

BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

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

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2020-042024
Article Type:	Original research
Date Submitted by the Author:	24-Jun-2020
Complete List of Authors:	Bunge, Eveline; Pallas health research and consultancy van Hout, Ben; University of Sheffield, HEDS, SchARR Haas, Sylvia ; Formerly Technical University Of Munich Spentzouris, Georgios; Daiichi Sankyo Europe GmbH Cohen, Alexander; Guys and St Thomas' NHS Foundation Trust, King's College
Keywords:	Adult cardiology < CARDIOLOGY, STATISTICS & RESEARCH METHODS, Cardiology < INTERNAL MEDICINE

SCHOLARONE™
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1
2
3 **CRITICAL APPRAISAL AND ISSUES REGARDING GENERALISABILITY OF COMPARATIVE**
4 **EFFECTIVENESS STUDIES OF NOACS IN ATRIAL FIBRILLATION AND THEIR RELATION TO CLINICAL**
5 **TRIAL DATA - A SYSTEMATIC REVIEW**
6
7
8
9

10 **Eveline M. Bunge¹, Ben van Hout², Sylvia Haas³, Georgios Spentzouris⁴, Alexander T. Cohen⁵**
11

12
13 ¹ Pallas, health research and consultancy BV, P.O. Box 21238, 3001 AE Rotterdam, the Netherlands
14

15
16 ² University of Sheffield, HEDS, ScHARR, Sheffield, UK
17

18
19 ³ Formerly Technical University of Munich, Munich, Germany
20

21
22 ⁴ Daiichi Sankyo Europe, Munich, Germany
23

24
25 ⁵ Guys and St Thomas' NHS Foundation Trust, King's College London, London, UK
26

27 * Corresponding author:
28

29 bunge@pallashrc.com
30

31 Word count: 4196
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

KEY WORDS

atrial fibrillation; non-vitamin K antagonist oral anticoagulants (NOAC); comparative effectiveness research;
systematic review

For peer review only

ABSTRACT

Objective: To critically appraise the published comparative effectiveness studies on non-vitamin K antagonist oral anticoagulants (NOAC) in nonvalvular atrial fibrillation (NVAf).

Materials and Methods: We performed a systematic literature review in Medline and EMBase to investigate the way comparisons were made. Results were also compared with expectations formulated on the basis of trial results with specific attention to the patient years in each study.

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, eight studies weighted patients based on the inverse probability of treatment, one study used propensity score stratification and ten studies applied a proportional hazards model. These studies have some important limitations, 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 7,700 and 59,500 patients are needed to confirm the observed differences in bleedings and between 51,800 and 7,994,300 to confirm differences in efficacy.

Conclusion: Meaningful comparisons 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 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 NOACs in atrial fibrillation patients 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.

For peer review only

INTRODUCTION

Guidelines state a preference for non-vitamin K antagonist oral anticoagulants (NOACs) above vitamin K antagonists (VKA) in patients with nonvalvular 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; i.e. 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 channeling 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. Cox proportional hazards (Cox PH) regression model adjustment can be used but large sample sizes are needed when 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 randomized clinical trial, one needs substantial numbers 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 propensity score is the probability of an individual receiving a specific treatment given a specific set of patient characteristics (e.g. age, gender, comorbidities).[9] Variables related to the outcome should be included in the propensity score 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

1
2
3 be matched to another patient will be excluded from the analyses, and in IPTW, when patients on one treatment
4 have a low propensity score and patients treated with the other treatment have a high propensity score, extreme
5 weights can occur which can bias the results.[12]
6
7
8
9

10 To gain more understanding in how the above described methodologies were applied in peer-reviewed CER on
11 effectiveness and safety in NOACs in NVAF patients, we conducted a systematic literature review. Within this we
12 compare the results with those from a naive analysis of the results of the four major trial for rivaroxaban
13 (Rivaroxaban Once Daily Oral Direct Factor Xa Inhibition Compared with Vitamin K Antagonism for Prevention of
14 Stroke and Embolism Trial in Atrial Fibrillation (ROCKET AF)), apixaban (Apixaban for Reduction in Stroke and
15 Other Thromboembolic Events in Atrial Fibrillation (ARISTOTLE)), dabigatran (Randomized Evaluation of Long-
16 Term Anticoagulation Therapy (RE-LY)) and edoxaban (Effective Anticoagulation with Factor Xa Next Generation
17 in Atrial Fibrillation–Thrombolysis in Myocardial Infarction 48 (ENGAGE AF)) and compare the results from the
18 various analyses with those from the trials.
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

METHODS

Information sources, search strategy and eligibility criteria

We performed a systematic literature review to identify peer-reviewed comparative effectiveness research 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 appendix 1 for the search strings). The search was conducted on 23-04-2019 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, in- and exclusion criteria and the following baseline characteristics: age, proportion males, CHA2DS2Vasc score and comorbidity index.

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

- The variables included in the propensity score model
- Explanation of the variable selection procedure for propensity score model
- Distribution of baseline characteristics for each group before propensity score analysis
- In case of PSM:
 - matching ratio,
 - distance metric,

- 1
- 2
- 3
 - with or without replacement,
- 4
 - comparability of baseline characteristics in the matched groups,
- 5
 - sample size before and after matching
- 6
- 7
- 8
 - In case of IPTW:
- 9
 - comparability of baseline characteristics in the weighted groups
- 10
 - extreme weights
- 11
- 12
- 13
 - In case of PS stratification:
- 14
 - number of strata, comparability of baseline characteristics
- 15
- 16

17 In case of analyses in which no PS was used in the main analyses, we evaluated whether the ratio number of
18 covariates to the number of events seemed sufficient to produce valid results.[8]
19
20

21 22 23 **Naïve trial analysis**

24 Trials are quite often designed with a null hypothesis and associated with a power calculation while real world
25 studies are often dictated by the number of observations available. To give the results from the real-world-
26 evidence some perspective we undertook a naïve trial analysis in which the risk reductions from each trial with
27 respect to efficacy and safety outcomes were applied to an average number of outcomes observed in the warfarin
28 arms in each trial. This leads to an estimate of the relevant rates for each drug and the differences are illustrated
29 by the number of patients (sample size) needed in a randomised clinical trial to confirm the estimated differences.
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

RESULTS

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

For peer review only

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, 2017 USA	OptumLabs Data Warehouse Oct 1, 2010 through Feb 28, 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 caliper of 0.01; After PSM, standardized differences of all baseline characteristics were <10%.	rivaroxaban: n=19,301 dabigatran: n=17,426 apixaban: n=6,576	rivaroxaban vs. dabigatran: n=31,574 apixaban vs. rivaroxaban: n=13,130 apixaban vs. dabigatran: n=13,084 (more than 90% of original smallest samples size)	Apixaban had the most favorable gastrointestinal bleeding profile and rivaroxaban had the least favorable safety profile. Apixaban had the most favorable gastrointestinal safety profile among all age groups.
Amin, 2018 (J Manag Care Spec Pharm) USA	Medicare & Medicaid Services Jan 1, 2012, to Dec 31, 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-month before study start.	Age: 77.2±7.0-78.4±7.4 Male: 47.4-50.6% CHA2DS2-VASc: 4.4-4.6 CCI: 2.5-2.7	Hospitalization 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 caliper of 0.01; After PSM, standardized differences of all baseline characteristics were <10%.	rivaroxaban: n=53,146 apixaban: n=20,853 dabigatran: n=16,743	rivaroxaban vs. apixaban: n=41,608 dabigatran vs. apixaban: n=30,836 (more than 90% of original smallest samples size)	Apixaban was associated with significantly lower risks of all-cause, stroke/SE-related, and MB-related hospitalizations compared with dabigatran, and rivaroxaban
Amin 2018 (J Med Econ) USA	OptumInsight research database Jan 1, 2012 – Sept 30, 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-month before study start.	NR	Hospitalization 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 caliper of 0.01; After PSM, standardized differences of all baseline characteristics were <10%.	rivaroxaban: n=14,163 apixaban: n=8,652 dabigatran: n=3,684	apixaban vs. rivaroxaban: N=16,880 apixaban vs. dabigatran: N=7,114 (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, 2018 Denmark	National patient register, Register of Medicinal Product Statistics July 1, 2013 – March 31, 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 (i.e. intracranial bleeding, gastrointestinal bleeding (bleeding ulcer, hematemesis or melena) or other serious bleeding (anemia 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 caliper of 0. (replacement yes or no not reported) All baseline characteristics were well balanced after matching, except for calendar year	apixaban: n=4,292 dabigatran: n=3,913 rivaroxaban: n=3,805	apixaban vs. rivaroxaban: N=7,352 apixaban vs. dabigatran: N=6,470 rivaroxaban vs. dabigatran: N=5,440	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.

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, 2019 France	French nationwide claims and hospitalization 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 three 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	Hospitalization with a main diagnosis of ischemic stroke or systemic embolism or major bleeding and all-cause death. (ICD-10) codes	1 matched cohort PSM method not reported. After PSM, standardized 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, 2018 USA	Medicare and Medicaid Services Jan 1, 2010 - Dec 31, 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 ischemic stroke or transient ischemic 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; 3-way propensity matching (VKA was one of the groups, but not further discussed here) After PSM, standardized 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
Deitelzweig, 2017 USA	Humana Research Database (Medicare coverage) Jan 2013 - 30 Sept 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: 2.7-3.0	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>.05).	rivaroxaban: n=11,082 apixaban: n=8,250 dabigatran: n=2,474	apixaban vs. rivaroxaban: n=13,620 apixaban vs. dabigatran: n= 4,654	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, 2018 USA	Department of Defence data Jan 1, 2012, to Sept 30, 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, standardized differences of all baseline characteristics were <10%.	rivaroxaban: n=15,680 apixaban: n=11,754 dabigatran: n=4,312	rivaroxaban vs. apixaban: n=22,568 dabigatran vs. apixaban: n=8,258	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

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, 2017 Taiwan	National Health Insurance program 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 standardized difference	dabigatran: n=10,625; rivaroxaban: n=4,609	dabigatran vs. rivaroxaban: N=9,200	Rivaroxaban therapy was associated with a statistically significant increase in all-cause death compared with dabigatran
Lin, 2017 USA	IMS Pharmetrics Plus database Jan 2013 – Sept 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=8,124 apixaban vs. dabigatran: N=5,368	Apixaban is associated with reduced risk of hospitalisation compared with dabigatran and rivaroxaban.
Lip, 2016 (Thromb Haemost) USA	Truven MarketScan® Commercial Claims and Encounter and Medicare Supplemental and Coordination of Benefits Databases Jan 2012 to Dec 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, standardized differences of all baseline characteristics were <10%.	rivaroxaban: n=17,801 apixaban: n=7,438 dabigatran: n=4,661	apixaban vs dabigatran: n=14,798 rivaroxaban vs dabigatran: n=9,314 apixaban vs rivaroxaban: n=8,814	Compared to apixaban, rivaroxaban initiation was associated with significantly higher risk of major bleeding. The difference for dabigatran was not statistically significant
Lip, 2018 USA	Medicare and Medicaid Services Medicare; Truven MarketScan, IMS PharMetrics Plus Database, Optum Clinformatics Data Mart, and the Humana Research Database Jan 1, 2013, to Sept 30, 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	Hospitalizations 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, standardized 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.

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
Lutsey, 2018 USA	MarketScan Commercial Database Jan 1, 2010 through Sept 30, 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 caliper 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=6,200	Risk of VTE was lowest among those prescribed apixaban and dabigatran
Mentias, 2018 USA	Medicare & Medicaid Services Jan 1, 2010, to Dec 31, 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 ischemic stroke or major bleeding as defined by Rothendler ⁶ and Suh based on the primary ICD-9-CM diagnosis on inpatient standard analytical files claims for acute care stays.	1 cohort 3-way PSM. (VKA was one of the groups, but not further discussed here) After PSM, standardized 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
Norby, 2017 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	ischemic stroke (primary discharge), intracranial bleeding (primary discharge), myocardial infarction (1st or 2 nd 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 caliper 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, 2016 USA	Optum Labs Data Warehouse Oct 1, 2010 - Feb 28, 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 caliper of 0.01. A standardized 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
Shantha, 2017 USA	Medicare and Medicaid Nov 1, 2011 - Dec 31, 2013	Newly diagnosed NVAF patients and initiated OAC use.	Males: Age: 74.7±5.9-74.9±6. CHADS2-Vasc: 3.7-3.8 Gagne score: 2.9-2.9 Women: Age: 76.6±6.6- 76.9±6.6 CHADS2-Vasc: 4.8-4.9	inpatient admissions for acute ischemic 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)	rivaroxaban: n=23,177 dabigatran: n=21,979	dabigatran vs. rivaroxaban: n= 37,298	The reduced risk of ischemic stroke in patients taking rivaroxaban, compared with dabigatran, seems to be limited to men, whereas the higher risk of

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
			Gagne: 3.0-3.1		A standardized difference < 10% was considered acceptable			bleeding seems to be limited to women
Villines, 2019 USA	US Department of Defence Military Health System database 1 July 2010 to 30 June 2016 for the dabigatran vs. rivaroxaban cohort, and 28 Dec 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=9,604	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; Gagne: Gagne comorbidity score;

For peer review only

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, 2017 USA	HealthCore integrated research environment Nov 1, 2009 - Jan 31, 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	Hospitalization 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 $\leq 10\%$.	dabigatran: n=8,539 rivaroxaban: n=8,398 apixaban: n=3,689	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, 2018 Taiwan	Taiwan National Health Insurance Research June 1, 2012 - Dec 31, 2016	NVAF patients with their first prescription of OAC	Age: 75 \pm 10- 76 \pm 10 Male: 55-60% CHA2DS2-VASc: 3.7-3.9 Comorbidity index: NR	Hospitalization for ischemic stroke/systemic embolism, intracranial hemorrhage, 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 $\leq 10\%$.	rivaroxaban: n=27,777 dabigatran: n=20,079 apixaban: n=5,843	Three low-dose NOACs showed similar performance as without subgrouping
Charlton, 2018 USA	HealthCore Integrated Research Environment database Nov 1, 2010 - March 31, 2014	NVAF patients hospitalized for bleeding after starting OAC (AF diagnosis 6 months before starting one of the index drugs).	Age: 68.0 \pm 12.5- 69.6 \pm 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- 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-day
Graham, 2016 USA	Medicare Nov 4, 2011 - June 30, 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 y: 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 hospitalization 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 $\leq 10\%$.	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

