Comparison of oral anticoagulants for stroke prevention in atrial fibrillation using the UK Clinical Practice Research Datalink Aurum: A reference trial (ARISTOTLE) emulation study

Emma Maud Powell1*, Usha Gungabissoon2, John Tazare3, Liam Smeeth1, Paris J Baptiste4, Turki M Bin Hammad1,5, Angel YS Wong1, Ian J Douglas1, Kevin Wing1

1 Department of Non-communicable Disease Epidemiology, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom
2 Epidemiology, GSK, London, United Kingdom
3 Department of Medical Statistics, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom
4 Clinical Effectiveness Group, Centre for Primary Care, Wolfson Institute of Population Health, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom
5 Methodology and biostatistics team, Department of Efficacy and Safety, Drug sector, Saudi Food and Drug Authority, Riyadh, Saudi Arabia

The views expressed in this paper are those of the author and not necessarily reflect those of the SFDA or its stakeholders. Guaranteeing the accuracy and the validity of the data is a sole responsibility of the research team.

* Corresponding author

E-mail maud.teoh@lshtm.ac.uk (EMP)

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
ABSTRACT

Background

Stroke prevention treatment guidance for patients with atrial fibrillation (AF) uses evidence generated from randomised controlled trials (RCTs). However, applicability to patient groups excluded from trials remains unknown. Real-world patient data provides an opportunity to evaluate outcomes in a trial analogous population of direct oral anticoagulants (DOACs) users and in patients otherwise excluded from RCTs, however there remains uncertainty on the validity of the methods and suitability of the data.

Successful reference trial emulation can support the generation of evidence around treatment effects in groups excluded or underrepresented in the original trials.

Methods and findings

Patients with AF in a UK primary care database Clinical Practice Research Datalink (CPRD Aurum) prescribed apixaban or warfarin from 1 Jan 2013 to 31 Jul 2019 were selected. ARISTOTLE eligibility criteria were applied to this population and matched to the RCT apixaban arm on baseline characteristics creating a trial-analagous apixaban cohort; this was propensity-score matched to warfarin users in the CPRD Aurum. ARISTOTLE outcomes were assessed using Cox proportional hazards regression stratified by prior warfarin exposure status during 2.5 years of patient follow-up and results benchmarked against the trial results before treatment effectiveness was further evaluated based on (warfarin) time in therapeutic range (TTR).

The analysis sample comprised 8734 apixaban users and propensity-score matched 8734 warfarin users in CPRD. Results [Hazard Ratio (95% Confidence Interval)] confirmed apixaban non-inferiority
for stroke or systemic embolism (SE) [CPRD 0.98 (0.82,1.19) vs trial 0.79 (0.66,0.95)] and death from any cause [CPRD 1.03 (0.93,1.14) vs trial 0.89 (0.80,0.998)] but did not indicate apixaban superiority.

Absolute event rates for Stroke/SE were similar for apixaban in CPRD Aurum and ARISTOTLE (1.27%/year) whereas a lower event rate was observed for warfarin (CPRD Aurum 1.29%/year, ARISTOTLE 1.60%/year)

Analysis by TTR suggested non-inferiority of apixaban in those with TTR < 0.75 [Stroke/SE 0.94 (0.75,1.19), all-cause death 0.99 (0.87,1.12)]. However, apixaban was associated with increased hazards compared with well-controlled warfarin treatment (TTR ≥ 0.75) [Stroke/SE 1.49 (1.13,1.97), all-cause death 1.75 (1.49,2.06)]. The main limitation of the study’s methodology are the risk of residual confounding, channelling bias and attrition bias in the warfarin arm.

Conclusions

Analysis of non-interventional data generated results demonstrating non-inferiority of apixaban vs warfarin consistent with the pre-specified benchmarking criteria. Unlike in ARISTOTLE superiority of apixaban vs warfarin was not seen which may be linked to the lower proportion of Asian patients and higher proportion of patients with well-controlled warfarin compared to ARISTOTLE. The methodological template developed can be used to investigate treatment effects of oral anticoagulants in patient groups excluded from or under-represented in trials and also provides a framework which can be adapted to investigate treatment effects for other conditions.
Why Was This Study Done?

- Stroke prevention treatment guidelines for patients with atrial fibrillation (AF) are based on results from randomised controlled trials (RCTs), we do not know if these results are relevant to patients that would not have been eligible to be included in the RCTs.

- This study used routinely collected health data from the UK to emulate an RCT that compared apixaban to warfarin, ARISTOTLE, and also looked at whether the benefit of apixaban compared with warfarin was impacted by the quality of warfarin therapy (measured by time in therapeutic range, TTR).

- Emulating an RCT for stroke prevention in patients with AF should help to understand how transferable RCT results are to ‘real-world’ practices and whether this methodological approach can help to improve treatment options and outcomes for patient groups currently underrepresented in clinical trials.

What Did the Researchers Do and Find?

- The researchers looked at patients with AF in a UK primary care data prescribed apixaban or warfarin and applied a “reference trial emulation” approach, in which the ARISTOTLE trial eligibility, selection and analysis approaches were applied to UK primary care data and results benchmarked against those of ARISTOTLE.

- Patients prescribed apixaban had similar rates of outcomes to those prescribed warfarin in our cohort and our results were successfully benchmarked against ARISTOTLE. Unlike ARISTOTLE we did not see superiority of apixaban vs warfarin [Hazard ratio (95% confidence interval)] for time to stroke or systemic embolism: 0.98 (0.82,1.19) in our cohort vs 0.79 (0.66,0.95) in ARISTOTLE.

- We also found the benefit of apixaban vs warfarin differed depending on the quality of warfarin therapy with apixaban superior to poorly controlled warfarin therapy (TTR < 0.75).
[Stroke/SE 0.94 (0.75,1.19), Death 0.99 (0.87,1.12)] whereas apixaban was not superior to well-controlled warfarin (TTR ≥ 0.75) [Stroke/SE 1.49 (1.13,1.97), Death 1.75 (1.49,2.06)].

What Do These Findings Mean?

- Our results support the NICE guidelines on selecting treatment for stroke prevention in patients with AF and also provide reassurance on continuing warfarin in patients with high TTR.

- We can use UK primary health care data to emulate a reference trial of treatments for the prevention of stroke in AF.

- We can use the data and methods to look at how well treatments work in patients that would not have been included in RCTs such as those with multimorbidity or patient groups under-represented in RCTs such as ethnic minority groups and older patients.

- Study limitations include the possibility of residual confounding, a risk patients doing well on warfarin were over-represented in our cohort, and a lower proportion of Asian participants in our cohort compared with ARISTOTLE.

Introduction

Atrial fibrillation (AF) is a common type of cardiac arrhythmia with an estimated prevalence of 3.3% in UK adults aged ≥ 35 years [1]. AF is a risk factor for stroke; patients with AF have a five-fold increased risk of stroke compared with people without AF [2] and around a quarter of all strokes are attributed to this arrhythmia [3]. In addition, increased levels of mortality, morbidity and disability with longer hospital stays are observed in stroke patients with AF compared with stroke patients without AF [4, 5].

Pharmacological therapy recommended to reduce the risk of stroke in AF includes the use of oral anticoagulants (OACs). The introduction of direct oral anticoagulants (DOACs) for AF since 2012 in the UK provided a choice of treatment alongside the older OAC class of vitamin K antagonists (VKA),
such as warfarin which has been available for over 60 years. The VKA OACs require regular
monitoring of international normalised ratio (INR) to keep patients in the optimal therapeutic range
(typically 2.0 to 3.0) in which risk of both ischemic and bleeding events are minimised [6]. A patient
may require dose adjustments to stay within their INR target range. A key measure of quality of
warfarin treatment is therefore the time in therapeutic range (TTR) which estimates the proportion
of time a patient has spent with INR within optimal range. A TTR of 0.75 or greater is often
considered as indicating optimal INR control and suggests a patient is spending a high proportion of
their time in their INR target range.

ARISTOTLE was a pivotal RCT of the DOAC apixaban designed to demonstrate non-inferiority
compared with warfarin in the prevention of stroke or systemic embolism (SE) in patients with AF.
The results demonstrated superiority of apixaban over warfarin for both prevention of stroke/SE and
safety (major bleeding) [7]. Results in the EU patient subset from the trial suggested the observed
superiority of apixaban might be dependent on how well warfarin therapy was managed in the
comparator group [8], an analysis that has not yet been performed outside of trial settings. In the
NICE review of ARISTOTLE, several professional groups noted the TTR of warfarin users in ARISTOTLE
may be lower than what is typical in UK clinical practice [9].

