BMJ Open

Critical appraisal and issues regarding generalisability of comparative effectiveness studies of NOACs in atrial fibrillation and their relation to clinical trial data: a systematic review

Eveline M Bunge ,[1] Ben van Hout,[2] Sylvia Haas,[3] Georgios Spentzouris,[4] Alexander Cohen[5]

To cite: Bunge EM, van Hout B, Haas S, *et al*. Critical appraisal and issues regarding generalisability of comparative effectiveness studies of NOACs in atrial fibrillation and their relation to clinical trial data: a systematic review. *BMJ Open* 2021;**11**:e042024. doi:10.1136/bmjopen-2020-042024

► Prepublication history and additional material for this paper is available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2020-042024).

Received 24 June 2020
Revised 14 January 2021
Accepted 18 January 2021

[1]Pallas health research and consultancy BV, Rotterdam, The Netherlands
[2]School of Health and Related Research, University of Sheffield, Sheffield, UK
[3]Formerly Technical University of Munich, Munich, Germany
[4]Daiichi Sankyo Europe, Munchen, Germany
[5]Department of Haematology, Guys and St Thomas' NHS Foundation Trust, King's College London, London, UK

Correspondence to
Ms Eveline M Bunge;
bunge@pallashrc.com

ABSTRACT

Objective To critically appraise the published comparative effectiveness studies on non-vitamin K antagonist oral anticoagulants (NOACs) in non-valvular atrial fibrillation (NVAF). Results were compared with expectations formulated on the basis of trial results with specific attention to the patient years in each study.

Methods All studies that compared the effectiveness or safety between at least two NOACs in patients with NVAF were eligible. We performed a systematic literature review in Medline and EMbase to investigate the way comparisons between NOACs were made, search date 23 April 2019. Critical appraisal of the studies was done using among others ISPOR Good Research Practices for comparative effectiveness research.

Results We included 39 studies in which direct comparison between at least two NOACs were made. Almost all studies concerned patient registries, pharmacy or prescription databases and/or health insurance database studies using a cohort design. Corrections for differences in patient characteristics was applied in all but two studies. Eighteen studies matched using propensity scores (PS), 8 studies weighted patients based on the inverse probability of treatment, 1 study used PS stratification and 10 studies applied a proportional hazards model. These studies have some important limitations regarding unmeasured confounders and channelling bias, even though the larger part of the studies were well conducted technically. On the basis of trial results, expected differences are small and a naïve analysis suggests trials with between 7200 and 56 500 patients are needed to confirm the observed differences in bleedings and between 51 800 and 7 994 300 to confirm differences in efficacy.

Discussion Comparisons regarding effectiveness and safety between NOACs on the basis of observational data, even after correction for baseline characteristics, may not be reliable due to unmeasured confounders, channelling bias and insufficient sample size. These limitations should be kept in mind when results of these studies are used to decide on ranking NOAC treatment options.

Strengths and limitations of this study

- To our knowledge, this is the first systematic review that critically appraised the quality and generalisability of the comparative effectiveness studies on non-vitamin K antagonist oral anticoagulants (NOACs) in patients with atrial fibrillation and to relate this to clinical trial data.
- A naïve trial analysis was conducted to estimate the number of patients needed in a randomised clinical trial to confirm the differences in efficacy and bleeding.
- Thirty-nine articles were included, of which only one included all four NOACs.

INTRODUCTION

Guidelines state a preference for non-vitamin K antagonist oral anticoagulants (NOACs) above vitamin K antagonists (VKAs) in patients with non-valvular atrial fibrillation (NVAF) requiring prevention of stroke and systemic embolism.[1 2] However, no recommendation for a specific NOAC is made in these guidelines, and in daily practice, physicians have to make a choice which of the four available NOACs (dabigatran, rivaroxaban, apixaban, edoxaban) they prescribe for a particular patient.[3–6]

In the absence of head-to-head trials, comparative effectiveness research (CER) has been conducted to compare the NOACs with regard to effectiveness and safety. This is also described as real-world evidence; that is, the data will come from patients treated in daily practice. Comparisons on effectiveness and safety between NOACs are however not easy to make, as patients will not be prescribed one of the NOACs at random. The choice of a certain NOAC for a patient will at least partly be driven by patient characteristics,

such as age, concomitant medications, and the risk of stroke and/or bleeding. This can lead to systematic differences between the treatment groups, which is known as channelling bias.[7] In order to make a valid comparison on effectiveness and safety between the NOACs, adjusting for these characteristics is necessary when these characteristics are also related to the outcome (confounding variables).

Several techniques exist to correct for imbalances in risks, but there is no gold standard and all methods have advantages and disadvantages. Cox proportional hazards (Cox PH) regression model adjustment can be used but large sample sizes are needed when the number of events is relatively low and the number of covariates is high (as a rule of thumb, about 10 events per predictor variable[8]) and these large sample sizes are not always available. Event rates are low, around 1 per 100 patient years for efficacy outcomes and to detect differences, even in a randomised clinical trial, one needs substantial number of patients. This number would only increase when the results are contaminated by a lack of balance between the patients' groups. Another method to adjust for confounding is using propensity scores (PS) to create comparable patient groups before the analysis. A PS is the probability of an individual receiving a specific treatment given a specific set of patient characteristics (eg, age, gender, comorbidities).[9] Variables related to the outcome should be included in the PS despite their strength of association on treatment (exposure) selection. This will increase the precision of the estimated exposure effect, while bias will not be increased. Variables that are related to the exposure but not the outcome will decrease the precision of the estimated exposure effect without decreasing bias.[10] Adjustment for confounding using PS can be done by matching the treatment groups on the PS, by weighing treatment groups based on the PS inverse probability of treatment weighting (IPTW), by PS stratification or by covariate adjustment using the PS.[9 11] Well-conducted PS methods will lead to treatment groups that are very well comparable regarding important confounders, which increases the confidence in the results; however, there are also some disadvantages. For instance, in PS matching studies, patients who cannot be matched to another patient will be excluded from the analyses, and in IPTW, when patients on one treatment have a low PS and patients treated with the other treatment have a high PS, extreme weights can occur which can bias the results.[12]

To gain more understanding in how the above described methodologies were applied in peer-reviewed CER on effectiveness and safety in NOACs in patients with NVAF, we conducted a systematic literature review. Within this, we compare the results with those from a naïve analysis of the results of the four major trial for rivaroxaban, apixaban, dabigatran and edoxaban, and compare the results from the various analyses with those from the trials.

METHODS

Information sources, search strategy and eligibility criteria

We performed a systematic literature review to identify peer-reviewed CER on NOACs in patients with atrial fibrillation. A search in Medline (access through PubMed) and EMbase was performed combining search strings on NOAC, VKA and atrial fibrillation (see online supplemental appendix 1 for the search strings). The search was conducted on 23 April 2019 and we checked all articles published in English language. The title and abstract selection was done in duplicate by two independent researchers.

The following inclusion criteria were used:

- Population: patients with NVAF.
- Intervention: NOAC (dabigatran, rivaroxaban, apixaban or edoxaban).
- Comparator: other NOAC(s) (dabigatran, rivaroxaban, apixaban and/or edoxaban).
- Outcomes: effectiveness and safety.
- Study type: comparative effectiveness studies with a cohort design.

The following exclusion criteria were applied:

- Studies on only one NOAC.
- Studies in which VKA is the comparator for the NOACs, and NOACs are not compared against each other.
- Studies on cost-effectiveness and healthcare resources use.
- Studies on adherence or persistence.

Critical appraisal

We checked the setting, inclusion and exclusion criteria, and the following baseline characteristics: age, proportion males, CHA2DS2-VASc (Congestive heart failure, Hypertension, Age ≥75 years, Diabetes Mellitus, Prior Stroke or TIA or thromboembolism, Vascular disease, Age 65–74 years, Sex) score and comorbidity index.

We used the criteria suggested by ISPOR, Yao *et al*[13] and Austin *et al* as a guidance to critically appraise the articles in which PS were used.[12 14 15] The criteria we checked concerned:

- The variables included in the PS model.
- Explanation of the variable selection procedure for PS model.
- Distribution of baseline characteristics for each group before PS analysis.
- In case of PSM:
  - Matching ratio.
  - Distance metric.
  - With or without replacement.
  - Comparability of baseline characteristics in the matched groups.
  - Sample size before and after matching.
- In case of IPTW:
  - Comparability of baseline characteristics in the weighted groups.
  - Extreme weights.
- In case of PS stratification:

- Number of strata, comparability of baseline characteristics.

- In case of analyses in which no PS was used in the main analyses:
  - We evaluated whether the ratio number of covariates to the number of events seemed sufficient to produce valid results.[8]
- Sensitivity analyses to further explore the magnitude of residual confounding (ie, case–cross-over study designs; clinical details in a subsample; proxy measures; or instrumental variable techniques).

Naïve trial analysis

Trials are quite often designed with a null hypothesis and associated with a power calculation, while real-world studies are often dictated by the number of observations available. To give the results from the real-world evidence some perspective, we undertook a naïve trial analysis in which the risk reductions from each trial with respect to efficacy and safety outcomes were applied to an average number of outcomes observed in the warfarin arms in each trial. This leads to an estimate of the relevant rates for each drug and the differences are illustrated by the number of patients (sample size) needed in a randomised clinical trial to confirm the estimated differences.

RESULTS

In total, we found 1302 unique articles in our search, of which 39 articles fulfilled the inclusion and exclusion criteria and were included for data extraction (see figure 1). In tables 1–5, study characteristics are presented. The most important differences between the studies are outlined in table 6.

More than 50% of the studies were conducted in the USA (n=24),[16–39] five were conducted in Denmark,[40–44] four in Taiwan,[45–48] and one in France,[49] Sweden,[50] Scotland,[51] the UK,[52] Spain[53] and China.[54] Dabigatran and rivaroxaban were included in all 39 studies, apixaban was included in 26 studies and edoxaban was included in 1 study. Next to these NOACs, VKA was included in 25 of these studies as one of the comparators. The results below focus on the NOAC to NOAC comparisons only.