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, 2019 USA	fee-for-service Medicare Part A (hospitalization), Part B (office-based care), and Part D (prescription drug coverage) Oct, 2010 - Sept, 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	Hospitalized 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.
Hernandez, 2017 USA	Medicare Nov 4, 2011 -Dec 31, 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	ischemic 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 hemorrhage and gastrointestinal bleeding, not further described.	Extreme weights: not reported Balanced if the absolute value of the STD was ≤10%.	dabigatran n=9,138 rivaroxaban n=8,367	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, 2016 Denmark	Danish national prescription registry, Danish national patient register, Danish civil registration system August, 2011 -Oct, 2015	NVAF patients who were naïve to oral anticoagulants (no use of oral anticoagulant within one year)	Age (median, IQR): 67.6 (62.0-72.4)-71.8 (65.7-78.9) Male: 56.9-66.1% CHA2DS2VASc: 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=7,192 apixaban: n=6,349	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
Meng, 2019 Taiwan	National Health Insurance claims database June 1, 2012 - May 31, 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, ischemic stroke, intracranial hemorrhage, gastrointestinal hemorrhage needing transfusion, ICD-10 codes, whether primary and	Extreme weights: not reported Balanced if the absolute value of the STD was ≤10%.	dabigatran: n=13,505 rivaroxaban: n=6,551 Weighted pseudo-cohort	Rivaroxaban seemed to be associated with an increased risk of all-cause death compared with dabigatran

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46

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
				secondary codes were used is not described		dabigatran: n=13,508; rivaroxaban: n=6,547	

For peer review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46

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-Khahili, 2016 Sweden	tertiary referral cardiology outpatient clinic (the Stockholm Heart Center) Dec, 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 rates owing mainly to the highest number of minor bleedings, and apixaban had the lowest bleeding rates and side effects
Alonso, 2017 USA	Truven Health MarketScan® Commercial Claims and Encounter Database and the Medicare Supplemental and Coordination of Benefits Database Jan 1, 2007 - Dec 31, 2014	NVAF patients with a first prescription of OAC after Nov 2, 2011.	Age: 67.2±12.4- 69.3±12.5 Male: 60.1-65.1% CHA2DS2-VASc: 2.9-3.6 Comorbidity index: NR	Hospitalization for liver injury potentially related to drug hepatotoxicity, ICD-9-CM codes in any position	rivaroxaban: n=30,347; dabigatran: n=17,286; apixaban: n=9,205	Risk of liver disease hospitalization was higher in rivaroxaban users compared to dabigatran and apixaban users
Chan, 2016 Taiwan	Taiwan National Health Insurance Research Database. Jan 1, 1996 - Dec 31, 2013	NVAF patients newly diagnosed	Age: 75±9- 76 ±9 Male: 54-58 CHA2DS2-VASc: 4.1-4.1 Comorbidity index: NR	Ischemic stroke or systemic embolism, ICH, hospitalization for GI bleeding, acute myocardial infarction (AMI), all hospitalizations 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= 5,921 rivaroxaban 10 mg: n=3,916	No differences were found between rivaroxaban and dabigatran in risk for thromboembolic events, intracranial haemorrhage, critical gastrointestinal bleeding, or all-cause mortality. However, rivaroxaban was associated with a higher risk for noncritical gastrointestinal bleeding than dabigatran

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
Hernandez, 2017 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	Ischemic stroke, death, bleeding events, gastrointestinal bleeding, treatment persistence. ICD-9 codes, whether primary and secondary codes were used is not described	rivaroxaban: n=5,139; apixaban: n=2,358; dabigatran: 1,415;	Apixaban had the most favourable effectiveness and safety profile
Lamberts, 2017 Denmark	Danish national patient registry, Danish national prescription registry, Danish civil personal registry up to December 31, 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=7,963; rivaroxaban: n=6,715;	Apixaban had a lower adjusted major bleeding risk compared with rivaroxaban and dabigatran
Lip, 2016 (Int J Clin Pract) USA	Truven MarketScan® Commercial & Medicare supplemental US database Jan 1, 2013 - Dec 31, 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: 1.7-1.9	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=4,173 apixaban: n=2,402	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, 2019 Scotland	Prescribing Information System, the Scottish Morbidity Records/ Hospital Inpatients and Outpatient attendance datasets; National Records of Scotland Drug's approval date – Dec 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: 1.1-1.4	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=7,265 apixaban: n=6,200; dabigatran: n=1,112;	All NOACs were similarly effective in preventing strokes and systemic embolisms, while patients being treated with rivaroxaban exhibited the highest bleeding risks.
Staerk, 2018 Denmark	Danish national patient registry, Danish national prescription registry,	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	stroke/thromboembolism (TE), ischaemic stroke, major bleeding, intracranial bleeding and gastrointestinal bleeding, ICD-10 codes whether primary and	dabigatran: n=11,492 apixaban: n=11,064 rivaroxaban: n=8966	Rivaroxaban was associated with higher bleeding risk

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
	Danish civil registration system March 1, 2012 - Dec 31, 2016		Comorbidity index: NR	secondary codes were used is not described		compared with dabigatran and apixaban and dabigatran was associated with lower intracranial bleeding risk compared with rivaroxaban and apixaban.
Tepper, 2018 USA	Truven MarketScan Commercial Claims and Encounter and Medicare Supplemental & Coordination of Benefits Early View Database Jan 1, 2013 - Oct 31, 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: 1.6-1.8	Bleeding, ICD-9-CM codes, whether primary and secondary codes were used is not described	rivaroxaban: n=30,529 dabigatran: n=20,963 apixaban: n=8,785;;	Rivaroxaban appeared to have an increased risk of any bleeding, clinically relevant non-major bleeding, and major inpatient bleeding, compared to 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 2018 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= 16,547 apixaban: n= 10,601 dabigatran: n=5,537	Apixaban was associated with a lower risk of major bleed than rivaroxaban. Rivaroxaban was associated with a higher risk of intracranial bleed compared to apixaban. rivaroxaban was associated with higher risks compared with apixaban for haematuria, all gastrointestinal bleed and upper gastrointestinal bleed.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46

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
						The risk of primary ischaemic stroke did not differ between any of the anticoagulants

For peer review only

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
Cerda, 2019 Spain	Oral Anticoagulant Treatment Unit of the Hemostasis and Thrombosis Department of the University Hospital Vall d'Hebron from Barcelona (Spain) Jan, 2015 - Sept, 2017	NVAF patients with nonvalvular 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 ischemic stroke and intracranial hemorrhage were similar among different NOACs, but rates of major bleeding were higher with dabigatran and apixaban and lower with rivaroxaban.
Li, 2017 China	Queen Mary Hospital, Hong Kong Jan, 2008 - Dec, 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 ischemic 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 ischemic stroke risk compared with patients on rivaroxaban. There was no significant difference in ischemic stroke risk between those on rivaroxaban and dabigatran.

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, 2016 Denmark	Danish national prescription registry, Danish national patient register, Danish civil registration system Feb. 1, 2012 - July 31, 2014	NVAF patients who were new-users of OAC (no OAC treatment in at least the last two 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	ischemic stroke/systemic embolism/transient ischemic 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=8,908 rivaroxaban: n=1,405;	Rivaroxaban and dabigatran had similar stroke rates. Bleeding and mortality rates were higher in rivaroxaban versus dabigatran.

Peer review only

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 (e.g. 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 - N=265,583
Study results	Of the 26 studies in which apixaban, rivaroxaban and dabigatran were included: - apixaban was favourable compared to dabigatran and rivaroxaban: n=13 - no single favourable NOAC: n=13

* about equal distribution between dabigatran, rivaroxaban and apixaban. Edoxaban is not included in these studies.

1
2
3 More than 50% of the studies were conducted in the USA (n=24),[16-39] five were conducted in Denmark,[40-44]
4 four in Taiwan,[45-48] and one in France,[49] Sweden,[50] Scotland,[51] the UK,[52] Spain,[53] and China.[54]
5
6 Dabigatran and rivaroxaban were included in all 39 studies, apixaban was included in 26 studies and edoxaban
7
8 was included in 1 study. Next to these NOACs, VKA was included in 25 of these studies as one of the
9
10 comparators. The results below focus on the NOAC to NOAC comparisons only.

11
12 In the studies that included apixaban, dabigatran and rivaroxaban, rivaroxaban was most dominantly used in the
13
14 USA, UK, Scotland, and Taiwan, while dabigatran was the most prescribed NOAC in Denmark. In three other
15
16 European studies the distribution was about equal between the three NOACs. In none of the included studies,
17
18 apixaban was the most dominantly prescribed NOAC.

21 **Setting**

22 Most studies concerned patient registries, pharmacy or prescription databases and/or health insurance databases
23
24 (n=39), while there were three clinical practice based studies.[50, 53, 54]

27 **Study population**

28
29 All studies included only NVAF patients. In seven studies, it was specifically described that patients were newly
30
31 diagnosed with NVAF and initiated NOAC treatment during study period.[21, 27, 34, 37, 40, 45, 54] None of the
32
33 other studies included prevalent users of (N)OAC, but included e.g. 'newly treated', 'initiating treatment', 'new
34
35 users', 'first-time prescription' of NVAF patients who were prescribed (N)OAC. In some studies (N)OAC use in the
36
37 past (between 3 months and 2 years before index date) was allowed, while this seemed not be allowed in some
38
39 other studies, or it was not described.

41 *Inclusion criteria*

42
43 Five studies concerned elderly patients specifically (i.e. ≥ 65 years old),[19, 21, 23-25] two included adults ≥ 45
44
45 years old,[33, 40] and one study included patients between 30 and 100 years of age.[44] The other studies
46
47 included all adults with atrial fibrillation (it was assumed that if no further age specification was provided, 'adults'
48
49 meant that all >18 years old were included). In one study only patients who were hospitalised for bleeding after
50
51 start with OAC treatment were included.[22] No other focus on a specific group of AF patients was found.

54 *Exclusion criteria*

55
56 NOAC use that could be related to other disorders, such as transient AF, major knee or hip surgery, venous
57
58 thromboembolism or pulmonary embolism, were specifically described as exclusion criteria in most studies,
59
60

1
2
3 except in ten studies.[16, 27, 28, 33-35, 50, 52-54] In one study patients with liver injury before their first oral
4 anticoagulant (OAC) prescription were specifically excluded.[18]
5
6
7

8 *Baseline characteristics*

9
10 Baseline characteristics of the NVAF patients differed between studies. Mean age ranged from 65-84 years
11 between the studies. The percentage of males ranged from 39-73%, and the mean CHA2DS2-Vasc Score ranged
12 from 2.1-4.9. Excluding the five studies that specifically focussed on an elderly population of ≥ 65 years old and
13 the two additional studies that used the Medicare database (only patients of 65 years or older are in Medicare),
14 the mean age ranged from 65-78 years old. Different measures were used to assess the comorbidity index:
15 Charlson comorbidity index, Charlson-Deyo index and Gagne comorbidity score, while in 30 of the 43 studies no
16 comorbidity index was presented.
17
18
19
20
21
22

23 **Selection of covariates**

24
25 Most studies (n=34) did not provide a rationale for the selection of covariates that were included in the PS model
26 or in adjusted analysis. However, in one of the articles an extensive rationale and selection procedure of co-
27 variates that were included in the analysis was provided.[33] In three other studies, the authors selected
28 covariates based on medical knowledge on risk factors with reference to earlier published studies.[31, 39, 52] In
29 one other study it was reported that sociodemographic and clinical characteristics that were associated with
30 treatment initiation and the risk of major bleeding were included in the model to adjust for differences across
31 cohorts, without further explanation or reference.[30]
32
33
34
35
36
37
38

39 **Definition primary study outcomes**

40
41 Primary outcomes differed between the studies. Effectiveness outcomes included in the studies included stroke,
42 systemic embolism, (or composite of stroke/systemic embolism), all-cause death, myocardial infarction, venous
43 thromboembolism and safety outcomes included major bleeding, or a specific type of bleeding (e.g. intracranial
44 haemorrhage, gastrointestinal bleeding etc.) and liver injury. In most studies, ICD-9 or ICD-10 codes were used,
45 but whether this concerned a primary diagnosis only or whether it could be either a primary or a second diagnosis
46 differed between the studies. In some studies it was not described whether the ICD codes referred to primary
47 diagnosis only or to a primary or secondary diagnosis.
48
49
50
51
52
53

54 **Statistical approaches to adjust for confounding (primary analysis)**

55
56 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
57 studies.[17, 22, 24, 25, 28, 43, 46, 48] PS stratified analyses was done in one study.[41] In twelve studies, the
58 primary analyses utilised a Cox PH regression model in which adjustment for confounding was done.[18, 27, 31,
59
60

1
2
3 38, 42, 44, 45, 50-52] Finally, in two studies no adjustment for differences in baseline characteristics was
4 performed.[53, 54]
5
6
7

8 *PS matching*

9 Co-variables

10 Creatinine clearance was not included as a covariate in any of the 18 studies. All 18 studies took the following
11 covariates into account: age, sex, CHA2DS2-VASc score and/or the individual comorbidities included in this
12 score, HAS-BLED score and/or the individual conditions included in this score (except alcohol use in Lai et al.
13 [47]), renal disease, and co-medication use such as antiplatelets. Some included other comorbidities, such as
14 cancer, rheumatic disease, specific heart diseases, COPD, HIV, dementia, depression, neurological disorders,
15 and/or a various list of co-medications as well.
16
17
18
19
20
21
22

23 Matching method

24 In one study the matching method was not described.[49] In two studies, the calliper used was not described.[23,
25 29] In seven studies 1:1 PS matching without replacement was used and a calliper of 0.01 was applied.[16, 19,
26 20, 26, 30, 32, 36] Five other studies also matched 1:1 without replacement but used another calliper: in three
27 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,
28 three-way matching was used.[21, 34, 37]
29
30
31
32
33
34

35 Balance co-variables

36 In two studies it was not described how the balance between covariates was evaluated.[33, 35] In two studies the
37 balance was evaluated using $p < 0.05$ (of which one also used standardized difference of <10%),[23, 47] and in
38 another study it was stated that the groups were comparable even though a p value of > 0.05 was found.[29]
39 Balance was checked with an absolute standardized difference of <10% in 13 studies.[16, 19-21, 26, 30, 32, 34,
40 36, 37, 39, 40, 47, 49] Balance was reached in all studies after matching.
41
42
43
44
45
46
47

48 Sample size

49 In four studies the sample size before matching was not reported[29, 35, 36, 39] and in one study the sample size
50 after matching was not reported.[34] At study start (before PSM), sample size between the NOACs differed
51 greatly, except in three studies.[21, 37, 40]
52
53
54