Treatment guidelines for DOACs are based on evidence from randomised controlled trials (RCTs),
however, it is unclear whether these results extend to patient groups typically excluded from trials
such as those with increased bleeding risk or severe comorbidities. Whilst there have been a number
of previous studies of DOAC effectiveness using non-interventional data, there remains uncertainty
on whether the data sources and methods used have fully accounted for the lack of treatment
randomisation and issues such as selection bias and confounding. Comparing results from real-world
studies with RCT results is challenging due to differences in patient populations, treatment
adherence, and study design. However, reference trial emulation involves use of an existing named
RCT to (1) inform observational study design and (2) benchmark results against, providing
confidence in validity of the selected observational methods and data. [10-13]. The non-interventional analysis methods can then be applied, under a set of assumptions, to reliably estimate effects in groups of patients with AF who would have been excluded from (or underrepresented in) the reference trial [14] such as patients aged > 80 that were under-represented in ARISTOTLE compared with patients with AF in UK clinical practice and patients with increased bleeding risk that were excluded by the trial eligibility criteria.

There is increasing interest in trial emulation using observation data, and in the application of recent developments in pharmacoepidemiology methods involving the inclusion of prevalent users. This study used a framework which involved coarsened exact matching to select patients matching the trial population on aggregate, and sampling of prevalent users in a way that avoids selection bias and emulates the process of screening into an RCT, to construct a cohort of patients similar to the target trial population which included both new and prevalent users. This methodological approach could be adapted to a variety of treatments and different therapeutic areas.

This study sought to (1) create an ARISTOTLE-analogous cohort using routinely collected primary and secondary care data in the UK (2) benchmark results obtained in the ARISTOTLE-analogous cohort with ARISTOTLE results and (3) explore whether apixaban treatment-effects in clinical practice are influenced by how well warfarin therapy is controlled.

Materials and methods

This study is reported as per the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guideline (S1 Checklist).

Study design

A propensity score matched cohort study with emulation of a reference trial (ARISTOTLE).

Setting/data sources
This study used non-interventional data from UK Clinical Practice Research Datalink (CPRD) Aurum, a database containing anonymised data from 738 primary care practices across England (approximately 13% of the population of England with 19 million patient records and 7 million active as of September 2018 [15]. CPRD Aurum contains information on clinical diagnoses, prescribing, referrals, tests and demographic/lifestyle factors and is representative of the population of England in geographical spread, social deprivation, age and sex [15]. This study also used 2 additional data sources linked to CPRD Aurum: Hospital Episodes Statistics (HES) data, which contains data on patients admitted to NHS hospitals including diagnoses, admission and discharge, and Office of National Statistics (ONS) mortality data.

The reference trial (ARISTOTLE)

ARISTOTLE was a randomised, double-blind trial completed in 2011, comparing apixaban with warfarin in the prevention of stroke and SE. The trial included 18201 patients with AF and at least one additional risk factor for stroke. The trial was designed to test for non-inferiority of apixaban compared with warfarin (non-inferiority margin of 1.38 for the upper limit of the 95% CI of the hazard ratio for the primary outcome), and showed apixaban superiority for (1) the primary outcome of stroke or SE (HR 0.79; 95% CI 0.66, 0.95), (2) the safety endpoint of major bleeding (HR 0.69; 95% CI 0.60, 0.80), and (3) death from any cause (HR 0.89; 95% CI 0.80, 0.99). The ARISTOTLE findings led to the National Institute for Health and Care Excellence (NICE) guidelines on stroke prophylaxis in patients with AF recommending apixaban as a treatment.

ARISTOTLE eligibility criteria and summary baseline patient characteristics were used to select a cohort of patients from CPRD Aurum analogous to the ARISTOTLE participants.
The use of CPRD and ARISTOTLE are described in a previous publication [14] and use of CPRD for this project was approved by the MHRA Independent Scientific Advisory Committee [ISAC protocol in S2]. All data used in this study were anonymised.

Diagnostic and therapeutic codelists

All diagnostic and therapeutic codelist files used are available at https://datacompass.lshtm.ac.uk/id/eprint/3590/.

Patient Selection

Step 1: application of trial eligibility criteria to patients in CPRD

We first selected HES-linked patients registered in CPRD Aurum between January 1, 2013 and July 31, 2019, who had at least 6 months between registration and the index date. ARISTOTLE recruited both new (warfarin-naïve) and prevalent (warfarin-experienced) users of warfarin with randomisation stratified on prior warfarin (or other VKA) exposure status (warfarin naïve or experienced). To be classified as warfarin-naïve patients were required to have no evidence of exposure to warfarin or other VKA in the 5 years prior to the index date. To enable selection of a similar cohort of patients in CPRD Aurum (including both new and prevalent users of warfarin), the following process was used in determining index date:

- apixaban users
  index date = first prescription of apixaban in the study period
  apixaban user classified as warfarin-naïve or warfarin-experienced at this date;

- warfarin users
  for new users of warfarin: index date = first prescription of warfarin in the study period;
  for prevalent users of warfarin: a pool of potential index dates was selected containing all
prescription dates in the study period, with index date selected at the later treatment-history sampling stage (see step 3).

ARISTOTLE eligibility criteria (supplementary table A2) [7] were applied giving a trial-eligible cohort for apixaban users, a trial-eligible cohort of new users of warfarin, and a pool of potential index dates (with all potential index dates kept in regardless of ARISTOTLE eligibility at this stage) for warfarin continuers (prevalent warfarin users).

**Step 2: selection of apixaban trial-analogous patients in CPRD**

We selected a subset of the CPRD Aurum trial-eligible apixaban patients that better matched the ARISTOTLE apixaban participants based on aggregate summaries for the following key ARISTOTLE baseline characteristics:

- Age
- Sex
- Congestive heart failure or left ventricular systolic dysfunction
- Hypertension requiring treatment
- Diabetes mellitus
- Prior stroke/transient ischaemic attack (TIA)/systemic embolism (SE)
- Level of renal impairment
- Prior VKA/warfarin exposure

To characterise the baseline patient characteristics of ARISTOTLE, we used the key publication of the trial results [7], discussion of trial results by regulatory bodies [8, 9, 16] and publications on the trial presenting cross-tabulations on key characteristics [17,18].

An ARISTOTLE-analogous cohort of CPRD Aurum apixaban patients was then selected using a modified form of coarsened exact matching [19] (see Appendix for details).
Step 3: matching of apixaban trial-analogous patients to warfarin trial-eligible patients in CPRD

To emulate ARISTOTLE which stratified randomisation on prior VKA exposure status, patients in the CPRD cohort were matched separately within the VKA-naïve and VKA-experienced strata. A 3-step procedure, based on methods proposed by Suissa et al [20] and Webster Clark et al [21], was used to select and match patients in the VKA-experienced strata whilst avoiding selection bias; this procedure is summarised in Figure 1 and described in in S3 Appendix.

The trial-analogous CPRD Aurum apixaban patients were matched to warfarin CPRD Aurum patients using greedy nearest-neighbour matching on the logit of the propensity score (PS); a caliper of 0.2 times the standard deviation of the logit of the propensity score was used for matching as recommended by Austin [22].

The covariates included in the propensity score models are detailed Table 1.

Table 1: Covariates Included in the Propensity Score Models

<table>
<thead>
<tr>
<th>Category</th>
<th>Variable List</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographics</td>
<td>age, sex, ethnicity</td>
</tr>
<tr>
<td>CHADS$_2$ stroke risk factors</td>
<td>congestive heart failure or left ventricular systolic dysfunction, hypertension requiring treatment, diabetes mellitus, prior stroke/TIA/systemic embolism</td>
</tr>
<tr>
<td>Vascular stroke risk factors</td>
<td>prior myocardial infarction, peripheral artery disease, aortic plaque, history of pulmonary embolism or deep vein thrombosis</td>
</tr>
<tr>
<td>Other risk factors</td>
<td>body mass index, systolic blood pressure, history of bleeding, smoking status, alcohol consumption, socioeconomic status (imd2105_5), ethnicity</td>
</tr>
<tr>
<td>Concomitant medications</td>
<td>aspirin, clopidogrel, NSAIDs, antacids, statins, angiotensin-converting enzyme inhibitors (ACEIs) or angiotensin receptor blockers (ARBs), beta blockers, calcium channel blockers, statins, amiodarone, digoxin, proton pump inhibitors, H2 receptor antagonist</td>
</tr>
<tr>
<td>Comorbidities</td>
<td>renal function, history of fall, Charlson comorbidity components [COPD, connective tissue disease, peptic ulcer disease, liver disease, hemiplegia, cancer, haematological cancer], healthcare utilization [number of GP consults in the prior year, number of hospitalizations in the prior year]</td>
</tr>
</tbody>
</table>
The model resulting in the most balanced cohort was chosen with balance assessed by looking at standardised differences across all variables after matching using a target threshold of 0.05 for the maximum difference allowed for any individual variable. Balance of covariates considered to be most important in predicting outcome were prioritised namely age, sex, and stroke risk factors.

<table>
<thead>
<tr>
<th>AF factors</th>
<th>time since AF diagnosis, history of valvular disease, history of valvular surgery</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthcare utilisation</td>
<td>number of GP consults in the prior year, number of hospitalizations in the prior year</td>
</tr>
</tbody>
</table>

AF=atrial fibrillation; COPD=chronic obstructive pulmonary disease; GP=general practitioner; NSAIDs=non-steroidal anti-inflammatory drugs; TIA=transient ischemic attack;
Figure 1: Matching of apixaban trial-analogous patients to warfarin trial-eligible patients

* This method has been found in a simulation study (Webster-Clark et al [21]) to give unbiased results

CPRD=Clinical Practice Research Datalink; RCT=randomised controlled trial;
VKA=vitamin K antagonist.
Exposures and outcomes

Exposures

Exposure to apixaban (5mg/2.5mg) or warfarin was determined using CPRD prescribing records with no restrictions on the dose prescribed.

