 1898.
Time horizon	8	State the time horizon(s) over which costs and consequences are being evaluated and say why appropriate.	Page 1901.
Discount rate	9	Report the choice of discount rate(s) used for costs and outcomes and say why appropriate.	Page 1901, method section. Reported but not justified.
Choice of health outcomes	10	Describe what outcomes were used as the measure(s) of benefit in the evaluation and their relevance for the type of analysis performed.	Page 1901, method section. Reported but not justified.
Measurement of effectiveness	11a	Single study-based estimates: Describe fully the design features of the single effectiveness study and why the single study was a sufficient source of clinical effectiveness data.	Not applicable.
Measurement of effectiveness	11b	Synthesis-based estimates: Describe fully the methods used for identification of included studies and synthesis of clinical effectiveness data.	Appendices SA and SB.
Measurement and valuation of preference-based outcomes	12	If applicable, describe the population and methods used to elicit preferences for outcomes.	Not applicable.
Estimating resources and costs	13a	Single study-based economic evaluation: Describe approaches used to estimate resource use associated with the alternative interventions. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs.	Not applicable.
Estimating resources and costs	13b	Model-based economic evaluation: Describe approaches and data sources used to estimate resource use associated with model health states. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to cost opportunity cost.	Appendices SC, page 25-30.

Table XVIII-1 CHEERS Checklist (Dorian, 2014), continued

Section/item	Item No	Recommendation	Reported on page No/line No
Currency, price date, and conversion	14	Report the dates of the estimated resource quantities and unit costs. Describe methods for adjusting estimated unit costs to the year of reported costs if necessary. Describe methods for converting costs into a common currency base and the exchange rate.	Costs were inflated to 2011 values, Appendix SC, page 25. Methods for adjusting estimated unit cost are not reported. Methods for converting costs into a common currency is not applicable.
Choice of model	15	Describe and give reasons for the specific type of decision analytical model used. Providing a figure to show model structure is strongly recommended.	Page 1898.
Assumptions	16	Describe all structural or other assumptions underpinning the decision-analytical model.	Appendix SC, page 5-17.
Analytical methods	17	Describe all analytical methods supporting the evaluation. This could include methods for dealing with skewed, missing, or censored data; extrapolation methods; methods for pooling data; approaches to validate or make adjustments (such as half cycle corrections) to a model; and methods for handling population heterogeneity and uncertainty.	Appendix SC, page 33 (model validation). Appendix SC, page 36-51 (dealing with uncertainty).
Results			
Study parameters	18	Report the values, ranges, references, and, if used, probability distributions for all parameters. Report reasons or sources for distributions used to represent uncertainty where appropriate. Providing a table to show the input values is strongly recommended.	Appendices SC, page 49-65.
Incremental costs and outcomes	19	For each intervention, report mean values for the main categories of estimated costs and outcomes of interest, as well as mean differences between the comparator groups. If applicable, report incremental cost-effectiveness ratios.	Page 1902-1903
Characterising uncertainty	20a	Single study-based economic evaluation: Describe the effects of sampling uncertainty for the estimated incremental cost and incremental effectiveness parameters, together with the impact of methodological assumptions (such as discount rate, study perspective).	Not applicable
Characterising uncertainty	20b	Model-based economic evaluation: Describe the effects on the results of uncertainty for all input parameters, and uncertainty related to the structure of the model and assumptions.	Page 1902 (one-way sensitivity analysis), page 1904 (probabilistic sensitivity analysis).
Characterising heterogeneity	21	If applicable, report differences in costs, outcomes, or cost-effectiveness that can be explained by variations between subgroups of patients with different baseline characteristics or other observed variability in effects that are not reducible by more information.	Appendices SC, page 70-87.
Discussion			
	22	Summarise key study findings and describe how they support the conclusions reached. Discuss limitations and the generalisability of the findings and how the findings fit with current knowledge.	Page 1904-1905.
Other			
Source of funding	23	Describe how the study was funded and the role of the funder in the identification, design, conduct, and reporting of the analysis. Describe other non-monetary sources of support.	Page 1905.
Conflicts of interest	24	Describe any potential for conflict of interest of study contributors in accordance with journal policy. In the absence of a journal policy, we recommend authors comply with International Committee of Medical Journal Editors recommendations.	Page 1905.

Table XVIII-2 CHEERS Checklist (Lip, 2014)

Section/item	Item No	Recommendation	Reported on page No/line No
Title and abstract			
Title	1	Identify the study as an economic evaluation or use more specific terms such as "cost-effectiveness analysis", and describe the interventions compared.	Page 192.
Abstract	2	Provide a structured summary of objectives, perspective, methods (including study design and inputs), results (including base-case and uncertainty analyses), and conclusions.	Page 192.
Introduction			
Background and objectives	3	Provide an explicit statement of the broader context for the study. Present the study question and its relevance for health policy or practice decisions.	Page 193.
Methods			
Target population and subgroups	4	Describe characteristics of the base-case population and subgroups analysed, including why they were chosen.	Page 196-197, Table I.
Setting and location	5	State relevant aspects of the system(s) in which the decision(s) need(s) to be made.	Page 193.
Study perspective	6	Describe the perspective of the study and relate this to the costs being evaluated.	Page 193.
Comparators	7	Describe the interventions or strategies being compared and state why they were chosen.	Page 193.
Time horizon	8	State the time horizon(s) over which costs and consequences are being evaluated and say why appropriate.	Page 193.
Discount rate	9	Report the choice of discount rate(s) used for costs and outcomes and say why appropriate.	Page 199, method section. Reported but not justified.
Choice of health outcomes	10	Describe what outcomes were used as the measure(s) of benefit in the evaluation and their relevance for the type of analysis performed.	Page 199, method section. Reported but not justified.
Measurement of effectiveness	11a	Single study-based estimates: Describe fully the design features of the single effectiveness study and why the single study was a sufficient source of clinical effectiveness data.	Not applicable.
Measurement of effectiveness	11b	Synthesis-based estimates: Describe fully the methods used for identification of included studies and synthesis of clinical effectiveness data.	Appendix A.
Measurement and valuation of preference-based outcomes	12	If applicable, describe the population and methods used to elicit preferences for outcomes.	Not applicable.
Estimating resources and costs	13a	Single study-based economic evaluation: Describe approaches used to estimate resource use associated with the alternative interventions. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs.	Not applicable.
Estimating resources and costs	13b	Model-based economic evaluation: Describe approaches and data sources used to estimate resource use associated with model health states. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to cost opportunity cost.	Page 200.

Table XVIII-2 CHEERS Checklist (Lip, 2014), continued

Section/item	Item No	Recommendation	Reported on page No/line No
Currency, price date, and conversion	14	Report the dates of the estimated resource quantities and unit costs. Describe methods for adjusting estimated unit costs to the year of reported costs if necessary. Describe methods for converting costs into a common currency base and the exchange rate.	It is only stated that the cost in the model are reflected in 2011 British pounds, but no additional info is provided.
Choice of model	15	Describe and give reasons for the specific type of decision analytical model used. Providing a figure to show model structure is strongly recommended.	Page 193.
Assumptions	16	Describe all structural or other assumptions underpinning the decision-analytical model.	Page 194-195.
Analytical methods	17	Describe all analytical methods supporting the evaluation. This could include methods for dealing with skewed, missing, or censored data; extrapolation methods; methods for pooling data; approaches to validate or make adjustments (such as half cycle corrections) to a model; and methods for handling population heterogeneity and uncertainty.	Page 200 (dealing with uncertainty).
Results			
Study parameters	18	Report the values, ranges, references, and, if used, probability distributions for all parameters. Report reasons or sources for distributions used to represent uncertainty where appropriate. Providing a table to show the input values is strongly recommended.	Not reported.
Incremental costs and outcomes	19	For each intervention, report mean values for the main categories of estimated costs and outcomes of interest, as well as mean differences between the comparator groups. If applicable, report incremental cost-effectiveness ratios.	Page 201
Characterising uncertainty	20a	Single study-based economic evaluation: Describe the effects of sampling uncertainty for the estimated incremental cost and incremental effectiveness parameters, together with the impact of methodological assumptions (such as discount rate, study perspective).	Not applicable
Characterising uncertainty	20b	Model-based economic evaluation: Describe the effects on the results of uncertainty for all input parameters, and uncertainty related to the structure of the model and assumptions.	Page 202 (one-way sensitivity analysis), page 205 (probabilistic sensitivity analysis).
Characterising heterogeneity	21	If applicable, report differences in costs, outcomes, or cost-effectiveness that can be explained by variations between subgroups of patients with different baseline characteristics or other observed variability in effects that are not reducible by more information.	Not reported.
Discussion			
	22	Summarise key study findings and describe how they support the conclusions reached. Discuss limitations and the generalisability of the findings and how the findings fit with current knowledge.	Page 205-207.
Other			
Source of funding	23	Describe how the study was funded and the role of the funder in the identification, design, conduct, and reporting of the analysis. Describe other non-monetary sources of support.	Page 207.
Conflicts of interest	24	Describe any potential for conflict of interest of study contributors in accordance with journal policy. In the absence of a journal policy, we recommend authors comply with International Committee of Medical Journal Editors recommendations.	Page 207.

Table XVIII-3 CHEERS Checklist (Sterne, 2017)

Section/item	Item No	Recommendation	Reported on page No/line No
Title and abstract			
Title	1	Identify the study as an economic evaluation or use more specific terms such as "cost-effectiveness analysis", and describe the interventions compared.	Page VII.
Abstract	2	Provide a structured summary of objectives, perspective, methods (including study design and inputs), results (including base-case and uncertainty analyses), and conclusions.	Page VII.
Introduction			
Background and objectives	3	Provide an explicit statement of the broader context for the study. Present the study question and its relevance for health policy or practice decisions.	Page XIV.
Methods			
Target population and subgroups	4	Describe characteristics of the base-case population and subgroups analysed, including why they were chosen.	Brief description of the base population in page 16. No subgroup analysis considered.
Setting and location	5	State relevant aspects of the system(s) in which the decision(s) need(s) to be made.	Page 15.
Study perspective	6	Describe the perspective of the study and relate this to the costs being evaluated.	Page 15.
Comparators	7	Describe the interventions or strategies being compared and state why they were chosen.	Page 16.
Time horizon	8	State the time horizon(s) over which costs and consequences are being evaluated and say why appropriate.	Page 24.
Discount rate	9	Report the choice of discount rate(s) used for costs and outcomes and say why appropriate.	Page 22.
Choice of health outcomes	10	Describe what outcomes were used as the measure(s) of benefit in the evaluation and their relevance for the type of analysis performed.	Page 22, method section. Reported but not justified.
Measurement of effectiveness	11a	Single study-based estimates: Describe fully the design features of the single effectiveness study and why the single study was a sufficient source of clinical effectiveness data.	Not applicable.
Measurement of effectiveness	11b	Synthesis-based estimates: Describe fully the methods used for identification of included studies and synthesis of clinical effectiveness data.	Page 7-13.
Measurement and valuation of preference-based outcomes	12	If applicable, describe the population and methods used to elicit preferences for outcomes.	Not applicable.
Estimating resources and costs	13a	Single study-based economic evaluation: Describe approaches used to estimate resource use associated with the alternative interventions. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs.	Not applicable.
Estimating resources and costs	13b	Model-based economic evaluation: Describe approaches and data sources used to estimate resource use associated with model health states. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to cost opportunity cost.	Page 30-32