In the studies that included apixaban, dabigatran and rivaroxaban, rivaroxaban was most dominantly used in the USA, the UK, Scotland and Taiwan, while dabigatran was the most prescribed NOAC in Denmark. In

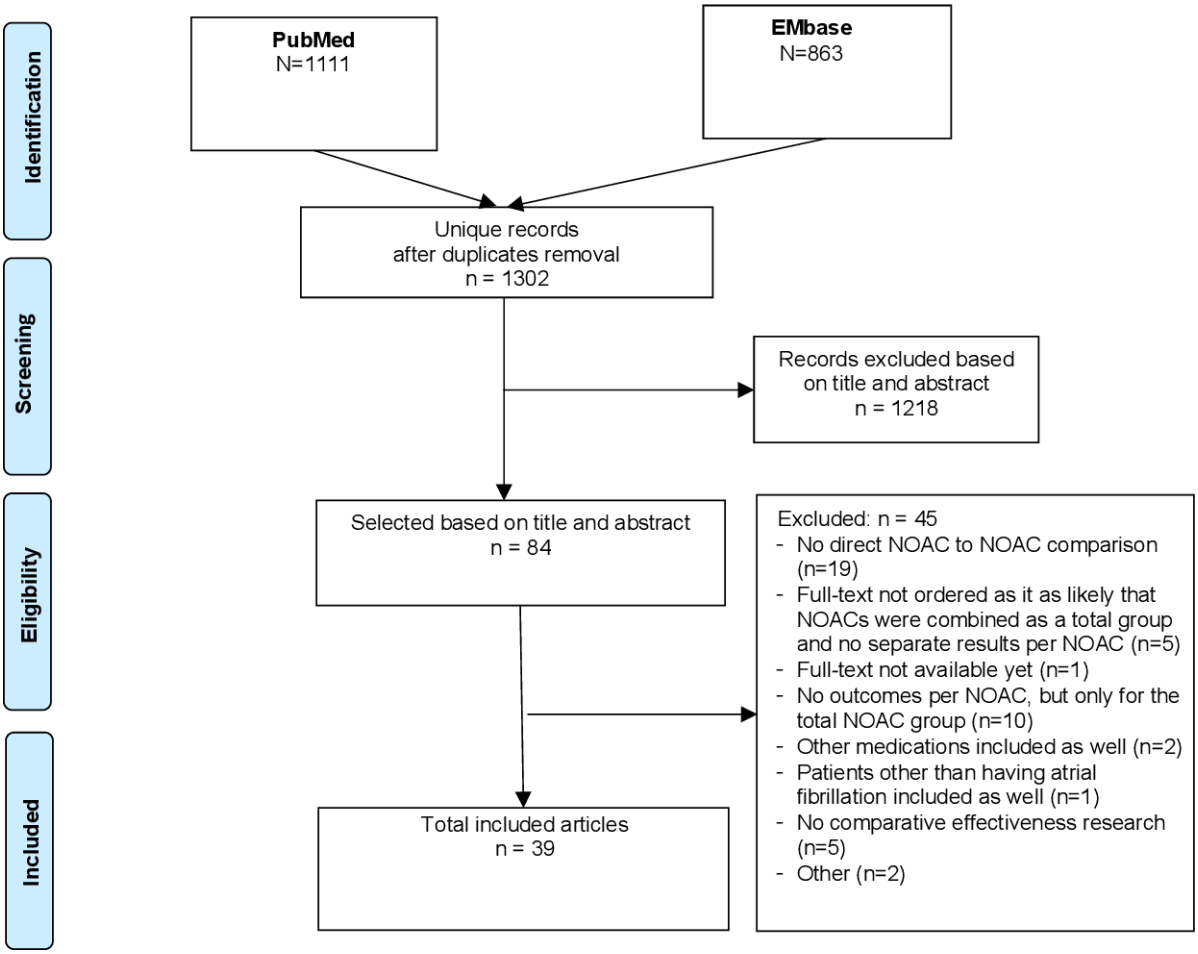

Figure 1 PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart. NOACs, non-vitamin K antagonist oral anticoagulants.

Table 1 Characteristics of the included articles that used propensity score matching (PSM) as primary analyses (n=18)

Author and country	Setting and study period	Study population	Patient characteristics before PSM (range between NOACs)	Primary outcome definition	PSM details	Sample size before matching	Sample size after matching	Results/conclusion as reported in the article
Abraham *et al*,[16] USA	OptumLabs Data Warehouse 1 October 2010 through 28 February 2015	NVAF patients, 18 years of age or older, identified by their index prescription of a NOAC during study period (excluded if NOAC prescribed during 12 months before index date). No reporting on earlier VKA use.	Age: 69.2±11.6–72.2±11.1 Male: 54.0%–60.5% CHA2DS2-VASc: 3.2–4.0 CDI: 2.3–2.7	Gastrointestinal bleeding: definition by Lewis *et al* 2002 using inpatient hospital claims for relevant primary and secondary discharge diagnoses	3 matched cohorts; 1:1 PSM without replacement and with a calliper of 0.01. After PSM, standardised differences of all baseline characteristics were <10%.	Rivaroxaban: n=19301 Dabigatran: n=17426 Apixaban: n=6576	Rivaroxaban vs dabigatran: n=31574 Apixaban vs rivaroxaban: n=13130 Apixaban vs dabigatran: n=13084 (more than 90% of original smallest samples size)	Apixaban had the most favourable gastrointestinal bleeding profile and rivaroxaban had the least favourable safety profile. Apixaban had the most favourable gastrointestinal safety profile among all age groups.
Amin *et al*,[19] (*J Manag Care Spec Pharm*) USA	Medicare & Medicaid Services 1 January 2012 to 31 December 2014	NVAF patients of at least 65 years old, OAC treatment-naïve, ≥1 prescription claim for OAC during study period. Excluded if OAC pharmacy claim during the 12 months before study start.	Age: 77.2±7.0–78.4±7.4 Male: 47.4%–50.6% CHA2DS2-VASc: 4.4–4.6 CCI: 2.5 to 2.7	Hospitalisation for stroke, systemic embolism and major bleeding: ICD-9 code as primary discharge diagnosis	2 matched cohorts; 1:1 PSM with nearest neighbour without replacement and with a calliper of 0.01. After PSM, standardised differences of all baseline characteristics were <10%.	Rivaroxaban: n=53146 Apixaban: n=20853 Dabigatran: n=16743	Rivaroxaban vs apixaban: n=41608 Dabigatran vs apixaban: n=30836 (more than 90% of original smallest samples size)	Apixaban was associated with significantly lower risks of all-cause, stroke/SE-related, and MB-related hospitalisations compared with dabigatran and rivaroxaban

Continued

Table 1 Continued

Author and country	Setting and study period	Study population	Patient characteristics before PSM (range between NOACs)	Primary outcome definition	PSM details	Sample size before matching	Sample size after matching	Results/conclusion as reported in the article
Amin *et al*,[19] (*J Med Econ Spec Pharm*) USA	OptumInsight research database 1 January 2012–30 September 2015	NVAF patients of at least 18 years old, OAC treatment-naïve, ≥1 prescription claim for OAC during study period. Excluded if OAC pharmacy claim during the 12 months before study start.	NR	Hospitalisation for stroke, systemic embolism and major bleeding: ICD-9 code as primary discharge diagnosis	2 matched cohorts; 1:1 PSM with nearest neighbour without replacement and with a calliper of 0.01. After PSM, standardised differences of all baseline characteristics were <10%.	Rivaroxaban: n=14 163 Apixaban: n=8652 Dabigatran: n=3684	Apixaban vs rivaroxaban: n=16 880 Apixaban vs dabigatran: n=7114 (more than 90% of original smallest samples size)	Rivaroxaban patients were associated with a significantly higher risk of all-cause and major bleeding-related hospitalisations and dabigatran patients were associated with a significantly higher risk of major bleeding hospitalisation compared with apixaban
Andersen *et al*,[40] Denmark	National patient register, Register of Medicinal Product Statistics 1 July 2013–31 March 2016	NVAF patients who were new users of NOAC aged 45 years of age or older, with a recent diagnosis of NVAF (received no OAC treatment in the 12 months before inclusion; 'recent diagnosis' is not defined)	Online material not available	Stroke, systemic embolism and major bleeding (ie, intracranial bleeding, gastrointestinal bleeding (bleeding ulcer, haematemesis ormelena) or other serious bleeding (anaemia caused by bleeding, bleeding of unknown origin, bleeding of the respiratory or urinary tract, peritoneal, retinal or orbital bleeding): hospital admission with a primary or secondary	3 matched cohorts; 1:1 PSM with nearest neighbour with a calliper of 0 (replacement yes or no not reported). All baseline characteristics were well balanced after matching, except for calendar year.	Apixaban: n=4292 Dabigatran: n=3913 Rivaroxaban: n=3805	Apixaban vs rivaroxaban: n=7352 Apixaban vs dabigatran: n=6470 Rivaroxaban vs dabigatran: n=5440	There were no statistically significant differences in risk of stroke or systemic embolism or major bleeding in propensity-matched comparisons between apixaban, dabigatran and rivaroxaban used in standard doses

Continued

Table 1 Continued

Author and country	Setting and study period	Study population	Patient characteristics before PSM (range between NOACs)	Primary outcome definition	PSM details	Sample size before matching	Sample size after matching	Results/conclusion as reported in the article
Blin *et al*,[49] France	French nationwide claims and hospitalisation database, Système National des Données de Santé 2013–2015	NVAF patients of at least 18 years old, all new users of standard or reduced doses of NOAC in (received no OAC treatment in the 3 years before the index date)	Age: 65.3±10.2–69.0±11.1 Male: 62.7%–68.3% Modified CHA2DS2-VASc≥2: 57.1%–67.4% Comorbidities: NR	Hospitalisation with a main diagnosis of ischaemic stroke or systemic embolism or major bleeding and all-cause death (ICD-10 codes)	1 matched cohort PSM method not reported. After PSM, standardised differences of all baseline characteristics were <10%.	Rivaroxaban: n=18 829 Dabigatran: n=10 847	Dabigatran vs rivaroxaban: n=16 580	Dabigatran had similar or better effectiveness than rivaroxaban but lower bleeding risk. Death rates were not different.
Briasoulis *et al*,[21] USA	Medicare and Medicaid Services 1 January 2010–31 December 2013	NVAF patients newly diagnosed of ≥65 years old and initiated OAC treatment during study period	Age: 75.4±6–75.5±6 Male: 50%–53% CHA2DS2-VASc: 4.1–4.1 Gagne: 2.7–2.7	All-cause mortality, stroke, including ischaemic stroke or transient ischaemic attack, gastrointestinal bleeding, any bleeding, non-gastrointestinal bleeding, acute myocardial infarction. ICD-9-CM reported in inpatient claims, whether primary and secondary codes were used is not described.	1 matched cohort; three-way propensity matching (VKA was one of the groups, but not further discussed here). After PSM, standardised differences of all baseline characteristics were <10%.	Rivaroxaban: n=14 257 Dabigatran: n=13 522	Dabigatran vs rivaroxaban: n=26 814	Rivaroxaban was associated with higher gastrointestinal bleeding rates than dabigatran

Continued

Table 1 Continued

Author and country	Setting and study period	Study population	Patient characteristics before PSM (range between NOACs)	Primary outcome definition	PSM details	Sample size before matching	Sample size after matching	Results/conclusion as reported in the article
Deitelzweig *et al*,[23] USA	Humana Research Database (Medicare coverage) January 2013–30 September 2015	NVAF patients age of ≥65 years, OAC treatment naïve (excluded if they had a pharmacy claim for OAC during the baseline period, which was 12 months before index date)	Age: 76.8±8.3–78.0±9.0 Male: 51.5%–55.1% CHA2DS2-VASc: 4.3–4.6 CCI: 2.7 to 3.0	Hospitalisation claims of stroke, systemic embolism and major bleeding: ICD-9 code as primary discharge diagnosis	2 matched cohorts; 1:1 PSM with nearest neighbour (replacement yes or no and calliper not reported) balanced with key patient characteristics not statistically different (p>0.05)	Rivaroxaban: n=11 082 Apixaban: n=8250 Dabigatran: n=2474	Apixaban vs rivaroxaban: n=13 620 Apixaban vs dabigatran: n=4654	Apixaban is associated with significantly lower risk of stroke/systemic embolism and major bleeding than rivaroxaban, and a trend towards better outcomes vs dabigatran
Gupta *et al*,[26] USA	Department of Defence data 1 January 2012 to 30 September 2015	NVAF patients, treatment-naïve (excluded if a pharmacy claim for an OAC during the baseline period)	NR	Inpatient claim of stroke, systemic embolism or major bleeding as primary or secondary diagnosis based on validated administrative claims-based algorithms	2 matched cohorts; 1:1 PSM with nearest neighbour without replacement with a calliper of 0.01. After PSM, standardised differences of all baseline characteristics were <10%.	Rivaroxaban: n=15 680 Apixaban: n=11 754 Dabigatran: n=4312	Rivaroxaban vs apixaban: n=22 568 Dabigatran vs apixaban: n=8258	Rivaroxaban was associated with a significantly higher risk of stroke/systemic embolism and major bleeding compared with apixaban. Dabigatran use was associated with a numerically higher risk of stroke/systemic embolism and a significantly higher risk of major bleeding compared with apixaban.