55 *IPTW*

56
57
58
59
60

1
2
3 In one study, balance was tested using ANOVAs for significant differences.[22] Balance was checked with an
4 absolute standardized difference of <10% in the other nine studies.[17, 24, 25, 28, 43, 46, 48] Balance was
5 reached in all studies after IPTW.
6
7

8 There was no reporting on extreme weights in the eight included studies.[17, 22, 24, 25, 28, 43, 46, 48]
9

10 11 *PS stratification*

12
13 In one study, asymmetric trimming of the PS was done, which resulted in a small part of both treatment groups
14 being removed in order to gain in comparability. Balance in co-variables was reached with standardized difference
15 of <10%. In a Cox model this trimmed PS was used in 10 deciles as strata.[41]
16
17

18 19 20 Cox HP regression models

21
22 In ten studies, Cox HP regression models were applied with adjustment for a number of confounders.[18, 27, 31,
23 38, 42, 44, 45, 50-52] In one of these studies, the number of events per variable was not sufficient for such an
24 analyses.[50] The ratio was acceptable in the other studies for at least some of the outcomes.[18, 28, 31, 38, 42,
25 44, 45, 51, 52]
26
27
28
29

30 31 Unadjusted analysis

32
33 In two studies no adjustment for confounding factors seemed to have been done, even though significant
34 differences between treatment groups existed at baseline. Cerda et al. presented events per 100 patient-years
35 and used a log-rank test to determine whether outcomes differed between the NOACs.[53] Li et al. conducted a
36 Cox proportional hazard model, likely unadjusted, but this was not clearly described in the article.[54]
37
38
39
40

41 **Study results**

42
43 Which NOAC performed best differed between the included studies. We found only one study that included all
44 four NOACs, in which no preference for one specific NOAC was found, except that rates of major bleeding were
45 lower with rivaroxaban.[53] Of the 26 studies in which apixaban, rivaroxaban and dabigatran were included,
46 apixaban was favourable compared to dabigatran and rivaroxaban in 13 studies, of which 10 were from the USA,
47 two from Europe and one from Asia,[16, 17, 19, 20, 23, 26, 28, 29, 32, 36, 42, 50, 52] while dabigatran and
48 rivaroxaban were not found to be the single most favourable NOAC in any of the remaining 13 studies. Results for
49 these 13 studies were mixed, with either no favourable NOAC at all, or one NOAC was selected as the least
50 favourable, while the other two NOACs did not differ.
51
52
53
54
55
56
57
58
59
60

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 required to confirm that difference.

For peer review only

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

For peer review only

1
2
3 The primary safety endpoint was major bleeding for RE-LY, ARISTOTLE, and ENGAGE AF and major bleeding
4 plus clinically relevant non-major bleeding for ROCKET AF, but data on major bleeds only for ROCKET-AF are
5 available as well. Major bleeds in the warfarin arms were estimated at 3.36% (RE-LY),^[3] 3.4% (ROCKET),^[6]
6 3.09% (ARISTOTLE),^[5] and 3.43% (ENGAGE).^[4] From this range we choose a relatively arbitrary base rate of
7 3.2% and applied the observed risk reduction to estimate comparable base rates of 2.21% for apixaban 2.57% for
8 edoxaban, 2.96% for dabigatran and 3.29% for rivaroxaban. Using the sample size calculator,^[55] the biggest
9 expected difference was between rivaroxaban and apixaban and it was estimated that a trial with 7,196 patients
10 would be needed to confirm this difference. A much smaller difference is between edoxaban and apixaban which
11 would require a trial of 56,512 patients to confirm that difference.
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

DISCUSSION

In total, we found 39 studies directly comparing the effectiveness and/or safety of at least two NOACs in NVAF patients. 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 sufficiently quality based on the technical aspects of the studies, there are some issues that can hamper the generalisability of the results. These issues concern remaining confounding, the use of a smaller or broader calliper, differences in baseline characteristics between studies, channelling bias and change in treatment paradigm.

Balance in baseline characteristics between NOACs was checked with p-values or a standardized 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, and balance was reached for all of these variables, one should keep in mind that balance between unmeasured or unmeasurable factors cannot be assumed.[14] Therefore, there is always a possibility of residual confounding. This possibility was acknowledged in all included studies. 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, 57] 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.

In general, a calliper of <0.2 of the standard deviation of the logit of the PS is considered to be 'optimal'. [58] 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, e.g. on the baseline risk for stroke.

All included studies focused on NVAF patients only. NOAC use that could have been related to other conditions was excluded specifically in 34 of the 36 included studies. 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 focussed on an elderly population of NVAF patients of ≥ 65 years old. Besides applying specific inclusion criteria regarding age in some studies, these differences also depended on the specific registry

1
2
3 or database that was used, e.g. Medicare is for people of 65 years old or older. Even though only five of the
4 included studies focused on an elderly NVAF population, and the others applied broad age ranges, there were
5 differences in mean age, proportion of males and mean CHA2DS2Vasc score between the studies, which can
6 have an impact on the results and jeopardize the generalisability of the results.
7
8
9

10
11 Rivaroxaban was the most prescribed NOAC in almost all included studies from the USA. However, in the first
12 quarter of 2017, apixaban was the most prescribed NOAC in NVAF in the USA (i.e. in 50% of new OAC
13 prescriptions). Especially older patients, women, increased stroke or bleeding risk and having comorbidities was
14 associated with prescription of apixaban versus other NOACs.[59] Rivaroxaban was also the most prescribed
15 NOAC in the included studies from the UK and Scotland. Based on the CPRD, 56.5% of the OAC prescriptions
16 concerned a NOAC, of which rivaroxaban was still described most often in 2015.[60] Dabigatran was described
17 most often in the studies from Denmark. Haastrup et al. described that most AF patients that initiated NOAC
18 received dabigatran between 2008 and 2016, but a trend was observed that per 1000 person-years the number of
19 patients described dabigatran decreased and the number of patients receiving rivaroxaban and apixaban
20 increased.[61] This shows that the treatment paradigm changed over time, and might still be changing, and this
21 pattern differs between the USA, Europe and Asia. Channelling bias therefore likely occurs.
22
23
24
25
26
27
28
29

30 Our naïve analysis predicts that in terms of the primary efficacy outcome observational studies will need a
31 relatively high number of patients to be able to demonstrate the differences between the NOACs and a small
32 sample size will not allow robust comparison to be made.
33
34
35
36

37 The pattern of major bleeding events seen in the included observational studies, confirms the expectation from
38 our naïve analysis of the pivotal clinical trials that rivaroxaban seems to have the least favourable safety profile
39 among apixaban and dabigatran. The findings are not consistent to allow for a robust conclusion between
40 apixaban and dabigatran which confirms the need for a high number of patients, although a trend for a slight
41 better safety profile of apixaban can be observed.
42
43
44
45

46 The requirement for a high number of patients to compare NOACs both in terms of efficacy and safety as
47 predicted by the pivotal trial results is confirmed by the findings of the observational studies. This finding may
48 support the claim that the differences between the NOACs are relatively small.
49
50
51

52
53 In summary, even though the larger part of these studies are technically well conducted, these studies have some
54 important limitations regarding the generalisability of the study results especially given the relatively high patient
55 number required for a meaningful comparison between NOACs. Most studies included all NVAF patients on
56 NOAC available in the registry/database during the study period and did not apply further specific in- and
57 exclusion criteria, but differences between studies regarding baseline characteristics existed. Mean age at study
58
59
60

1
2
3 start and baseline risk for stroke (CHA2DS2-VASc score) differed between the studies. As channelling bias
4 cannot be ruled out, the result of these studies might not be generalisable. Furthermore, results from the PS
5 studies are only applicable to the patients that were kept in the analyses as patients excluded from the analysis
6 likely differ from the ones that were included in the analysis. The 1:1 matched cohorts depended on the sample
7 size of the NOAC with the least number of patients and as a result many patients from the larger of the two NOAC
8 groups were excluded as they could not be matched. Besides these study specific limitations, differences in
9 reimbursement and ICD-coding exist between the USA and Europe, and also within Europe. Differences in
10 reimbursement may lead to differences in adherence and non-adherence could lead to worse outcomes. These
11 limitations should be kept in mind when results of these studies are used to decide what NOAC should be
12 prescribed for a certain patient. Finally, given the small differences between efficacy and safety outcomes
13 between NOACs, the element of patient preference should be taken into consideration,[62] as tailoring anti-
14 coagulation treatment towards patient preferences can promote adherence to treatment.
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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.

For peer review only

COMPETING INTEREST

EB reports grants from Daiichi Sankyo, during the conduct of the study; grants from Daiichi Sankyo, outside the submitted work. BvH 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 Inc, personal fees from Janssen, personal fees from ONO Pharmaceuticals, from AbbVie, outside the submitted work.

AUTHORS' CONTRIBUTION

EB: Conceptualization (support); Methodology (equal); Writing – Original Draft Preparation; Writing – Review & Editing (equal). BvH: Conceptualization (support); Methodology (equal); Writing – Review & Editing (equal). SH: Conceptualization (support); Writing – Review & Editing (equal). GS: Conceptualization (support), Supervision; Writing – Review & Editing (equal). ATC: Conceptualization (lead); Writing – Review & Editing (equal)

For peer review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

DATA SHARING STATEMENT

No new data were generated or analysed in support of this research.

For peer review only

FUNDING

This work was supported by Daiichi Sankyo Europe.

For peer review only

1
2
3 PATIENT AND PUBLIC INVOLVEMENT
4

5 This research was done without patient involvement. Patients were not invited to comment on the study design
6 and were not consulted to develop patient relevant outcomes or interpret the results. Patients were not invited to
7 contribute to the writing or editing of this document for readability or accuracy.
8
9

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

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(38):2893-962.10.1093/eurheartj/ehw210
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(2):e125-e51.10.1161/CIR.0000000000000665
3. Connolly SJ, Ezekowitz MD, Yusuf S, et al. Dabigatran versus warfarin in patients with atrial fibrillation. *N Engl J Med*. 2009;361(12):1139-51.10.1056/NEJMoa0905561
4. Giugliano RP, Ruff CT, Braunwald E, et al. Edoxaban versus warfarin in patients with atrial fibrillation. *N Engl J Med*. 2013;369(22):2093-104.10.1056/NEJMoa1310907
5. Granger CB, Alexander JH, McMurray JJ, et al. Apixaban versus warfarin in patients with atrial fibrillation. *N Engl J Med*. 2011;365(11):981-92.10.1056/NEJMoa1107039
6. Patel MR, Mahaffey KW, Garg J, et al. Rivaroxaban versus warfarin in nonvalvular atrial fibrillation. *N Engl J Med*. 2011;365(10):883-91.10.1056/NEJMoa1009638
7. Petri H, Urquhart J. Channeling bias in the interpretation of drug effects. *Stat Med*. 1991;10(4):577-81.10.1002/sim.4780100409
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(12):1503-10.10.1016/0895-4356(95)00048-8
9. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrical*. 1983;70(1):41-55
10. Brookhart MA, Schneeweiss S, Rothman KJ, et al. Variable selection for propensity score models. *Am J Epidemiol*. 2006;163(12):1149-56.10.1093/aje/kwj149
11. D'Agostino RB, Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med*. 1998;17(19):2265-81.10.1002/(sici)1097-0258(19981015)17:19<2265::aid-sim918>3.0.co;2-b
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(4):1654-70.10.1177/0962280215584401
13. Austin PC. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Stat Med*. 2008;27(12):2037-49.10.1002/sim.3150

- 1
2
3 14. Johnson ML, Crown W, Martin BC, et al. Good research practices for comparative effectiveness
4 research: analytic methods to improve causal inference from nonrandomized studies of treatment effects using
5 secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force
6 Report--Part III. *Value Health*. 2009;12(8):1062-73.10.1111/j.1524-4733.2009.00602.x
7
8
9
10 15. Yao XI, Wang X, Speicher PJ, et al. Reporting and Guidelines in Propensity Score Analysis: A
11 Systematic Review of Cancer and Cancer Surgical Studies. *J Natl Cancer Inst*. 2017;109(8).10.1093/jnci/djw323
12
13 16. Abraham NS, Noseworthy PA, Yao X, et al. Gastrointestinal Safety of Direct Oral Anticoagulants: A
14 Large Population-Based Study. *Gastroenterology*. 2017;152(5):1014-22.e1.10.1053/j.gastro.2016.12.018
15
16 17. Adeboyeje G, Sylwestrzak G, Barron JJ, et al. Major Bleeding Risk During Anticoagulation with Warfarin,
17 Dabigatran, Apixaban, or Rivaroxaban in Patients with Nonvalvular Atrial Fibrillation. *J Manag Care Spec Pharm*.
18 2017;23(9):968-78.10.18553/jmcp.2017.23.9.968
19
20 18. Alonso A, MacLehose RF, Chen LY, et al. Prospective study of oral anticoagulants and risk of liver injury
21 in patients with atrial fibrillation. *Heart*. 2017;103(11):834-9.10.1136/heartjnl-2016-310586
22
23 19. Amin A, Keshishian A, Trocio J, et al. A Real-World Observational Study of Hospitalization and Health
24 Care Costs Among Nonvalvular Atrial Fibrillation Patients Prescribed Oral Anticoagulants in the U.S. Medicare
25 Population. *J Manag Care Spec Pharm*. 2018;24(9):911-20.10.18553/jmcp.2018.24.9.911
26
27 20. Amin A, Keshishian A, Vo L, et al. Real-world comparison of all-cause hospitalizations, hospitalizations
28 due to stroke and major bleeding, and costs for non-valvular atrial fibrillation patients prescribed oral
29 anticoagulants in a US health plan. *J Med Econ*. 2018;21(3):244-53.10.1080/13696998.2017.1394866
30
31 21. Briasoulis A, Inampudi C, Akintoye E, et al. Safety and Efficacy of Novel Oral Anticoagulants Versus
32 Warfarin in Medicare Beneficiaries With Atrial Fibrillation and Valvular Heart Disease. *J Am Heart Assoc*.
33 2018;7(8).10.1161/jaha.118.008773
34
35 22. Charlton B, Adeboyeje G, Barron JJ, et al. Length of hospitalization and mortality for bleeding during
36 treatment with warfarin, dabigatran, or rivaroxaban. *PLoS One*.
37 2018;13(3):e0193912.10.1371/journal.pone.0193912
38
39 23. Deitelzweig S, Luo X, Gupta K, et al. Comparison of effectiveness and safety of treatment with apixaban
40 vs. other oral anticoagulants among elderly nonvalvular atrial fibrillation patients. *Curr Med Res Opin*.
41 2017;33(10):1745-54.10.1080/03007995.2017.1334638
42
43 24. Graham DJ, Baro E, Zhang R, et al. Comparative Stroke, Bleeding, and Mortality Risks in Older
44 Medicare Patients Treated with Oral Anticoagulants for Nonvalvular Atrial Fibrillation. *Am J Med*.
45 2019.10.1016/j.amjmed.2018.12.023
46
47 25. Graham DJ, Reichman ME, Wernecke M, et al. Stroke, Bleeding, and Mortality Risks in Elderly Medicare
48 Beneficiaries Treated With Dabigatran or Rivaroxaban for Nonvalvular Atrial Fibrillation. *JAMA Intern Med*.
49 2016;176(11):1662-71.10.1001/jamainternmed.2016.5954
50
51
52
53
54
55
56
57
58
59
60