Table XVIII-3 CHEERS Checklist (Sterne, 2017), continued

Section/item	Item No	Recommendation	Reported on page No/line No
Currency, price date, and conversion	14	Report the dates of the estimated resource quantities and unit costs. Describe methods for adjusting estimated unit costs to the year of reported costs if necessary. Describe methods for converting costs into a common currency base and the exchange rate.	Costs were inflated to 2013-2014 values, page 30. Methods for adjusting estimated unit cost are not reported. Methods for converting costs into a common currency is not applicable.
Choice of model	15	Describe and give reasons for the specific type of decision analytical model used. Providing a figure to show model structure is strongly recommended.	Page 24-26.
Assumptions	16	Describe all structural or other assumptions underpinning the decision-analytical model.	Page 99, Table 68.
Analytical methods	17	Describe all analytical methods supporting the evaluation. This could include methods for dealing with skewed, missing, or censored data; extrapolation methods; methods for pooling data; approaches to validate or make adjustments (such as half cycle corrections) to a model; and methods for handling population heterogeneity and uncertainty.	Page 33-34.
Results			
Study parameters	18	Report the values, ranges, references, and, if used, probability distributions for all parameters. Report reasons or sources for distributions used to represent uncertainty where appropriate. Providing a table to show the input values is strongly recommended.	Page 30-34.
Incremental costs and outcomes	19	For each intervention, report mean values for the main categories of estimated costs and outcomes of interest, as well as mean differences between the comparator groups. If applicable, report incremental cost-effectiveness ratios.	Page 90 (no ICERs were reported, INB was reported instead).
Characterising uncertainty	20a	Single study-based economic evaluation: Describe the effects of sampling uncertainty for the estimated incremental cost and incremental effectiveness parameters, together with the impact of methodological assumptions (such as discount rate, study perspective).	Not applicable
Characterising uncertainty	20b	Model-based economic evaluation: Describe the effects on the results of uncertainty for all input parameters, and uncertainty related to the structure of the model and assumptions.	Page 92-93 (probabilistic sensitivity analysis).
Characterising heterogeneity	21	If applicable, report differences in costs, outcomes, or cost-effectiveness that can be explained by variations between subgroups of patients with different baseline characteristics or other observed variability in effects that are not reducible by more information.	Not reported
Discussion			
	22	Summarise key study findings and describe how they support the conclusions reached. Discuss limitations and the generalisability of the findings and how the findings fit with current knowledge.	Page 261-262
Other			
Source of funding	23	Describe how the study was funded and the role of the funder in the identification, design, conduct, and reporting of the analysis. Describe other non-monetary sources of support.	Page VIII
Conflicts of interest	24	Describe any potential for conflict of interest of study contributors in accordance with journal policy. In the absence of a journal policy, we recommend authors comply with International Committee of Medical Journal Editors recommendations.	Page III

Table XVIII-4 CHEERS Checklist (Verhoef, 2014)

Section/item	Item No	Recommendation	Reported on page No/line No
Title and abstract			
Title	1	Identify the study as an economic evaluation or use more specific terms such as "cost-effectiveness analysis", and describe the interventions compared.	Page 451.
Abstract	2	Provide a structured summary of objectives, perspective, methods (including study design and inputs), results (including base-case and uncertainty analyses), and conclusions.	Page 451.
Introduction			
Background and objectives	3	Provide an explicit statement of the broader context for the study. Present the study question and its relevance for health policy or practice decisions.	Page 452.
Methods			
Target population and subgroups	4	Describe characteristics of the base-case population and subgroups analysed, including why they were chosen.	Not reported.
Setting and location	5	State relevant aspects of the system(s) in which the decision(s) need(s) to be made.	Page 452.
Study perspective	6	Describe the perspective of the study and relate this to the costs being evaluated.	Page 452.
Comparators	7	Describe the interventions or strategies being compared and state why they were chosen.	Page 452.
Time horizon	8	State the time horizon(s) over which costs and consequences are being evaluated and say why appropriate.	Page 452.
Discount rate	9	Report the choice of discount rate(s) used for costs and outcomes and say why appropriate.	Page 455.
Choice of health outcomes	10	Describe what outcomes were used as the measure(s) of benefit in the evaluation and their relevance for the type of analysis performed.	Page 454.
Measurement of effectiveness	11a	Single study-based estimates: Describe fully the design features of the single effectiveness study and why the single study was a sufficient source of clinical effectiveness data.	Not applicable.
Measurement of effectiveness	11b	Synthesis-based estimates: Describe fully the methods used for identification of included studies and synthesis of clinical effectiveness data.	Page 453.
Measurement and valuation of preference-based outcomes	12	If applicable, describe the population and methods used to elicit preferences for outcomes.	Not applicable.
Estimating resources and costs	13a	Single study-based economic evaluation: Describe approaches used to estimate resource use associated with the alternative interventions. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs.	Not applicable.
Estimating resources and costs	13b	Model-based economic evaluation: Describe approaches and data sources used to estimate resource use associated with model health states. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to cost opportunity cost.	Page 455.

Table XVIII-4 CHEERS Checklist (Verhoef, 2014), continued

Section/item	Item No	Recommendation	Reported on page No/line No
Currency, price date, and conversion	14	Report the dates of the estimated resource quantities and unit costs. Describe methods for adjusting estimated unit costs to the year of reported costs if necessary. Describe methods for converting costs into a common currency base and the exchange rate.	Costs were inflated to 2012 values, page 455. Methods for adjusting estimated unit cost are not reported. Methods for converting costs into a common currency is not applicable.
Choice of model	15	Describe and give reasons for the specific type of decision analytical model used. Providing a figure to show model structure is strongly recommended.	Page 453.
Assumptions	16	Describe all structural or other assumptions underpinning the decision-analytical model.	Page 452-454.
Analytical methods	17	Describe all analytical methods supporting the evaluation. This could include methods for dealing with skewed, missing, or censored data; extrapolation methods; methods for pooling data; approaches to validate or make adjustments (such as half cycle corrections) to a model; and methods for handling population heterogeneity and uncertainty.	Page 455 (dealing with uncertainty).
Results			
Study parameters	18	Report the values, ranges, references, and, if used, probability distributions for all parameters. Report reasons or sources for distributions used to represent uncertainty where appropriate. Providing a table to show the input values is strongly recommended.	Page 454-455.
Incremental costs and outcomes	19	For each intervention, report mean values for the main categories of estimated costs and outcomes of interest, as well as mean differences between the comparator groups. If applicable, report incremental cost-effectiveness ratios.	Page 456-457.
Characterising uncertainty	20a	Single study-based economic evaluation: Describe the effects of sampling uncertainty for the estimated incremental cost and incremental effectiveness parameters, together with the impact of methodological assumptions (such as discount rate, study perspective).	Not applicable.
Characterising uncertainty	20b	Model-based economic evaluation: Describe the effects on the results of uncertainty for all input parameters, and uncertainty related to the structure of the model and assumptions.	Page 457 (one-way sensitivity analysis), page 457(probabilistic sensitivity analysis).
Characterising heterogeneity	21	If applicable, report differences in costs, outcomes, or cost-effectiveness that can be explained by variations between subgroups of patients with different baseline characteristics or other observed variability in effects that are not reducible by more information.	Page 456-457.
Discussion			
	22	Summarise key study findings and describe how they support the conclusions reached. Discuss limitations and the generalisability of the findings and how the findings fit with current knowledge.	Page 457-459.
Other			
Source of funding	23	Describe how the study was funded and the role of the funder in the identification, design, conduct, and reporting of the analysis. Describe other non-monetary sources of support.	Page 460.
Conflicts of interest	24	Describe any potential for conflict of interest of study contributors in accordance with journal policy. In the absence of a journal policy, we recommend authors comply with International Committee of Medical Journal Editors recommendations.	Not reported.

Appendix XIX: Cumulative incidence curves (cost-effectiveness)

Figure XIX-1 a (stroke-ischaemic)

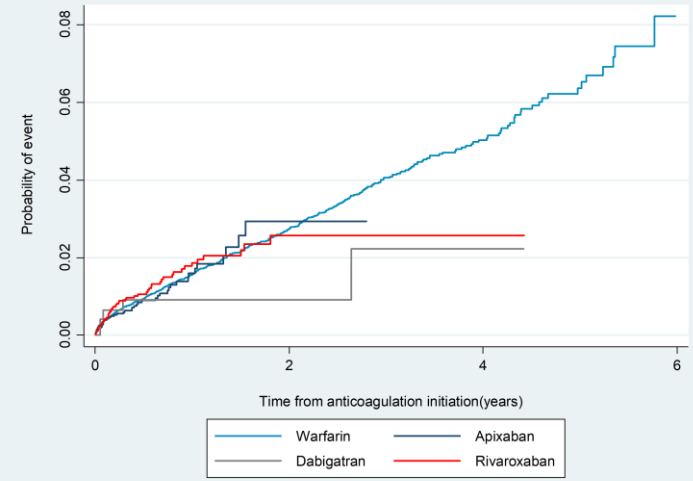


Figure XIX-I b (TIA)

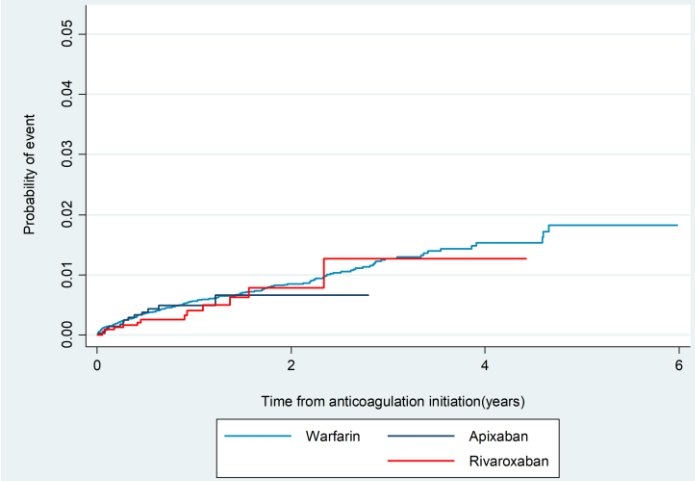


Figure XIX-1 c (SE)

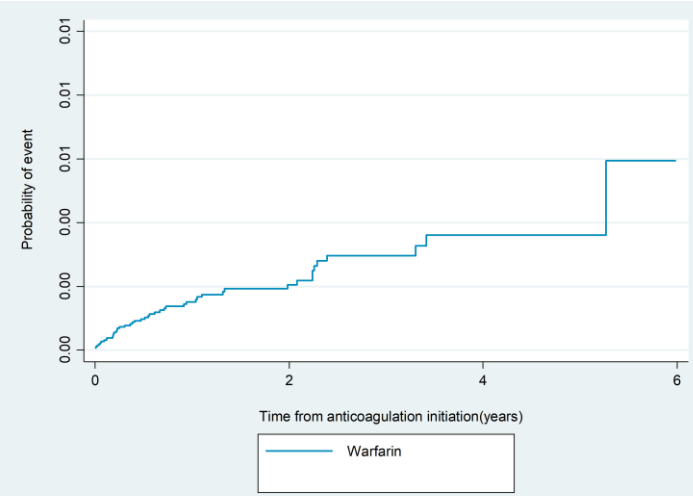


Figure XIX-1 d (MI)