Continued

Table 1 Continued

Author and country	Setting and study period	Study population	Patient characteristics before PSM (range between NOACs)	Primary outcome definition	PSM details	Sample size before matching	Sample size after matching	Results/conclusion as reported in the article
Lai *et al*,[47] Taiwan	National Health Insurance programme 2011 to 2014	NVAF and flutter patients, ≥20 years, new-users (new users not further defined)	Age: 75.1±9.7–75.4±9.6 Male: 54.7%–56.7% CHA2DS2-VASc: 3.3–3.3 Comorbidity index: NR	All-cause death	1 matched cohort; 1:1 PSM with calliper <0.2 (neighbour and replacement not reported). Balance checked with p values and standardised difference.	Dabigatran: n=10625; Rivaroxaban: n=4609	Dabigatran vs rivaroxaban: n=9200	Rivaroxaban therapy was associated with a statistically significant increase in all-cause death compared with dabigatran
Lin *et al*,[29] USA	IMS Pharmetrics Plus database January 2013–September 2015	NVAF patients of at least 18 years old who initiated OAC (received no OAC treatment received 12 months before the index date)	NR	Major bleeding first listed in ICD-9 diagnosis or procedure codes	2 cohorts; 1:1 PSM with nearest neighbour (replacement and calliper not reported). Patient key characteristic being similar with p>0.05.	NR	Apixaban vs rivaroxaban: n=8124 Apixaban vs dabigatran: n=5368	Apixaban is associated with reduced risk of hospitalisation compared with dabigatran and rivaroxaban
Lip *et al*,[30] (*Thromb Haemost*) USA	Truven MarketScan Commercial Claims and Encounter and Medicare Supplemental and Coordination of Benefits Databases January 2012 to December 2014	NVAF patients ≥18 years who newly initiated OACs (patients with a prescription claim for OAC prior to the index date were excluded)	Age: 66.5±12.4–68.5±12.4 Male: 61.4%–65.0% CHA2DS2-VASc: 2.6–2.8 CDI: 1.6–1.8	Major bleeding listed first primary ICD-9 code	3 cohorts; 1:1 PSM with nearest neighbour without replacement with a maximum calliper of 0.01. After PSM, standardised differences of all baseline characteristics were <10%.	Rivaroxaban: n=17801 Apixaban: n=7438 Dabigatran: n=4661	Apixaban vs dabigatran: n=14798 Rivaroxaban vs dabigatran: n=9314 Apixaban vs rivaroxaban: n=8814	Compared with apixaban, rivaroxaban initiation was associated with significantly higher risk of major bleeding. The difference for dabigatran was not statistically significant.

Continued

Table 1 Continued

Author and country	Setting and study period	Study population	Patient characteristics before PSM (range between NOACs)	Primary outcome definition	PSM details	Sample size before matching	Sample size after matching	Results/conclusion as reported in the article
Lip *et al*,[32] USA	Medicare and Medicaid Services Medicare; Truven MarketScan, IMS PharMetrics Plus Database, Optum Clinformatics Data Mart, and the Humana Research Database 1 January 2013 to 30 September 2015	NVAF patients newly prescribed OAC, (received no OAC treatment in the 12 months before the index date)	Age: 71.4±11.4–73.1±11.6 Male: 55.0%–59.6% CHA2DS2-VASc: 3.3–3.6 CDI: 2.4–2.8	Hospitalisations with stroke, systemic embolism or major bleeding as the principal or first-listed diagnosis	3 cohorts 1:1 PSM with nearest neighbour without replacement with a maximum calliper of 0.01. After PSM, standardised differences of all baseline characteristics were <10%.	Rivaroxaban: n=103 477 Apixaban: n=63 484 Dabigatran: n=27 571	Apixaban–rivaroxaban: n=125 238 Dabigatran–rivaroxaban: n=55 076 Apixaban–dabigatran: n=54 192	Apixaban was associated with a lower rate of stroke/systemic embolism and major bleeding compared with dabigatran and rivaroxaban. Dabigatran was associated with a lower rate of major bleeding compared with rivaroxaban, with similar rates of stroke/systemic embolism.
Lutsey, 2019 USA	MarketScan Commercial Database 1 January 2010 through 30 September 2015	NVAF patients aged 45 and older with at least one prescription for OAC after their first AF claim (de novo patients or first initiation of treatment)	Age: 69.1±11.4–69.9±11.7 Male: 59.4–63.7 CHA2DS2-VASc: 3.3–3.6 Comorbidity index: NR	Venous thromboembolism: at least one inpatient ICD-9 claim (first listed or not is not specified)	3 cohorts 1:1 PSM with a maximum calliper of 0.25 (neighbour and replacement not reported). Balance not described.	Rivaroxaban: n=31 119 Dabigatran: n=28 089 Apixaban: n=17 112	Rivaroxaban vs apixaban: n=32 468 Dabigatran vs rivaroxaban: n=21 160 Dabigatran vs apixaban: n=6200	Risk of VTE was lowest among those prescribed apixaban and dabigatran
Mentias *et al*,[34] USA	Medicare & Medicaid Services 1 January 2010 to 31 December 2013	NVAF patients, newly diagnosed who initiated an OAC within 90 days of diagnosis	Age: 75.8±6.4–75.8±6.4 Male: 48.9%–50.1% CHA2DS2-VASc: 4.3–4.3 Gagne: 3.0–3.0	Inpatient admission for acute ischaemic stroke or major bleeding as defined by Rothendler* and Suh based on the primary ICD-9-CM diagnosis on inpatient standard analytical files claims for acute care stays	1 cohort three-way PSM (VKA was one of the groups, but not further discussed here). After PSM, standardised differences of all baseline characteristics were <10%.	Rivaroxaban: n=23 177 Dabigatran: n=21 979	NR	Rivaroxaban users had significantly higher major bleeding risk compared with dabigatran users in the medium and high comorbidity groups

Continued

Table 1 Continued

Author and country	Setting and study period	Study population	Patient characteristics before PSM (range between NOACs)	Primary outcome definition	PSM details	Sample size before matching	Sample size after matching	Results/conclusion as reported in the article
Norby *et al*,[35] *USA*	Truven Health MarketScan Commercial Claims and Encounters Database and the Medicare Supplemental and Coordination of Benefits Database	NVAF patients with at least one prescription of NOAC after their first AF claim (first prescription of OAC)	Age: 67.2±12.0–68.1±12.3 Male: 60.6–62.7 CHA2DS2-VASc: 2.6–2.9 Comorbidity index: NR	Ischaemic stroke (primary discharge), intracranial bleeding (primary discharge), myocardial infarction (first or second position of an inpatient discharge diagnosis) and gastrointestinal bleeding (primary and secondary diagnoses), presence of transfusion codes, and presence/absence of trauma codes to exclude trauma-related bleeding based on ICD-9 codes	1 cohort; 1:1 PSM, greedy matching technique with a calliper of 0.25	NR	Rivaroxaban vs dabigatran: n=16 957	Endpoint rates were similar when comparing anticoagulant-naïve rivaroxaban and dabigatran initiators, with the exception of higher gastrointestinal bleeding risk in rivaroxaban users
Noseworthy *et al*,[36] USA	OptumLabs Data Warehouse 1 October 2010–28 February 2015	NVAF patients ≥18 years, who were OAC users during study period	NR	Inpatient admission for stroke or systemic embolism or major bleeding (ICD-9 codes in the primary or secondary diagnosis positions of inpatient claims)	3 cohorts; 1:1 PSM without replacement and with a calliper of 0.01. A standardised difference <10% was considered acceptable.	NR	Rivaroxaban vs dabigatran: n=31 574 Apixaban vs rivaroxaban: n=13 130 Apixaban vs dabigatran: n=13 084	Dabigatran, rivaroxaban and apixaban appear to have similar effectiveness, although apixaban may be associated with a lower bleeding risk and rivaroxaban may be associated with an elevated bleeding risk

Continued

Table 1 Continued

Author and country	Setting and study period	Study population	Patient characteristics before PSM (range between NOACs)	Primary outcome definition	PSM details	Sample size before matching	Sample size after matching	Results/conclusion as reported in the article
Shantha *et al*,[37] USA	Medicare and Medicaid 1 November 2011–31 December 2013	Newly diagnosed NVAF patients and initiated OAC use	Male: Age: 74.7±5.9–74.9±6. CHADS2-VASc: 3.7–3.8 Gagne score: 2.9–2.9 Female: Age: 76.6±6.6–76.9±6.6 CHADS2-VASc: 4.8–4.9 Gagne: 3.0–3.1	Inpatient admissions for acute ischaemic stroke or major bleeding (primary ICD-9-CM diagnosis on inpatient standard analytical files claims for acute care stays)	1 cohort; three-way PSM (VKA was one of the groups, but not further discussed here). A standardised difference <10% was considered acceptable.	Rivaroxaban: n=23 177 Dabigatran: n=21 979	Dabigatran vs rivaroxaban: n=37 298	The reduced risk of ischaemic stroke in patients taking rivaroxaban, compared with dabigatran, seems to be limited to men, whereas the higher risk of bleeding seems to be limited to women
Villines *et al*,[39] USA	US Department of Defence Military Health System database 1 July 2010 to 30 June 2016 for the dabigatran vs rivaroxaban cohort, and 28 December 2011 to 30 June 2016 for the dabigatran vs apixaban cohort	NVAF patients ≥18 years newly initiated on standard-dose NOAC (first initiation of treatment, AF diagnosis in the 12 months before the index date or on the index date)	Age (mean): 70.9–71.3 Male: 60%–62% CHA2DS2-VASc: 3.1–3.1 CCI score: 4.3–4.3	Stroke or major bleeding, ICD-9 or 10 codes, whether primary and secondary codes were used is not described	2 cohorts 1:1 PSM nearest neighbour with a calliper of 0.20 (replacement not reported). Balanced if the absolute value of the STD was ≤10%.	NR	Dabigatran vs rivaroxaban: n=25 526 Dabigatran vs apixaban: n=9604	Dabigatran was associated with significantly lower major bleeding risk vs rivaroxaban, and no significant difference in stroke risk. For dabigatran vs apixaban, the reduced sample size limited the ability to draw definitive conclusions.

Age: mean, SD unless stated otherwise.
CCI, Charlson Comorbidity Index; CDI, Charlson-Deyo Index; CHA2DS2-VASc, Congestive heart failure, Hypertension, Age ≥75 years, Diabetes Mellitus, Prior Stroke or TIA or thromboembolism, Vascular disease, Age 65–74 years, Sex; Gagne, Gagne Comorbidity Score; NOACs, non-vitamin K antagonist oral anticoagulants; NVAF, non-valvular atrial fibrillation.