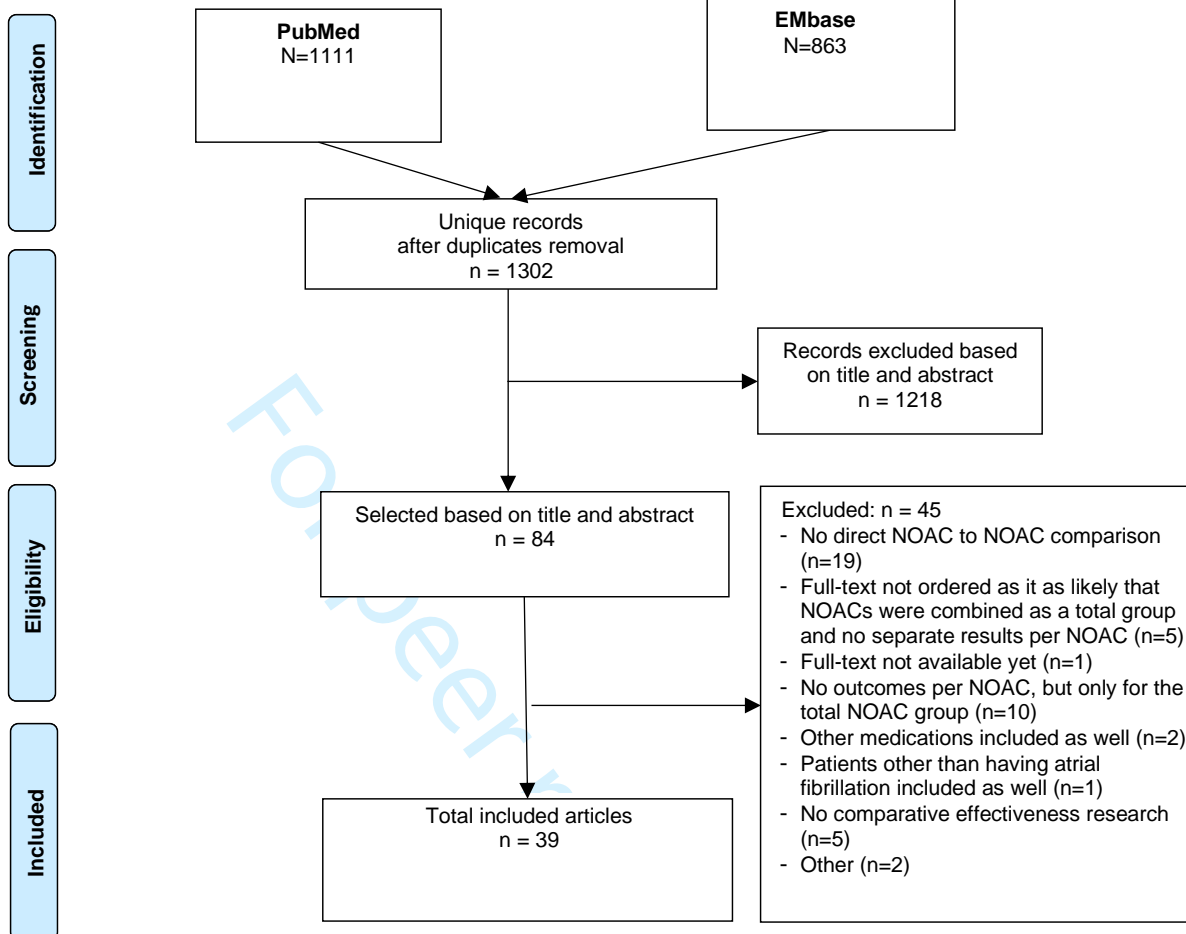
- 1
2
3 26. Gupta K, Trocio J, Keshishian A, et al. Real-World Comparative Effectiveness, Safety, and Health Care
4 Costs of Oral Anticoagulants in Nonvalvular Atrial Fibrillation Patients in the U.S. Department of Defense
5 Population. *J Manag Care Spec Pharm*. 2018;24(11):1116-27.10.18553/jmcp.2018.17488
6
7
8 27. Hernandez I, Zhang Y, Saba S. Comparison of the Effectiveness and Safety of Apixaban, Dabigatran,
9 Rivaroxaban, and Warfarin in Newly Diagnosed Atrial Fibrillation. *Am J Cardiol*. 2017;120(10):1813-
10 9.10.1016/j.amjcard.2017.07.092
11
12
13 28. Hernandez IZY. Comparing Stroke and Bleeding with Rivaroxaban and Dabigatran in Atrial Fibrillation:
14 Analysis of the US Medicare Part D Data. *Am J Cardiovasc Drugs*. 2017;17(1):37-47.10.1007/s40256-016-0189-9
15
16
17 29. Lin J, Trocio J, Gupta K, et al. Major bleeding risk and healthcare economic outcomes of non-valvular
18 atrial fibrillation patients newly-initiated with oral anticoagulant therapy in the real-world setting. *J Med Econ*.
19 2017;20(9):952-61.10.1080/13696998.2017.1341902
20
21
22 30. Lip GY, Keshishian A, Kamble S, et al. Real-world comparison of major bleeding risk among non-valvular
23 atrial fibrillation patients initiated on apixaban, dabigatran, rivaroxaban, or warfarin. A propensity score matched
24 analysis. *Thromb Haemost*. 2016;116(5):975-86.10.1160/th16-05-0403
25
26
27 31. Lip GY, Pan X, Kamble S, et al. Major bleeding risk among non-valvular atrial fibrillation patients initiated
28 on apixaban, dabigatran, rivaroxaban or warfarin: a "real-world" observational study in the United States. *Int J Clin*
29 *Pract*. 2016;70(9):752-63.10.1111/ijcp.12863
30
31
32 32. Lip GYH, Keshishian A, Li X, et al. Effectiveness and Safety of Oral Anticoagulants Among Nonvalvular
33 Atrial Fibrillation Patients. *Stroke*. 2018;49(12):2933-44.10.1161/STROKEAHA.118.020232
34
35
36 33. Lutsey PL, Norby FL, Zakai NA, et al. Oral anticoagulation therapy and subsequent risk of venous
37 thromboembolism in atrial fibrillation patients. *Curr Med Res Opin*. 2019;35(5):837-
38 45.10.1080/03007995.2018.1541445
39
40
41 34. Mentias A, Shantha G, Chaudhury P, et al. Assessment of Outcomes of Treatment With Oral
42 Anticoagulants in Patients With Atrial Fibrillation and Multiple Chronic Conditions: A Comparative Effectiveness
43 Analysis. *JAMA Netw Open*. 2018;1(5):e182870.10.1001/jamanetworkopen.2018.2870
44
45
46 35. Norby FL, Bengtson LGS, Lutsey PL, et al. Comparative effectiveness of rivaroxaban versus warfarin or
47 dabigatran for the treatment of patients with non-valvular atrial fibrillation. *BMC Cardiovasc Disord*.
48 2017;17(1):238.10.1186/s12872-017-0672-5
49
50
51 36. Noseworthy PA, Yao X, Abraham NS, et al. Direct Comparison of Dabigatran, Rivaroxaban, and
52 Apixaban for Effectiveness and Safety in Nonvalvular Atrial Fibrillation. *Chest*. 2016;150(6):1302-
53 12.10.1016/j.chest.2016.07.013
54
55
56 37. Shantha PSG, Bhave PD, Girotra S, et al. Sex-Specific Comparative Effectiveness of Oral
57 Anticoagulants in Elderly Patients With Newly Diagnosed Atrial Fibrillation. *Circ Cardiovasc Qual Outcomes*.
58 2017;10(4).10.1161/circoutcomes.116.003418
59
60

- 1
2
3 38. Tepper PG, Mardekian J, Masseria C, et al. Real-world comparison of bleeding risks among non-valvular
4 atrial fibrillation patients prescribed apixaban, dabigatran, or rivaroxaban. *PLoS One*.
5
6 2018;13(11):e0205989.10.1371/journal.pone.0205989
7
8 39. Villines TCAAPMTWEARTTDOKSE. Comparative safety and effectiveness of dabigatran vs. rivaroxaban
9 and apixaban in patients with non-valvular atrial fibrillation: A retrospective study from a large healthcare system.
10
11 *Eur Heart J - Cardiovascular Pharmacotherapy*. 2019;5(2):80-90.10.1093/ehjcvp/pvy044 FULL TEXT LINK
12 <http://dx.doi.org/10.1093/ehjcvp/pvy044>
13
14 40. Andersson NW, Svanstrom H, Lund M, et al. Comparative effectiveness and safety of apixaban,
15 dabigatran, and rivaroxaban in patients with non-valvular atrial fibrillation. *Int J Cardiol*. 2018;268:113-
16
17 9.10.1016/j.ijcard.2018.03.047
18
19 41. Gorst-Rasmussen A, Lip GY, Bjerregaard Larsen T. Rivaroxaban versus warfarin and dabigatran in atrial
20 fibrillation: comparative effectiveness and safety in Danish routine care. *Pharmacoepidemiol Drug Saf*.
21
22 2016;25(11):1236-44.10.1002/pds.4034
23
24 42. Lamberts M, Staerk L, Olesen JB, et al. Major Bleeding Complications and Persistence With Oral
25 Anticoagulation in Non-Valvular Atrial Fibrillation: Contemporary Findings in Real-Life Danish Patients. *J Am*
26
27 *Heart Assoc*. 2017;6(2).10.1161/jaha.116.004517
28
29 43. Larsen TB, Skjoth F, Nielsen PB, et al. Comparative effectiveness and safety of non-vitamin K antagonist
30 oral anticoagulants and warfarin in patients with atrial fibrillation: propensity weighted nationwide cohort study.
31
32 *BMJ*. 2016;353:i3189.10.1136/bmj.i3189
33
34 44. Staerk L, Gerds TA, Lip GYH, et al. Standard and reduced doses of dabigatran, rivaroxaban and
35 apixaban for stroke prevention in atrial fibrillation: a nationwide cohort study. *J Intern Med*. 2018;283(1):45-
36
37 55.10.1111/joim.12683
38
39 45. Chan YH, Kuo CT, Yeh YH, et al. Thromboembolic, Bleeding, and Mortality Risks of Rivaroxaban and
40 Dabigatran in Asians With Nonvalvular Atrial Fibrillation. *J Am Coll Cardiol*. 2016;68(13):1389-
41
42 401.10.1016/j.jacc.2016.06.062
43
44 46. Chan YH, See LC, Tu HT, et al. Efficacy and Safety of Apixaban, Dabigatran, Rivaroxaban, and Warfarin
45 in Asians With Nonvalvular Atrial Fibrillation. *J Am Heart Assoc*. 2018;7(8).10.1161/jaha.117.008150
46
47 47. Lai CL, Chen HM, Liao MT, et al. Comparative Effectiveness and Safety of Dabigatran and Rivaroxaban
48 in Atrial Fibrillation Patients. *J Am Heart Assoc*. 2017;6(4).10.1161/jaha.116.005362
49
50 48. Meng SW, Lin TT, Liao MT, et al. Direct Comparison of Low-Dose Dabigatran and Rivaroxaban for
51 Effectiveness and Safety in Patients with Non-Valvular Atrial Fibrillation. *Acta Cardiol Sin*. 2019;35(1):42-
52
53 54.10.6515/acs.201901_35(1).20180817a
54
55
56
57
58
59
60

- 1
2
3 49. Blin PD-PCCYBJMPAALRDCMN. Comparative Effectiveness and Safety of Standard or Reduced Dose
4 Dabigatran vs. Rivaroxaban in Nonvalvular Atrial Fibrillation. *Clin Pharm Therapeutics*. 2019;105(6):1439-
5 55.10.1002/cpt.1318
6
7
8 50. Al-Khalili F, Lindstrom C, Benson L. The safety and persistence of non-vitamin-K-antagonist oral
9 anticoagulants in atrial fibrillation patients treated in a well structured atrial fibrillation clinic. *Curr Med Res Opin*.
10 2016;32(4):779-85.10.1185/03007995.2016.1142432
11
12
13 51. Mueller T, Alvarez-Madrado S, Robertson C, et al. Comparative safety and effectiveness of direct oral
14 anticoagulants in patients with atrial fibrillation in clinical practice in Scotland. *Br J Clin Pharmacol*.
15 2019;85(2):422-31.10.1111/bcp.13814
16
17
18 52. Vinogradova Y, Coupland C, Hill T, et al. Risks and benefits of direct oral anticoagulants versus warfarin
19 in a real world setting: cohort study in primary care. *BMJ*. 2018;362:k2505.10.1136/bmj.k2505
20
21
22 53. Cerda M, Cerezo-Manchado JJ, Johansson E, et al. Facing real-life with direct oral anticoagulants in
23 patients with nonvalvular atrial fibrillation: outcomes from the first observational and prospective study in a
24 Spanish population. *J Comp Eff Res*. 2019;8(3):165-78.10.2217/cer-2018-0134
25
26
27 54. Li WH, Huang D, Chiang CE, et al. Efficacy and safety of dabigatran, rivaroxaban, and warfarin for stroke
28 prevention in Chinese patients with atrial fibrillation: the Hong Kong Atrial Fibrillation Project. *Clin Cardiol*.
29 2017;40(4):222-9.10.1002/clc.22649
30
31
32 55. <https://clincalc.com/stats/samplesize.aspx> [accessed May 2020]
33
34 56. Ho DE, Imai K, King G, et al. Matching as nonparametric preprocessing for reducing model dependence
35 in parametric causal inference. *Political Analysis*. 2007;15:199-236
36
37 57. Bohula EA, Giugliano RP, Ruff CT, et al. Impact of Renal Function on Outcomes With Edoxaban in the
38 ENGAGE AF-TIMI 48 Trial. *Circulation*. 2016;134(1):24-36.10.1161/circulationaha.116.022361
39
40
41 58. Austin PC, Stuart EA. Estimating the effect of treatment on binary outcomes using full matching on the
42 propensity score. *Stat Methods Med Res*. 2017;26(6):2505-25.10.1177/0962280215601134
43
44 59. Zhu J, Alexander GC, Nazarian S, et al. Trends and Variation in Oral Anticoagulant Choice in Patients
45 with Atrial Fibrillation, 2010-2017. *Pharmacotherapy*. 2018;38(9):907-20.10.1002/phar.2158
46
47
48 60. Loo SY, Dell'Aniello S, Huiart L, et al. Trends in the prescription of novel oral anticoagulants in UK
49 primary care. *Br J Clin Pharmacol*. 2017;83(9):2096-106.10.1111/bcp.13299
50
51 61. Hastrup SB, Hellfritsch M, Rasmussen L, et al. Use of Non-Vitamin K Antagonist Oral Anticoagulants
52 2008-2016: A Danish Nationwide Cohort Study. *Basic Clin Pharmacol Toxicol*. 2018;123(4):452-
53 63.10.1111/bcpt.13024
54
55
56 62. Vaanholt MCW, Weernink MGM, von Birgelen C, et al. Perceived advantages and disadvantages of oral
57 anticoagulants, and the trade-offs patients make in choosing anticoagulant therapy and adhering to their drug
58 regimen. *Patient Educ Couns*. 2018;101(11):1982-9.10.1016/j.pec.2018.06.019
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only



The following search string was used for PubMed, and adapted for Cochrane and EMBase.