Outcomes

The primary effectiveness outcome was the composite of stroke (ischemic or haemorrhagic) or systemic embolism (SE); individual components of this outcome (stroke, ischemic or uncertain type of stroke, haemorrhagic stroke, SE) and death from any cause were the key secondary effectiveness outcomes. Secondary effectiveness outcomes included myocardial infarction (MI), pulmonary embolism or DVT, and composite endpoints of effectiveness outcomes. The primary safety outcome was major bleeding (including by location – intracranial, gastrointestinal, or other location such as urinary or gynaecological). All outcomes involved hospitalisation or death and were ascertained using HES and ONS data. The ICD-10 codes used in ascertaining stroke occurrence have been recommended as having high positive predictive value [23].

Statistical analysis

Methods of Analysis

A prospective protocol was published prior to the analysis detailing the planned analyses [14, also in Appendix].

Changes from the planned protocol are described in Table 2.

Table 2: Changes from Planned Analyses

<table>
<thead>
<tr>
<th>Original Planned Analysis</th>
<th>Updated Analysis</th>
<th>Reason for change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients to be selected from both CPRD GOLD and CPRD Aurum.</td>
<td>Only CPRD Aurum used.</td>
<td>There was a much larger sample size available in CPRD Aurum meaning combining of the 2 data sources was not required.</td>
</tr>
<tr>
<td>Censoring scheme to censor at 5 years after index date.</td>
<td>Censoring scheme censored at 2.5 years after index date.</td>
<td>The ARISTOTLE trial had median duration of follow-up of 1.8 years (IQR</td>
</tr>
</tbody>
</table>
1.4, 2.3) therefore a 2.5 year cut-off gives a more similar duration of follow-up than 5 years.

### Adherence of apixaban users to be measured by proportion of days covered by prescriptions.
- Treatment persistence measured instead (proportion of patients still on index treatment at date of censoring).
- Repeat prescriptions are often issued automatically meaning comparing number of days covered by prescribed pills to the number of days in the treatment period did not provide useful insight on adherence.

### Supplementary analysis in patients deemed adherent (PDC ≥ 80%, ARISTOTLE compliance limit).
- Analysis by TTR only.
- Unable to ascertain useful measure of adherence in the apixaban users.

### Non-inferiority will be concluded when the upper limit of the 95% CI for the hazard ratio must be less than 1.52 (upper limit in the EU subgroup of ARISTOTLE).
- Non-inferiority will be concluded when the upper limit of the 95% CI for the hazard ratio is less than 1.38 (same non-inferiority margin of ARISTOTLE).
- The non-inferiority margin used in ARISTOTLE was the one agreed by regulators to represent the maximum acceptable clinical difference. By applying the same margin, we ensure that the conclusion is based on more rigorous criteria.

### Aim to include prior INR control in propensity model for vitamin K antagonist-experienced patients.
- Primary analysis does not include prior INR control. Post hoc sensitivity analysis performed including prior INR control in the propensity score model.
- High rate of missing data for prior INR control made it not advisable to include this variable in the propensity score model for the main analysis. Other variables predictive of poor INR control such as age are already included. Post hoc sensitivity analysis including INR control in the propensity score model performed to assess the potential impact of not including this variable following question in peer review on the omission of this variable.

### Post hoc analysis assessing apixaban dose-adjustment in CPRD Aurum
- Suggested by peer review to provide evidence on the quality of dose adjustment in CPRD Aurum and how this may impact the results in the trial-analogous cohort.

---

All time-to-event endpoints were analysed using a Cox proportional hazards model, stratified by prior VKA status (experienced, naive). The effectiveness outcomes were analysed using the intention-to-treat principle and major bleeding was analysed using an on-treatment censoring scheme. Patients were censored at 2.5 years after index date reflecting typical maximum duration of follow-up in ARISTOTLE. Cluster-robust standard errors were used with pair membership as the clustering variable [24,25]. The proportional hazards assumption was assessed by looking at the log-
log of the Kaplan-Meier survival curves and inspection of scaled Schoenfeld residuals plotted against time. Analyses were performed using SAS version 9.4 and R version 4.2.1.

**Supplementary analyses**

A protocol planned analysis in the subset of patients deemed adherent (with adherence measured by TTR in the warfarin users and by proportion of days covered by prescriptions in the apixaban users) was planned to assess the impact of adherence on outcomes. The planned analysis was not possible due to the apixaban prescription data not providing a useful measure of adherence. An analysis by INR TTR was performed instead to assess the impact of warfarin control on results with all outcomes analysed by TTR (TTR < 0.75 and TTR ≥ 0.75). In order to perform the TTR analysis whilst maintaining balance in the baseline covariates, inverse probability treatment weighting (IPTW) was used to rebalance the baseline characteristics, applying stabilised weights to the ARISTOTLE-analogous apixaban users. A similar approach to the main analysis was used with propensity score models constructed separately for the new users and warfarin-experienced users.

An additional post hoc analysis was performed looking at the proportion of apixaban patients prescribed reduced-dose apixaban along with a comparison of the patients meeting the criteria for dose-reduction against the dose actually prescribed. In this analysis apixaban dose in the ARISTOTLE-analogous CPRD cohort was assessed and compared against the ARISTOTLE protocol-specified criteria and NICE criteria for reduced apixaban dose. ARISTOTLE specified that participants meeting any 2 of the following criteria assessed at the time of randomisation should have their apixaban dose reduced to 2.5mg BID: age ≥ 80 years, body weight ≤ 60 kg, or serum creatinine ≥ 1.5 mg/dL. These criteria are equivalent to the NICE guidelines for dose reduction with NICE having an additional criterion indicating reduced dose in those with creatinine clearance 15–29 mL/minute.

In addition, to assess the impact of the quality of dose-adjustment in the CPRD cohort on the observed effectiveness of apixaban relative to warfarin, a supplementary post hoc analysis was
performed looking at the results in the subset of apixaban patients prescribed the correct dose compared with IPTW re-balanced warfarin comparators.

**Sensitivity analyses**

Primary and secondary effectiveness outcomes were also analysed using the on-treatment censoring scheme to investigate whether treatment discontinuation compromises confidence in the effectiveness analyses.

Treatment persistence was defined by looking at longitudinal prescription data for OACs; OAC treatment windows were derived in which gaps >= 6 months between prescription dates were considered as distinct treatment windows. The end of each OAC treatment window was derived as the date of the last prescription of index OAC + the number of days supply given in the last prescription + a grace period of 30 days. In cases of overlapping OAC treatment windows the date of the first prescription of the subsequent OAC treatment window was used to define the end of the prior OAC window. A prescription for a different OAC from the index OAC treatment was considered as a treatment switch. An ending of index OAC treatment with no subsequent prescription for any other OAC recorded was considered as treatment stop. Gaps of >= 6 months with no subsequent OAC prescriptions recorded were categorised as having stopped OAC treatment.

The set of patients who switched or discontinued treatment during follow-up were examined to ascertain whether selection bias due to attrition may have affected the on-treatment analyses (Table A9 in Appendix).

Apixaban was first launched for AF in the UK in January 2013, with relatively few patients receiving a prescription in the first year it was available; we therefore performed a sensitivity analysis with the start of study period shifted forwards a year to investigate the impact of inclusion of early adopters who may differ from later adopters of a new drug.
Confounding and bias

In the study period apixaban was a newly available treatment leading to the possibility of channelling bias [26]. By applying trial eligibility criteria to both treatment cohorts and matching using baseline covariates we aimed to minimise channelling bias. To handle confounding, treatment arms were matched using PSM [27].

Benchmarking results against ARISTOTLE

The study hypothesis was that results in the CPRD ARISTOTLE-analogous cohort would be comparable to the ARISTOTLE results, as defined by the pre-specified benchmarking criteria. A slightly weaker benefit of apixaban vs warfarin was expected based on the weaker benefit seen in the EU subgroup of ARISTOTLE and an expectation that the quality of warfarin control in UK patients may be higher than that observed in ARISTOTLE.

The benchmarking criteria for considering the results in the trial-analogous CPRD cohort to be comparable with ARISTOTLE were pre-specified and published previously [14]:

1. The effect size must be clinically comparable with the ARISTOTLE findings; the HR for time to stroke/SE with the HR must be between 0.69 and 0.99. This range is not symmetrical around the ARISTOTLE estimate of 0.79 as it is anticipated the treatment effect in routine clinical care may be weaker than that seen in the optimised setting of a clinical trial.

2. The upper limit of the 95% CI for the HR for time to stroke/SE must be less than 1.38 (non-inferiority margin used in ARISTOTLE, updated since protocol – see Table 2).