Table 2 Characteristics of the included articles that used inverse probability of treatment weighting as primary analyses (n=8)

Author and country	Setting and study period	Study population	Patient characteristics (range between NOACs)	Primary outcome definition	IPWT details	Sample size	Result/conclusion as reported in the article
Adeboyeje *et al*,[17] USA	HealthCore Integrated Research Environment 1 November 2009–31 January 2016	NVAF patients newly prescribed OAC (no prescriptions for any anticoagulant in the 6-month period preceding their index dates)	Age (mean): 66–69 Male: 59.1%–65.5% CHA2DS2-VASc: 2.7–3.2 Comorbidity index: NR	Hospitalisation for major bleeding (ICD-9 CM codes; whether primary and secondary codes were used is not described	Extreme weights: not reported. Balanced if the absolute value of the STD was ≤10%.	Dabigatran: n=8539 Rivaroxaban: n=8398 Apixaban: n=3689	Apixaban and dabigatran were associated with lower major bleeding risk compared with rivaroxaban; however, apixaban had a lower risk of major gastrointestinal bleeding than dabigatran
Chan *et al*,[46] Taiwan	Taiwan National Health Insurance Research 1 June 2012–31 December 2016	NVAF patients with their first prescription of OAC	Age: 75±10–76±10 Male: 55%–60% CHA2DS2-VASc: 3.7–3.9 Comorbidity index: NR	Hospitalisation for ischaemic stroke/ systemic embolism, intracranial haemorrhage, major gastrointestinal bleeding, acute myocardial infarction, all major bleeding events and all-cause mortality. ICD-9 and 10 codes, whether primary and secondary codes were used is not described.	Extreme weights: not reported. Balanced if the absolute value of the STD was ≤10%.	Rivaroxaban: n=27 777 Dabigatran: n=20 079 Apixaban: n=5843	Three low-dose NOACs showed similar performance as without subgrouping
Charlton *et al*,[22] USA	HealthCore Integrated Research Environment database 1 November 2010–31 March 2014	NVAF patients hospitalised for bleeding after starting OAC (AF diagnosis 6 months before starting one of the index drugs)	Age: 68.0±12.5–69.6±12.6 Male: 61.8–62.9 CHA2DS2-VASc: 3.8–3.8 CDI: 2.0–2.3	Total length of hospital stay, proportion of patients admitted to the ICU, mean length of ICU stay, and all-cause 30-day and 90-day mortality, ICD-9 codes, whether primary and secondary codes were used is not described	Extreme weights: not reported. Balance was tested using ANOVAs for significant differences.	Dabigatran: n=442 Rivaroxaban n=256	There were no significant differences in relative risk of all-cause 30 or 90 days

Continued

Table 2 Continued

Author and country	Setting and study period	Study population	Patient characteristics (range between NOACs)	Primary outcome definition	IPWT details	Sample size	Result/conclusion as reported in the article
Graham *et al*,[25] USA	Medicare 4 November 2011–30 June 2014	NVAF patients, at least 65 years old, initiating OAC at standard doses (first treatment, received no NOAC treatment for other indications in the last 6 months before the index date)	Age: 65–74 years: 50%–51% Age: 75–84: 40%–40% Age≥85: 9%–10% Male: 53%–53% CHADS2 ≥2: 66%–67% Comorbidity index: NR	Thromboembolic stroke, ICH, major extracranial bleeding events and mortality (as the first study outcome or within 30 days after hospitalisation for another primary outcome event), ICD-9 codes, whether primary and secondary codes were used is not described.	Extreme weights: not reported. Balanced if the absolute value of the STD was ≤10%.	Rivaroxaban: n=66 651 Dabigatran: n=52 240 Weighted cohorts Rivaroxaban: n=66 630 Dabigatran: n=52 264	Treatment with rivaroxaban was associated with statistically significant increases in intracranial bleeding and major extracranial bleeding, including major gastrointestinal bleeding, compared with dabigatran
Graham *et al*,[24] USA	Fee-for-service Medicare Part A (hospitalisation), Part B (office-based care), and Part D (prescription drug coverage) October 2010–September 2015	NVAF patients of ≥65 years old (first initiation of treatment)	Age (mean): 74.9–75.5 Male: 52.2%–59.3% CHA2DS2-VASc≥2: 96.6%–97.4% Comorbidity index: NR	Hospitalised due to thromboembolic stroke, intracranial haemorrhage, major extracranial bleeding and all-cause mortality. ICD codes from the first hospital discharge diagnosis position.	Not described how weighted cohort was composed. Balanced if the absolute value of the STD was ≤10%.	Rivaroxaban: n=106 389 Dabigatran: n=86 198 Apixaban: n=73 039 Weighted cohort Rivaroxaban: n=106 369 Dabigatran: n=86 293 Apixaban: n=72 921	Dabigatran and apixaban were associated with a more favourable benefit–harm profile than rivaroxaban

Continued

Table 2 Continued

Author and country	Setting and study period	Study population	Patient characteristics (range between NOACs)	Primary outcome definition	IPWT details	Sample size	Result/conclusion as reported in the article
Hernandez *et al*,[27] USA	Medicare 4 November 2011–31 December 2013	NVAF patients (at any time before the index date; no NOAC treatment at least 3 months before the index date)	High dose: Age: <65: 5.0%–6.3% Age: 65–74: 38.4%–39.3% Age: ≥75: 55.3–55.7 Male: 45.9–49.5 CHADS2: 3.3–3.3 Comorbidity index: NR	Ischaemic stroke (inpatient, emergency room, or outpatient claim with primary or secondary, ICD-9 codes), other thromboembolic events, and all-cause mortality; ICD-9 codes, whether primary and secondary codes were used is not described. Any bleeding event and major bleeding; intracranial haemorrhage and gastrointestinal bleeding, not further described.	Extreme weights: not reported. Balanced if the absolute value of the STD was ≤10%.	Dabigatran n=9138 Rivaroxaban n=8367	There was no difference in stroke prevention between rivaroxaban and dabigatran; however, rivaroxaban was associated with a higher risk of thromboembolic events other than stroke, death and bleeding.
Larsen *et al*,[43] Denmark	Danish National Prescription Registry, Danish National Patient Register, Danish Civil Registration System August 2011–October 2015	NVAF patients who were naïve to oral anticoagulants (no use of oral anticoagulant within 1 year)	Age (median, IQR): 67.6 (62.0–72.4)–71.8 (65.7–78.9) Male: 56.9%–66.1% CHA2DS2-VASc: 2.2–2.8 Comorbidity index: NR	Ischaemic stroke or systemic embolism, ICD-10 codes whether primary and secondary codes were used is not described	Extreme weights: not reported. Balanced if the absolute value of the STD was ≤10%.	Dabigatran: n=12 701 Rivaroxaban: n=7192 Apixaban: n=6349	Apixaban and dabigatran were associated with a significantly lower risk of death compared with rivaroxaban. Risk of any bleeding or major bleeding were significantly lower for apixaban and dabigatran than for rivaroxaban.

Continued

Table 2 Continued

Author and country	Setting and study period	Study population	Patient characteristics (range between NOACs)	Primary outcome definition	IPWT details	Sample size	Result/conclusion as reported in the article
Meng *et al*,[48] Taiwan	National Health Insurance claims database 1 June 2012–31 May 2015	All NVAF patients aged ≥20 years who initiated NOACs during study period	Age <65: 11.8%–13.5% Age 65–74: 29.7%–32.7% Age ≥75: 53.8%–58.4% Male: 54.6%–56.2% CHA2DS2-VASc: 3.2–3.3 Comorbidity index: NR	All-cause death, ischaemic stroke, intracranial haemorrhage, gastrointestinal haemorrhage needing transfusion, ICD-10 codes, whether primary and secondary codes were used is not described	Extreme weights: not reported. Balanced if the absolute value of the STD was ≤10%.	Dabigatran: n=13 505 Rivaroxaban: n=6551 Weighted pseudo-cohort Dabigatran: n=13 508; Rivaroxaban: n=6547	Rivaroxaban seemed to be associated with an increased risk of all-cause death compared with dabigatran

ANOVAs, analyses of variance; CHA2DS2-VASc, Congestive heart failure, Hypertension, Age ≥75 years, Diabetes Mellitus, Prior Stroke or TIA or thromboembolism, Vascular disease, Age 65–74 years, Sex; NOACs, non-vitamin K antagonist oral anticoagulants; NVAF, non-valvular atrial fibrillation.

three other European studies, the distribution was about equal between the three NOACs. In none of the included studies apixaban was the most dominantly prescribed NOAC.

Setting

Most studies concerned patient registries, pharmacy or prescription databases and/or health insurance databases (n=39), while there were three clinical practice-based studies.[50 53 54]

Study population

All studies included only patients with NVAF. In seven studies, it was specifically described that patients were newly diagnosed with NVAF and initiated NOAC treatment during study period.[21 27 34 37 40 45 54] None of the other studies included prevalent users of (N)OAC, but included, for example, 'newly treated', 'initiating treatment', 'new users', 'first-time prescription' of NVAF patients who were prescribed (N)OAC. In some studies, (N)OAC use in the past (between 3 months and 2 years before index date) was allowed, while this seemed not be allowed in some other studies, or it was not described.

Inclusion criteria

Five studies concerned elderly patients specifically (ie, ≥65 years old),[19 21 23–25] two included adults ≥45 years old[33 40] and one study included patients between 30 and 100 years of age.[44] The other studies included all adults with atrial fibrillation (it was assumed that if no further age specification was provided, 'adults' meant that all >18 years old were included). In one study, only patients who were hospitalised for bleeding after start with OAC treatment were included.[22] No other focus on a specific group of patients with AF was found.

Exclusion criteria

NOAC use that could be related to other disorders, such as transient AF, major knee or hip surgery, venous thromboembolism or pulmonary embolism, were specifically described as exclusion criteria in most studies, except in 10 studies.[16 27 28 33–35 50 52–54] In one study, patients with liver injury before their first oral anticoagulant (OAC) prescription were specifically excluded.[18]

Baseline characteristics

Baseline characteristics of patients with NVAF differed between studies. Mean age ranged from 65 to 84 years between the studies. The percentage of males ranged from 39% to 73%, and the mean CHA2DS2-VASc score ranged from 2.1 to 4.9. Excluding the five studies that specifically focused on an elderly population of ≥65 years old and the two additional studies that used the Medicare database (only patients of 65 years or older are in Medicare), the mean age ranged from 65 to 78 years. Different measures were used to assess the comorbidity index: Charlson Comorbidity Index, Charlson-Deyo Index and Gagne Comorbidity Score, while in 30 of the 43 studies no comorbidity index was presented.