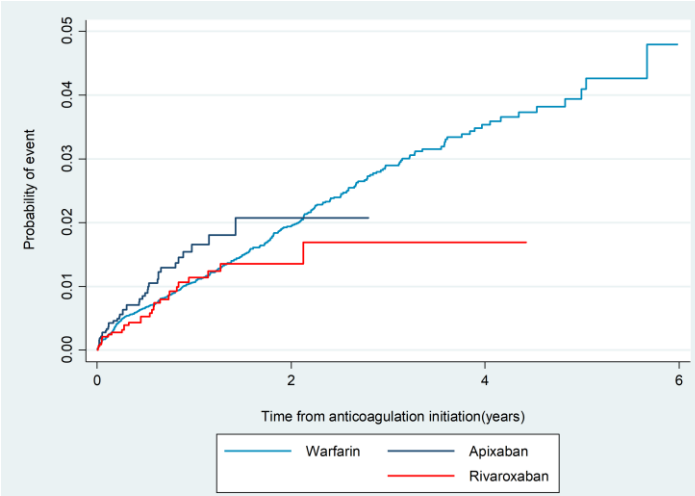


Figure XIX-1 Stroke, TIA, SE, MI (any dose)

Note: due to disclosure restrictions, in the case of fewer than five events, the cumulative incidence curve was not reported.
Abbreviations: TIA=transient ischaemic attack, SE=systemic embolism, MI=myocardial

Figure XIX-2 a (major bleeding)

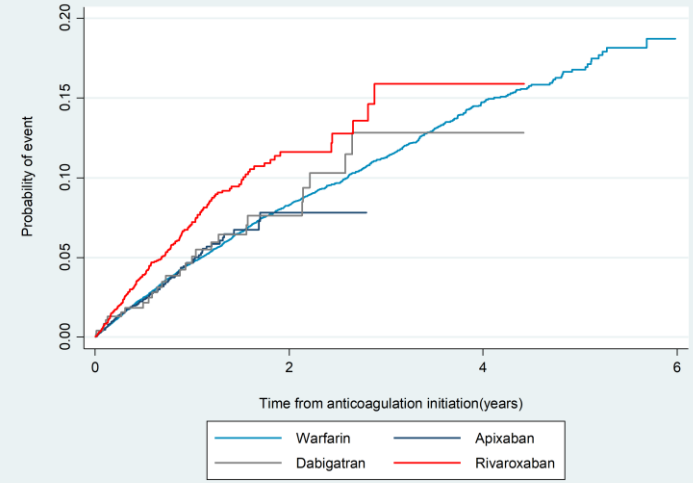


Figure XIX-2 b (ICH)

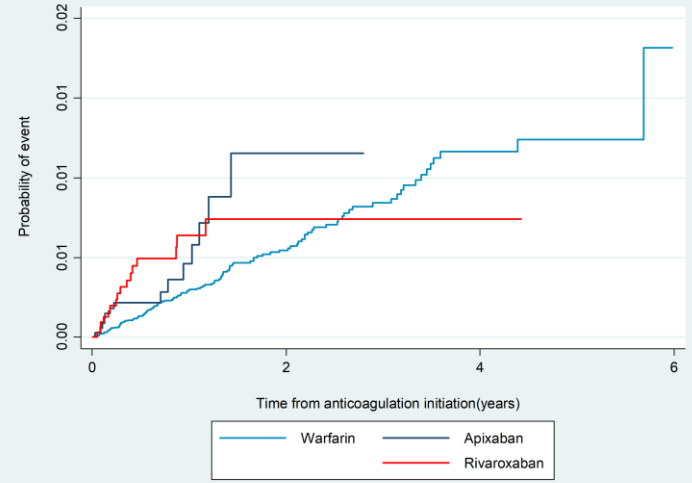


Figure XIX-2 c (mortality-all-cause)

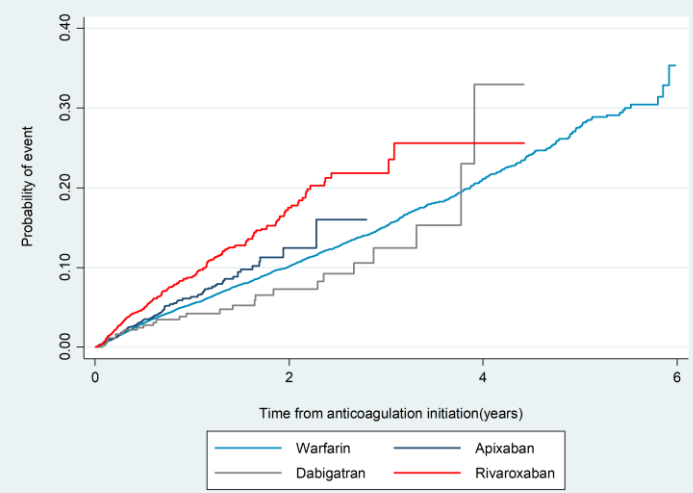


Figure XIX-2 Major bleeding, ICH, mortality (any dose)

Note: due to disclosure restrictions, in the case of fewer than five events, the cumulative incidence curve was not reported

Abbreviations: ICH= intracranial haemorrhage

Figure XIX-3 a (stroke-ischaemic)

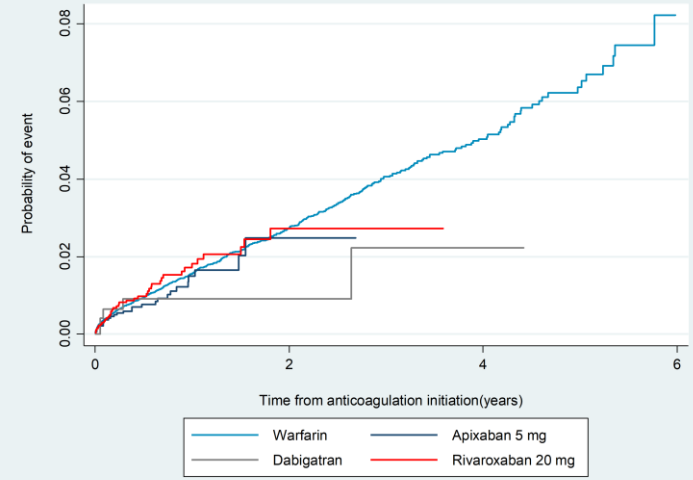


Figure XIX-3 b (TIA)

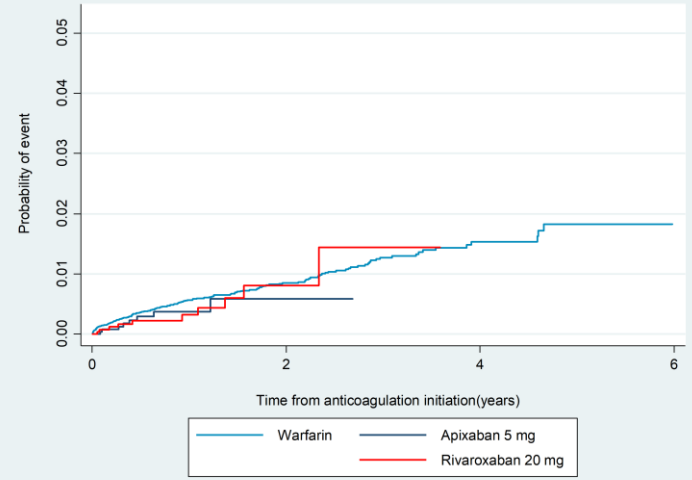


Figure XIX-3 c (SE)

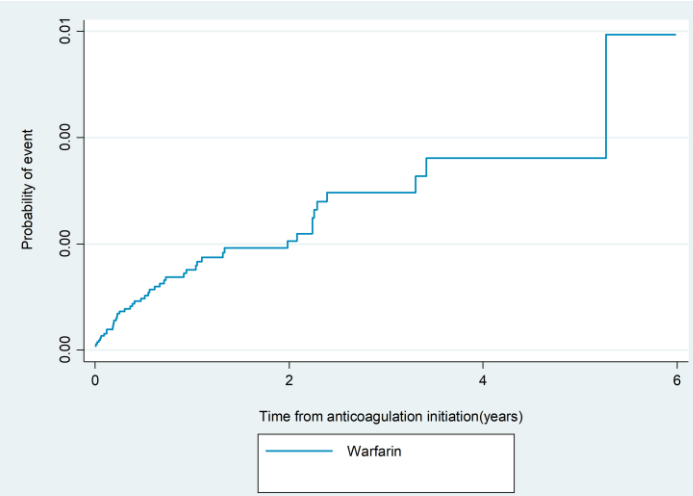


Figure XIX-3 d (MI)

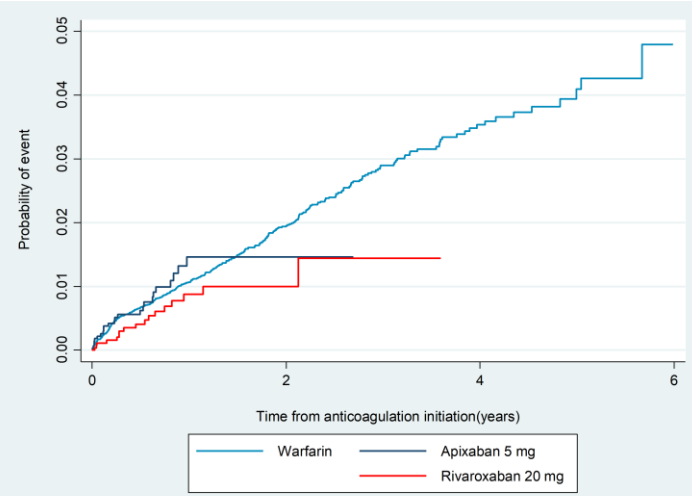


Figure XIX-3 Stroke, TIA, SE, MI (standard dose)

Note: due to disclosure restrictions, in the case of fewer than five events, the cumulative incidence curve was not reported

Abbreviations: TIA=transient ischaemic attack, SE=systemic embolism, MI=myocardial

Figure XIX-4 a (major bleeding)

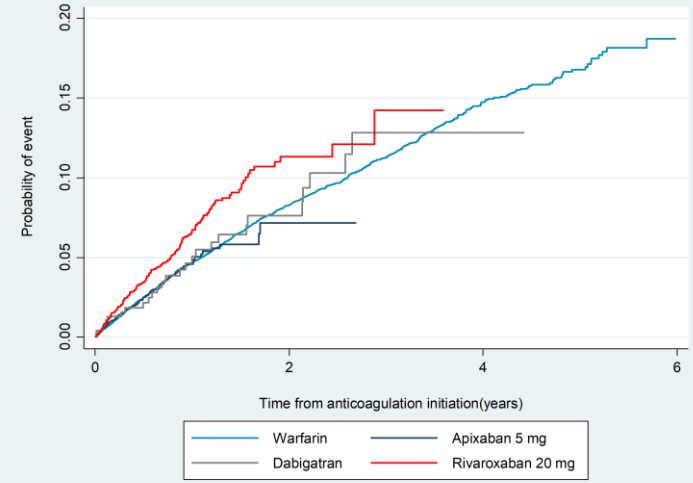


Figure XIX-4 b (ICH)

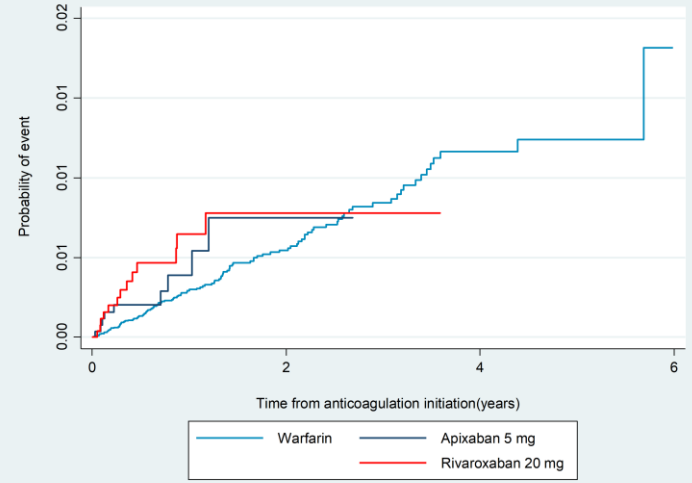


Figure XIX-4 c (mortality-all-cause)