The benchmarking step applied only to the primary effectiveness outcome in the trial-analogous CPRD cohort; results in other groups such as patients underrepresented or excluded from the trial would not necessarily be expected to remain consistent to the RCT results given the relative risks may differ in these groups. Comparability of other outcomes was to be assessed descriptively with no formal criteria or hypothesis testing used.
Patients with missing systolic blood pressure (0.1%), body mass index (3.3%), smoking status (0.1%),
or socioeconomic status (0.1%) were excluded from the trial-eligible cohort as the proportion of
patients with these missing was low. Patients with missing renal function (1.3%), ethnicity (0.4%), or
alcohol use (5.6%) were kept in the cohort through a missing indicator approach; this approach is
valid under the assumption that these variables act as confounders and influence clinician
prescribing decisions only when observed [28]. A total of 1176 (13.3%) warfarin users in the CPRD
cohort did not have INR measurements in the data during their treatment period and were not
included in the analysis by TTR.

**Ethics**

Scientific approval was provided by the London School of Hygiene and Tropical Medicine research
ethics committee (ref 17682) and the independent scientific advisory committee of the Medicines
and Healthcare Products Regulatory Agency (protocol no. 19_066R). CPRD data are already
approved via a national research ethics committee for purely non-interventional research of this
type. CPRD data are analysed anonymously therefore individual patient consent is not sought by
contributing medical practices when data is shared with CPRD; however, patients are able to opt out
of their patient information being shared for research.

**Results**

**Participants**

Between January 1, 2013 and July 31, 2019 there were 86,888 people with AF prescribed apixaban
and 159,632 prescribed warfarin in HES-linked CPRD Aurum practices (Figure 2). Application of
minimum registration period and ARISTOTLE inclusion criteria reduced this to 67,539 apixaban and
139,527 warfarin patients. After applying ARISTOTLE exclusion criteria there were 41,487 apixaban
and 101,159 warfarin patients.
Selecting apixaban patients to match ARISTOTLE on key baseline characteristics yielded 9,120 apixaban patients (3,912 new users and 5,208 prevalent users) available for propensity score matching to 101,159 warfarin patients. For 274 apixaban patients no match could be found giving a propensity score matched cohort of 8846 apixaban and 8846 warfarin patients.

Figure 2: Selection of ARISTOTLE-analogous CPRD Aurum Cohort

Flow of number of individuals included in the analysis. AF = atrial fibrillation; ALT = alanine transaminase; AST = aspartate transaminase; BMI = body mass index; BP = blood pressure; CPRD = Clinical Practice Research Datalink; HES: Hospital Episodes Statistics; Rx = Prescription; SES = socioeconomic status; ULN = upper limit of normal; VKA = vitamin K antagonist.

a Severe comorbid condition with life expectancy <1 year or reasons making participation impractical; b ALT or AST > 2X ULN or Total Bilirubin ≥ 1.5X ULN; c Pregnant or breastfeeding within 3 years prior

See supplementary table A1 in S3 Appendix for detailed list of inclusion and exclusion criteria.

Note: For prevalent warfarin users trial eligibility only revealed at point of random selection into the cohort for prevalent users. Numbers in figure show maximum theoretical number of warfarin users available should they be selected only at a time they were eligible for the trial.

Application of ARISTOTLE inclusion/exclusion criteria and matching to ARISTOTLE

Applying the ARISTOTLE inclusion/exclusion criteria and matching to ARISTOTLE baseline patient characteristics resulted in a cohort similar to the ARISTOTLE apixaban participants (Table 3); for example median age was 78 and mean CHADS₂ score 2.4 in CPRD Aurum before applying trial criteria and matching whereas the median age of 71 and mean CHADS₂ score 2.1 after these steps matched the ARISTOTLE apixaban participants. The ARISTOTLE-analogous apixaban arm matched the trial arm on prior VKA exposure, age, sex, stroke risk factors and CHADS₂ score, and proportion of patients with moderate or severe renal impairment.

Differences remained on baseline characteristics it was not feasible to match on namely: ethnicity (95.2% white, 2.4% Asian in CPRD Aurum apixaban vs 82.6% white, 14.4% Asian in ARISTOTLE) and concomitant medications (amiodarone 3.8%, aspirin 5.8%, digoxin 13.9% in CPRD Aurum apixaban users vs amiodarone 11.1%, aspirin 31.3%, digoxin 32.0% in ARISTOTLE apixaban arm). See Appendix for details on matching feasibility.
Propensity score matching of CPRD Aurum trial-analogous apixaban users to CPRD Aurum warfarin users

Results of Propensity score matching

Before propensity score matching, differences between treatment groups were evident for most baseline variables including age (median age 71 in apixaban vs 78 in warfarin), sex (apixaban 35.6% female vs warfarin 43.6%), and stroke risk factors [see Table 3]. After propensity score matching all baseline characteristics were well balanced (maximum standardised difference .031). From 9120 apixaban users only 274 (3.0%) were dropped due to unsuccessful matching.
<table>
<thead>
<tr>
<th>Characteristic - n(%) unless specified</th>
<th>No ARISTOTLE criteria or matching</th>
<th>After applying ARISTOTLE criteria</th>
<th>After applying ARISTOTLE criteria, matching to the trial and PSM apixaban to warfarin</th>
<th>Standardized difference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Apixaban</strong> (N=73 843)</td>
<td><strong>Warfarin</strong> (N=146 332)</td>
<td><strong>Apixaban</strong> (N=41 487)</td>
<td><strong>Warfarin</strong> (N=101 159)</td>
<td><strong>Apixaban</strong> (N=8 846)</td>
</tr>
<tr>
<td>Age – years, median (IQR)</td>
<td>78 (70, 85)</td>
<td>78 (71, 84)</td>
<td>78 (72, 84)</td>
<td>71 (63, 77)</td>
</tr>
<tr>
<td><strong>Female sex</strong></td>
<td>34 430 (46.6)</td>
<td>63 321 (43.3)</td>
<td>19 591 (47.2)</td>
<td>44 197 (43.7)</td>
</tr>
<tr>
<td>Systolic blood pressure – mmHg, median (IQR)</td>
<td>130 (120, 140)</td>
<td>131 (120, 140)</td>
<td>130 (120, 140)</td>
<td>130 (120, 140)</td>
</tr>
<tr>
<td><strong>Missing</strong></td>
<td>132</td>
<td>267</td>
<td>60</td>
<td>125</td>
</tr>
<tr>
<td>Weight – kg, median (IQR)</td>
<td>79 (67, 92)</td>
<td>80 (68, 93)</td>
<td>80 (69, 94)</td>
<td>85 (73, 100)</td>
</tr>
<tr>
<td>Prior myocardial infarction</td>
<td>9 958 (13.5)</td>
<td>20 406 (13.9)</td>
<td>5 035 (12.1)</td>
<td>13 446 (13.3)</td>
</tr>
<tr>
<td>Prior clinically relevant or spontaneous bleeding</td>
<td>16 972 (23.0)</td>
<td>31 034 (21.2)</td>
<td>7 721 (18.6)</td>
<td>19 007 (18.8)</td>
</tr>
<tr>
<td>History of fall within previous year</td>
<td>2 443 (3.3)</td>
<td>2 688 (1.8)</td>
<td>1 093 (2.6)</td>
<td>1 561 (1.5)</td>
</tr>
<tr>
<td>Prior use VKA &gt;30 days</td>
<td>24 240 (32.8)</td>
<td>102 725 (70.2)</td>
<td>12 558 (30.3)</td>
<td>75 787 (74.9)</td>
</tr>
<tr>
<td><strong>Qualifying risk factors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age ≥ 75 years</td>
<td>45 762 (62.0)</td>
<td>93 436 (63.9)</td>
<td>26 730 (64.4)</td>
<td>68 197 (67.4)</td>
</tr>
<tr>
<td>Prior stroke, TIA, or SE</td>
<td>20 713 (28.1)</td>
<td>38 132 (26.1)</td>
<td>11 422 (27.5)</td>
<td>25 898 (25.6)</td>
</tr>
<tr>
<td>Heart failure or reduced LVEF</td>
<td>22 329 (30.2)</td>
<td>50 480 (34.5)</td>
<td>11 650 (28.1)</td>
<td>33 422 (33.0)</td>
</tr>
<tr>
<td>Diabetes</td>
<td>20 104 (27.2)</td>
<td>40 103 (27.4)</td>
<td>11 630 (28.0)</td>
<td>28 496 (28.2)</td>
</tr>
<tr>
<td>Hypertension req. treatment</td>
<td>52 406 (71.0)</td>
<td>105 097 (71.8)</td>
<td>31 780 (76.6)</td>
<td>76 923 (76.0)</td>
</tr>
</tbody>
</table>