Table 3 Characteristics of the included articles that used adjusted Cox proportional hazard models as primary analyses (n=10)

Author and country	Setting and study period	Study population	Patient characteristics (range between NOACs)	Primary outcome definition	Sample size	Results/conclusion as reported in the article
Al-Khalili, 2016 Sweden	Tertiary referral cardiology outpatient clinic (the Stockholm Heart Center) December 2011–May 2014	NVAF patients from a single cardiology outpatient clinic incorporating the AF unit (initiate NOAC treatment)	Age: 72±8–73±8 Male: 50%–51% CHA2DS2-VASc: 3–3 Comorbidity index: NR	Major bleeding was defined according to the criteria of the International Society of Thrombosis and Hemostasis	Rivaroxaban: n=282; Apixaban: n=251 Dabigatran: n=233;	Rivaroxaban was associated with the highest bleeding rate owing mainly to the highest number of minor bleedings, and apixaban had the lowest bleeding rates and side effects
Alonso *et al*,[18] USA	Truven Health MarketScan Commercial Claims and Encounter Database and the Medicare Supplemental and Coordination of Benefits Database 1 January 2007–31 December 2014	NVAF patients with a first prescription of OAC after 2 November 2011	Age: 67.2±12.4–69.3±12.5 Male: 60.1%–65.1% CHA2DS2-VASc: 2.9–3.6 Comorbidity index: NR	Hospitalisation for liver injury potentially related to drug hepatotoxicity, ICD-9-CM codes in any position	Rivaroxaban: n=30347; dabigatran: n=17286; Apixaban: n=9205	Risk of liver disease hospitalisation was higher in rivaroxaban users compared with dabigatran and apixaban users
Chan *et al*,[45] Taiwan	Taiwan National Health Insurance Research Database. 1 January 1996–31 December 2013	NVAF patients newly diagnosed	Age: 75±9–76±9 Male: 54–58 CHA2DS2-VASc: 4.1–4.1 Comorbidity index: NR	Ischaemic stroke or systemic embolism, ICH, hospitalisation for gastrointestinal (GI) bleeding, acute myocardial infarction (AMI), all hospitalisations for bleeding and all-cause mortality. All discharge diagnosis according to the ICD, whether primary and secondary codes were used is not described.	Dabigatran 110 mg: n=5921 Rivaroxaban 10 mg: n=3916	No differences were found between rivaroxaban and dabigatran in risk for thromboembolic events, intracranial haemorrhage, critical GI bleeding or all-cause mortality. However, rivaroxaban was associated with a higher risk for noncritical gastrointestinal bleeding than dabigatran.
Hernandez *et al*[27], USA	Medicare database Jan 1, 2013 - Dec 31, 2014	NVAF patients newly diagnosed	Age: 74.9±8.7–77.4±8.6 Male: 42.5%–47.0% CHA2DS2-VASc: 4.3–4.7 Comorbidity index: NR	Ischaemic stroke, death, bleeding events, gastrointestinal bleeding, treatment persistence. ICD-9 codes, whether primary and secondary codes, were used is not described.	Rivaroxaban: n=5139; Apixaban: n=2358; Dabigatran: 1415;	Apixaban had the most favourable effectiveness and safety profile

Continued

Table 3 Continued

Author and country	Setting and study period	Study population	Patient characteristics (range between NOACs)	Primary outcome definition	Sample size	Results/conclusion as reported in the article
Lamberts *et al*,[42] Denmark	Danish National Patient Registry, Danish National Prescription Registry, Danish Civil Personal Registry up to 31 December 2015	NVAF patients ≥18 years, with newly prescribed OAC (no prescription at least 6 months before inclusion)	Age: 71.5±11.0–75.4±11.10 Male: 50.8%–56.7% CHA2DS2-VASc: 2.7–3.2 Comorbidity index: NR	Major bleeding events requiring hospitalisation, ICD-10 codes, whether primary and secondary codes were used is not described	Dabigatran: n=15 413; Apixaban: n=7963; Rivaroxaban: n=6715	Apixaban had a lower adjusted major bleeding risk compared with rivaroxaban and dabigatran
Lip *et al*,[30] (*Int J Clin Pract*) USA	Truven MarketScan Commercial & Medicare supplemental US database 1 January 2013–31 December 2013	NVAF patients ≥18 years with newly prescribed OAC (no OACs received at least 1 year before the start of the OAC treatment)	Age: 66.8±12.2–69.3±12.3 Male: 63.1%–65.8% CHA2DS2-VASc: 2.6–2.8 CCI: 1.7 to 1.9	Major bleeding was identified using hospital claims, which had a bleeding diagnosis code as the first listed primary ICD-9 diagnosis code	Rivaroxaban: n=10 050 Dabigatran: n=4173 Apixaban: n=2402	Initiation with rivaroxaban was associated with a significantly greater risk of major bleeding compared with initiation on apixaban. There was no significant difference in the risk of major bleeding among patients newly initiated on dabigatran compared with apixaban.
Mueller *et al*,[51] Scotland	Prescribing Information System, the Scottish Morbidity Records/ Hospital Inpatients and Outpatient attendance datasets; National Records of Scotland Drug's approval date—December 2015	NVAF patients who initiated NOAC treatment	Age: 71.1±12.0–74.8±11.0 Male: 53.5%–73.1% CHA2DS2-VASc: 2.5–3.0 CCI: 1.1 to 1.4	Strokes, systemic embolism, death due to cardiovascular, pulmonary embolism, bleeding events, clinical endpoints, according to ICD-10 codes whether primary and secondary codes were used is not described	Rivaroxaban: n=7265 Apixaban: n=6200 Dabigatran: n=1112	All NOACs were similarly effective in preventing strokes and systemic embolisms, while patients being treated with rivaroxaban exhibited the highest bleeding risks
Staerk *et al*,[44] Denmark	Danish National Patient Registry, Danish National Prescription Registry, Danish civil registration system 1 March 2012–31 December 2016	NVAF patients, first-time OAC users (no previous OAC use), between 30 and 100 years old	Standard dose: Age (median, IQR): 67 (61, 71)–71 (65, 78) Male: 55.4%–63.7% CHA2DS2-VASc (median); 2–3 Comorbidity index: NR	Stroke/ thromboembolism (TE), ischaemic stroke, major bleeding, intracranial bleeding and gastrointestinal bleeding, ICD-10 codes whether primary and secondary codes were used is not described	Dabigatran: n=11 492 Apixaban: n=11 064 Rivaroxaban: n=8966	Rivaroxaban was associated with higher bleeding risk compared with dabigatran and apixaban and dabigatran was associated with lower intracranial bleeding risk compared with rivaroxaban and apixaban.

Continued

Table 3 Continued

Author and country	Setting and study period	Study population	Patient characteristics (range between NOACs)	Primary outcome definition	Sample size	Results/conclusion as reported in the article
Tepper *et al*,[38] USA	Truven MarketScan Commercial Claims and Encounter and Medicare Supplemental & Coordination of Benefits Early View Database 1 January 2013–31 October 2014	NVAF patients aged ≥18 years with new initiators of NOACs or switched from warfarin to a NOAC	Age: 68±12–70±12 Male: 65.3–62.7 CHA2DS2-VASc: 2.4–2.5 CCI: 1.6 to 1.8	Bleeding, ICD-9-CM codes, whether primary and secondary codes were used is not described	Rivaroxaban: n=30529 Dabigatran: n=20963 Apixaban: n=8785	Rivaroxaban appeared to have an increased risk of any bleeding, clinically relevant non-major bleeding and major inpatient bleeding, compared with apixaban patients. There was no significant difference in any bleeding, clinically relevant non-major bleeding or inpatient major bleeding risks between patients treated with dabigatran and apixaban.
Vinogradova *et al*,[52] UK	UK general practices contributing to QResearch or Clinical Practice Research Datalink 2011–2016	NVAF patients, new NOAC (received no OAC treatment in at least the last 12 months)	QResearch: Age: 74.7±10.7–76.5±10.9 Male: 51.8%–58.0% CHA2DS2-VASc: NR Comorbidity index: NR	Major bleeding after entry to the study which led to a hospital admission or death, based on linked hospital or mortality records	Rivaroxaban: n=16547 Apixaban: n=10601 Dabigatran: n=5537	Apixaban was associated with a lower risk of major bleed than rivaroxaban. Rivaroxaban was associated with a higher risk of intracranial bleed compared with apixaban. rivaroxaban was associated with higher risks compared with apixaban for haematuria, all gastrointestinal bleed and upper gastrointestinal bleed. The risk of primary ischaemic stroke did not differ between any of the anticoagulants.

AF, atrial fibrillation; CHA2DS2-VASc, Congestive heart failure, Hypertension, Age ≥75 years, Diabetes Mellitus, Prior Stroke or TIA or thromboembolism, Vascular disease, Age 65–74 years, Sex; NOAC, non-vitamin K antagonist oral anticoagulants; NVAF, non-valvular atrial fibrillation.

Table 4 Characteristics of the included articles that used unadjusted primary analysis (n=2)

Author and country	Setting and study period	Study population	Patient characteristics (range between NOACs)	Primary outcome definition	Primary analysis	Sample size	Results/conclusion as reported in the article
Cerdá *et al*,[53] Spain	Oral Anticoagulant Treatment Unit of the Hemostasis and Thrombosis Department of the University Hospital Vall d'Hebron from Barcelona (Spain) January 2015–September 2017	NVAF patients with non-valvular AF, with or without prior stroke, that had started treatment with any NOAC for the prevention of stroke	Age: 73.1±15.2–78.9±8.7 Male: 45.1%–63.4% CHA2DS2-VASc: 3.9–4.4 Comorbidity index: NR	Major bleeding according to ISTH 2005	Log-rank test	Rivaroxaban: n=663 Dabigatran: n=352 Apixaban: n=325 Edoxaban: n=103	Rates of ischaemic stroke and intracranial haemorrhage (ICH) were similar among different NOACs, but rates of major bleeding were higher with dabigatran and apixaban and lower with rivaroxaban
Li *et al*,[54] China	Queen Mary Hospital, Hong Kong January 2008–December 2014	NVAF patients diagnosed during study period	Age: 71.9±11.1–73.3±12.1 Male: 53.1%–59.8% CHA2DS2-VASc: 3.6–3.7 Comorbidity index: NR	The primary outcome was a composite of hospital admission with ischaemic stroke or ICH, or death during the follow-up period. ICD-10 codes in medical records, and discharge summaries, whether primary and secondary codes were used is not described.	Cox proportional hazard model (likely unadjusted, but this is not clearly described in the article)	Rivaroxaban: n=669 Dabigatran: n=467	Dabigatran had a lower ischaemic stroke risk compared with patients on rivaroxaban. There was no significant difference in ischaemic stroke risk between those on rivaroxaban and dabigatran.

CHA2DS2-VASc, Congestive heart failure, Hypertension, Age ≥75 years, Diabetes Mellitus, Prior Stroke or TIA or thromboembolism, Vascular disease, Age 65–74 years, Sex; NOACs, non-vitamin K antagonist oral anticoagulants; NVAF, non-valvular atrial fibrillation.

Table 5 Characteristics of the included articles that used propensity score stratification as primary analyses (n=1)

Author and country	Setting and study period	Study population	Patient characteristics (range between NOACs)	Primary outcome definition	PS details	Sample size	Results/conclusion as reported in the article
Gorst-Rasmussen *et al*,[41] Denmark	Danish National Prescription Registry, Danish National Patient Register, Danish Civil Registration System 1 February 2012–31 July 2014	NVAF patients who were new users of OAC (no OAC treatment in at least the last 2 years)	Standard dose: Age: 66.0±8.5–72.8±9.9 Male: 51.1%–63.5% CHA2DS2-VASc: 2.1–3.0 Comorbidity index: NR	Ischaemic stroke/systemic embolism/transient ischaemic attack, any bleeding and all-cause death. ICD-10 codes, whether primary and secondary codes were used is not described	Asymmetric trimming of the propensity score. Trimmed propensity score was used in 10 deciles as strata. Balanced if the absolute value of the STD was ≤10%.	Dabigatran: n=8908 Rivaroxaban: n=1405;	Rivaroxaban and dabigatran had similar stroke rates. Bleeding and mortality rates were higher in rivaroxaban vs dabigatran.