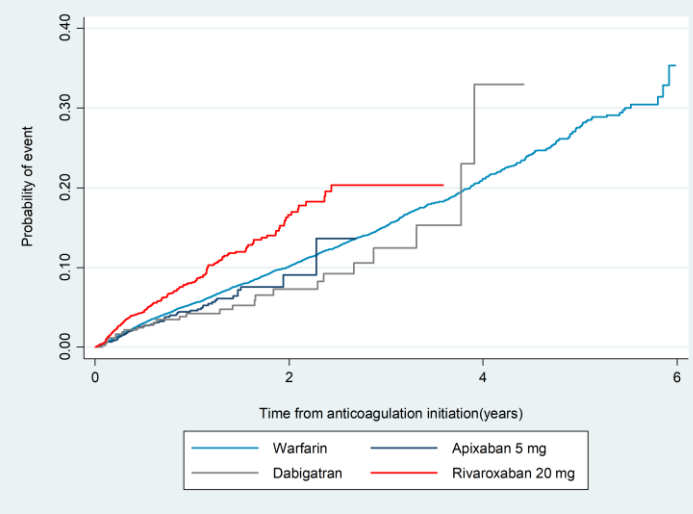


Figure XIX-4 Major bleeding, ICH, mortality (standard dose)

Note: due to disclosure restrictions, in the case of fewer than five events, the cumulative incidence curve was not reported

Abbreviations: ICH= intracranial haemorrhage

Figure XIX-5 a (stroke-ischaemic)

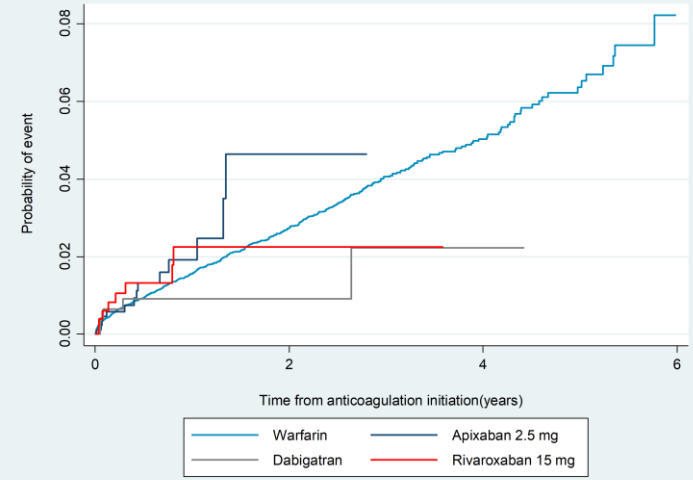


Figure XIX-5 b (TIA)

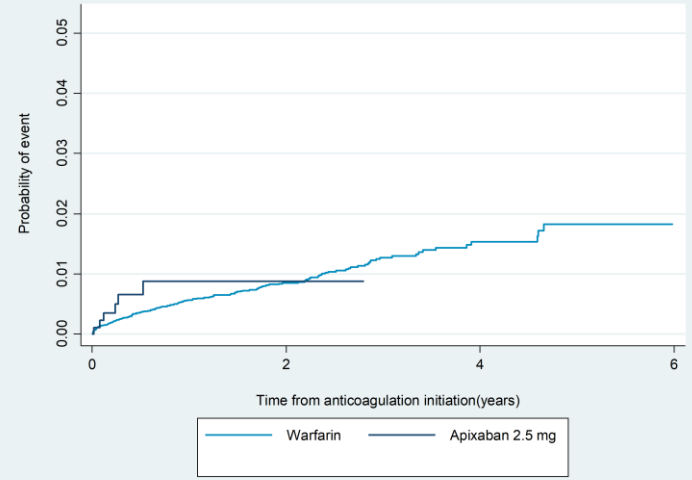


Figure XIX-5 c (SE)

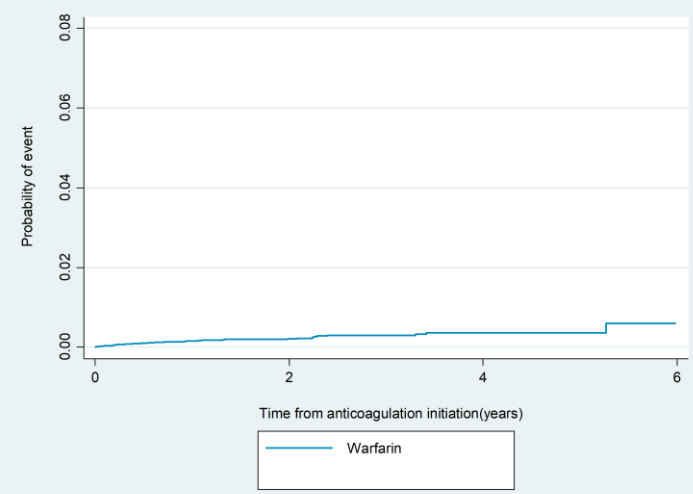


Figure XIX-5 d (MI)

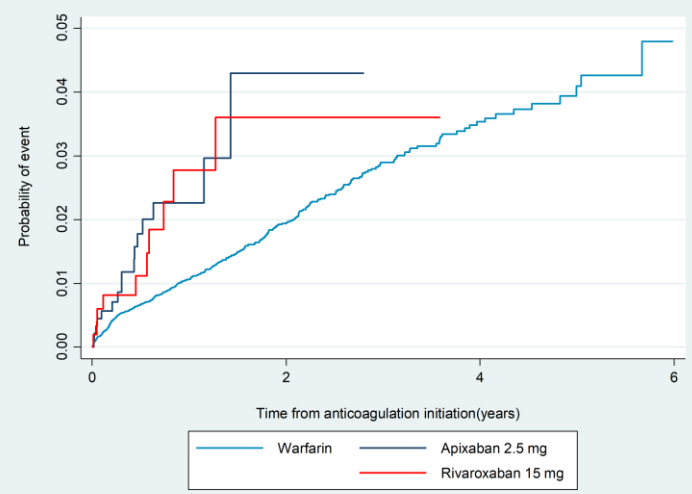


Figure XIX-5 Stroke, TIA, SE, MI (reduced dose)

Note: due to disclosure restrictions, in the case of fewer than five events, the cumulative incidence curve was not reported

Abbreviations: TIA=transient ischaemic attack, SE=systemic embolism, MI=myocardial

Figure XIX-6 a (major bleeding)

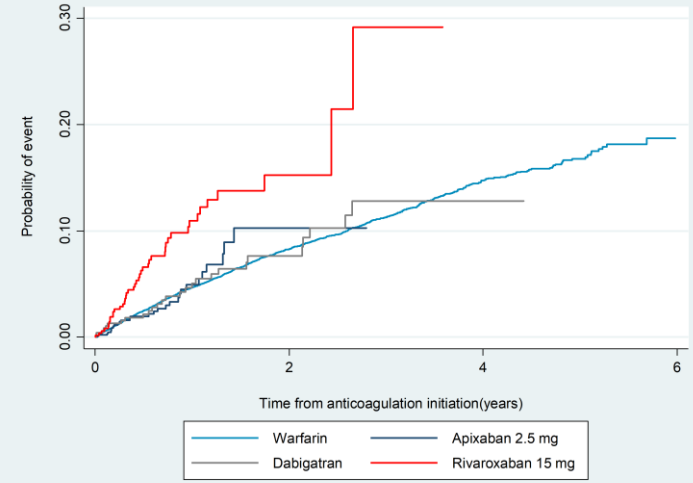


Figure XIX-6 b (ICH)

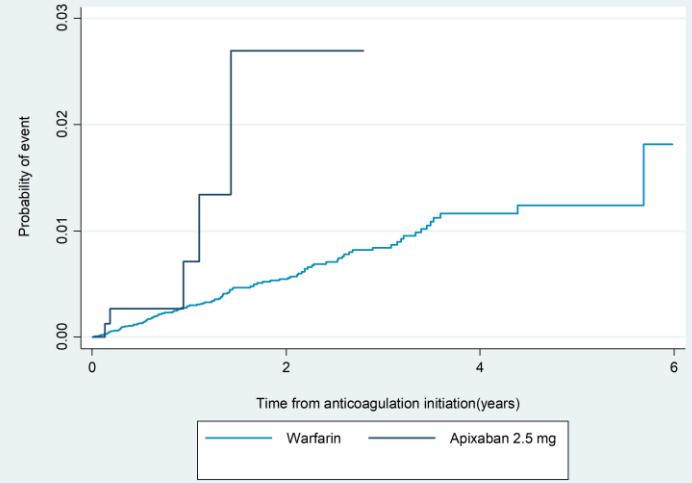


Figure XIX-6 c (mortality-all-cause)

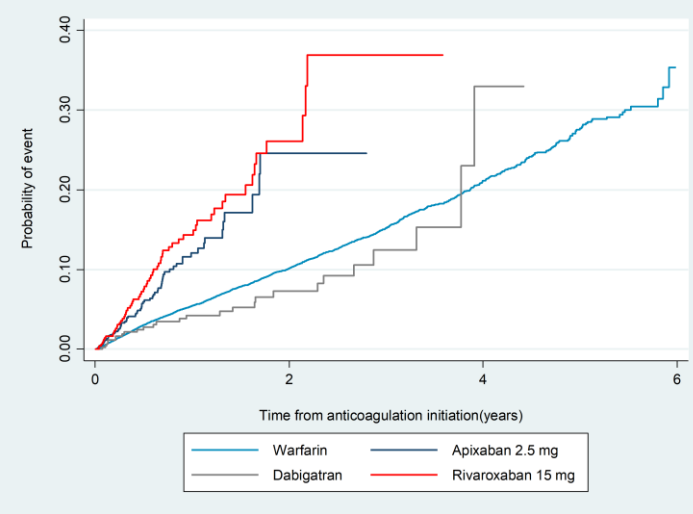


Figure XIX-6 Major bleeding, ICH, mortality (reduced dose)

Note: due to disclosure restrictions, in the case of fewer than five events, the cumulative incidence curve was not reported

Abbreviations: ICH= intracranial haemorrhage

Appendix XX: DOACs HRs for AF patients 70 years or older (Table I)