#1. NOAC

direct oral anticoagulant*[tiab] OR direct oral anti-coagulant*[tiab] OR direct oral anticoagulation[tiab] OR direct oral anti-coagulation[tiab] OR direct-acting oral anticoagulant*[tiab] OR direct-acting oral anti-coagulant*[tiab] OR direct-acting oral anticoagulation[tiab] OR direct-acting oral anti-coagulation[tiab] OR DOAC[tiab] OR novel oral anticoagulant*[tiab] OR novel oral anti-coagulant*[tiab] OR Novel oral anticoagulation[tiab] OR Novel oral anti-coagulation[tiab] OR NOAC[tiab] OR Rivaroxaban[tiab] OR Apixaban[tiab] OR Edoxaban[tiab] OR Dabigatran[tiab] OR "Non VKA Oral Anticoagulant"[tiab] OR "Non Vitamin K Antagonist Oral Anticoagulant"[tiab]

#2. Comparative effectiveness studies

comparative effectiveness research[mesh] OR comparative effectiveness[tiab] OR real-world[tiab] OR real-life[tiab] OR cohort studies[mesh] OR cohort[tiab]

#3. Atrial fibrillation

atrial fibrillation[tiab]

Limits:

Language: English



PRISMA 2009 Checklist

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	3,4
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	4
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	NA
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	5,6
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	5,6
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	S1_Table
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	5
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	5
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	5
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	5,6
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	NA
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I ²) for each meta-analysis.	NA



PRISMA 2009 Checklist

Page 1 of 2

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	NA
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	NA
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	S2_Fig
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	8-21
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	8-27
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	NA
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	NA
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	NA
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	NA
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	30
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	30,31
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	31,32
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: www.prisma-statement.org.

Page 2 of 2

For peer review only - <http://bmjopen.bmj.com/site/about/guidelines.xhtml>

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

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2020-042024.R1
Article Type:	Original research
Date Submitted by the Author:	14-Jan-2021
Complete List of Authors:	Bunge, Eveline; Pallas Health Research and Consultancy van Hout, Ben; University of Sheffield, HEDS, SchARR Haas, Sylvia ; Formerly Technical University Of Munich Spentzouris, Georgios; Daiichi Sankyo Europe GmbH Cohen, Alexander; Guys and St Thomas' NHS Foundation Trust, King's College
Primary Subject Heading:	Cardiovascular medicine
Secondary Subject Heading:	Research methods
Keywords:	Adult cardiology < CARDIOLOGY, STATISTICS & RESEARCH METHODS, Cardiology < INTERNAL MEDICINE

SCHOLARONE™
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1
2
3 **CRITICAL APPRAISAL AND ISSUES REGARDING GENERALISABILITY OF COMPARATIVE**
4 **EFFECTIVENESS STUDIES OF NOACS IN ATRIAL FIBRILLATION AND THEIR RELATION TO CLINICAL**
5 **TRIAL DATA - A SYSTEMATIC REVIEW**
6
7
8
9

10 **Eveline M. Bunge¹, Ben van Hout², Sylvia Haas³, Georgios Spentzouris⁴, Alexander T. Cohen⁵**
11

12
13 ¹ Pallas, health research and consultancy BV, P.O. Box 21238, 3001 AE Rotterdam, the Netherlands
14

15
16 ² University of Sheffield, HEDS, ScHARR, Sheffield, UK
17

18
19 ³ Formerly Technical University of Munich, Munich, Germany
20

21
22 ⁴ Daiichi Sankyo Europe, Munich, Germany
23

24
25 ⁵ Guys and St Thomas' NHS Foundation Trust, King's College London, London, UK
26

27 * Corresponding author:
28

29 bunge@pallashrc.com
30

31 Word count: 4406
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 **KEY WORDS**

4 atrial fibrillation; non-vitamin K antagonist oral anticoagulants (NOAC); comparative effectiveness research;
5
6 systematic review
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

ABSTRACT

Objective: To critically appraise the published comparative effectiveness studies on non-vitamin K antagonist oral anticoagulants (NOAC) in nonvalvular 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 23-04-2019. Critical appraisal of the studies was done using amongst other ISPOR checklists 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, eight studies weighted patients based on the inverse probability of treatment, one study used propensity score stratification and ten 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 7,700 and 59,500 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 NOACs in atrial fibrillation patients 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.

For peer review only

INTRODUCTION

Guidelines state a preference for non-vitamin K antagonist oral anticoagulants (NOACs) above vitamin K antagonists (VKA) in patients with nonvalvular 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; i.e. 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 channeling 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 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 randomized clinical trial, one needs substantial numbers 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 propensity score is the probability of an individual receiving a specific treatment given a specific set of patient characteristics (e.g. age, gender, comorbidities).[9] Variables related to the outcome should be included in the propensity score 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

1
2
3 results, however, there are also some disadvantages. For instance, in PS matching studies, patients who cannot
4 be matched to another patient will be excluded from the analyses, and in IPTW, when patients on one treatment
5 have a low propensity score and patients treated with the other treatment have a high propensity score, extreme
6 weights can occur which can bias the results.[12]
7
8
9

10
11 To gain more understanding in how the above described methodologies were applied in peer-reviewed CER on
12 effectiveness and safety in NOACs in NVAf patients, we conducted a systematic literature review. Within this we
13 compare the results with those from a naïve analysis of the results of the four major trial for rivaroxaban,
14 apixaban, dabigatran and edoxaban, and compare the results from the various analyses with those from the trials.
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

METHODS

Information sources, search strategy and eligibility criteria

We performed a systematic literature review to identify peer-reviewed comparative effectiveness research 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 appendix 1 for the search strings). The search was conducted on 23-04-2019 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, in- and exclusion criteria and the following baseline characteristics: age, proportion males, CHA2DS2Vasc score and comorbidity index.

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

- The variables included in the propensity score model
- Explanation of the variable selection procedure for propensity score model
- Distribution of baseline characteristics for each group before propensity score 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 (i.e. case-crossover study designs; clinical details in a subsample; proxy measures; or instrumental variable (IV) 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 in- and exclusion criteria and were included for data extraction, see figure 1. In table 1 to 5, study characteristics are presented. The most important differences between the studies are outlined in table 6.

For peer review only

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, 2017 USA	OptumLabs Data Warehouse Oct 1, 2010 through Feb 28, 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 caliper of 0.01; After PSM, standardized differences of all baseline characteristics were <10%.	rivaroxaban: n=19,301 dabigatran: n=17,426 apixaban: n=6,576	rivaroxaban vs. dabigatran: n=31,574 apixaban vs. rivaroxaban: n=13,130 apixaban vs. dabigatran: n=13,084 (more than 90% of original smallest samples size)	Apixaban had the most favorable gastrointestinal bleeding profile and rivaroxaban had the least favorable safety profile. Apixaban had the most favorable gastrointestinal safety profile among all age groups.
Amin, 2018 (J Manag Care Spec Pharm) USA	Medicare & Medicaid Services Jan 1, 2012, to Dec 31, 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-month before study start.	Age: 77.2±7.0-78.4±7.4 Male: 47.4-50.6% CHA2DS2-VASc: 4.4-4.6 CCI: 2.5-2.7	Hospitalization 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 caliper of 0.01; After PSM, standardized differences of all baseline characteristics were <10%.	rivaroxaban: n=53,146 apixaban: n=20,853 dabigatran: n=16,743	rivaroxaban vs. apixaban: n=41,608 dabigatran vs. apixaban: n=30,836 (more than 90% of original smallest samples size)	Apixaban was associated with significantly lower risks of all-cause, stroke/SE-related, and MB-related hospitalizations compared with dabigatran, and rivaroxaban
Amin 2018 (J Med Econ) USA	OptumInsight research database Jan 1, 2012 – Sept 30, 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-month before study start.	NR	Hospitalization 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 caliper of 0.01; After PSM, standardized differences of all baseline characteristics were <10%.	rivaroxaban: n=14,163 apixaban: n=8,652 dabigatran: n=3,684	apixaban vs. rivaroxaban: N=16,880 apixaban vs. dabigatran: N=7,114 (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, 2018 Denmark	National patient register, Register of Medicinal Product Statistics July 1, 2013 – March 31, 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 (i.e, intracranial bleeding, gastro-intestinal bleeding (bleeding ulcer, hematemesis or melena) or other serious bleeding (anemia 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 caliper of 0. (replacement yes or no not reported) All baseline characteristics were well balanced after matching, except for calendar year	apixaban: n=4,292 dabigatran: n=3,913 rivaroxaban: n=3,805	apixaban vs. rivaroxaban: N=7,352 apixaban vs. dabigatran: N=6,470 rivaroxaban vs. dabigatran: N=5,440	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.

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, 2019 France	French nationwide claims and hospitalization 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 three 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	Hospitalization with a main diagnosis of ischemic stroke or systemic embolism or major bleeding and all-cause death. (ICD-10) codes	1 matched cohort PSM method not reported. After PSM, standardized 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, 2018 USA	Medicare and Medicaid Services Jan 1, 2010 - Dec 31, 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 ischemic stroke or transient ischemic 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; 3-way propensity matching (VKA was one of the groups, but not further discussed here) After PSM, standardized 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
Deitelzweig, 2017 USA	Humana Research Database (Medicare coverage) Jan 2013 - 30 Sept 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: 2.7-3.0	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>.05).	rivaroxaban: n=11,082 apixaban: n=8,250 dabigatran: n=2,474	apixaban vs. rivaroxaban: n=13,620 apixaban vs. dabigatran: n= 4,654	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, 2018 USA	Department of Defence data Jan 1, 2012, to Sept 30, 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, standardized differences of all baseline characteristics were <10%.	rivaroxaban: n=15,680 apixaban: n=11,754 dabigatran: n=4,312	rivaroxaban vs. apixaban: n=22,568 dabigatran vs. apixaban: n=8,258	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

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, 2017 Taiwan	National Health Insurance program 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 standardized difference	dabigatran: n=10,625; rivaroxaban: n=4,609	dabigatran vs. rivaroxaban: N=9,200	Rivaroxaban therapy was associated with a statistically significant increase in all-cause death compared with dabigatran
Lin, 2017 USA	IMS Pharmetrics Plus database Jan 2013 – Sept 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=8,124 apixaban vs. dabigatran: N=5,368	Apixaban is associated with reduced risk of hospitalisation compared with dabigatran and rivaroxaban.
Lip, 2016 (Thromb Haemost) USA	Truven MarketScan® Commercial Claims and Encounter and Medicare Supplemental and Coordination of Benefits Databases Jan 2012 to Dec 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, standardized differences of all baseline characteristics were <10%.	rivaroxaban: n=17,801 apixaban: n=7,438 dabigatran: n=4,661	apixaban vs dabigatran: n=14,798 rivaroxaban vs dabigatran: n=9,314 apixaban vs rivaroxaban: n=8,814	Compared to apixaban, rivaroxaban initiation was associated with significantly higher risk of major bleeding. The difference for dabigatran was not statistically significant
Lip, 2018 USA	Medicare and Medicaid Services Medicare; Truven MarketScan, IMS PharMetrics Plus Database, Optum Clinformatics Data Mart, and the Humana Research Database Jan 1, 2013, to Sept 30, 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	Hospitalizations 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, standardized 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.