**CPRD Aurum**

**ARISTOTLE Trial**
<table>
<thead>
<tr>
<th>CHADS&lt;sub&gt;2&lt;/sub&gt; score, mean ± SD</th>
<th>2.4 ± 1.5</th>
<th>2.4 ± 1.4</th>
<th>2.5 ± 1.3</th>
<th>2.5 ± 1.2</th>
<th>2.1 ± 1.1</th>
<th>2.1 ± 1.1</th>
<th>2.1 ± 1.1</th>
<th>0.003</th>
<th>2.1 ± 1.1</th>
<th>2.1 ± 1.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHADS&lt;sub&gt;2&lt;/sub&gt;=0</td>
<td>6.494 (8.8)</td>
<td>10.240 (7.0)</td>
<td>134 (0.3)</td>
<td>356 (0.4)</td>
<td>52 (0.6)</td>
<td>55 (0.6)</td>
<td>0.004</td>
<td>54 (0.6)</td>
<td>58 (0.6)</td>
<td></td>
</tr>
<tr>
<td>CHADS&lt;sub&gt;2&lt;/sub&gt;=1</td>
<td>14.860 (20.1)</td>
<td>28.124 (19.2)</td>
<td>10.602 (25.6)</td>
<td>23.539 (23.3)</td>
<td>2.971 (33.6)</td>
<td>2.912 (32.9)</td>
<td>0.014</td>
<td>3.046 (33.4)</td>
<td>3.025 (33.3)</td>
<td></td>
</tr>
<tr>
<td>CHADS&lt;sub&gt;2&lt;/sub&gt;=2</td>
<td>19.844 (26.9)</td>
<td>43.294 (29.6)</td>
<td>12.969 (31.3)</td>
<td>32.980 (32.6)</td>
<td>3.157 (35.7)</td>
<td>3.239 (36.6)</td>
<td>0.019</td>
<td>3.262 (35.8)</td>
<td>3.254 (35.8)</td>
<td></td>
</tr>
<tr>
<td>CHADS&lt;sub&gt;2&lt;/sub&gt; ≥3</td>
<td>32.645 (44.2)</td>
<td>64.674 (44.2)</td>
<td>17.783 (42.9)</td>
<td>44.284 (43.8)</td>
<td>2.666 (30.1)</td>
<td>2.640 (29.8)</td>
<td>0.006</td>
<td>2.758 (30.2)</td>
<td>2.744 (30.2)</td>
<td></td>
</tr>
</tbody>
</table>

**Medications at index date**

<table>
<thead>
<tr>
<th>Medication</th>
<th>Index Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACE inhibitor or ARB</td>
<td>34,899 (47.3)</td>
</tr>
<tr>
<td>Amiodarone</td>
<td>1,903 (2.6)</td>
</tr>
<tr>
<td>Beta-blocker</td>
<td>46,173 (62.5)</td>
</tr>
<tr>
<td>Aspirin</td>
<td>32,645 (44.2)</td>
</tr>
<tr>
<td>Clopidogrel</td>
<td>6,464 (70.9)</td>
</tr>
<tr>
<td>Digoxin</td>
<td>9,771 (13.2)</td>
</tr>
<tr>
<td>Calcium channel blocker</td>
<td>19,659 (26.6)</td>
</tr>
<tr>
<td>Statin</td>
<td>39,027 (52.9)</td>
</tr>
<tr>
<td>Non-steroidal anti-inflammatory</td>
<td>4,953 (6.7)</td>
</tr>
<tr>
<td>Gastric antacid drugs</td>
<td>1,833 (2.5)</td>
</tr>
<tr>
<td>Proton pump inhibitor</td>
<td>28,444 (38.0)</td>
</tr>
<tr>
<td>H2 receptor antagonist</td>
<td>3,188 (4.3)</td>
</tr>
<tr>
<td>Renal function, creatinine clearance</td>
<td>21 591 (29.2)</td>
</tr>
<tr>
<td>------------------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Mild imp., &gt;50 to 80 ml/min</td>
<td>28 976 (39.2)</td>
</tr>
<tr>
<td>Moderate imp. (&gt;30 to 50 ml/min)</td>
<td>17 007 (23.0)</td>
</tr>
<tr>
<td>Severe imp. (≤30 ml/min)</td>
<td>4 317 (5.8)</td>
</tr>
<tr>
<td>Not reported</td>
<td>1 952 (2.6)</td>
</tr>
<tr>
<td>Peripheral artery disease</td>
<td>5 984 (8.1)</td>
</tr>
<tr>
<td>Aortic plaque</td>
<td>17 919 (24.3)</td>
</tr>
<tr>
<td>Smoking status</td>
<td></td>
</tr>
<tr>
<td>Non-smoker</td>
<td>27 568 (37.3)</td>
</tr>
<tr>
<td>Ex-smoker</td>
<td>40 815 (55.3)</td>
</tr>
<tr>
<td>Current smoker</td>
<td>5 236 (7.1)</td>
</tr>
<tr>
<td>Missing</td>
<td>224</td>
</tr>
<tr>
<td>Alcohol consumption</td>
<td></td>
</tr>
<tr>
<td>Non-drinker</td>
<td>27 185 (36.8)</td>
</tr>
<tr>
<td>Light, 1 to 14 units per week</td>
<td>32 190 (43.6)</td>
</tr>
<tr>
<td>Moderate, 15 to 42 units per week</td>
<td>8 950 (12.1)</td>
</tr>
<tr>
<td>Heavy, &gt; 42 units per week</td>
<td>1 488 (2.0)</td>
</tr>
<tr>
<td></td>
<td>3 901</td>
</tr>
<tr>
<td>-------------------------</td>
<td>-------</td>
</tr>
<tr>
<td><strong>Socioeconomic status</strong></td>
<td></td>
</tr>
<tr>
<td>(England IMD2015 Quintile)</td>
<td></td>
</tr>
<tr>
<td>1 (least deprived)</td>
<td>18 893 (25.6)</td>
</tr>
<tr>
<td>2</td>
<td>17 203 (23.3)</td>
</tr>
<tr>
<td>3</td>
<td>14 591 (19.8)</td>
</tr>
<tr>
<td>4</td>
<td>12 283 (16.6)</td>
</tr>
<tr>
<td>5 (most deprived)</td>
<td>10 804 (14.6)</td>
</tr>
<tr>
<td>Missing</td>
<td>69</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>70 703 (95.7)</td>
</tr>
<tr>
<td>Black</td>
<td>714 (1.0)</td>
</tr>
<tr>
<td>Asian</td>
<td>1 371 (1.9)</td>
</tr>
<tr>
<td>Other</td>
<td>198 (0.3)</td>
</tr>
<tr>
<td>Mixed</td>
<td>152 (0.2)</td>
</tr>
<tr>
<td>Unknown</td>
<td>385 (0.5)</td>
</tr>
<tr>
<td><strong>Charlson comorbidity components</strong></td>
<td></td>
</tr>
<tr>
<td>Chronic obstructive pulmonary disease</td>
<td>10 324 (14.0)</td>
</tr>
<tr>
<td>Connective tissue disease</td>
<td>5 377 (7.3)</td>
</tr>
<tr>
<td>Condition</td>
<td>Before ARISTOTLE</td>
</tr>
<tr>
<td>---------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Peptic ulcer</td>
<td>4,400 (6.0)</td>
</tr>
<tr>
<td>Liver disease</td>
<td>761 (1.0)</td>
</tr>
<tr>
<td>Hemiplegia</td>
<td>265 (0.4)</td>
</tr>
<tr>
<td>Non-haematological Cancer</td>
<td>12,567 (17.0)</td>
</tr>
<tr>
<td>Haematological cancer</td>
<td>1,966 (2.7)</td>
</tr>
<tr>
<td>BMI - kg/m², median (IQR)</td>
<td>28 (24, 32)</td>
</tr>
<tr>
<td>Missing</td>
<td>2,270</td>
</tr>
</tbody>
</table>

Table 3: Baseline characteristics of patients with Atrial Fibrillation prescribed apixaban and warfarin in CPRD Aurum compared with ARISTOTLE participants: i) before and ii) after applying ARISTOTLE inclusion and exclusion criteria and iii) after matching to the trial participants.

ACE = angiotensin-converting enzyme; ARB = angiotensin-receptor blocker; BMI=body mass index; CHADS² = stroke risk factor score based on Congestive heart failure, Hypertension, Age ≥ 75 years, Diabetes, prior Stroke; CPRD = Clinical Practice Research Datalink; IMD2015= Index of Multiple Deprivation 2015; imp.=impairment; IQR=interquartile range; LVEF=left ventricular ejection fraction; PSM = propensity score matching; SD=standard deviation; SE=systemic embolism;TIA=transient ischemic attack; VKA = vitamin K antagonist;
Main results

The hazard ratio (HR) for stroke/systemic embolism (SE) in the propensity score matched groups was 0.98 (95% CI 0.82, 1.19) (Figure 3 and Table A3 in S3 Appendix). This association was consistent with the non-inferiority margin (upper limit of the 95% CI less than 1.38) [7] but did not show superiority as predicted by ARISTOTLE [HR 0.79 (95% CI 0.66, 0.95)] (Figure 3 and Appendix Table A2 in S3 Appendix). The outcome of all-cause mortality also showed non-inferiority [Aurum 1.03 (0.93, 1.14) vs trial 0.89 (0.80, 0.998)] but did not indicate apixaban superiority. Absolute event rates for the primary outcome and components were close to the trial for apixaban – for example [comparing Aurum vs trial] stroke/SE event rate of 1.27%/year vs. 1.27% whereas the warfarin group had a lower event rate compared with ARISTOTLE (stroke/SE event rate of 1.29%/year vs. 1.60% and hemorrhagic stroke 0.33%/yr vs 0.47%/yr) (Figure 3). Mean duration of follow-up in the cohort was 1.8 years in the apixaban arm and 2.2 years in the warfarin arm.