Intervention (compared to warfarin)	Outcome	HR 95 % CI	Standard error	Events and (event rates)
Apixaban				
	TIA	0.75 (0.36, 1.53)	0.272	29 (1.29)
	SE	0.62 (0.13, 3.00)	0.499	10 (0.58)
	Stroke	0.65 (0.42, 0.98)	0.138	<5 (0.29)
	MI	1.41 (0.93, 2.14)	0.298	35 (2.04)
	Major bleeding	0.80 (0.63, 1.02)	0.098	88 (5.13)
	ICH	1.20 (0.55, 2.64)	0.481	9 (0.52)
	Mortality (all-cause)	1.01 (0.83, 1.23)	0.100	143 (8.34)
Apixaban (standard dose)				
	TIA	0.53 (0.18, 1.50)	0.281	17 (1.46)
	SE	0.42 (0.05, 3.45)	0.451	<5 (0.43)
	Stroke	0.64 (0.38, 1.08)	0.171	<5 (0.43)
	MI	1.39 (0.83, 2.34)	0.369	19 (1.63)
	Major bleeding	0.85 (0.64, 1.13)	0.122	60 (5.16)
	ICH	0.86 (0.29, 2.54)	0.476	<5 (0.43)
	Mortality (all-cause)	0.97 (0.75, 1.25)	0.125	75 (6.45)
Apixaban (reduced dose)				
	TIA	1.05 (0.43, 2.58)	0.482	12 (2.17)
	SE	1.02 (0.12, 8.51)	1.104	6 (1.09)
	Stroke	0.65 (0.35, 1.20)	0.204	<5 (0.90)
	MI	1.59 (0.90, 2.81)	0.462	16 (2.90)
	Major bleeding	0.77 (0.52, 1.14)	0.155	28 (5.07)
	ICH	1.89 (0.67, 5.33)	0.999	5 (0.90)
	Mortality (all-cause)	1.02 (0.78, 1.34)	0.141	68 (12.30)
Dabigatran				
	TIA	0.96 (0.24, 3.89)	0.686	<5 (1.52)
	SE*	1.00 (1.00, 1.00)	0.000	<5 (1.52)
	Stroke	1.00 (1.00, 1.00)	0.167	<5 (1.52)
	MI	1.00 (1.00, 1.00)	0.236	<5 (1.52)
	Major bleeding	1.54 (1.04, 2.28)	0.307	26 (7.88)
	ICH	1.00 (1.00, 1.00)	0.712	<5 (1.52)
	Mortality (all-cause)	0.82 (0.53, 1.28)	0.186	20 (6.06)
Rivaroxaban				
	TIA	0.73 (0.39, 1.36)	0.231	41 (1.92)
	SE	0.29 (0.04, 2.23)	0.303	<5 (0.25)
	Stroke	0.85 (0.60, 1.19)	0.148	12 (0.56)
	MI	0.81 (0.53, 1.26)	0.182	24 (1.12)
	Major bleeding	1.29 (1.08, 1.54)	0.117	162 (7.57)
	ICH	1.61 (0.90, 2.88)	0.478	16 (0.75)
	Mortality (all-cause)	1.42 (1.23, 1.65)	0.106	245 (11.45)
Rivaroxaban (standard dose)				
	TIA	0.77 (0.38, 1.54)	0.274	31 (1.95)
	SE	1.00 (1.00, 1.00)	0.000	not estimated
	Stroke	0.90 (0.61, 1.32)	0.177	9 (0.57)
	MI	0.71 (0.41, 1.24)	0.202	14 (0.88)
	Major bleeding	1.26 (1.03, 1.55)	0.131	115 (7.24)
	ICH	1.76 (0.93, 3.31)	0.568	13 (0.82)
	Mortality (all-cause)	1.49 (1.26, 1.76)	0.127	177 (11.14)
Rivaroxaban (reduced dose)				
	TIA	0.52 (0.13, 2.17)	0.380	8 (1.95)
	SE	1.50 (0.19, 11.70)	1.570	<5 (1.22)
	Stroke	0.70 (0.34, 1.41)	0.257	<5 (1.22)
	MI	1.26 (0.63, 2.51)	0.443	9 (2.20)
	Major bleeding	1.63 (1.18, 2.24)	0.267	42 (10.25)
	ICH	1.04 (0.24, 4.38)	0.763	<5 (1.22)
	Mortality (all-cause)	1.29 (0.98, 1.70)	0.180	59 (14.40)

* Assumed no treatment effect difference between warfarin and dabigatran for SE and ICH.
Abbreviations: TIA= transient ischaemic attack, SE=systemic embolism, MI=myocardial infarction, ICH=intracranial haemorrhage.
Note: due to disclosure restrictions, in the case of fewer than five events, "<5" was reported

References

1. Dal-Ré, R., P. Janiaud, and J.P.A. Ioannidis, *Real-world evidence: How pragmatic are randomized controlled trials labeled as pragmatic?* BMC Medicine, 2018. **16**(1): p. 49.

2. Sherman, R.E., et al., *Real-World Evidence — What Is It and What Can It Tell Us?* New England Journal of Medicine, 2016. **375**(23): p. 2293-2297.

3. Albertson, T.E., et al., *The Salford Lung Study: a pioneering comparative effectiveness approach to COPD and asthma in clinical trials.* Pragmatic and observational research, 2017. **8**: p. 175-181.

4. McDonald, L., et al., *Real-world data in the United Kingdom: opportunities and challenges.* BMC Medicine, 2016. **14**: p. 97.

5. Euser, A.M., et al., *Cohort studies: prospective versus retrospective.* Nephron Clin Pract, 2009. **113**(3): p. c214-7.

6. The International Network of Agencies for Health Technology Assessment. *What is Health Technology Assessment (HTA)?* 2018 [cited 2018 June 2018]; Available from: http://www.inahta.org/.

7. Garrison, L.P., Jr., et al., *Using real-world data for coverage and payment decisions: the ISPOR Real-World Data Task Force report.* Value Health, 2007. **10**(5): p. 326-35.

8. London School of Economics. *The use of Real World Evidence in the European context: An analysis of key expert opinion*. 2016 [cited 2016 November 2015]; Available from: http://eprints.lse.ac.uk/68442/.

9. Griffiths, E.A. and N.K. Vadlamudi, *Not ready for the real world? The role of non-RCT evidence in health technology assessment.* Value in Health, 2016. **19**(3): p. A286.

10. S., W.N., *Clinical Epidemiology: The Study of the Outcome of Illness. 2nd ed*. 2006: Oxford University Press.

11. Sox, H.C. and S.N. Goodman, *The methods of comparative effectiveness research.* Annu Rev Public Health, 2012. **33**: p. 425-45.

12. Blonde, L., et al., *Interpretation and Impact of Real-World Clinical Data for the Practicing Clinician.* Advances in therapy, 2018. **35**(11): p. 1763-1774.

13. The Farr Institue of Health Informatics Research. *The Farr Institue of Health Informatics Research*. 2017 [cited 2017 December 2017]; Available from: http://farrinstitute.org/.

14. Fleming, M., B. Kirby, and K.I. Penny, *Record linkage in Scotland and its applications to health research.* Journal of Clinical Nursing, 2012. **21**(19pt20): p. 2711-2721.

15. Information Services Division Scotland. *Coding & Terminology Systems*. 2015 [cited 2016 June 2016]; Available from: http://www.isdscotland.org/Products-and-Services/Terminology-Services/Coding-and-Terminology-Systems/.

16. Scottish Medicines Consortium. *Scottish Medicines Consortium. apixaban (Eliquis)*. 2013 [cited 2015 November 2015]; Available from: https://www.scottishmedicines.org.uk/SMC_Advice/Advice/836_13_apixiban_Eliquis/apixaban_Eliquis.

17. Scottish Medicines Consortium. *Scottish Medicines Consortium. rivaroxaban (Xarelto) atrial*. 2012 [cited 2015 November 2015]; Available from: https://www.scottishmedicines.org.uk/SMC_Advice/Advice/756_12_rivaroxaban_Xarelto_atrial/rivaroxaban_Xarelto_atrial.

18. Scottish Medicines Consortium. *Scottish Medicines Consortium. dabigatran (Pradaxa)*. 2011 [cited 2015 November 2015]; Available from: https://www.scottishmedicines.org.uk/SMC_Advice/Advice/672_11_dabigatran_Pradaxa/dabigatran_Pradaxa.

19. Lane, D.A., et al., *Temporal Trends in Incidence, Prevalence, and Mortality of Atrial Fibrillation in Primary Care.* J Am Heart Assoc, 2017. **6**(5).

20. Zoni-Berisso, M., et al., *Epidemiology of atrial fibrillation: European perspective.* Clin Epidemiol, 2014. **6**: p. 213-20.

21. Investigators, W.M.P.P., *The world health organization monica project (monitoring trends and determinants in cardiovascular disease): A major international collaboration.* Journal of Clinical Epidemiology, 1988. **41**(2): p. 105-114.

22. Martinez, C., et al., *Increasing incidence of non-valvular atrial fibrillation in the UK from 2001 to 2013.* Heart, 2015. **101**(21): p. 1748-1754.

23. Public Health England. *Atrial fibrillation prevalence estimates in England*. 2017 [cited 2018 May 2018]; Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/644869/atrial_fibrillation_AF_briefing.pdf

24. Hinton, W., et al., *Incidence and prevalence of cardiovascular disease in English primary care: a cross-sectional and follow-up study of the Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC).* BMJ Open, 2018. **8**(8): p. e020282.

25. Cross-Party Group on Heart Disease and Stroke. *A Focus on Atrial Fibrillation in Scotland*. 2017 [cited 2018 November 2018]; Available from: https://www.stroke.org.uk/sites/default/files/a_focus_on_atrial_fibrillation_in_scotland.pdf.

26. Markides, V. and R.J. Schilling, *Atrial fibrillation: classification, pathophysiology, mechanisms and drug treatment.* Heart, 2003. **89**(8): p. 939-943.

27. Levy, J.H., et al., *Direct Oral Anticoagulants: New Drugs and New Concepts.* JACC: Cardiovascular Interventions, 2014. **7**(12): p. 1333-1351.

28. Ueberham, L., et al., *Pharmacological and Non-pharmacological Treatments for Stroke Prevention in Patients with Atrial Fibrillation.* Advances in Therapy, 2017. **34**(10): p. 2274-2294.

29. National Institute for Health and Care Excellence. *Preventing stroke in people with atrial fibrillation*. 2018 [cited 2018 April 2018]; Available from: https://pathways.nice.org.uk/pathways/atrial-fibrillation/preventing-stroke-in-people-with-atrial-fibrillation.

30. Melgaard, L., 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): p. 1030-8.

31. Pisters, 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): p. 1093-1100.

32. SCHULMAN, S., et al., *Definition of major bleeding in clinical investigations of antihemostatic medicinal products in non-surgical patients.* Journal of Thrombosis and Haemostasis, 2005. **3**(4): p. 692-694.

33. NHS choices. *Warfarin*. 2015 [cited 2016 November 2016]; Available from: http://www.nhs.uk/conditions/Anticoagulants-warfarin-/Pages/Introduction.aspx.

34. Ganetsky, M., et al., *Dabigatran: Review of Pharmacology and Management of Bleeding Complications of This Novel Oral Anticoagulant.* Journal of Medical Toxicology, 2011. **7**(4): p. 281-287.

35. National Institute for Health and Care Excellence. *NICE technology appraisal guidance 275. Apixaban for preventing stroke and systemic embolism in people with non-valvular atrial fibrillation*. 2013 [cited 2015 November 2015]; Available from: http://www.nice.org.uk/guidance/ta275.

36. National Institute for Health and Care Excellence. *NICE technology appraisal guidance 256. Rivaroxaban for the prevention of stroke and systemic embolism in people with atrial fibrillation*. 2012 [cited 2015 November 2015]; Available from: http://www.nice.org.uk/guidance/ta256.

37. National Institute for Health and Care Excellence. *NICE technology appraisal guidance 249.Dabigatran Etexilate for the prevention of stroke and systemic embolism in atrial fibrillation*. 2012 [cited 2015 November 2015]; Available from: http://www.nice.org.uk/guidance/ta249.

38. Mueller, T., et al., *Use of direct oral anticoagulants in patients with atrial fibrillation in Scotland: Applying a coherent framework to drug utilisation studies.* Pharmacoepidemiol Drug Saf, 2017. **26**(11): p. 1378-1386.