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
Lutsey, 2018 USA	MarketScan Commercial Database Jan 1, 2010 through Sept 30, 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 caliper 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=6,200	Risk of VTE was lowest among those prescribed apixaban and dabigatran
Mentias, 2018 USA	Medicare & Medicaid Services Jan 1, 2010, to Dec 31, 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 ischemic stroke or major bleeding as defined by Rothendler ⁶ and Suh based on the primary ICD-9-CM diagnosis on inpatient standard analytical files claims for acute care stays.	1 cohort 3-way PSM. (VKA was one of the groups, but not further discussed here) After PSM, standardized 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
Norby, 2017 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	ischemic stroke (primary discharge), intracranial bleeding (primary discharge), myocardial infarction (1st or 2 nd 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 caliper 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, 2016 USA	Optum Labs Data Warehouse Oct 1, 2010 - Feb 28, 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 caliper of 0.01. A standardized 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
Shantha, 2017 USA	Medicare and Medicaid Nov 1, 2011 - Dec 31, 2013	Newly diagnosed NVAF patients and initiated OAC use.	Males: Age: 74.7±5.9-74.9±6. CHADS2-Vasc: 3.7-3.8 Gagne score: 2.9-2.9 Women: Age: 76.6±6.6- 76.9±6.6 CHADS2-Vasc: 4.8-4.9	inpatient admissions for acute ischemic 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)	rivaroxaban: n=23,177 dabigatran: n=21,979	dabigatran vs. rivaroxaban: n= 37,298	The reduced risk of ischemic stroke in patients taking rivaroxaban, compared with dabigatran, seems to be limited to men, whereas the higher risk of

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
			Gagne: 3.0-3.1		A standardized difference < 10% was considered acceptable			bleeding seems to be limited to women
Villines, 2019 USA	US Department of Defence Military Health System database 1 July 2010 to 30 June 2016 for the dabigatran vs. rivaroxaban cohort, and 28 Dec 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=9,604	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; Gagne: Gagne comorbidity score;

For peer review only

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, 2017 USA	HealthCore integrated research environment Nov 1, 2009 - Jan 31, 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	Hospitalization 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 $\leq 10\%$.	dabigatran: n=8,539 rivaroxaban: n=8,398 apixaban: n=3,689	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, 2018 Taiwan	Taiwan National Health Insurance Research June 1, 2012 - Dec 31, 2016	NVAF patients with their first prescription of OAC	Age: 75 \pm 10- 76 \pm 10 Male: 55-60% CHA2DS2-VASc: 3.7-3.9 Comorbidity index: NR	Hospitalization for ischemic stroke/systemic embolism, intracranial hemorrhage, 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 $\leq 10\%$.	rivaroxaban: n=27,777 dabigatran: n=20,079 apixaban: n=5,843	Three low-dose NOACs showed similar performance as without subgrouping
Charlton, 2018 USA	HealthCore Integrated Research Environment database Nov 1, 2010 - March 31, 2014	NVAF patients hospitalized for bleeding after starting OAC (AF diagnosis 6 months before starting one of the index drugs).	Age: 68.0 \pm 12.5- 69.6 \pm 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- 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-day
Graham, 2016 USA	Medicare Nov 4, 2011 - June 30, 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 y: 50-51% Age: 75-84: 40-40% Age \geq 85: 9-10% Male: 53-53% CHADS2 \geq 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 hospitalization 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 $\leq 10\%$.	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

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, 2019 USA	fee-for-service Medicare Part A (hospitalization), Part B (office-based care), and Part D (prescription drug coverage) Oct, 2010 - Sept, 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	Hospitalized 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.
Hernandez, 2017 USA	Medicare Nov 4, 2011 -Dec 31, 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	ischemic 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 hemorrhage and gastrointestinal bleeding, not further described.	Extreme weights: not reported Balanced if the absolute value of the STD was ≤10%.	dabigatran n=9,138 rivaroxaban n=8,367	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, 2016 Denmark	Danish national prescription registry, Danish national patient register, Danish civil registration system August, 2011 -Oct, 2015	NVAF patients who were naïve to oral anticoagulants (no use of oral anticoagulant within one year)	Age (median, IQR): 67.6 (62.0-72.4)-71.8 (65.7-78.9) Male: 56.9-66.1% CHA2DS2VASc: 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=7,192 apixaban: n=6,349	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
Meng, 2019 Taiwan	National Health Insurance claims database June 1, 2012 - May 31, 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, ischemic stroke, intracranial hemorrhage, gastrointestinal hemorrhage needing transfusion, ICD-10 codes, whether primary and	Extreme weights: not reported Balanced if the absolute value of the STD was ≤10%.	dabigatran: n=13,505 rivaroxaban: n=6,551 Weighted pseudo-cohort	Rivaroxaban seemed to be associated with an increased risk of all-cause death compared with dabigatran

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46

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
				secondary codes were used is not described		dabigatran: n=13,508; rivaroxaban: n=6,547	

For peer review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46

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-Khahili, 2016 Sweden	tertiary referral cardiology outpatient clinic (the Stockholm Heart Center) Dec, 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 rates owing mainly to the highest number of minor bleedings, and apixaban had the lowest bleeding rates and side effects
Alonso, 2017 USA	Truven Health MarketScan® Commercial Claims and Encounter Database and the Medicare Supplemental and Coordination of Benefits Database Jan 1, 2007 - Dec 31, 2014	NVAF patients with a first prescription of OAC after Nov 2, 2011.	Age: 67.2±12.4- 69.3±12.5 Male: 60.1-65.1% CHA2DS2-VASc: 2.9-3.6 Comorbidity index: NR	Hospitalization for liver injury potentially related to drug hepatotoxicity, ICD-9-CM codes in any position	rivaroxaban: n=30,347; dabigatran: n=17,286; apixaban: n=9,205	Risk of liver disease hospitalization was higher in rivaroxaban users compared to dabigatran and apixaban users
Chan, 2016 Taiwan	Taiwan National Health Insurance Research Database. Jan 1, 1996 - Dec 31, 2013	NVAF patients newly diagnosed	Age: 75±9- 76 ±9 Male: 54-58 CHA2DS2-VASc: 4.1-4.1 Comorbidity index: NR	Ischemic stroke or systemic embolism, ICH, hospitalization for GI bleeding, acute myocardial infarction (AMI), all hospitalizations 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= 5,921 rivaroxaban 10 mg: n=3,916	No differences were found between rivaroxaban and dabigatran in risk for thromboembolic events, intracranial haemorrhage, critical gastrointestinal bleeding, or all-cause mortality. However, rivaroxaban was associated with a higher risk for noncritical gastrointestinal bleeding than dabigatran

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
Hernandez, 2017 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	Ischemic stroke, death, bleeding events, gastrointestinal bleeding, treatment persistence. ICD-9 codes, whether primary and secondary codes were used is not described	rivaroxaban: n=5,139; apixaban: n=2,358; dabigatran: 1,415;	Apixaban had the most favourable effectiveness and safety profile
Lamberts, 2017 Denmark	Danish national patient registry, Danish national prescription registry, Danish civil personal registry up to December 31, 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=7,963; rivaroxaban: n=6,715;	Apixaban had a lower adjusted major bleeding risk compared with rivaroxaban and dabigatran
Lip, 2016 (Int J Clin Pract) USA	Truven MarketScan® Commercial & Medicare supplemental US database Jan 1, 2013 - Dec 31, 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: 1.7-1.9	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=4,173 apixaban: n=2,402	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, 2019 Scotland	Prescribing Information System, the Scottish Morbidity Records/ Hospital Inpatients and Outpatient attendance datasets; National Records of Scotland Drug's approval date – Dec 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: 1.1-1.4	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=7,265 apixaban: n=6,200; dabigatran: n=1,112;	All NOACs were similarly effective in preventing strokes and systemic embolisms, while patients being treated with rivaroxaban exhibited the highest bleeding risks.
Staerk, 2018 Denmark	Danish national patient registry, Danish national prescription registry,	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	stroke/thromboembolism (TE), ischaemic stroke, major bleeding, intracranial bleeding and gastrointestinal bleeding, ICD-10 codes whether primary and	dabigatran: n=11,492 apixaban: n=11,064 rivaroxaban: n=8966	Rivaroxaban was associated with higher bleeding risk

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
	Danish civil registration system March 1, 2012 - Dec 31, 2016		Comorbidity index: NR	secondary codes were used is not described		compared with dabigatran and apixaban and dabigatran was associated with lower intracranial bleeding risk compared with rivaroxaban and apixaban.
Tepper, 2018 USA	Truven MarketScan Commercial Claims and Encounter and Medicare Supplemental & Coordination of Benefits Early View Database Jan 1, 2013 - Oct 31, 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: 1.6-1.8	Bleeding, ICD-9-CM codes, whether primary and secondary codes were used is not described	rivaroxaban: n=30,529 dabigatran: n=20,963 apixaban: n=8,785;;	Rivaroxaban appeared to have an increased risk of any bleeding, clinically relevant non-major bleeding, and major inpatient bleeding, compared to 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 2018 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= 16,547 apixaban: n= 10,601 dabigatran: n=5,537	Apixaban was associated with a lower risk of major bleed than rivaroxaban. Rivaroxaban was associated with a higher risk of intracranial bleed compared to apixaban. rivaroxaban was associated with higher risks compared with apixaban for haematuria, all gastrointestinal bleed and upper gastrointestinal bleed.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46

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
						The risk of primary ischaemic stroke did not differ between any of the anticoagulants

For peer review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46

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
Cerda, 2019 Spain	Oral Anticoagulant Treatment Unit of the Hemostasis and Thrombosis Department of the University Hospital Vall d'Hebron from Barcelona (Spain) Jan, 2015 - Sept, 2017	NVAF patients with nonvalvular 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 ischemic stroke and intracranial hemorrhage were similar among different NOACs, but rates of major bleeding were higher with dabigatran and apixaban and lower with rivaroxaban.
Li, 2017 China	Queen Mary Hospital, Hong Kong Jan, 2008 - Dec, 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 ischemic 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 ischemic stroke risk compared with patients on rivaroxaban. There was no significant difference in ischemic stroke risk between those on rivaroxaban and dabigatran.

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, 2016 Denmark	Danish national prescription registry, Danish national patient register, Danish civil registration system Feb. 1, 2012 - July 31, 2014	NVAF patients who were new-users of OAC (no OAC treatment in at least the last two 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	ischemic stroke/systemic embolism/transient ischemic 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=8,908 rivaroxaban: n=1,405;	Rivaroxaban and dabigatran had similar stroke rates. Bleeding and mortality rates were higher in rivaroxaban versus dabigatran.

Peer review only

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 (e.g. 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 - N=265,583
Study results	Of the 26 studies in which apixaban, rivaroxaban and dabigatran were included: - apixaban was favourable compared to dabigatran and rivaroxaban: n=13 - no single favourable NOAC: n=13

* about equal distribution between dabigatran, rivaroxaban and apixaban. Edoxaban is not included in these studies.

1
2
3 More than 50% of the studies were conducted in the USA (n=24),[16-39] five were conducted in Denmark,[40-44]
4 four in Taiwan,[45-48] and one in France,[49] Sweden,[50] Scotland,[51] the UK,[52] Spain,[53] and China.[54]
5
6 Dabigatran and rivaroxaban were included in all 39 studies, apixaban was included in 26 studies and edoxaban
7
8 was included in 1 study. Next to these NOACs, VKA was included in 25 of these studies as one of the
9
10 comparators. The results below focus on the NOAC to NOAC comparisons only.

11
12 In the studies that included apixaban, dabigatran and rivaroxaban, rivaroxaban was most dominantly used in the
13
14 USA, UK, Scotland, and Taiwan, while dabigatran was the most prescribed NOAC in Denmark. In three other
15
16 European studies the distribution was about equal between the three NOACs. In none of the included studies,
17
18 apixaban was the most dominantly prescribed NOAC.

21 **Setting**

22 Most studies concerned patient registries, pharmacy or prescription databases and/or health insurance databases
23
24 (n=39), while there were three clinical practice based studies.[50, 53, 54]

27 **Study population**

28
29 All studies included only NVAF patients. In seven studies, it was specifically described that patients were newly
30
31 diagnosed with NVAF and initiated NOAC treatment during study period.[21, 27, 34, 37, 40, 45, 54] None of the
32
33 other studies included prevalent users of (N)OAC, but included e.g. 'newly treated', 'initiating treatment', 'new
34
35 users', 'first-time prescription' of NVAF patients who were prescribed (N)OAC. In some studies (N)OAC use in the
36
37 past (between 3 months and 2 years before index date) was allowed, while this seemed not be allowed in some
38
39 other studies, or it was not described.

41 *Inclusion criteria*

42
43 Five studies concerned elderly patients specifically (i.e. ≥ 65 years old),[19, 21, 23-25] two included adults ≥ 45
44
45 years old,[33, 40] and one study included patients between 30 and 100 years of age.[44] The other studies
46
47 included all adults with atrial fibrillation (it was assumed that if no further age specification was provided, 'adults'
48
49 meant that all >18 years old were included). In one study only patients who were hospitalised for bleeding after
50
51 start with OAC treatment were included.[22] No other focus on a specific group of AF patients was found.

54 *Exclusion criteria*

55
56 NOAC use that could be related to other disorders, such as transient AF, major knee or hip surgery, venous
57
58 thromboembolism or pulmonary embolism, were specifically described as exclusion criteria in most studies,
59
60

1
2
3 except in ten studies.[16, 27, 28, 33-35, 50, 52-54] In one study patients with liver injury before their first oral
4 anticoagulant (OAC) prescription were specifically excluded.[18]
5
6
7

8 *Baseline characteristics*

9
10 Baseline characteristics of the NVAf patients differed between studies. Mean age ranged from 65-84 years
11 between the studies. The percentage of males ranged from 39-73%, and the mean CHA2DS2-Vasc Score ranged
12 from 2.1-4.9. Excluding the five studies that specifically focussed on an elderly population of ≥ 65 years old and
13 the two additional studies that used the Medicare database (only patients of 65 years or older are in Medicare),
14 the mean age ranged from 65-78 years old. Different measures were used to assess the comorbidity index:
15 Charlson comorbidity index, Charlson-Deyo index and Gagne comorbidity score, while in 30 of the 43 studies no
16 comorbidity index was presented.
17
18
19
20
21
22

23 **Selection of covariates**

24
25 Most studies (n=34) did not provide a rationale for the selection of covariates that were included in the PS model
26 or in adjusted analysis. However, in one of the articles an extensive rationale and selection procedure of co-
27 variates that were included in the analysis was provided.[33] In three other studies, the authors selected
28 covariates based on medical knowledge on risk factors with reference to earlier published studies.[31, 39, 52] In
29 one other study it was reported that sociodemographic and clinical characteristics that were associated with
30 treatment initiation and the risk of major bleeding were included in the model to adjust for differences across
31 cohorts, without further explanation or reference.[30]
32
33
34
35
36
37
38

39 **Definition primary study outcomes**

40
41 Primary outcomes differed between the studies. Effectiveness outcomes included in the studies included stroke,
42 systemic embolism, (or composite of stroke/systemic embolism), all-cause death, myocardial infarction, venous
43 thromboembolism and safety outcomes included major bleeding, or a specific type of bleeding (e.g. intracranial
44 haemorrhage, gastrointestinal bleeding etc.) and liver injury. In most studies, ICD-9 or ICD-10 codes were used,
45 but whether this concerned a primary diagnosis only or whether it could be either a primary or a second diagnosis
46 differed between the studies. In some studies it was not described whether the ICD codes referred to primary
47 diagnosis only or to a primary or secondary diagnosis.
48
49
50
51
52
53

54 **Statistical approaches to adjust for confounding (primary analysis)**

55
56 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
57 studies.[17, 22, 24, 25, 28, 43, 46, 48] PS stratified analyses was done in one study.[41] In twelve studies, the
58 primary analyses utilised a Cox PH regression model in which adjustment for confounding was done.[18, 27, 31,
59
60

1
2
3 38, 42, 44, 45, 50-52] Finally, in two studies no adjustment for differences in baseline characteristics was
4 performed.[53, 54]
5
6
7

8 *PS matching*

9 Co-variables

10 Creatinine clearance was not included as a covariate in any of the 18 studies. All 18 studies took the following
11 covariates into account: age, sex, CHA2DS2-VASc score and/or the individual comorbidities included in this
12 score, HAS-BLED score and/or the individual conditions included in this score (except alcohol use in Lai et al.
13 [47]), renal disease, and co-medication use such as antiplatelets. Some included other comorbidities, such as
14 cancer, rheumatic disease, specific heart diseases, COPD, HIV, dementia, depression, neurological disorders,
15 and/or a various list of co-medications as well.
16
17
18
19
20
21
22

23 Matching method

24 In one study the matching method was not described.[49] In two studies, the calliper used was not described.[23,
25 29] In seven studies 1:1 PS matching without replacement was used and a calliper of 0.01 was applied.[16, 19,
26 20, 26, 30, 32, 36] Five other studies also matched 1:1 without replacement but used another calliper: in three
27 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,
28 three-way matching was used.[21, 34, 37]
29
30
31
32
33
34

35 Balance co-variables

36 In two studies it was not described how the balance between covariates was evaluated.[33, 35] In two studies the
37 balance was evaluated using $p < 0.05$ (of which one also used standardized difference of <10%),[23, 47] and in
38 another study it was stated that the groups were comparable even though a p value of > 0.05 was found.[29]
39 Balance was checked with an absolute standardized difference of <10% in 13 studies.[16, 19-21, 26, 30, 32, 34,
40 36, 37, 39, 40, 47, 49] Balance was reached in all studies after matching.
41
42
43
44
45
46
47

48 Sample size

49 In four studies the sample size before matching was not reported[29, 35, 36, 39] and in one study the sample size
50 after matching was not reported.[34] At study start (before PSM), sample size between the NOACs differed
51 greatly, except in three studies.[21, 37, 40]
52
53
54
55
56
57
58
59
60

IPTW

In one study, balance was tested using ANOVAs for significant differences.[22] Balance was checked with an absolute standardized 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 co-variables was reached with standardized difference of <10%. In a Cox model this trimmed PS was used in 10 deciles as strata.[41]

Cox HP regression models

In ten 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 analyses.[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. Cerda et al. presented events per 100 patient-years and used a log-rank test to determine whether outcomes differed between the NOACs.[53] Li et al. conducted a Cox proportional hazard model, likely unadjusted, but this was not clearly described in the article.[54]

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 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 to dabigatran and rivaroxaban in 13 studies, of which 10 were from the USA, two from Europe and one 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 required to confirm that difference.