Figure 3: Forest plot showing hazard ratios (dots) and 95% confidence intervals (lines) for apixaban vs warfarin. Absolute event rates (%/year) and Hazard Ratio (95% Confidence Intervals) are presented for key effectiveness outcomes in i) ARISTOTLE, ii) CPRD Aurum trial-matched cohort, iii) CPRD Aurum trial-matched with TTR < 0.75, and iv) CPRD Aurum trial-matched with TTR ≥ 0.75. Dashed line shows non-inferiority margin 1.38 for the upper bound of the 95% CI of the hazard ratio used in ARISTOTLE for the primary outcome of stroke or systemic embolism. For the analysis by TTR inverse probability of treatment weighting was applied to the apixaban users targeting the treatment effect in the warfarin users with TTR < 0.75 and TTR ≥ 0.75.

CI=confidence interval; CPRD=Clinical Practice Research Datalink; HR=hazard ratio; TTR=time in therapeutic range.

Analysis of impact of warfarin time in therapeutic range (TTR)

TTR was higher in the CPRD cohort than in ARISTOTLE (mean 0.73 vs. 0.62, median 0.76 vs 0.66).

Analysis by TTR suggested non-inferiority of apixaban vs warfarin in those with TTR < 0.75 [Stroke/SE 0.94 (0.75, 1.19), all-cause death 0.99 (0.87, 1.12)] (Figure 3). Apixaban was associated with increased
hazards of both outcomes compared to warfarin in those with well-controlled warfarin treatment

(TTR ≥ 0.75) [Stroke/SE 1.49 (1.13,1.97), all-cause death 1.75 (1.49,2.06)] (Figure 3).

**Analysis of apixaban dose-adjustment**

The proportion of patients meeting the criteria for reduced dose apixaban was similar between the CPRD ARISTOTLE-analogous apixaban, warfarin, and RCT apixaban groups (4.9%, 4.9%, and 4.7% respectively). When including the additional NICE criteria of creatinine clearance 5.1% of apixaban users in the ARISTOTLE-analogous cohort had an indication for reduced-dose apixaban yet a larger proportion (14.3%) were prescribed reduced dose apixaban implying some patients in CPRD Aurum may have been prescribed the wrong dose and/or information on criteria for dose reduction may have been missing from CPRD Aurum.

**Table 4: Apixaban Dose-adjustment in CPRD Aurum compared with ARISTOTLE**

<table>
<thead>
<tr>
<th></th>
<th>CPRD Aurum ARISTOTLE-analogous Apixaban (N = 8846)</th>
<th>CPRD Aurum ARISTOTLE-analogous Warfarin (N = 8846)</th>
<th>ARISTOTLE RCT Apixaban (N=9120)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard 5.0 mg BID dose</td>
<td>7580 (85.7%)</td>
<td>N/A</td>
<td>8692 (95.3%)</td>
</tr>
<tr>
<td>Reduced 2.5mg BID dose</td>
<td>1266 (14.3%)</td>
<td>N/A</td>
<td>428 (4.7%)</td>
</tr>
<tr>
<td>Reduced dose indicated per ARISTOTLE criteria</td>
<td>434 (4.9%)</td>
<td>436 (4.9%)</td>
<td>428 (4.7%)</td>
</tr>
<tr>
<td>Reduced dose indicated per NICE criteria</td>
<td>454 (5.1%)</td>
<td>459 (5.2%)</td>
<td>NR</td>
</tr>
</tbody>
</table>

NICE criteria for dose-adjustment included additional criteria of creatinine clearance 15–29 mL/minute. N/A=Not applicable. NR=Not reported.

A further analysis of the quality of dose-adjustment in patients in CPRD Aurum indicated 10.5% of patients may have been prescribed an incorrect dose of apixaban at the index prescription based on the data contained in their EHRs. The majority of incorrect dose relating to patients being prescribed reduced-dose apixaban despite not meeting the criteria for dose reduction. A large proportion of patients prescribed an incorrect dose had only 1 dose adjustment criteria (59.6% of those with incorrect dose) suggesting some prescribers may have thought a dose reduction was warranted.
when only 1 criteria was present. Other possible reasons for the incorrect dose-adjustment observed here may be data on the criteria missing from the EHR record (ie incorrect ascertainment) or consideration of other medical history which made a prescriber adjust the dose.

Table 5: Quality of apixaban dose-adjustment in CPRD Aurum ARISTOTLE-analogous cohort

<table>
<thead>
<tr>
<th>Dose Status Against NICE Criteria For Dose-adjustment at Index Date</th>
<th>CPRD Aurum ARISTOTLE-analogous Apixaban (N = 8846)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients on correct dose</td>
<td>7921 (89.5%)</td>
</tr>
<tr>
<td>Patients on incorrect dose</td>
<td>925 (10.5%)</td>
</tr>
<tr>
<td>Standard 5.0 mg BID dose despite meeting criteria for dose reduction</td>
<td>59 (0.7%)</td>
</tr>
<tr>
<td>Reduced 2.5mg BID dose despite not meeting criteria for dose reduction</td>
<td>866 (9.8%)</td>
</tr>
<tr>
<td>0 dose adjustment criteria recorded in EHR</td>
<td>313 (3.5%)</td>
</tr>
<tr>
<td>1 dose adjustment criteria recorded in EHR</td>
<td>553 (6.3%)</td>
</tr>
<tr>
<td>Age &gt; 80 years</td>
<td>389 (4.4%)</td>
</tr>
<tr>
<td>Body weight ≤ 60 kg</td>
<td>57 (0.6%)</td>
</tr>
<tr>
<td>Serum creatinine ≥ 1.5 mg/dL</td>
<td>107 (1.2%)</td>
</tr>
</tbody>
</table>

To assess the impact of the quality of dose-adjustment in the CPRD cohort on the effectiveness of apixaban a supplementary post hoc analysis was performed looking at the results in the subset of apixaban patients prescribed the correct dose (N=7921) compared with IPTW re-balanced warfarin comparators. The results in this subset were consistent with the primary results showing apixaban to be non-inferior to warfarin (Stroke/SE 0.96 [0.78,1.17], death 0.97 [0.87,1.09]) with the results moving slightly closer to those observed in ARISTOTLE.

Safety results

The analysis for safety outcomes is presented in Figure 4 and Table A5 in S3 Appendix; patients on apixaban had a lower risk of major bleeding compared with those on warfarin, HR (95% CI) 0.88 (0.77,1.00), consistent with ARISTOTLE. Analysis by TTR suggested superiority of apixaban for major bleeding in those with TTR <0.75 [0.70 (0.59,0.82)] whereas apixaban users had a higher risk of major bleeding compared with those with optimal warfarin control (TTR ≥ 0.75) [1.31 (1.08,1.59)].
Figure 4: Forest plot showing hazard ratios (dots) and 95% confidence intervals (lines) for apixaban vs warfarin. Absolute event rates (%/year) and Hazard Ratio (95% Confidence Intervals) are presented for key safety outcomes in i) ARISTOTLE, ii) CPRD Aurum trial-matched cohort, iii) CPRD Aurum trial-matched with TTR<0.75, and iv) CPRD Aurum trial-matched with TTR≥0.75.

For the analysis by TTR inverse probability of treatment weighting was applied to the apixaban users targeting the treatment effect in the warfarin users with TTR <0.75 and TTR ≥0.75.

CI=confidence interval; CPRD=Clinical Practice Research Datalink; HR=hazard ratio; TTR=time in therapeutic range.

Sensitivity analyses

Table A7 in S3 Appendix shows the proportion of patients switching treatment. A higher proportion of patients on warfarin switched to an alternative OAC during follow-up compared with those on apixaban (16.3% vs 6.1%).

Comparing patients who switched treatment during follow-up with those that continued on index treatment (Table A8 in S3 Appendix) suggests possible selection bias due to attrition in on-treatment analyses with median TTR markedly lower in warfarin users who switched treatments compared with persistent warfarin users (median TTR 0.64 vs 0.78). On-treatment analyses would likely be biased against apixaban since patients doing badly on warfarin (i.e., with low TTR) who would be more likely to experience events in the warfarin arm would be censored at treatment switch.

On-treatment analyses censoring around treatment switch or discontinuation are presented for the effectiveness analyses in the appendix (Table A6 in S3 Appendix); the results show evidence of the expected attrition bias against apixaban when compared with the ITT results in Figure 2, for example HR for stroke/SE is 1.04 (0.86, 1.25) in the on-treatment compared with 0.98 (95% CI 0.82, 1.19) in the ITT analysis.

Repeating the analysis with start of study period shifted forwards a year to investigate the impact of inclusion of early adopters yielded similar results to the primary analysis (Table A9 in S3 Appendix).