39. Scottish Medicines Consortium. *Scottish Medicines Consortium. edoxaban (Lixiana)*. 2015 [cited 2017 May 2017]; Available from: https://www.scottishmedicines.org.uk/medicines-advice/edoxaban-lixiana-fullsubmission-109515/.

40. Scottish Medicines Consortium. *Scottish Medicines Consortium*. 2018 [cited 2018 May 2018]; Available from: https://www.scottishmedicines.org.uk/.

41. Connolly , S.J., et al., *Dabigatran versus Warfarin in Patients with Atrial Fibrillation.* New England Journal of Medicine, 2009. **361**(12): p. 1139-1151.

42. Granger, C.B., et al., *Apixaban versus Warfarin in Patients with Atrial Fibrillation.* New England Journal of Medicine, 2011. **365**(11): p. 981-992.

43. Patel, M.R., et al., *Rivaroxaban versus warfarin in nonvalvular atrial fibrillation.* N Engl J Med, 2011. **365**(10): p. 883-91.

44. Al-Khatib, S.M., et al., *Outcomes of apixaban vs. warfarin by type and duration of atrial fibrillation: results from the ARISTOTLE trial.* European Heart Journal, 2013. **34**(31): p. 2464-2471.

45. Chung, N., et al., *Safety of edoxaban, an oral factor Xa inhibitor, in Asian patients with non-valvular atrial fibrillation.* Thromb Haemost, 2011. **105**(3): p. 535-44.

46. Connolly, S.J., et al., *Betrixaban compared with warfarin in patients with atrial fibrillation: results of a phase 2, randomized, dose-ranging study (Explore-Xa).* Eur Heart J, 2013. **34**(20): p. 1498-505.

47. Eikelboom, J.W., et al., *Rationale and design of AVERROES: apixaban versus acetylsalicylic acid to prevent stroke in atrial fibrillation patients who have failed or are unsuitable for vitamin K antagonist treatment.* Am Heart J, 2010. **159**(3): p. 348-353.e1.

48. Ezekowitz, M.D., et al., *Dabigatran with or without concomitant aspirin compared with warfarin alone in patients with nonvalvular atrial fibrillation (PETRO Study).* Am J Cardiol, 2007. **100**(9): p. 1419-26.

49. Hori, M., et al., *Rivaroxaban vs. warfarin in Japanese patients with atrial fibrillation - the J-ROCKET AF study.* Circ J, 2012. **76**(9): p. 2104-11.

50. Ogawa, S., Y. Shinohara, and K. Kanmuri, *Safety and efficacy of the oral direct factor xa inhibitor apixaban in Japanese patients with non-valvular atrial fibrillation. -The ARISTOTLE-J study.* Circ J, 2011. **75**(8): p. 1852-9.

51. Ruff, C.T., et al., *Evaluation of the novel factor Xa inhibitor edoxaban compared with warfarin in patients with atrial fibrillation: design and rationale for the Effective aNticoaGulation with factor xA next GEneration in Atrial Fibrillation-Thrombolysis In Myocardial Infarction study 48 (ENGAGE AF-TIMI 48).* Am Heart J, 2010. **160**(4): p. 635-41.

52. Weitz, J.I., et al., *Randomised, parallel-group, multicentre, multinational phase 2 study comparing edoxaban, an oral factor Xa inhibitor, with warfarin for stroke prevention in patients with atrial fibrillation.* Thromb Haemost, 2010. **104**(3): p. 633-41.

53. Yamashita, T., et al., *Randomized, multicenter, warfarin-controlled phase II study of edoxaban in Japanese patients with non-valvular atrial fibrillation.* Circ J, 2012. **76**(8): p. 1840-7.

54. Boehringer Ingelheim. *A Dose Response Study of Dabigatran Etexilate(BIBR 1048) in Pharmacodynamics and Safety in Patients With Non-valvular Atrial Fibrillation in Comparison to Warfarin*. 2011 [cited 2017 November 2017]; Available from: https://clinicaltrials.gov/ct2/show/NCT01136408.

55. López-López, J.A., et al., *Oral anticoagulants for prevention of stroke in atrial fibrillation: systematic review, network meta-analysis, and cost effectiveness analysis.* BMJ, 2017. **359**.

56. Li, T., et al., *Network meta-analysis-highly attractive but more methodological research is needed.* BMC Medicine, 2011. **9**(1): p. 79.

57. Riley, R.D., J.P.T. Higgins, and J.J. Deeks, *Interpretation of random effects meta-analyses.* BMJ, 2011. **342**: p. d549.

58. Ntaios, G., et al., *Real-World Setting Comparison of Nonvitamin-K Antagonist Oral Anticoagulants Versus Vitamin-K Antagonists for Stroke Prevention in Atrial Fibrillation: A Systematic Review and Meta-Analysis.* Stroke, 2017. **48**(9): p. 2494-2503.

59. Austin, P.C., *An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies.* Multivariate Behavioral Research, 2011. **46**(3): p. 399-424.

60. Gorst-Rasmussen, A., G.Y. Lip, and T. Bjerregaard Larsen, *Rivaroxaban versus warfarin and dabigatran in atrial fibrillation: comparative effectiveness and safety in Danish routine care.* Pharmacoepidemiol Drug Saf, 2016. **25**(11): p. 1236-1244.

61. Nielsen, P.B., et al., *Effectiveness and safety of reduced dose non-vitamin K antagonist oral anticoagulants and warfarin in patients with atrial fibrillation: propensity weighted nationwide cohort study.* BMJ, 2017. **356**.

62. Vinogradova, Y., et al., *Risks and benefits of direct oral anticoagulants versus warfarin in a real world setting: cohort study in primary care*. BMJ, 2018. **362**.

63. Larsen, T.B., et al., *Comparative effectiveness and safety of non-vitamin K antagonist oral anticoagulants and warfarin in patients with atrial fibrillation: propensity weighted nationwide cohort study.* Bmj, 2016. **353**: p. i3189.

64. Proietti, M., et al., *Real-World Use of Apixaban for Stroke Prevention in Atrial Fibrillation: A Systematic Review and Meta-Analysis.* Stroke, 2018. **49**(1): p. 98-106.

65. Aldridge, R.W., et al., *Accuracy of Probabilistic Linkage Using the Enhanced Matching System for Public Health and Epidemiological Studies.* PLoS One, 2015. **10**(8): p. e0136179.

66. Dusetzina, S.B., et al., *AHRQ Methods for Effective Health Care*, in *Linking Data for Health Services Research: A Framework and Instructional Guide*. 2014, Agency for Healthcare Research and Quality (US): Rockville (MD).

67. Kendrick, S. and J. Clarke, *The Scottish Record Linkage System.* Health Bull (Edinb), 1993. **51**(2): p. 72-9.

68. Information Services Division Scotland. *General Acute Inpatient and Day Case - Scottish Morbidity Record (SMR01)*. 2015 [cited 2017 November 2017]; Available from: http://www.ndc.scot.nhs.uk/National-Datasets/data.asp?SubID=2.

69. Manchikanti, L., F.J. Falco, and J.A. Hirsch, *Necessity and implications of ICD-10: facts and fallacies.* Pain Physician, 2011. **14**(5): p. E405-25.

70. O'Malley, K.J., et al., *Measuring Diagnoses: ICD Code Accuracy.* Health Services Research, 2005. **40**(5 Pt 2): p. 1620-1639.

71. Health & Social Care Information Centre. *OPCS-4 Classification*. 2015 [cited 2015 December 2015]; Available from: http://systems.hscic.gov.uk/data/clinicalcoding/codingstandards/opcs4.

72. Information Services Division Scotland. *Outpatient Attendance (SMR00)*. 2015 [cited 2015 January 2015]; Available from: http://www.ndc.scot.nhs.uk/National-Datasets/data.asp?SubID=2.

73. Scotland, I.S.D. *Prescribing and Medicines*. 2015 [cited 2017; Available from: http://www.isdscotland.org/Health-Topics/Prescribing-and-Medicines/PRISMS/.

74. Alvarez-Madrazo, S., et al., *Data Resource Profile: The Scottish National Prescribing Information System (PIS).* International Journal of Epidemiology, 2016.

75. Information Service Division Scotland. *Care Home Census 2015*. 2015 [cited 2015 November 2015]; Available from: http://www.isdscotland.org/Health-Topics/Health-and-Social-Community-Care/Care-Homes/Previous-Publications/.

76. Kind P, H.G., Macran S.,. *UK population norms for EQ-5D: Centre for Health Economics, University of York*. 1999 [cited 2018 June 2018]; Available from: https://www.york.ac.uk/che/pdf/DP172.pdf.

77. National Records of Scotland. *National Records of Scotland*. 2015 [cited 2015 November 2015]; Available from: http://www.gro-scotland.gov.uk/statistics-and-data.

78. Charlson, M.E., et al., *A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation.* Journal of Chronic Diseases, 1987. **40**(5): p. 373-383.

79. Heath Improvement Scotland. *Prevention of stroke in patients with atrial fibrillation*. 2013 [cited 2016 January 2016]; Available from: http://www.sign.ac.uk/assets/af_publication.pdf.

80. Bennell, M.C., et al., *Identifying predictors of cumulative healthcare costs in incident atrial fibrillation: a population-based study.* J Am Heart Assoc, 2015. **4**(4).

81. Hallinen, T., et al., *Direct costs of warfarin treatment among patients with atrial fibrillation in a Finnish health care setting.* Curr Med Res Opin, 2006. **22**(4): p. 683-92.

82. Nelson, W.W., et al., *Out-of-range international normalized ratio values and healthcare cost among new warfarin patients with non-valvular atrial fibrillation.* J Med Econ, 2015. **18**(5): p. 333-40.

83. Ghate, S.R., et al., *All-cause and bleeding-related health care costs in warfarin-treated patients with atrial fibrillation.* J Manag Care Pharm, 2011. **17**(9): p. 672-84.

84. Rohrbacker, N.J., et al., *The burden of atrial fibrillation and other cardiac arrhythmias in an employed population: associated costs, absences, and objective productivity loss.* J Occup Environ Med, 2010. **52**(4): p. 383-91.

85. Wu, E.Q., et al., *Economic burden and co-morbidities of atrial fibrillation in a privately insured population.* Curr Med Res Opin, 2005. **21**(10): p. 1693-9.

86. Holstenson, E., et al., *Predictors of costs related to cardiovascular disease among patients with atrial fibrillation in five European countries.* Europace, 2011. **13**(1): p. 23-30.

87. Reynolds, M.R., et al., *Healthcare resource utilization and costs associated with recurrent episodes of atrial fibrillation: the FRACTAL registry*. J Cardiovasc Electrophysiol, 2007. **18**(6): p. 628-33.

88. Jonsson, L., et al., *Cost of illness and drivers of cost in atrial fibrillation in Sweden and Germany.* Appl Health Econ Health Policy, 2010. **8**(5): p. 317-25.

89. Sankaranarayanan, R., et al., *Comparison of Atrial Fibrillation in the Young versus That in the Elderly: A Review.* Cardiology Research and Practice, 2013. **2013**: p. 16.

90. Johnsen, S.P., et al., *Cost of illness of atrial fibrillation: a nationwide study of societal impact.* BMC Health Services Research, 2017. **17**(1): p. 714.

91. Stewart, S., et al., *Cost of an emerging epidemic: an economic analysis of atrial fibrillation in the UK.* Heart, 2004. **90**(3): p. 286-292.

92. Casciano, J.P., et al., *The costs of warfarin underuse and nonadherence in patients with atrial fibrillation: a commercial insurer perspective.* J Manag Care Pharm, 2013. **19**(4): p. 302-16.