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

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 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 7,196 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 NVAF patients. 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 sufficiently 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 standardized 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. [14] 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.[14] Therefore, due to the lack of randomization, 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 (i.e. will the differences in effectiveness and safety between NOACs be smaller or larger) of this potential bias due to lack of randomization 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 standard deviation 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

1
2
3 patients from the analysis will limit the generalisability of the results to the total patient population, especially when
4 the excluded patients differ from the included patients, e.g. on the baseline risk for stroke.
5
6
7

8 All included studies focused on NVAF patients only.. In eight studies, inclusion criteria regarding age were
9 applied. Three of these will likely still cover the largest part of NOAC users as they set relatively broad age
10 ranges. The other five focussed on an elderly population of NVAF patients of ≥ 65 years old. Besides applying
11 specific inclusion criteria regarding age in some studies, these differences also depended on the specific registry
12 or database that was used, e.g. Medicare is for people of 65 years old or older. Even though only five of the
13 included studies focused on an elderly NVAF population, and the others applied broad age ranges, there were
14 differences in mean age, proportion of males and mean CHA₂DS₂Vasc score between the studies, which can
15 have an impact on the results and jeopardize the generalisability of the results.
16
17
18
19
20
21
22

23 Rivaroxaban was the most prescribed NOAC in almost all included studies from the USA. However, in the first
24 quarter of 2017, apixaban was the most prescribed NOAC in NVAF in the USA (i.e. in 50% of new OAC
25 prescriptions). Especially older patients, women, increased stroke or bleeding risk and having comorbidities was
26 associated with prescription of apixaban versus other NOACs.[60] Rivaroxaban was also the most prescribed
27 NOAC in the included studies from the UK and Scotland. Based on the CPRD, 56.5% of the OAC prescriptions
28 concerned a NOAC, of which rivaroxaban was still described most often in 2015.[61] Dabigatran was described
29 most often in the studies from Denmark. Haastrup et al. described that most AF patients that initiated NOAC
30 received dabigatran between 2008 and 2016, but a trend was observed that per 1000 person-years the number of
31 patients described dabigatran decreased and the number of patients receiving rivaroxaban and apixaban
32 increased.[62] This shows that the treatment paradigm changed over time, and might still be changing, and this
33 pattern differs between the USA, Europe and Asia. Channelling bias therefore likely occurs and might shift
34 between the NOACs. Although in a few studies it was mentioned that selective prescriptions were noticed and
35 that these might have changed over time, none of the included studies dealt with temporal trends in prescription
36 patterns.
37
38
39
40
41
42
43
44
45
46
47
48
49

50 Our naïve analysis predicts that in terms of the primary efficacy outcome observational studies will need a
51 relatively high number of patients to be able to demonstrate the differences between the NOACs and a small
52 sample size will not allow robust comparison to be made.
53
54
55
56

57 The pattern of major bleeding events seen in the included observational studies, confirms the expectation from
58 our naïve analysis of the pivotal clinical trials that rivaroxaban seems to have the least favourable safety profile
59 among apixaban and dabigatran. The findings are not consistent to allow for a robust conclusion between
60

1
2
3 apixaban and dabigatran which confirms the need for a high number of patients, although a trend for a slight
4 better safety profile of apixaban can be observed.

5
6 The requirement for a high number of patients to compare NOACs both in terms of efficacy and safety as
7 predicted by the pivotal trial results is confirmed by the findings of the observational studies. This finding may
8 support the claim that the differences between the NOACs are relatively small.
9

10
11
12
13 In the process of conducting systematic reviews it is inevitable that the review will never be completely up to date
14 with the most recent published evidence. Even though our search ended in April 2019, recently published studies
15 will have encountered the same issues as described above. Residual confounding and channelling bias cannot
16 have been ruled out in newer publications. Ideally, head-to-head trials should be conducted to compare the
17 efficacy/effectiveness and safety of the four NOACs to overcome the methodological issues in the comparative
18 effectiveness studies. To our knowledge, one head to head trial including all four NOACs is currently running. This
19 nationwide cluster randomized cross-over study aims to compare efficacy and safety of the four NOACs
20 (clinicaltrials.gov; NCT03129490)
21
22
23
24
25
26
27

28
29 In conclusion, even though the larger part of these studies are conducted as well as possible considering what
30 data are available, there are some important limitations regarding the generalisability of the study results
31 especially given the relatively high patient number required for a meaningful comparison between NOACs. Most
32 studies included all NVAf patients on NOAC available in the registry/database during the study period and did not
33 apply further specific in- and exclusion criteria, but differences between studies regarding baseline characteristics
34 existed. Mean age at study start and baseline risk for stroke (CHA₂DS₂-VASc score) differed between the
35 studies. As channelling bias cannot be ruled out, the result of these studies might not be generalisable.
36
37 Furthermore, results from the PS studies are only applicable to the patients that were kept in the analyses as
38 patients excluded from the analysis likely differ from the ones that were included in the analysis. The 1:1 matched
39 cohorts depended on the sample size of the NOAC with the least number of patients and as a result many
40 patients from the larger of the two NOAC groups were excluded as they could not be matched. In clinical practice,
41 these limitations should be kept in mind when results of these studies are used to decide what NOAC should be
42 prescribed for a certain patient. Given the small differences between efficacy and safety outcomes between
43 NOACs, the element of patient preference should be taken into consideration,^[63] as tailoring anti-coagulation
44 treatment towards patient preferences can promote adherence to treatment.
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

FIGURE LEGEND

Figure 1: PRISMA flowchart

For peer review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 **ACKNOWLEDGEMENTS**
4

5 The authors would like to thank Pearl Gumbs for the initiation of this project and her input with regard to study
6 design and interpretation of the data. Pearl Gumbs was working at Daiichi Sankyo Europe at that time.
7
8
9

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

COMPETING INTEREST

EB reports grants from Daiichi Sankyo, during the conduct of the study; grants from Daiichi Sankyo, outside the submitted work. BvH 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 Inc, personal fees from Janssen, personal fees from ONO Pharmaceuticals, from AbbVie, outside the submitted work.

1
2
3 **AUTHORS' CONTRIBUTION**

4 EB: Conceptualization (support); Methodology (equal); Writing – Original Draft Preparation; Writing – Review &
5 Editing (equal). BvH: Conceptualization (support); Methodology (equal); Writing – Review & Editing (equal). SH:
6 Conceptualization (support); Writing – Review & Editing (equal). GS: Conceptualization (support), Supervision;
7 Writing – Review & Editing (equal). ATC: Conceptualization (lead); Writing – Review & Editing (equal)
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

1
2
3 **DATA SHARING STATEMENT**
4

5 No new data were generated or analysed in support of this research.
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

FUNDING

This work was supported by Daiichi Sankyo Europe. GRANT / AWARD NUMBER: N/A

For peer review only

PATIENT AND PUBLIC INVOLVEMENT

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.

For peer review only

REFERENCES

1. Kirchhof P, Benussi S, Kotecha D, et al. 2016 ESC Guidelines for the management of atrial fibrillation developed in collaboration with EACTS. *European heart journal*. 2016;37(38):2893-962 10.1093/eurheartj/ehw210.
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(2):e125-e51 10.1161/CIR.0000000000000665.
3. Connolly SJ, Ezekowitz MD, Yusuf S, et al. Dabigatran versus warfarin in patients with atrial fibrillation. *The New England journal of medicine*. 2009;361(12):1139-51 10.1056/NEJMoa0905561.
4. Giugliano RP, Ruff CT, Braunwald E, et al. Edoxaban versus warfarin in patients with atrial fibrillation. *The New England journal of medicine*. 2013;369(22):2093-104 10.1056/NEJMoa1310907.
5. Granger CB, Alexander JH, McMurray JJ, et al. Apixaban versus warfarin in patients with atrial fibrillation. *The New England journal of medicine*. 2011;365(11):981-92 10.1056/NEJMoa1107039.
6. Patel MR, Mahaffey KW, Garg J, et al. Rivaroxaban versus warfarin in nonvalvular atrial fibrillation. *The New England journal of medicine*. 2011;365(10):883-91 10.1056/NEJMoa1009638.
7. Petri H, Urquhart J. Channeling bias in the interpretation of drug effects. *Stat Med*. 1991;10(4):577-81 10.1002/sim.4780100409.
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. *Journal of clinical epidemiology*. 1995;48(12):1503-10 10.1016/0895-4356(95)00048-8.
9. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70(1):41-55.
10. Brookhart MA, Schneeweiss S, Rothman KJ, et al. Variable selection for propensity score models. *Am J Epidemiol*. 2006;163(12):1149-56 10.1093/aje/kwj149.
11. D'Agostino RB, Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med*. 1998;17(19):2265-81 10.1002/(sici)1097-0258(19981015)17:19<2265::aid-sim918>3.0.co;2-b.
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(4):1654-70 10.1177/0962280215584401.
13. Austin PC. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Stat Med*. 2008;27(12):2037-49 10.1002/sim.3150.
14. 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 in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research*. 2009;12(8):1062-73 10.1111/j.1524-4733.2009.00602.x.
15. 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(8) 10.1093/jnci/djw323.
16. Abraham NS, Noseworthy PA, Yao X, et al. Gastrointestinal Safety of Direct Oral Anticoagulants: A Large Population-Based Study. *Gastroenterology*. 2017;152(5):1014-22.e1 10.1053/j.gastro.2016.12.018.
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. *Journal of managed care & specialty pharmacy*. 2017;23(9):968-78 10.18553/jmcp.2017.23.9.968.