CHA2DS2-Vasc, Congestive heart failure, Hypertension, Age ≥75 years, Diabetes Mellitus, Prior Stroke or TIA or thromboembolism, Vascular disease, Age 65–74 years, Sex; NOACs, non-vitamin K antagonist oral anticoagulants; NVAF, non-valvular atrial fibrillation; OAC, oral anticoagulants.

Table 6 Main differences between the included studies (n=39)

Study item	Range, total number of studies or description
Country	USA: n=24 Denmark: n=5 Taiwan: n=4 China: n=1 France: n=1 Scotland: n=1 Sweden: n=1 Spain: n=1 UK: n=1
NOAC included in included studies	Dabigatran: n=39 Rivaroxaban: n=39 Apixaban: n=26 Edoxaban: n=1
Most prescribed NOAC in included studies per country	Dabigatran: Denmark Rivaroxaban: USA, UK, China, Scotland and Taiwan Apixaban: In none of the included studies Edoxaban: In none of the included studies About equal*: France, Spain, Sweden
Baseline characteristics	Mean age, years: 65–84 % males: 39–73 Mean CHA2DS2-VASc: 2.1–4.9
Primary study outcomes	Effectiveness outcomes: ► Stroke ► Systemic embolism or composite of stroke/systemic embolism ► All-cause death ► Myocardial infarction ► Venous thromboembolism Safety outcomes: ► Major bleeding ► A specific type of bleeding (eg, intracranial haemorrhage, gastrointestinal bleeding etc) ► Liver injury
Statistical approaches	PS matching: n=18 IPTW: n=8 PS stratification: n=1 Cox PH regression model: n=10 Unadjusted analyses: n=2
Sample size	n=698–265 583
Study results	Of the 26 studies in which apixaban, rivaroxaban and dabigatran were included: ► Apixaban was favourable compared with dabigatran and rivaroxaban: n=13 ► No single favourable NOAC: n=13

Continued

Table 6 Continued

Study item	Range, total number of studies or description

*About equal distribution between dabigatran, rivaroxaban and apixaban. Edoxaban is not included in these studies.
CHA2DS2-Vasc, Congestive heart failure, Hypertension, Age ≥75 years, Diabetes Mellitus, Prior Stroke or TIA or thromboembolism, Vascular disease, Age 65–74 years, Sex; NOACs, non-vitamin K antagonist oral anticoagulants.

Selection of covariates

Most studies (n=34) did not provide a rationale for the selection of covariates that were included in the PS model or in adjusted analysis. However, in one of the articles, an extensive rationale and selection procedure of covariates that were included in the analysis was provided.[33] In three other studies, the authors selected covariates based on medical knowledge on risk factors with reference to earlier published studies.[31 39 52] In one other study, it was reported that sociodemographic and clinical characteristics that were associated with treatment initiation and the risk of major bleeding were included in the model to adjust for differences across cohorts, without further explanation or reference.[30]

Definition of primary study outcomes

Primary outcomes differed between the studies. Effectiveness outcomes included in the studies included stroke, systemic embolism (or composite of stroke/systemic embolism), all-cause death, myocardial infarction, venous thromboembolism and safety outcomes included major bleeding, or a specific type of bleeding (eg, intracranial haemorrhage, gastrointestinal bleeding) and liver injury. In most studies, ICD-9 or ICD-10 codes were used, but whether this concerned a primary diagnosis only or whether it could be either a primary or a second diagnosis differed between the studies. In some studies, it was not described whether the ICD codes referred to primary diagnosis only or to a primary or secondary diagnosis.

Statistical approaches to adjust for confounding (primary analysis)

In 18 studies, PS matching was done.[16 19–21 23 26 29 30 32–37 39 40 47 49] IPTW was used in eight studies.[17 22 24 25 28 43 46 48] PS-stratified analyses was done in one study.[41] In 12 studies, the primary analyses used a Cox PH regression model in which adjustment for confounding was done.[18 27 31 38 42 44 45 50–52] Finally, in two studies no adjustment for differences in baseline characteristics was performed.[53 54]

PS matching

Covariates

Creatinine clearance was not included as a covariate in any of the 18 studies. All 18 studies took the following covariates into account: age, sex, CHA2DS2-VASc score and/or the individual comorbidities included in this score, HAS-BLED score (Hypertension, Abnormal renal and liver function, Stroke, Bleeding, Labile INR, Elderly, Drugs or alcohol) and/or the individual conditions included in this score (except alcohol use in Lai *et al*[47]), renal disease and co-medication use such as antiplatelets. Some included other comorbidities, such as cancer, rheumatic disease, specific heart diseases, Chronic Obstructive Pulmonary Disease (COPD), HIV, dementia, depression, neurological disorders and/or a various list of co-medications as well.

Matching method

In one study, the matching method was not described.[49] In two studies, the calliper used was not described.[23 29] In seven studies, 1:1 PS matching without replacement was used and a calliper of 0.01 was applied.[16 19 20 26 30 32 36] Five other studies also matched 1:1 without replacement but used another calliper: in three studies, a calliper of 0.2 was used,[39 40 47] while two others used a calliper of <0.25.[33 35] In three studies, three-way matching was used.[21 34 37]

Balance covariates

In two studies, it was not described how the balance between covariates was evaluated.[33 35] In two studies, the balance was evaluated using p<0.05 (of which one also used standardised difference of <10%),[23 47] and in another study, it was stated that the groups were comparable even though a p value of >0.05 was found.[29] Balance was checked with an absolute standardised difference of <10% in 13 studies.[16 19–21 26 30 32 34 36 37 39 40 47 49] Balance was reached in all studies after matching.

Sample size

In four studies, the sample size before matching was not reported,[29 35 36 39] and in one study, the sample size after matching was not reported.[34] At study start (before PSM), sample size between the NOACs differed greatly, except in three studies.[21 37 40]

IPTW

In one study, balance was tested using analysis of variance (ANOVAs) for significant differences.[22] Balance was checked with an absolute standardised difference of <10% in the other nine studies.[17 24 25 28 43 46 48] Balance was reached in all studies after IPTW.

There was no reporting on extreme weights in the eight included studies.[17 22 24 25 28 43 46 48]

PS stratification

In one study, asymmetric trimming of the PS was done, which resulted in a small part of both treatment groups being removed in order to gain in comparability. Balance in covariates was reached with standardised difference of <10%. In a Cox model, this trimmed PS was used in 10 deciles as strata.[41]

Cox HP regression models

In 10 studies, Cox HP regression models were applied with adjustment for a number of confounders.[18 27 31 38 42 44 45 50–52] In one of these studies, the number of events per variable was not sufficient for such an analysis.[50] The ratio was

acceptable in the other studies for at least some of the outcomes.[18 28 31 38 42 44 45 51 52]

Unadjusted analysis

In two studies, no adjustment for confounding factors seemed to have been done, even though significant differences between treatment groups existed at baseline. Cerdá *et al*[53] presented events per 100 patient-years and used a log-rank test to determine whether outcomes differed between the NOACs. Li *et al*[54] conducted a Cox proportional hazard model, likely unadjusted, but this was not clearly described in the article.

Sensitivity analyses

Although in some articles sensitivity analyses were done, none of the included studies further explored the magnitude of residual confounding in their sensitivity analyses using one of the approaches recommended by IPSOR (see the Methods section).

Study results

Which NOAC performed best differed between the included studies. We found only one study that included all four NOACs, in which no preference for one specific NOAC was found, except that rates of major bleeding were lower with rivaroxaban.[53] Of the 26 studies in which apixaban, rivaroxaban and dabigatran were included, apixaban was favourable compared with dabigatran and rivaroxaban in 13 studies, of which 10 were from the USA, 2 from Europe and 1 from Asia,[16 17 19 20 23 26 28 29 32 36 42 50 52] while dabigatran and rivaroxaban were not found to be the single most favourable NOAC in any of the remaining 13 studies. Results for these 13 studies were mixed, with either no favourable NOAC at all or one NOAC was selected as the least favourable, while the other two NOACs did not differ.

Naïve trial analysis

The primary efficacy endpoint (strokes/SE) in the warfarin arms were estimated at 1.69% (RE-LY),[3] 2.2% (ROCKET),[6] 1.60% (ARISTOTLE)[5] and 1.50% (ENGAGE)[4] (see table 7). From this range, we chose a relatively arbitrary base rate of 1.6% and applied the observed risk reduction to estimate comparable base rates of 1.05% for dabigatran, 1.24% for rivaroxaban, 1.26% for edoxaban and 1.27% for apixaban. Using the sample size calculator,[55] the biggest expected difference was between dabigatran and apixaban, and it was estimated that a trial sample size with 51 847 patients would be needed to confirm this difference. The smallest difference was between edoxaban and apixaban, and a trial of 7 994 340 patients is required to confirm that difference.

The primary safety endpoint was major bleeding for RE-LY, ARISTOTLE, and ENGAGE AF and major bleeding plus clinically relevant non-major bleeding for ROCKET AF, but data on major bleeds only for ROCKET-AF are available as well. Major bleeds in the warfarin arms were estimated at 3.36% (RE-LY),[3] 3.4% (ROCKET),[6] 3.09% (ARISTOTLE)[5] and 3.43% (ENGAGE).[4] From this range,

Table 7 Primary efficacy and safety endpoints of the four pivotal trials

	RE-LY[3]			ROCKET-AF[6]		ARISTOTLE[5]			ENGAGE-AF[4]		
	Dabigatran 150 mg n=6076	Dabigatran 110 mg n=6015	Warfarin n=6022	Rivaroxaban n=7131	Warfarin n=7133	Apixaban n=9120	Warfarin n=9081	Edoxaban 60 mg n=7035	Edoxaban 30 mg n=7034	Warfarin n=7036
Stroke/SE (%/year)	1.11	1.53	1.69	1.7	2.2	1.27	1.60	1.18	1.61	1.50
Major bleeding (%/year)	3.11	2.71	3.36	3.6	3.4	2.13	3.09	2.75	1.61	3.43

Bunge EM, *et al*. *BMJ Open* 2021;**11**:e042024. doi:10.1136/bmjopen-2020-042024

we choose a relatively arbitrary base rate of 3.2% and applied the observed risk reduction to estimate comparable base rates of 2.21% for apixaban 2.57% for edoxaban, 2.96% for dabigatran and 3.29% for rivaroxaban. Using the sample size calculator,[55] the biggest expected difference was between rivaroxaban and apixaban, and it was estimated that a trial with 7196 patients would be needed to confirm this difference. A much smaller difference is between edoxaban and apixaban which would require a trial of 56 512 patients to confirm that difference.