93. Manning, W.G., A. Basu, and J. Mullahy, *Generalized modeling approaches to risk adjustment of skewed outcomes data.* Journal of Health Economics, 2005. **24**(3): p. 465-488.

94. Burton, T.M., et al., *One-year follow-up healthcare costs of patients hospitalized for transient ischemic attack or ischemic stroke and discharged with aspirin plus extended-release dipyridamole or clopidogrel.* J Med Econ, 2012. **15**(6): p. 1217-25.

95. Duan, N., *Smearing Estimate: A Nonparametric Retransformation Method.* Journal of the American Statistical Association, 1983. **78**(383): p. 605-610.

96. Henry, A.G., et al., *Economic Evaluation in Clinical Trials*. 'Oxford University Press': Oxford, UK.

97. Tarricone, R., *Cost-of-illness analysis. What room in health economics?* Health Policy, 2006. **77**(1): p. 51-63.

98. Mogyorosy, Z. and P. Smith, *The main methodological issues in costing health care services: A literature review*. 2005.

99. de Lagasnerie, G., et al., *The economic burden of diabetes to French national health insurance: a new cost-of-illness method based on a combined medicalized and incremental approach.* Eur J Health Econ, 2018. **19**(2): p. 189-201.

100. Information Service Division Scotland. *Scottish National Tariff*. 2015 [cited 2015 December 2015]; Available from: http://www.isdscotland.org/Health-Topics/Finance/Scottish-National-Tariff/.

101. Healthcare Resource Groups. *Healthcare Resource Groups*. 2015 [cited 2015 November 2015]; Available from: http://www.hscic.gov.uk/hrg.

102. The Scottish Government. *Scottish Index of Multiple Deprivation*. 2015 [cited 2015 November 2015]; Available from: http://www.gov.scot/Topics/Statistics/SIMD.

103. The Scottish Government. *NHS Boards*. 2015 [cited 2015 January 2015]; Available from: http://www.gov.scot/Topics/Health/NHS-Workforce/NHS-Boards.

104. The Scottish Government. *Scottish Government Urban Rural Classification*. 2015 [cited 2015 January 2015]; Available from: http://www.gov.scot/Topics/Statistics/About/Methodology/UrbanRuralClassification.

105. Ahlbom, A. and L. Alfredsson, *Interaction: A word with two meanings creates confusion.* European Journal of Epidemiology, 2005. **20**(7): p. 563-564.

106. Leyland, A.H., et al., *Cause-specific inequalities in mortality in Scotland: two decades of change. A population-based study.* BMC Public Health, 2007. **7**: p. 172.

107. Zweifel, P., S. Felder, and M. Meiers, *Ageing of population and health care expenditure: a red herring?* Health Econ, 1999. **8**(6): p. 485-96.

108. Woodfield, R., I. Grant, and C.L. Sudlow, *Accuracy of Electronic Health Record Data for Identifying Stroke Cases in Large-Scale Epidemiological Studies: A Systematic Review from the UK Biobank Stroke Outcomes Group.* PLoS One, 2015. **10**(10): p. e0140533.

109. Camm, A.J., et al., *2012 focused update of the ESC Guidelines for the management of atrial fibrillation: an update of the 2010 ESC Guidelines for the management of atrial fibrillation. Developed with the special contribution of the European Heart Rhythm Association.* Eur Heart J, 2012. **33**(21): p. 2719-47.

110. Hahn, O.M. and R.L. Schilsky, *Randomized Controlled Trials and Comparative Effectiveness Research.* Journal of Clinical Oncology, 2012. **30**(34): p. 4194-4201.

111. Patel, M.R., et al., *Rivaroxaban versus Warfarin in Nonvalvular Atrial Fibrillation.* New England Journal of Medicine, 2011. **365**(10): p. 883-891.

112. Fernandez Y Garcia, E., et al., *Assessing Heterogeneity of Treatment Effects: Are Authors Misinterpreting Their Results?* Health services research, 2010. **45**(1): p. 283-301.

113. Kyriacou, D.N. and R.J. Lewis, *Confounding by Indication in Clinical ResearchConfounding by Indication in Clinical ResearchConfounding by Indication in Clinical Research.* JAMA, 2016. **316**(17): p. 1818-1819.

114. Faria, R., et al., *The use of observational data to inform estimates of treatment effectiveness in technology appraisal: methods for comparative individual patient data*. NICE DSU technical support document. 2015: NICE.

115. Greenland, S., *An introduction to instrumental variables for epidemiologists.* Int J Epidemiol, 2000. **29**(4): p. 722-9.

116. Craig, P., et al., *Natural Experiments: An Overview of Methods, Approaches, and Contributions to Public Health Intervention Research.* Annu Rev Public Health, 2017. **38**: p. 39-56.

117. Kreif, N., et al., *Regression-adjusted matching and double-robust methods for estimating average treatment effects in health economic evaluation.* Health Services and Outcomes Research Methodology, 2013. **13**(2): p. 174-202.

118. Rosenbaum, P.R. and D.B. Rubin, *The central role of the propensity score in observational studies for causal effects.* Biometrika, 1983. **70**(1): p. 41-55.

119. Normand, S.-L.T., et al., *Validating recommendations for coronary angiography following acute myocardial infarction in the elderly: A matched analysis using propensity scores.* Journal of Clinical Epidemiology, 2001. **54**(4): p. 387-398.

120. Elze, M.C., et al., *Comparison of Propensity Score Methods and Covariate Adjustment: Evaluation in 4 Cardiovascular Studies.* J Am Coll Cardiol, 2017. **69**(3): p. 345-357.

121. Kreif, N., et al., *Evaluating treatment effectiveness under model misspecification: A comparison of targeted maximum likelihood estimation with bias-corrected matching.* Stat Methods Med Res, 2014.

122. Austin, P.C., P. Grootendorst, and G.M. Anderson, *A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study.* Stat Med, 2007. **26**(4): p. 734-53.

123. Austin, P.C. and E.A. Stuart, *The performance of inverse probability of treatment weighting and full matching on the propensity score in the presence of model misspecification when estimating the effect of treatment on survival outcomes.* Statistical Methods in Medical Research, 2015.

124. Caliendo, M. and S. Kopeinig, *SOME PRACTICAL GUIDANCE FOR THE IMPLEMENTATION OF PROPENSITY SCORE MATCHING.* Journal of Economic Surveys, 2008. **22**(1): p. 31-72.

125. Rubin, D.B., *Estimating causal effects from large data sets using propensity scores.* Ann Intern Med, 1997. **127**(8 Pt 2): p. 757-63.

126. Rubin, D.B. and N. Thomas, *Combining Propensity Score Matching with Additional Adjustments for Prognostic Covariates.* Journal of the American Statistical Association, 2000. **95**(450): p. 573-585.

127. Austin, P.C., *Goodness-of-fit diagnostics for the propensity score model when estimating treatment effects using covariate adjustment with the propensity score.* Pharmacoepidemiol Drug Saf, 2008. **17**(12): p. 1202-17.

128. Cole, S.R. and M.A. Hernan, *Constructing inverse probability weights for marginal structural models.* Am J Epidemiol, 2008. **168**(6): p. 656-64.

129. Curtis, L.H., et al., *Using inverse probability-weighted estimators in comparative effectiveness analyses with observational databases.* Med Care, 2007. **45**(10 Supl 2): p. S103-7.

130. Glynn, R.J., S. Schneeweiss, and T. Stürmer, *Indications for Propensity Scores and Review of their Use in Pharmacoepidemiology.* Basic & Clinical Pharmacology & Toxicology, 2006. **98**(3): p. 253-259.

131. Austin, P.C., *Using the Standardized Difference to Compare the Prevalence of a Binary Variable Between Two Groups in Observational Research.* Communications in Statistics - Simulation and Computation, 2009. **38**(6): p. 1228-1234.

132. Sturmer, T., et al., *Propensity scores for confounder adjustment when assessing the effects of medical interventions using nonexperimental study designs.* J Intern Med, 2014. **275**(6): p. 570-80.

133. Latimer, N.R., et al., *Adjusting survival time estimates to account for treatment switching in randomized controlled trials--an economic evaluation context: methods, limitations, and recommendations.* Med Decis Making, 2014. **34**(3): p. 387-402.

134. Latimer, N.R., *Treatment switching in oncology trials and the acceptability of adjustment methods.* Expert Review of Pharmacoeconomics & Outcomes Research, 2015. **15**(4): p. 561-564.

135. Gerhard, T., et al., *Comparing marginal structural models to standard methods for estimating treatment effects of antihypertensive combination therapy.* BMC Med Res Methodol, 2012. **12**: p. 119.

136. Robins, J.M., M.A. Hernan, and B. Brumback, *Marginal structural models and causal inference in epidemiology.* Epidemiology, 2000. **11**(5): p. 550-60.

137. Hade, E.M. and B. Lu, *Bias associated with using the estimated propensity score as a regression covariate.* Statistics in medicine, 2014. **33**(1): p. 74-87.

138. Brookhart, M.A., J.A. Rassen, and S. Schneeweiss, *Instrumental variable methods in comparative safety and effectiveness research.* Pharmacoepidemiology and drug safety, 2010. **19**(6): p. 537-554.

139. Craig, P., et al., *Using natural experiments to evaluate population health interventions: new Medical Research Council guidance.* Journal of Epidemiology and Community Health, 2012.

140. Zhou, H., et al., *Difference-in-Differences Method in Comparative Effectiveness Research: Utility with Unbalanced Groups.* Applied Health Economics and Health Policy, 2016. **14**(4): p. 419-429.

141. Velentgas P, D.N., Nourjah P, et al., editors. , *Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide.* 2013.

142. Oldenburg, C.E., E. Moscoe, and T. Bärnighausen, *Regression Discontinuity for Causal Effect Estimation in Epidemiology.* Current epidemiology reports, 2016. **3**: p. 233-241.

143. Lip, G.Y., et al., *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): p. 263-72.

144. Poli, D., et al., *Comparison of HAS-BLED and HAS-BED Versus CHADS2 and CHA2DS2VASC Stroke and Bleeding Scores in Patients With Atrial Fibrillation.* The American Journal of Cardiology, 2017. **119**(7): p. 1012-1016.

145. Schneeweiss, S., et al., *Assessing the comparative effectiveness of newly marketed medications: methodological challenges and implications for drug development.* Clin Pharmacol Ther, 2011. **90**(6): p. 777-90.

146. Mueller, T., et al., *Comparative safety and effectiveness of direct oral anticoagulants in patients with atrial fibrillation in clinical practice in Scotland.* Br J Clin Pharmacol, 2019. **85**(2): p. 422-431.

147. Breslow, N.E., *Analysis of Survival Data under the Proportional Hazards Model.* International Statistical Review / Revue Internationale de Statistique, 1975. **43**(1): p. 45.

148. Hiller, L., A. Marshall, and J. Dunn, *Assessing violations of the proportional hazards assumption in Cox regression: does the chosen method matter?* Trials, 2015. **16**(Suppl 2): p. P134-P134.

149. Andersen, P.K. and R.D. Gill, *Cox's Regression Model for Counting Processes: A Large Sample Study.* The Annals of Statistics, 1982. **10**(4): p. 1100-1120.

150. Latimer, N.R., *NICE Decision Support Unit Technical Support Documents*, in *Survival Analysis For Economic Evaluations Alongside Clinical Trials - Extrapolation with Patient-Level Data*. 2013, National Institute for Health and Care Excellence (NICE)

151. Briggs, A., Claxton,K, Sculpher M., *Decision Modelling for Health Economic Evaluation*. 2006: Oxford University Press.