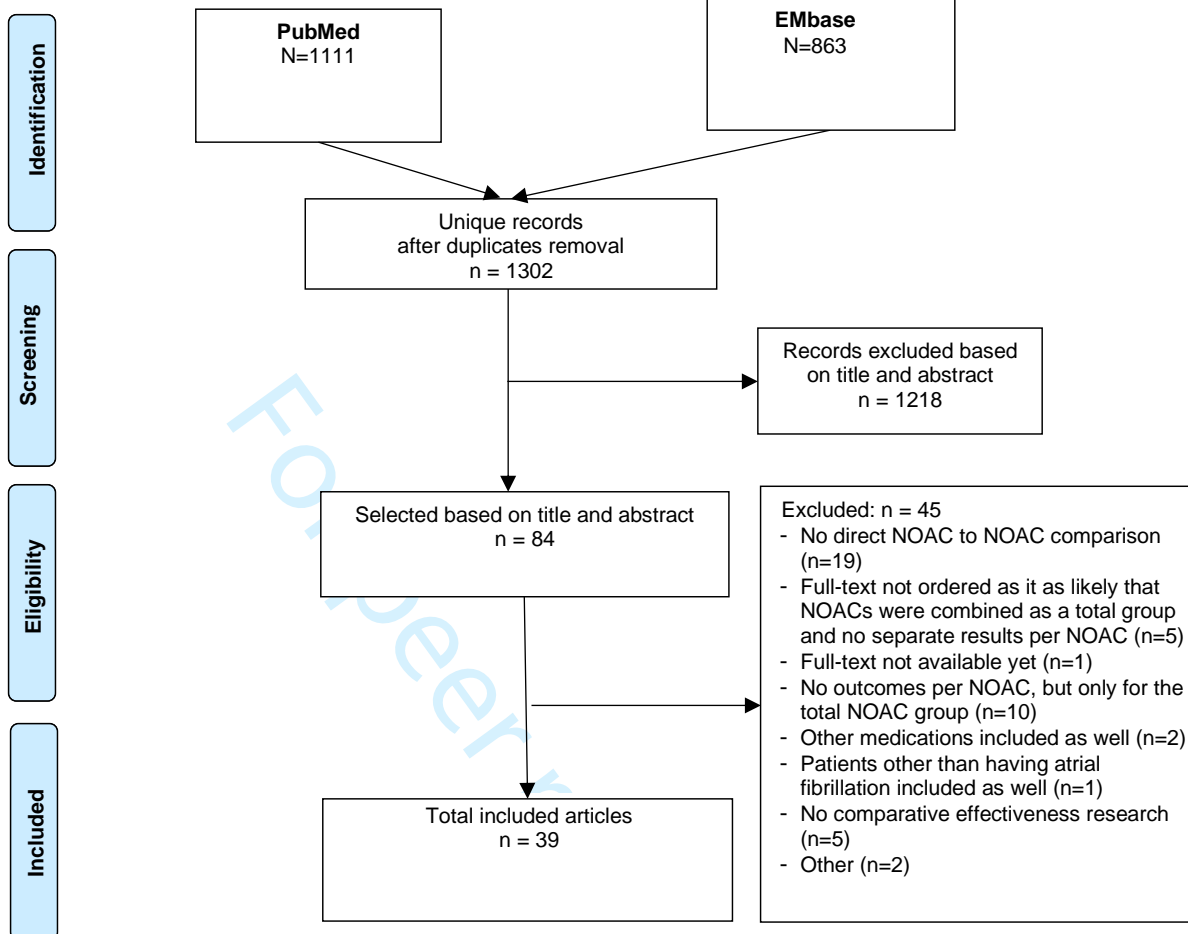
18. Alonso A, MacLehose RF, Chen LY, et al. Prospective study of oral anticoagulants and risk of liver injury in patients with atrial fibrillation. *Heart (British Cardiac Society)*. 2017;103(11):834-9 10.1136/heartjnl-2016-310586.
19. Amin A, Keshishian A, Trocio J, et al. A Real-World Observational Study of Hospitalization and Health Care Costs Among Nonvalvular Atrial Fibrillation Patients Prescribed Oral Anticoagulants in the U.S. Medicare Population. *Journal of managed care & specialty pharmacy*. 2018;24(9):911-20 10.18553/jmcp.2018.24.9.911.
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. *Journal of medical economics*. 2018;21(3):244-53 10.1080/13696998.2017.1394866.
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. *Journal of the American Heart Association*. 2018;7(8) 10.1161/jaha.118.008773.
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(3):e0193912 10.1371/journal.pone.0193912.
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. *Current medical research and opinion*. 2017;33(10):1745-54 10.1080/03007995.2017.1334638.
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. *The American journal of medicine*. 2019 10.1016/j.amjmed.2018.12.023.
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 internal medicine*. 2016;176(11):1662-71 10.1001/jamainternmed.2016.5954.
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. *Journal of managed care & specialty pharmacy*. 2018;24(11):1116-27 10.18553/jmcp.2018.17488.
27. Hernandez I, Zhang Y, Saba S. Comparison of the Effectiveness and Safety of Apixaban, Dabigatran, Rivaroxaban, and Warfarin in Newly Diagnosed Atrial Fibrillation. *The American journal of cardiology*. 2017;120(10):1813-9 10.1016/j.amjcard.2017.07.092.
28. Hernandez IZY. Comparing Stroke and Bleeding with Rivaroxaban and Dabigatran in Atrial Fibrillation: Analysis of the US Medicare Part D Data. *American journal of cardiovascular drugs : drugs, devices, and other interventions*. 2017;17(1):37-47 10.1007/s40256-016-0189-9 FULL TEXT LINK <http://dx.doi.org/10.1007/s40256-016-0189-9>.
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. *Journal of medical economics*. 2017;20(9):952-61 10.1080/13696998.2017.1341902.
30. Lip GY, 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. *Thrombosis and haemostasis*. 2016;116(5):975-86 10.1160/th16-05-0403.
31. Lip GY, 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. *International journal of clinical practice*. 2016;70(9):752-63 10.1111/ijcp.12863.
32. Lip GYH, Keshishian A, Li X, et al. Effectiveness and Safety of Oral Anticoagulants Among Nonvalvular Atrial Fibrillation Patients. *Stroke*. 2018;49(12):2933-44 10.1161/STROKEAHA.118.020232.

- 1
2
3 33. Lutsey PL, Norby FL, Zakai NA, et al. Oral anticoagulation therapy and subsequent risk of
4 venous thromboembolism in atrial fibrillation patients. *Current medical research and opinion*.
5 2019;35(5):837-45 10.1080/03007995.2018.1541445.
- 6 34. Mentias A, Shantha G, Chaudhury P, et al. Assessment of Outcomes of Treatment With Oral
7 Anticoagulants in Patients With Atrial Fibrillation and Multiple Chronic Conditions: A Comparative
8 Effectiveness Analysis. *JAMA network open*. 2018;1(5):e182870
9 10.1001/jamanetworkopen.2018.2870.
- 10 35. Norby FL, Bengtson LGS, Lutsey PL, et al. Comparative effectiveness of rivaroxaban versus
11 warfarin or dabigatran for the treatment of patients with non-valvular atrial fibrillation. *BMC*
12 *cardiovascular disorders*. 2017;17(1):238 10.1186/s12872-017-0672-5.
- 13 36. Noseworthy PA, Yao X, Abraham NS, et al. Direct Comparison of Dabigatran, Rivaroxaban,
14 and Apixaban for Effectiveness and Safety in Nonvalvular Atrial Fibrillation. *Chest*. 2016;150(6):1302-
15 12 10.1016/j.chest.2016.07.013.
- 16 37. Shantha PSG, Bhave PD, Girotra S, et al. Sex-Specific Comparative Effectiveness of Oral
17 Anticoagulants in Elderly Patients With Newly Diagnosed Atrial Fibrillation. *Circulation Cardiovascular*
18 *quality and outcomes*. 2017;10(4) 10.1161/circoutcomes.116.003418.
- 19 38. Tepper PG, Mardekian J, Masseria C, et al. Real-world comparison of bleeding risks among
20 non-valvular atrial fibrillation patients prescribed apixaban, dabigatran, or rivaroxaban. *PloS one*.
21 2018;13(11):e0205989 10.1371/journal.pone.0205989.
- 22 39. Villines TCAAPMTWEARTDOKSE. Comparative safety and effectiveness of dabigatran vs.
23 rivaroxaban and apixaban in patients with non-valvular atrial fibrillation: A retrospective study from a
24 large healthcare system. *Eur Heart J - Cardiovascular Pharmacotherapy*. 2019;5(2):80-90
25 10.1093/ehjcvp/pvy044 FULL TEXT LINK <http://dx.doi.org/10.1093/ehjcvp/pvy044>.
- 26 40. Andersson NW, Svanstrom H, Lund M, et al. Comparative effectiveness and safety of
27 apixaban, dabigatran, and rivaroxaban in patients with non-valvular atrial fibrillation. *International*
28 *journal of cardiology*. 2018;268:113-9 10.1016/j.ijcard.2018.03.047.
- 29 41. Gorst-Rasmussen A, Lip GY, Bjerregaard Larsen T. Rivaroxaban versus warfarin and
30 dabigatran in atrial fibrillation: comparative effectiveness and safety in Danish routine care.
31 *Pharmacoepidemiology and drug safety*. 2016;25(11):1236-44 10.1002/pds.4034.
- 32 42. Lamberts M, Staerk L, Olesen JB, et al. Major Bleeding Complications and Persistence With
33 Oral Anticoagulation in Non-Valvular Atrial Fibrillation: Contemporary Findings in Real-Life Danish
34 Patients. *Journal of the American Heart Association*. 2017;6(2) 10.1161/jaha.116.004517.
- 35 43. Larsen TB, Skjoth F, Nielsen PB, et al. Comparative effectiveness and safety of non-vitamin K
36 antagonist oral anticoagulants and warfarin in patients with atrial fibrillation: propensity weighted
37 nationwide cohort study. *BMJ (Clinical research ed)*. 2016;353:i3189 10.1136/bmj.i3189.
- 38 44. Staerk L, Gerds TA, Lip GYH, et al. Standard and reduced doses of dabigatran, rivaroxaban
39 and apixaban for stroke prevention in atrial fibrillation: a nationwide cohort study. *Journal of internal*
40 *medicine*. 2018;283(1):45-55 10.1111/joim.12683.
- 41 45. Chan YH, Kuo CT, Yeh YH, et al. Thromboembolic, Bleeding, and Mortality Risks of
42 Rivaroxaban and Dabigatran in Asians With Nonvalvular Atrial Fibrillation. *Journal of the American*
43 *College of Cardiology*. 2016;68(13):1389-401 10.1016/j.jacc.2016.06.062.
- 44 46. Chan YH, See LC, Tu HT, et al. Efficacy and Safety of Apixaban, Dabigatran, Rivaroxaban, and
45 Warfarin in Asians With Nonvalvular Atrial Fibrillation. *Journal of the American Heart Association*.
46 2018;7(8) 10.1161/jaha.117.008150.
- 47 47. Lai CL, Chen HM, Liao MT, et al. Comparative Effectiveness and Safety of Dabigatran and
48 Rivaroxaban in Atrial Fibrillation Patients. *Journal of the American Heart Association*. 2017;6(4)
49 10.1161/jaha.116.005362.
- 50 48. Meng SW, Lin TT, Liao MT, et al. Direct Comparison of Low-Dose Dabigatran and Rivaroxaban
51 for Effectiveness and Safety in Patients with Non-Valvular Atrial Fibrillation. *Acta Cardiologica Sinica*.
52 2019;35(1):42-54 10.6515/acs.201901_35(1).20180817a.
- 53
54
55
56
57
58
59
60

- 1
2
3 49. Blin PD-PCCYBJMPAALRDCMN. Comparative Effectiveness and Safety of Standard or Reduced
4 Dose Dabigatran vs. Rivaroxaban in Nonvalvular Atrial Fibrillation. *Clin Pharm Therapeutics*.
5 2019;105(6):1439-55 10.1002/cpt.1318 FULL TEXT LINK <http://dx.doi.org/10.1002/cpt.1318>.
6
7 50. Al-Khalili F, Lindstrom C, Benson L. The safety and persistence of non-vitamin-K-antagonist
8 oral anticoagulants in atrial fibrillation patients treated in a well structured atrial fibrillation clinic.
9 *Current medical research and opinion*. 2016;32(4):779-85 10.1185/03007995.2016.1142432.
10
11 51. Mueller T, Alvarez-Madrado S, Robertson C, et al. Comparative safety and effectiveness of
12 direct oral anticoagulants in patients with atrial fibrillation in clinical practice in Scotland. *British*
13 *journal of clinical pharmacology*. 2019;85(2):422-31 10.1111/bcp.13814.
14
15 52. Vinogradova Y, Coupland C, Hill T, et al. Risks and benefits of direct oral anticoagulants versus
16 warfarin in a real world setting: cohort study in primary care. *BMJ (Clinical research ed)*.
17 2018;362:k2505 10.1136/bmj.k2505.
18
19 53. Cerda M, Cerezo-Manchado JJ, Johansson E, et al. Facing real-life with direct oral
20 anticoagulants in patients with nonvalvular atrial fibrillation: outcomes from the first observational
21 and prospective study in a Spanish population. *Journal of comparative effectiveness research*.
22 2019;8(3):165-78 10.2217/cer-2018-0134.
23
24 54. Li WH, Huang D, Chiang CE, et al. Efficacy and safety of dabigatran, rivaroxaban, and warfarin
25 for stroke prevention in Chinese patients with atrial fibrillation: the Hong Kong Atrial Fibrillation
26 Project. *Clinical cardiology*. 2017;40(4):222-9 10.1002/clc.22649.
27
28 55. <https://clincalc.com/stats/samplesize.aspx> [
29
30 56. D.E. H, K. I, G. K, et al. Matching as nonparametric preprocessing for reducing model
31 dependence in parametric causal inference. *Political Analysis*. 2007;15:199-236.
32
33 57. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*.
34 1999;10(1):37-48.
35
36 58. Bohula EA, Giugliano RP, Ruff CT, et al. Impact of Renal Function on Outcomes With
37 Edoxaban in the ENGAGE AF-TIMI 48 Trial. *Circulation*. 2016;134(1):24-36
38 10.1161/circulationaha.116.022361.
39
40 59. Austin PC, Stuart EA. Estimating the effect of treatment on binary outcomes using full
41 matching on the propensity score. *Stat Methods Med Res*. 2017;26(6):2505-25
42 10.1177/0962280215601134.
43
44 60. Zhu J, Alexander GC, Nazarian S, et al. Trends and Variation in Oral Anticoagulant Choice in
45 Patients with Atrial Fibrillation, 2010-2017. *Pharmacotherapy*. 2018;38(9):907-20
46 10.1002/phar.2158.
47
48 61. Loo SY, Dell'Aniello S, Huiart L, et al. Trends in the prescription of novel oral anticoagulants in
49 UK primary care. *British journal of clinical pharmacology*. 2017;83(9):2096-106 10.1111/bcp.13299.
50
51 62. Haastrup SB, Hellfritzsch M, Rasmussen L, et al. Use of Non-Vitamin K Antagonist Oral
52 Anticoagulants 2008-2016: A Danish Nationwide Cohort Study. *Basic & clinical pharmacology &*
53 *toxicology*. 2018;123(4):452-63 10.1111/bcpt.13024.
54
55 63. Vaanholt MCW, Weernink MGM, von Birgelen C, et al. Perceived advantages and
56 disadvantages of oral anticoagulants, and the trade-offs patients make in choosing anticoagulant
57 therapy and adhering to their drug regimen. *Patient Educ Couns*. 2018;101(11):1982-9
58 10.1016/j.pec.2018.06.019.
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only



The following search string was used for PubMed, and adapted for Cochrane and EMBase.

#1. NOAC

direct oral anticoagulant*[tiab] OR direct oral anti-coagulant*[tiab] OR direct oral anticoagulation[tiab] OR direct oral anti-coagulation[tiab] OR direct-acting oral anticoagulant*[tiab] OR direct-acting oral anti-coagulant*[tiab] OR direct-acting oral anticoagulation[tiab] OR direct-acting oral anti-coagulation[tiab] OR DOAC[tiab] OR novel oral anticoagulant*[tiab] OR novel oral anti-coagulant*[tiab] OR Novel oral anticoagulation[tiab] OR Novel oral anti-coagulation[tiab] OR NOAC[tiab] OR Rivaroxaban[tiab] OR Apixaban[tiab] OR Edoxaban[tiab] OR Dabigatran[tiab] OR "Non VKA Oral Anticoagulant"[tiab] OR "Non Vitamin K Antagonist Oral Anticoagulant"[tiab]

#2. Comparative effectiveness studies

comparative effectiveness research[mesh] OR comparative effectiveness[tiab] OR real-world[tiab] OR real-life[tiab] OR cohort studies[mesh] OR cohort[tiab]

#3. Atrial fibrillation

atrial fibrillation[tiab]

Limits:

Language: English



PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	3,4
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	4
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	NA
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	5,6
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	5,6
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	S1_Table
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	5
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	5
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	5
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	5,6
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	NA
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I ²) for each meta-analysis.	NA



PRISMA 2009 Checklist

Page 1 of 2

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	NA
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	NA
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	S2_Fig
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	8-21
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	8-27
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	NA
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	NA
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	NA
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	NA
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	30
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	30,31
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	31,32
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: www.prisma-statement.org.

Page 2 of 2

For peer review only - <http://bmjopen.bmj.com/site/about/guidelines.xhtml>