DISCUSSION

In total, we found 39 studies directly comparing the effectiveness and/or safety of at least two NOACs in patients with NVAF. Three studies can be considered to be of low quality due to insufficiently described methods and/or small sample size.[50 53 54]

Even though the remaining studies could be considered of sufficient quality based on the technical aspects of the studies, there are some issues that can hamper the generalisability of the results. These issues concern residual confounding, the use of a smaller or broader calliper, differences in baseline characteristics between studies, channelling bias and change in treatment paradigm, and the high number of patients needed.

Balance in baseline characteristics between NOACs was checked with p values or a standardised difference of <10%. Balance was well at baseline in some studies, or was reached after PS matching or IWTP.[56] Even though some studies included over 40 covariates in their PS, in most studies, it was not described how the covariates were selected. The ISPOR Good Research Practices for Retrospective Database Analysis recommends to include all factors that are theoretically related to outcome or treatment selection, even if the relation is weak or statistically non-significant.[15] Directed acyclic graphs might be helpful as well.[57] And even though balance was reached for all of these variables, one should keep in mind that balance between unmeasured or unmeasurable factors cannot be assumed.[15] Therefore, due to the lack of randomisation, there is always a possibility of residual confounding. This possibility was acknowledged in all included studies, and all studies have largely the same missing covariates. Hardly any laboratory results and lifestyle information were included, such as body mass index, smoking status and alcohol consumption, which are also risk factors for ischaemic stroke and bleeding events, respectively. Creatinine clearance, for instance, seems to be an important covariate as subgroup analyses from the pivotal trials suggest that renal clearance might be an effect modifier.[5 58] Only in one study, however, the authors were able to take renal clearance into account in the adjusted analyses.[50] Especially when prescription of a certain NOAC in daily practice is driven by creatinine clearance, not adjusting for this variable may lead to biased results. However, it is unknown what the magnitude and direction (ie, will the differences in effectiveness and safety between NOACs be smaller or larger) of this potential bias due to lack of randomisation would be. The magnitude of residual confounding was not further explored in the sensitivity of the included studies.

In general, a calliper of <0.2 of the SD of the logit of the PS is considered to be 'optimal'.[59] About half of the included PS matching studies used a smaller calliper, namely, of <0.1. This means that the matching is more precise in these studies, but the disadvantage is that possibly more patients cannot be matched to another patient due to this smaller allowed maximum differences, and thus will be excluded from the analysis. Excluding patients from the analysis will limit the generalisability of the results to the total patient population, especially when the excluded patients differ from the included patients, for example, on the baseline risk for stroke.

All included studies focused on patients with NVAF only. In eight studies, inclusion criteria regarding age were applied. Three of these will likely still cover the largest part of NOAC users as they set relatively broad age ranges. The other five focused on an elderly population of patients with NVAF aged ≥65 years. Besides applying specific inclusion criteria regarding age in some studies, these differences also depended on the specific registry or database that was used, for example, Medicare is for people of 65 years old or older. Even though only five of the included studies focused on an elderly NVAF population, and the others applied broad age ranges, there were differences in mean age, proportion of males and mean CHA2DS2-VASc score between the studies, which can have an impact on the results and jeopardise the generalisability of the results.

Rivaroxaban was the most prescribed NOAC in almost all included studies from the USA. However, in the first quarter of 2017, apixaban was the most prescribed NOAC in NVAF in the USA (ie, in 50% of new OAC prescriptions). Especially older patients, women, increased stroke or bleeding risk and having comorbidities was associated with prescription of apixaban versus other NOACs.[60] Rivaroxaban was also the most prescribed NOAC in the included studies from the UK and Scotland. Based on the Clinical Practice Research Datalink (CPRD), 56.5% of the OAC prescriptions concerned a NOAC, of which rivaroxaban was still described most often in 2015.[61] Dabigatran was described most often in the studies from Denmark. Haastrup *et al* described that most patients with AF that initiated NOAC received dabigatran between 2008 and 2016, but a trend was observed that per 1000 person-years the number of patients described dabigatran decreased and the number of patients receiving rivaroxaban and apixaban increased.[62] This shows that the treatment paradigm changed over time, and might still be changing, and this pattern differs between the USA, Europe and Asia. Channelling bias therefore likely occurs and might shift between the NOACs. Although in a few studies it was mentioned that selective prescriptions were noticed and that these might have changed over time, none of the

included studies dealt with temporal trends in prescription patterns.

Our naïve analysis predicts that in terms of the primary efficacy outcome, observational studies will need a relatively high number of patients to be able to demonstrate the differences between the NOACs and a small sample size will not allow robust comparison to be made.

The pattern of major bleeding events seen in the included observational studies confirms the expectation from our naïve analysis of the pivotal clinical trials that rivaroxaban seems to have the least favourable safety profile among apixaban and dabigatran. The findings are not consistent to allow for a robust conclusion between apixaban and dabigatran which confirms the need for a high number of patients, although a trend for a slight better safety profile of apixaban can be observed.

The requirement for a high number of patients to compare NOACs both in terms of efficacy and safety as predicted by the pivotal trial results is confirmed by the findings of the observational studies. This finding may support the claim that the differences between the NOACs are relatively small.

In the process of conducting systematic reviews, it is inevitable that the review will never be completely up to date with the most recent published evidence. Even though our search ended in April 2019, recently published studies will have encountered the same issues as described above. Residual confounding and channelling bias cannot have been ruled out in newer publications. Ideally, head-to-head trials should be conducted to compare the efficacy/effectiveness and safety of the four NOACs to overcome the methodological issues in the comparative effectiveness studies. To our knowledge, one head-to-head trial including all four NOACs is currently running. This nationwide cluster randomised cross-over study aims to compare efficacy and safety of the four NOACs (clinicaltrials.gov; NCT03129490).

In conclusion, even though the larger part of these studies are conducted as well as possible considering what data are available, there are some important limitations regarding the generalisability of the study results especially given the relatively high patient number required for a meaningful comparison between NOACs. Most studies included all patients with NVAF on NOAC available in the registry/database during the study period and did not apply further specific inclusion and exclusion criteria, but differences between studies regarding baseline characteristics existed. Mean age at study start and baseline risk for stroke (CHA2DS2-VASc score) differed between the studies. As channelling bias cannot be ruled out, the result of these studies might not be generalisable. Furthermore, results from the PS studies are only applicable to the patients that were kept in the analyses as patients excluded from the analysis likely differ from the ones that were included in the analysis. The 1:1 matched cohorts depended on the sample size of the NOAC with the least number of patients and as a result many patients from the larger of the two NOAC groups were excluded as they could not be matched. In clinical practice, these limitations should be kept in mind when results of these studies are used to decide what NOAC should be prescribed for a certain patient. Given the small differences between efficacy and safety outcomes between NOACs, the element of patient preference should be taken into consideration,[63] as tailoring anticoagulation treatment towards patient preferences can promote adherence to treatment.

Acknowledgements The authors would like to thank Pearl Gumbs for the initiation of this project and her input with regard to study design and interpretation of the data. Pearl Gumbs was working at Daiichi Sankyo Europe at that time.

Contributors EMB: conceptualisation (support); methodology (equal); writing—original draft preparation; writing—review and editing (equal). BAvH: conceptualisation (support); methodology (equal); writing—review and editing (equal). SH: conceptualisation (support); writing—review and editing (equal). GS: conceptualisation (support), supervision; writing—review and editing (equal). AC: conceptualisation (lead); writing—review and editing (equal).

Funding This work was supported by Daiichi Sankyo Europe.

Competing interests EMB reports grants from Daiichi Sankyo during the conduct of the study; grants from Daiichi Sankyo, outside the submitted work. BAvH reports grants from Daiichi Sankyo during the conduct of the study. SH reports personal fees from Aspen, personal fees from Bayer, personal fees from BMS/Pfizer, personal fees from Daiichi-Sankyo, personal fees from Portola, outside the submitted work. GS reports personal fees from Daiichi Sankyo Europe GmbH, outside the submitted work. AC reports personal fees from Daiichi Sankyo Europe during the conduct of the study; grants and personal fees from Bayer AG, personal fees from Boehringer Ingelheim, grants and personal fees from Bristol-Myers Squibb, grants and personal fees from Pfizer Limited, personal fees from Portola Pharmaceuticals, personal fees from Janssen, personal fees from ONO Pharmaceuticals, from AbbVie, outside the submitted work.

Patient and public involvement statement This research was done without patient involvement. Patients were not invited to comment on the study design and were not consulted to develop patient relevant outcomes or interpret the results. Patients were not invited to contribute to the writing or editing of this document for readability or accuracy.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information.

ORCID iD

Eveline M Bunge http://orcid.org/0000-0001-7988-909X

REFERENCES

1 Kirchhof P, Benussi S, Kotecha D, *et al*. 2016 ESC guidelines for the management of atrial fibrillation developed in collaboration with EACTS. *Eur Heart J* 2016;37:2893–962.
2 January CT, Wann LS, Calkins H, *et al*. 2019 AHA/ACC/HRS focused update of the 2014 AHA/ACC/HRS guideline for the