Prior INR control was not included in the propensity score models for the VKA-experienced due to a high rate of missing prior INR data (missing for 34% in the apixaban arm). A post-hoc sensitivity
analysis including a prior INR control variable in the PSM gave results consistent with the primary results [Stroke/SE HR 95%CI 1.02 (0.86,1.21)]. Details of this post hoc analysis are in S3 Appendix.

Discussion

In our emulation of ARISTOTLE using UK routinely-collected healthcare data we found results that met our predefined criteria for comparability with the trial. We saw non-inferiority of apixaban vs warfarin for prevention of stroke or systemic embolism, all-cause mortality, and major bleeding, but did not see superiority of apixaban vs warfarin for these outcomes as was seen in ARISTOTLE. We found higher TTR in the patients using warfarin in our cohort compared with the warfarin arm of ARISTOTLE (median 0.76 vs 0.66). Our analysis by TTR showed that in patients well-controlled on warfarin (TTR ≥ 0.75) warfarin is superior to apixaban whereas in patients not well-controlled on warfarin (TTR < 0.75) apixaban is non-inferior. We saw evidence suggesting sub-optimal dosing of apixaban in our cohort with approximately 10% of patients in the apixaban arm prescribed the reduced dose without meeting the criteria for the reduced dose.

We found the differences in the overall treatment-effect estimates between our cohort and ARISTOTLE may be explained by: the lower proportion of Asian patients in our cohort, differences in INR control in the warfarin arm of our cohort compared with ARISTOTLE, and the higher proportion of patients prescribed a reduced dose of apixaban in our cohort compared with ARISTOTLE.

Our findings are consistent with a UK study of ischemic stroke which compared DOACs with warfarin [29]. A Danish study found similar results to ours for stroke/SE [30] although they found apixaban users had a lower risk of death, a study of US claims data [31] also found apixaban was associated with a lower risk of death. A systematic review and meta-analysis of 16 studies [32] found pooled results for stroke and ICH that were consistent with ours. One study (in US claims data) also aimed to replicate ARISTOTLE [33, 34] and in contrast to our study found superiority for apixaban for stroke/SE, which may be linked to population differences such as lower TTR in US patients on warfarin [35] and differences in ethnicity. None of these studies matched to the ARISTOTLE trial.
participants, included prevalent users, or looked at how warfarin control impacted results. Further
details on these studies including design and key results are summarized in Table A10 in S3
Appendix.

A key strength of our study was the use of a framework which sampled prevalent users (the
continuing users of warfarin in this study) in a way that avoided selection bias facilitating the
construction of a cohort of patients similar to the target trial population, which included both new
users of apixaban and warfarin (VKA-naïve) and patients with prior VKA exposure (VKA-experienced)
that were randomised to stay on warfarin or switch to apixaban. The use of propensity score
matching, stratified by treatment history, enabled us to select a matched cohort well balanced on
important covariates. The successful emulation of ARISTOTLE by our study shows that valid
treatment effects can be obtained for important outcomes with OACs using non-interventional
methods with routinely collected clinical data. Having validated this framework, in future studies we
can look at the effectiveness of oral anticoagulants in AF patient groups not included or
underrepresented in the RCT such as elderly patients and those at increased bleeding risk. We also
recommend future analyses with an extended follow-up period compared with this study to
compare the long term outcomes seen in the non-interventional cohort with projected long-term
outcomes from the RCT.

An additional strength of our study was the ability to explore the quality of warfarin treatment in our
cohort and the impact of INR control on the treatment effect estimates. Our finding that the benefits
of apixaban vs warfarin depended on the quality of INR control in the warfarin arm answers
questions raised in the NICE premeeting briefing which looked at apixaban in the NVAF population
and noted the TTR seen in ARISTOTLE “may be lower than what is typical in UK clinical practice” and
“apixaban compared with well-controlled warfarin (TTR 75% or more) may not be superior in the
long term” [8]. ARISTOTLE presented outcomes by centre (for example hospital) TTR quartile and did
not show a signal of treatment efficacy differing by centre TTR quartile. We were able to use inverse
probability of treatment weighting to estimate the treatment effect in the different warfarin TTR groups. Our findings may help inform choice of treatment in the context of prior warfarin treatment and availability of INR measurements.

Whilst our study aimed to emulate ARISTOTLE using suitable methods there were several limitations. Some of the criteria assessed for ARISTOTLE eligibility may not be well recorded in CPRD leading to a risk of misclassification. Furthermore, misclassification of ARISTOTLE eligibility criteria and baseline covariates could be differential by treatment in the VKA-experienced patients if criteria such as renal function are more likely to be checked before changing treatment. However, the most important risk factors for the primary outcome of stroke (the components of CHA\textsubscript{2}-DS\textsubscript{2}-VASc stroke risk score) are mostly well recorded in CPRD Aurum and HES.

Our cohort did not attempt to match the trial on the use of concomitant medications in order for our cohort to reflect typical UK prescribing. In ARISTOTLE 31% of participants were using aspirin and 11% using amiodarone at baseline whereas, in our cohort only 6% were recorded as using aspirin and 4% amiodarone. Amiodarone potentiates the effects of warfarin and concomitant use of amiodarone with DOACs is associated with increased risk of major bleeding [36], whilst concomitant use of aspirin increases the risk of bleeding for both warfarin[37] and DOACs [38]. The difference in concomitant medication usage between our cohort and the trial population may explain some of the observed differences in treatment effects.

A key limitation of our study was the inability to match ARISOTLE on ethnicity meaning the CPRD Aurum cohort included a low number of patients from Asian and Hispanic groups when compared with the RCT (14.5% of participants in ARISTOTLE were Asian compared to 2.4% in our ARISTOTLE-analogous CPRD cohort). There are known racial differences in the treatment effects of OACs with Asian patients experiencing a higher risk of haemorrhagic stroke and intracranial haemorrhage compared with White patients; in ARISTOTLE Asian participants experienced double the risk of stroke or systemic embolism when on warfarin therapy when compared with White participants.
The reasons for the increased risk of bleeding associated with warfarin therapy in Asian patients is hypothesised to be associated with differences in drug metabolism and prevalence of cerebral microbleeds [38]. The difference in proportion of Asian patients between our cohort and ARISTOTLE is therefore likely to explain some of the differences in treatment effects seen and limits the generalisability of our study, with the results of our study of most relevance to White patients. This limitation on ethnicity arose from the data source used and time period studied (patients with AF in CPRD Aurum 2013-2019) which had a low proportion of Asian patients, likely due to AF being associated with older age combined with a lower prevalence of AF in Asian patients compared with White patients [38]. Whilst CPRD Aurum is largely representative of the UK population in relation to ethnicity [39], diversity is still limited for older individuals. Despite this, CPRD Aurum has shown to be a useful resource for investigating treatment effects in different ethnic groups for indications such as hypertension which is more prevalent and occurs at a younger age in ethnic minority groups, with similar trial replication methods used to compare antihypertensive treatment effects in underrepresented ethnic groups [40].

The approach our study used for handling missing data on baseline covariates relied on assumptions on the relationship between missingness, treatment, and outcomes which may not be valid; however the low proportion of missing data means that this is unlikely to have impacted the results. In the coarsened exact matching step the choice of variables will have an impact on the resulting cohort selected meaning a different combination of variables could lead to different results. There is a risk that residual confounding may be present despite the use of propensity score matching. The use of propensity score matching also has the potential to introduce bias by dropping patients from the cohort [19], however propensity score matching is well suited to the process of trial emulation including prevalent users and a low number of apixaban users were dropped due to unsuccessful matching. The inclusion of prevalent users of warfarin in the cohort risks the introduction of selection bias [20,21]; this was avoided by use of a method shown to produce unbiased estimates in a simulation study [21]. We found consistent results between our new and prevalent user strata.
across multiple outcomes providing reassurance the method used was likely to have successfully avoided selection bias.

Apixaban along with other DOACs were rapidly adopted as preferred first line OAC in AF during the study period; it was therefore not possible to match on calendar date leading to a difference in follow-up time between the treatment arms in our cohort. A higher proportion of warfarin users switched to alternative OAC during follow-up compared with those prescribed apixaban (16% vs 6%). The impact of this differential switching during follow-up was addressed in the sensitivity analyses.

The availability of new alternative treatments during the study period also means there is a risk of channelling bias in that over time the patients still on warfarin are more likely to be those doing well on warfarin. INR control prior to the index date was not included in the propensity score for the prevalent users due to a high rate of missing data, however, other variables associated with poor INR control were included in the models and an exploratory post-hoc analysis including a variable for poor INR control gave results consistent with the primary results.