152. Allison, P., D., *Multiple Regression: A Primer*. 1999: Thousand Oaks, CA: Pine Forge Press.

153. Hosmer, D., W., Lemeshow, S., May, S., *Applied Survival Analysis.* 2nd ed. Wiley Series in Probability Statistics. 2008: John Wiley & Sons,Inc.

154. Shields, A.M. and G.Y. Lip, *Choosing the right drug to fit the patient when selecting oral anticoagulation for stroke prevention in atrial fibrillation.* J Intern Med, 2015. **278**(1): p. 1-18.

155. Diener, H.C., et al., *Choosing a particular oral anticoagulant and dose for stroke prevention in individual patients with non-valvular atrial fibrillation: part 1.* Eur Heart J, 2017. **38**(12): p. 852-859.

156. Drummond, M., Sculpher, M.J.,Torrance, G.W., O'Brien, B.J.,Stoddart, G.L., *Methods for the economic evaluation of health care programmes*. 3rd ed. 2005: Oxford University Press.

157. Larsen, T.B., et al., *Efficacy and safety of dabigatran etexilate and warfarin in "real-world" patients with atrial fibrillation: a prospective nationwide cohort study.* J Am Coll Cardiol, 2013. **61**(22): p. 2264-73.

158. Dolan, P., et al., *QALY maximisation and people's preferences: a methodological review of the literature.* Health Econ, 2005. **14**(2): p. 197-208.

159. Weinstein, M.C., *A QALY is a QALY is a QALY — Or is it?* Journal of Health Economics, 1988. **7**(3): p. 289-290.

160. Berg, J., et al., *Factors determining utility measured with the EQ-5D in patients with atrial fibrillation.* Qual Life Res, 2010. **19**(3): p. 381-90.

161. Whitehead, S.J. and S. Ali, *Health outcomes in economic evaluation: the QALY and utilities.* Br Med Bull, 2010. **96**: p. 5-21.

162. Devlin, N.J. and R. Brooks, *EQ-5D and the EuroQol Group: Past, Present and Future.* Applied health economics and health policy, 2017. **15**(2): p. 127-137.

163. EuroQol. *EQ-5D*. 2017 [cited 2017 May 2017]; Available from: https://euroqol.org/.

164. Robinson, A., et al., *How patients with atrial fibrillation value different health outcomes: a standard gamble study.* J Health Serv Res Policy, 2001. **6**(2): p. 92-8.

165. Johanesson, M., *The relationship between cost-effectiveness analysis and cost-benefit analysis*. 1995: Basil Blackwell, Oxford.

166. Brazier, J., Ratcliffe, J, Salomon, J, Tsuchiya, A, *Measuring and Valuing Benefits for Economic Evaluation.* 2007: Oxford University Press.

167. Caro, J.J., J. Moller, and D. Getsios, *Discrete event simulation: the preferred technique for health economic evaluations?* Value Health, 2010. **13**(8): p. 1056-60.

168. Briggs, A.H. and A.M. Gray, *Handling uncertainty when performing economic evaluation of healthcare interventions.* Health Technol Assess, 1999. **3**(2): p. 1-134.

169. Stevenson, M.D., J. Oakley, and J.B. Chilcott, *Gaussian process modeling in conjunction with individual patient simulation modeling: a case study describing the calculation of cost-effectiveness ratios for the treatment of established osteoporosis.* Med Decis Making, 2004. **24**(1): p. 89-100.

170. O'Hagan, A., M. Stevenson, and J. Madan, *Monte Carlo probabilistic sensitivity analysis for patient level simulation models: efficient estimation of mean and variance using ANOVA.* Health Econ, 2007. **16**(10): p. 1009-23.

171. Brown, J.B., et al., *The Mt. Hood challenge: cross-testing two diabetes simulation models.* Diabetes Res Clin Pract, 2000. **50 Suppl 3**: p. S57-64.

172. Karnon, J., *Alternative decision modelling techniques for the evaluation of health care technologies: Markov processes versus discrete event simulation.* Health Econ, 2003. **12**(10): p. 837-48.

173. Limone, B.L., et al., *Novel anticoagulants for stroke prevention in atrial fibrillation: a systematic review of cost-effectiveness models.* PLoS One, 2013. **8**(4): p. e62183.

174. Limone, B.L., et al., *Common flaws exist in published cost-effectiveness models of pharmacologic stroke prevention in atrial fibrillation.* Journal of Clinical Epidemiology, 2014. **67**(10): p. 1093-1102.

175. Sorensen, S.V., et al., *Cost-effectiveness of dabigatran etexilate for the prevention of stroke and systemic embolism in atrial fibrillation: a Canadian payer perspective.* Thromb Haemost, 2011. **105**(5): p. 908-19.

176. Gage, B.F., et al., *Cost-effectiveness of warfarin and aspirin for prophylaxis of stroke in patients with nonvalvular atrial fibrillation.* Jama, 1995. **274**(23): p. 1839-45.

177. Husereau, D., et al., *Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement.* BMJ : British Medical Journal, 2013. **346**: p. f1049.

178. Dorian, P., et al., *Cost-effectiveness of apixaban vs. current standard of care for stroke prevention in patients with atrial fibrillation.* European Heart Journal, 2014. **35**(28): p. 1897-1906.

179. Lip, G.Y.H., et al., *Cost-effectiveness of apixaban versus other new oral anticoagulants for stroke prevention in atrial fibrillation.* Clinical Therapeutics, 2014. **36**(2): p. 192-210.e20.

180. Sterne JAC, B.P., Bryden PA, et al., *Oral anticoagulants for primary prevention, treatment and secondary prevention of venous thromboembolic disease, and for prevention of stroke in atrial fibrillation: systematic review, network meta-analysis and cost-effectiveness analysis.* . NIHR Journals Library, 2017((Health Technology Assessment, No. 21.9.)).

181. Verhoef, T.I., et al., *Cost Effectiveness of New Oral Anticoagulants for Stroke Prevention in Patients with Atrial Fibrillation in Two Different European Healthcare Settings.* American Journal of Cardiovascular Drugs, 2014. **14**(6): p. 451-462.

182. Edwards, S. *Apixaban for the prevention of stroke and systemic embolism in people with non-valvular atrial fibrillation STA REPORT*. 2012 [cited 2016 December 2016]; Available from: https://www.nice.org.uk/guidance/ta275/documents/stroke-and-systemic-embolism-prevention-nonvalvular-atrial-fibrillation-apixaban-evidence-review-group-report2.

183. National Institute for Health and Care Excellence. *Guide to the methods of technology appraisal 2013*. 2013 [cited 2017 May 2017]; Available from: https://www.nice.org.uk/process/pmg9/chapter/the-reference-case.

184. National Institute for Health and Care Excellence. *Guide to the processes of technology appraisal 2014*. 2013 [cited 2015 November 2015]; Available from: https://www.nice.org.uk/process/pmg19/resources/guide-to-the-processes-of-technology-appraisal-pdf-72286663351237.

185. Pink, J., et al., *Dabigatran etexilate versus warfarin in management of non-valvular atrial fibrillation in UK context: quantitative benefit-harm and economic analyses.* BMJ, 2011. **343**.

186. Friberg, L., M. Rosenqvist, and G.Y. Lip, *Evaluation of risk stratification schemes for ischaemic stroke and bleeding in 182 678 patients with atrial fibrillation: the Swedish Atrial Fibrillation cohort study.* Eur Heart J, 2012. **33**(12): p. 1500-10.

187. Andersen, K.K. and T.S. Olsen, *Reduced poststroke mortality in patients with stroke and atrial fibrillation treated with anticoagulants: results from a Danish quality-control registry of 22,179 patients with ischemic stroke.* Stroke, 2007. **38**(2): p. 259-63.

188. The UK Government. *NHS reference costs: financial year 2011 to 2012*. 2011 [cited 2018 April 2018]; Available from: https://www.gov.uk/government/publications/nhs-reference-costs-financial-year-2011-to-2012.

189. Luengo-Fernandez, R., et al., *Population-based study of acute- and long-term care costs after stroke in patients with AF.* Int J Stroke, 2013. **8**(5): p. 308-14.

190. Office for National Statistics (ONS). *Consumer price inflation time series*. 2017 [cited 2017 May 2017]; Available from: https://www.ons.gov.uk/economy/inflationandpriceindices/timeseries/dkc3/mm23.

191. National Institute for Health and Care Excellence. *BNF Apixaban* 2018 [cited 2018 May 2018]; Available from: https://bnf.nice.org.uk/medicinal-forms/apixaban.html.

192. National Institute for Health and Care Excellence. *BNF Dabigatran Etexilate*. 2018 [cited 2018 May 2018]; Available from: https://bnf.nice.org.uk/medicinal-forms/dabigatran-etexilate.html.

193. National Institute for Health and Care Excellence. *BNF Rivaroxaban*. 2018 [cited 2018 May 2018]; Available from: https://bnf.nice.org.uk/medicinal-forms/rivaroxaban.html.

194. National Institute for Health and Care Excellence. *Costing Report: Atrial Fibrillation*. 2006 [cited 2018 May 2018]; Available from: https://www.nice.org.uk/guidance/cg36/documents/atrial-fibrillation-nice-guideline-second-consultation2.

195. Haacke, C., et al., *Long-term outcome after stroke: evaluating health-related quality of life using utility measurements.* Stroke, 2006. **37**(1): p. 193-8.

196. Lacey, E.A. and S.J. Walters, *Continuing inequality: gender and social class influences on self perceived health after a heart attack.* Journal of Epidemiology and Community Health, 2003. **57**(8): p. 622-627.

197. Lenert, L.A. and R.M. Soetikno, *Automated computer interviews to elicit utilities: potential applications in the treatment of deep venous thrombosis.* J Am Med Inform Assoc, 1997. **4**(1): p. 49-56.

198. Claxton, K., et al., *Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold.* Health Technol Assess, 2015. **19**(14): p. 1-503, v-vi.

199. Kreif, N., et al., *Methods for estimating subgroup effects in cost-effectiveness analyses that use observational data.* Med Decis Making, 2012. **32**(6): p. 750-63.

200. Byford, S. and J. Raftery, *Perspectives in economic evaluation.* BMJ, 1998. **316**(7143): p. 1529-1530.

201. Eikelboom, J.W., et al., *Rivaroxaban with or without Aspirin in Stable Cardiovascular Disease.* New England Journal of Medicine, 2017. **377**(14): p. 1319-1330.

202. Scottish Medicines Consortium. *Scottish Medicines Consortium. rivaroxaban (Xarelto) for the prevention of atherothrombotic events*. 2012 [cited 2019 July 2019]; Available from: https://www.scottishmedicines.org.uk/medicines-advice/rivaroxaban-xarelto-fullsubmission-smc2128/.

203. Department of Health & Social Care. *The 2019 Voluntary Scheme for Branded Medicines Pricing and Access - Chapters and Glossary*. 2018 [cited 2019 January 2019]; Available from:https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/761834/voluntary-scheme-for-branded-medicines-pricing-and-access-chapters-and-glossary.pdf.

204. T P, A., et al., *Pharmacogenomics: the right drug to the right person.* Journal of clinical medicine research, 2009. **1**(4): p. 191-194.

205. Tseng, A.S., et al., *Clinical Review of the Pharmacogenomics of Direct Oral Anticoagulants.* Cardiovasc Drugs Ther, 2018. **32**(1): p. 121-126.