management of patients with atrial fibrillation: a report of the American College of Cardiology/American heart association Task force on clinical practice guidelines and the heart rhythm society in collaboration with the Society of thoracic surgeons. *Circulation* 2019;140:e125–51.
3 Connolly SJ, Ezekowitz MD, Yusuf S, *et al*. Dabigatran versus warfarin in patients with atrial fibrillation. *N Engl J Med* 2009;361:1139–51.
4 Giugliano RP, Ruff CT, Braunwald E, *et al*. Edoxaban versus warfarin in patients with atrial fibrillation. *N Engl J Med* 2013;369:2093–104.
5 Granger CB, Alexander JH, McMurray JJV, *et al*. Apixaban versus warfarin in patients with atrial fibrillation. *N Engl J Med* 2011;365:981–92.
6 Patel MR, Mahaffey KW, Garg J, *et al*. Rivaroxaban versus warfarin in nonvalvular atrial fibrillation. *N Engl J Med* 2011;365:883–91.
7 Petri H, Urquhart J. Channeling bias in the interpretation of drug effects. *Stat Med* 1991;10:577–81.
8 Peduzzi P, Concato J, Feinstein AR, *et al*. Importance of events per independent variable in proportional hazards regression analysis. II. accuracy and precision of regression estimates. *J Clin Epidemiol* 1995;48:1503–10.
9 Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983;70:41–55.
10 Brookhart MA, Schneeweiss S, Rothman KJ, *et al*. Variable selection for propensity score models. *Am J Epidemiol* 2006;163:1149–56.
11 D'Agostino RB. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med* 1998;17:2265–81.
12 Austin PC, Stuart EA. 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. *Stat Methods Med Res* 2017;26:1654–70.
13 Yao XI, Wang X, Speicher PJ, *et al*. Reporting and guidelines in propensity score analysis: a systematic review of cancer and cancer surgical studies. *J Natl Cancer Inst* 2017;109. doi:10.1093/jnci/djw323. [Epub ahead of print: 01 08 2017].
14 Austin PC. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Stat Med* 2008;27:2037–49.
15 Johnson ML, Crown W, Martin BC, *et al*. Good research practices for comparative effectiveness research: analytic methods to improve causal inference from nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report--Part III. *Value Health* 2009;12:1062–73.
16 Abraham NS, Noseworthy PA, Yao X, *et al*. Gastrointestinal safety of direct oral anticoagulants: a large population-based study. *Gastroenterology* 2017;152:1014–22.
17 Adeboyeje G, Sylwestrzak G, Barron JJ, *et al*. Major bleeding risk during anticoagulation with warfarin, dabigatran, apixaban, or rivaroxaban in patients with nonvalvular atrial fibrillation. *J Manag Care Spec Pharm* 2017;23:968–78.
18 Alonso A, MacLehose RF, Chen LY, *et al*. Prospective study of oral anticoagulants and risk of liver injury in patients with atrial fibrillation. *Heart* 2017;103:834–9.
19 Amin A, Keshishian A, Trocio J, *et al*. A real-world observational study of hospitalization and health care costs among nonvalvular atrial fibrillation patients prescribed oral anticoagulants in the U.S. Medicare population. *J Manag Care Spec Pharm* 2018;24:911–20.
20 Amin A, Keshishian A, Vo L, *et al*. Real-World comparison of all-cause hospitalizations, hospitalizations due to stroke and major bleeding, and costs for non-valvular atrial fibrillation patients prescribed oral anticoagulants in a US health plan. *J Med Econ* 2018;21:244–53.
21 Briasoulis A, Inampudi C, Akintoye E, *et al*. Safety and efficacy of novel oral anticoagulants versus warfarin in Medicare beneficiaries with atrial fibrillation and valvular heart disease. *J Am Heart Assoc* 2018;7. doi:10.1161/JAHA.118.008773. [Epub ahead of print: 05 04 2018].
22 Charlton B, Adeboyeje G, Barron JJ, *et al*. Length of hospitalization and mortality for bleeding during treatment with warfarin, dabigatran, or rivaroxaban. *PLoS One* 2018;13:e0193912.
23 Deitelzweig S, Luo X, Gupta K, *et al*. Comparison of effectiveness and safety of treatment with apixaban vs. other oral anticoagulants among elderly nonvalvular atrial fibrillation patients. *Curr Med Res Opin* 2017;33:1745–54.
24 Graham DJ, Baro E, Zhang R, *et al*. Comparative stroke, bleeding, and mortality risks in older Medicare patients treated with oral anticoagulants for nonvalvular atrial fibrillation. *Am J Med* 2019;132:596-604.e11.
25 Graham DJ, Reichman ME, Wernecke M, *et al*. Stroke, bleeding, and mortality risks in elderly Medicare beneficiaries treated with dabigatran or rivaroxaban for nonvalvular atrial fibrillation. *JAMA Intern Med* 2016;176:1662–71.
26 Gupta K, Trocio J, Keshishian A, *et al*. Real-World comparative effectiveness, safety, and health care costs of oral anticoagulants in nonvalvular atrial fibrillation patients in the U.S. department of defense population. *J Manag Care Spec Pharm* 2018;24:1116–27.
27 Hernandez I, Zhang Y, Saba S. Comparison of the effectiveness and safety of apixaban, dabigatran, rivaroxaban, and warfarin in newly diagnosed atrial fibrillation. *Am J Cardiol* 2017;120:1813–9.
28 Hernandez I, Zhang Y. Comparing stroke and bleeding with rivaroxaban and dabigatran in atrial fibrillation: analysis of the US Medicare Part D data. *Am J Cardiovasc Drugs* 2017;17:37–47.
29 Lin J, Trocio J, Gupta K, *et al*. Major bleeding risk and healthcare economic outcomes of non-valvular atrial fibrillation patients newly-initiated with oral anticoagulant therapy in the real-world setting. *J Med Econ* 2017;20:952–61.
30 Lip GYH, Keshishian A, Kamble S, *et al*. Real-World comparison of major bleeding risk among non-valvular atrial fibrillation patients initiated on apixaban, dabigatran, rivaroxaban, or warfarin. A propensity score matched analysis. *Thromb Haemost* 2016;116:975–86.
31 Lip GYH, Pan X, Kamble S, *et al*. Major bleeding risk among non-valvular atrial fibrillation patients initiated on apixaban, dabigatran, rivaroxaban or warfarin: a "real-world" observational study in the United States. *Int J Clin Pract* 2016;70:752–63.
32 Lip GYH, Keshishian A, Li X, *et al*. Effectiveness and safety of oral anticoagulants among nonvalvular atrial fibrillation patients. *Stroke* 2018;49:2933–44.
33 Lutsey PL, Norby FL, Zakai NA, *et al*. Oral anticoagulation therapy and subsequent risk of venous thromboembolism in atrial fibrillation patients. *Curr Med Res Opin* 2019;35:837–45.
34 Mentias A, Shantha G, Chaudhury P, *et al*. Assessment of outcomes of treatment with oral anticoagulants in patients with atrial fibrillation and multiple chronic conditions: a comparative effectiveness analysis. *JAMA Netw Open* 2018;1:e182870.
35 Norby FL, Bengtson LGS, Lutsey PL, *et al*. Comparative effectiveness of rivaroxaban versus warfarin or dabigatran for the treatment of patients with non-valvular atrial fibrillation. *BMC Cardiovasc Disord* 2017;17:238.
36 Noseworthy PA, Yao X, Abraham NS, *et al*. Direct comparison of dabigatran, rivaroxaban, and apixaban for effectiveness and safety in nonvalvular atrial fibrillation. *Chest* 2016;150:1302–12.
37 Palamaner Subash Shantha G, Bhave PD, Girotra S, *et al*. Sex-Specific comparative effectiveness of oral anticoagulants in elderly patients with newly diagnosed atrial fibrillation. *Circ Cardiovasc Qual Outcomes* 2017;10.
38 Tepper PG, Mardekian J, Masseria C, *et al*. Real-World comparison of bleeding risks among non-valvular atrial fibrillation patients prescribed apixaban, dabigatran, or rivaroxaban. *PLoS One* 2018;13:e0205989.
39 Villines TC, Ahmad A, Petrini M, *et al*. Comparative safety and effectiveness of dabigatran vs. rivaroxaban and apixaban in patients with non-valvular atrial fibrillation: a retrospective study from a large healthcare system. *Eur Heart J Cardiovasc Pharmacother* 2019;5:80–90.
40 Andersson NW, Svanström H, Lund M, *et al*. Comparative effectiveness and safety of apixaban, dabigatran, and rivaroxaban in patients with non-valvular atrial fibrillation. *Int J Cardiol* 2018;268:113–9.
41 Gorst-Rasmussen A, Lip GYH, Bjerregaard Larsen T. Rivaroxaban versus warfarin and dabigatran in atrial fibrillation: comparative effectiveness and safety in Danish routine care. *Pharmacoepidemiol Drug Saf* 2016;25:1236–44.
42 Lamberts M, Staerk L, Olesen JB, *et al*. Major bleeding complications and persistence with oral anticoagulation in Non-Valvular atrial fibrillation: contemporary findings in real-life Danish patients. *J Am Heart Assoc* 2017;6. doi:10.1161/JAHA.116.004517. [Epub ahead of print: 14 02 2017].
43 Larsen TB, Skjøth F, Nielsen PB, *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:i3189.
44 Staerk L, Gerds TA, Lip GYH, *et al*. Standard and reduced doses of dabigatran, rivaroxaban and apixaban for stroke prevention in atrial fibrillation: a nationwide cohort study. *J Intern Med* 2018;283:45–55.
45 Chan Y-H, Kuo C-T, Yeh Y-H, *et al*. Thromboembolic, Bleeding, and Mortality Risks of Rivaroxaban and Dabigatran in Asians With Nonvalvular Atrial Fibrillation. *J Am Coll Cardiol* 2016;68:1389–401.

46 Chan Y-H, See L-C, Tu H-T, *et al*. Efficacy and safety of apixaban, dabigatran, rivaroxaban, and warfarin in Asians with nonvalvular atrial fibrillation. *J Am Heart Assoc* 2018;7. doi:10.1161/JAHA.117.008150. [Epub ahead of print: 05 04 2018].
47 Lai CL, Chen HM, Liao MT. Comparative effectiveness and safety of dabigatran and rivaroxaban in atrial fibrillation patients. *J Am Heart Assoc* 2017;6.
48 Meng S-W, Lin T-T, Liao M-T, *et al*. Direct comparison of low-dose dabigatran and rivaroxaban for effectiveness and safety in patients with Non-Valvular atrial fibrillation. *Acta Cardiol Sin* 2019;35:42–54.
49 Blin P, Dureau-Pournin C, Cottin Y, *et al*. Comparative effectiveness and safety of standard or reduced dose dabigatran vs. rivaroxaban in nonvalvular atrial fibrillation. *Clin. Pharmacol. Ther.* 2019;105:1439–55. doi:10.1002/cpt.1318
50 Al-Khalili F, Lindström C, Benson L. The safety and persistence of non-vitamin-K-antagonist oral anticoagulants in atrial fibrillation patients treated in a well structured atrial fibrillation clinic. *Curr Med Res Opin* 2016;32:779–85.
51 Mueller T, Alvarez-Madrazo S, Robertson C, *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;85:422–31.
52 Vinogradova Y, Coupland C, Hill T, *et al*. Risks and benefits of direct oral anticoagulants versus warfarin in a real world setting: cohort study in primary care. *BMJ* 2018;362:k2505.
53 Cerdá M, Cerezo-Manchado JJ, Johansson E, *et al*. Facing real-life with direct oral anticoagulants in patients with nonvalvular atrial fibrillation: outcomes from the first observational and prospective study in a Spanish population. *J Comp Eff Res* 2019;8:165–78.
54 Li W-H, Huang D, Chiang C-E, *et al*. Efficacy and safety of dabigatran, rivaroxaban, and warfarin for stroke prevention in Chinese patients with atrial fibrillation: the Hong Kong atrial fibrillation project. *Clin Cardiol* 2017;40:222–9.
55 ClinCalc. Sample size calculator, 2021. Available: https://clincalc.com/stats/samplesize.aspx
56 Ho DE, Imai K, King G, *et al*. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis* 2007;15:199–236. doi:10.1093/pan/mpl013
57 Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology* 1999;10:37–48.
58 Bohula EA, Giugliano RP, Ruff CT, *et al*. Impact of renal function on outcomes with edoxaban in the engage AF-TIMI 48 trial. *Circulation* 2016;134:24–36.
59 Austin PC, Stuart EA. Estimating the effect of treatment on binary outcomes using full matching on the propensity score. *Stat Methods Med Res* 2017;26:2505–25.
60 Zhu J, Alexander GC, Nazarian S, *et al*. Trends and variation in oral anticoagulant choice in patients with atrial fibrillation, 2010-2017. *Pharmacotherapy* 2018;38:907–20.
61 Loo SY, Dell'Aniello S, Huiart L, *et al*. Trends in the prescription of novel oral anticoagulants in UK primary care. *Br J Clin Pharmacol* 2017;83:2096–106.
62 Haastrup SB, Hellfritzsch M, Rasmussen L, *et al*. Use of non-vitamin K antagonist oral anticoagulants 2008-2016: a Danish nationwide cohort study. *Basic Clin Pharmacol Toxicol* 2018;123:452–63.
63 Vaanholt MCW, Weernink MGM, von Birgelen C, *et al*. Perceived advantages and disadvantages of oral anticoagulants, and the trade-offs patients make in choosing anticoagulant therapy and adhering to their drug regimen. *Patient Educ Couns* 2018;101:1982–9.