Adherence to treatment was difficult to assess in our study due to automatic repeat prescriptions; treatment persistence was more useful in providing a measure of pattern of medicine use over time. In the analysis by TTR the adherence of patients using apixaban was not accounted for, however, a previous UK study showed apixaban had higher adherence than VKAs [41] meaning we would expect to see better effectiveness outcomes in apixaban. Furthermore, the use of IPTW in the analysis by TTR means predictors of poor adherence are likely to have been balanced between treatments. The analysis of TTR is limited by this being a post-baseline measure available for only one treatment arm; this limitation was also evident in the RCTs of DOACs vs warfarin and is mitigated in our study through the use of IPTW. Sensitivity analyses in our cohort using an on-treatment censoring scheme showed evidence of attrition bias. The regular measurement of INR and availability of alternative anticoagulants makes warfarin therapy particularly prone to attrition bias since a patient may be
more likely to switch to a DOAC if their INR is frequently out of the optimal range or if they have not been adhering to scheduled INR testing.

To conclude, we found that applying a reference trial emulation approach allowed us to emulate a landmark randomized trial of apixaban versus warfarin using UK non-interventional data, with results meeting pre-specified benchmarking criteria based on the reference trial results. This trial emulation method provides valid treatment effect estimates for apixaban compared to warfarin and can be used to determine risks and benefits of AF medications in people treated in routine clinical care. This study demonstrates a successful real world application of novel methods that have been proposed for the inclusion of prevalent users in observational studies, with the application of an adaptation to mimic the screening process making the method suitable for emulation of RCTs that include prevalent users. These methods could be adapted for emulation of RCTs in other therapeutic areas and for looking at patient groups under-represented or excluded from RCTs.

We further showed that when compared to well-controlled warfarin therapy, apixaban was associated with harmful treatment effects. The weaker overall treatment benefit observed in our cohort appears to be due to a higher proportion of patients with well-controlled warfarin in the UK clinical context, compared with the trial. NICE guidance (2021) advises that for patients already on VKA their TTR should be taken into account when deciding whether to switch to a DOAC; our findings support the option of staying on VKA for those patients happy with their treatment and who have TTR $\geq 0.75$.

S1: STROBE checklist
S2: ISAC protocol for the ARISTOTLE emulation study
S3: Appendix containing supporting information

References


[16] Rose M, Beasley N. Center for Drug Evaluation and Research Application number: 202155Orig1s000 MEDICAL REVIEW(S). Available for download at: https://www.accessdata.fda.gov/drugsatfda_docs/nda/2012/202155Orig1s000TOC.cfm


i. **New users of apixaban (VKA-naïve) propensity score matched 1:1 to new users of warfarin.**

```
new users of apixaban  matched 1:1 using propensity score matching  new users of warfarin
```

ii. **Pool of prevalent warfarin users** containing all warfarin prescription dates in the study period for each warfarin user **sampled without replacement by treatment history** to ensure similar duration of prior VKA exposure for the prevalent users in the apixaban and warfarin CPRD Aurum arms:

```
switchers to apixaban

continuers on warfarin

not eligible at selected index date

selected index date

- 10 warfarin continuers selected for each VKA-experienced apixaban 'switcher'
- Matching started with 'earliest' (i.e. shortest treatment history) apixaban switcher to avoid selection bias
- Trial-eligibility of warfarin users hidden until the point of selection to avoid selection bias
- Any warfarin users selected and not trial-eligible at that timepoint (treatment history) were dropped from the pool of potential comparators and an alternative warfarin continuer sampled
- The sampling process can be seen as analogous to the screening process of a RCT with potential participants only having one chance to pass screening.
```

iii. **VKA-experienced apixaban users propensity score matched 1:1 to the treatment-history matched comparator pool.**

**Propensity scores were calculated stratified by treatment history strata and patients matched within treatment history strata.**

```
switchers to apixaban

continuers on warfarin

matched 1:1 using propensity score matching
```

**Fig1**
AURUM - Apixaban

People in CPRD Aurum with a prescription for apixaban and prior diagnosis of AF (acceptable and HES linked) = 86 888
Did not meet minimum registration period n=13 045
People that met minimum registration period n=73 843
People that met ARISTOTLE inclusion criteria due to:
- Age <18 years (n=4)
- No ARISTOTLE stroke risk factor (n=6 304)
People that met ARISTOTLE inclusion criteria = 67 539
Excluded due to ARISTOTLE exclusion criteria = 26 052:
- AF due to reversible causes (n=2 625)
- Moderate or severe mitral stenosis (n=585)
- Increased bleeding risk (n=8 463)
- Other condition requiring chronic anticoagulation (n=3 154)
- Persistent uncontrolled hypertension (n=1 602)
- Active infective endocarditis (n=52)
- Concomitant aspirin > 165 mg/day (n=31)
- Concomitant aspirin and thienopyridine (n=421)
- Severe comorbid conditiona (n=8 312)
- Alcohol or drug abuse (n=2 263)
- Recent ischemic stroke (within 7 days) (n=462)
- Severe renal insufficiency (n=2 969)
- Elevated ALT, AST, or Total Bilirubina (n=1 485)
- Platelets ≤ 100,000/ mm³ (n=536)
- Hemoglobin < 9 g/dL (n=1 095)
- Pregnant or breastfeedingb (n=48)
People after applying ARISTOTLE exclusion criteria = 41 487

AURUM - Warfarin

People in CPRD Aurum with a prescription for warfarin and prior diagnosis of AF (acceptable and HES linked) = 159 632
Did not meet minimum registration period n=13 300
People that met minimum registration period n=146 332
People that met ARISTOTLE inclusion criteria due to:
- Age <18 years (n=4)
- No ARISTOTLE stroke risk factor (n=6 801)
People that met ARISTOTLE inclusion criteria = 139 527
Excluded due to ARISTOTLE exclusion criteria = 38 368:
- AF due to reversible causes (n=6 012)
- Moderate or severe mitral stenosis (n=3 614)
- Increased bleeding risk (n=15 621)
- Other condition requiring chronic anticoagulation (n=10 085)
- Persistent uncontrolled hypertension (n=3 017)
- Active infective endocarditis (n=200)
- Concomitant aspirin > 165 mg/day (n=33)
- Concomitant aspirin and thienopyridine (n=469)
- Severe comorbid conditiona (n=13 324)
- Alcohol or drug abuse (n=2 727)
- Recent ischemic stroke (within 7 days) (n=154)
- Severe renal insufficiency (n=5 713)
- Elevated ALT, AST, or Total Bilirubina (n=3 029)
- Platelets ≤ 100,000/ mm³ (n=1 688)
- Hemoglobin < 9 g/dL (n=2 865)
- Pregnant or breastfeedingb (n=104)
People after applying ARISTOTLE exclusion criteria = 101 159 (4 764 109 Rx dates)

Fig 2
<table>
<thead>
<tr>
<th>Outcome</th>
<th>Rate (%/yr)</th>
<th>HR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Major bleeding</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARISTOTLE</td>
<td>2.13</td>
<td>0.69 (0.60, 0.80)</td>
</tr>
<tr>
<td>CPRD Aurum</td>
<td>2.45</td>
<td>0.88 (0.77, 1.00)</td>
</tr>
<tr>
<td>CPRD Aurum TTR &lt;0.75</td>
<td>2.69</td>
<td>0.70 (0.59, 0.82)</td>
</tr>
<tr>
<td>CPRD Aurum TTR ≥0.75</td>
<td>2.22</td>
<td>1.31 (1.08, 1.59)</td>
</tr>
<tr>
<td><strong>Intracranial bleeding</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARISTOTLE</td>
<td>0.33</td>
<td>0.42 (0.30, 0.58)</td>
</tr>
<tr>
<td>CPRD Aurum</td>
<td>0.35</td>
<td>0.71 (0.51, 1.00)</td>
</tr>
<tr>
<td>CPRD Aurum TTR &lt;0.75</td>
<td>0.36</td>
<td>0.55 (0.37, 0.83)</td>
</tr>
<tr>
<td>CPRD Aurum TTR ≥0.75</td>
<td>0.31</td>
<td>1.13 (0.69, 1.83)</td>
</tr>
<tr>
<td><strong>Other location bleeding</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARISTOTLE</td>
<td>1.79</td>
<td>0.79 (0.68, 0.93)</td>
</tr>
<tr>
<td>CPRD Aurum</td>
<td>0.60</td>
<td>0.93 (0.70, 1.22)</td>
</tr>
<tr>
<td>CPRD Aurum TTR &lt;0.75</td>
<td>0.64</td>
<td>0.70 (0.50, 0.99)</td>
</tr>
<tr>
<td>CPRD Aurum TTR ≥0.75</td>
<td>0.56</td>
<td>1.35 (0.91, 2.00)</td>
</tr>
<tr>
<td><strong>Gastrointestinal bleeding</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARISTOTLE</td>
<td>0.76</td>
<td>0.89 (0.70, 1.15)</td>
</tr>
<tr>
<td>CPRD Aurum</td>
<td>1.52</td>
<td>0.88 (0.74, 1.04)</td>
</tr>
<tr>
<td>CPRD Aurum TTR &lt;0.75</td>
<td>1.71</td>
<td>0.72 (0.58, 0.88)</td>
</tr>
<tr>
<td>CPRD Aurum TTR ≥0.75</td>
<td>1.39</td>
<td>1.35 (1.05, 1.73)</td>
</tr>
</tbody>
</table>

Fig4