openheart

Sex-specific comparative outcomes between oral anticoagulants in patients with atrial fibrillation: a systematic review and meta-analysis

Jan D Chobanov ,[1,2] Zixuan Wang,[3,4,5] Kenneth K C Man,[1,2,5,6] Edil Dayib,[1] Gregory Y H Lip,[7,8] Aroon D Hingorani,[9] Wai K Leung,[10] Ian C K Wong,[1,2,5,6,11] Pajaree Mongkhon,[12,13] Wallis C Y Lau [1,2,5,6]

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/openhrt-2024-002792).

To cite: Chobanov JD, Wang Z, Man KKC, *et al*. Sex-specific comparative outcomes between oral anticoagulants in patients with atrial fibrillation: a systematic review and meta-analysis. *Open Heart* 2024;**11**:e002792. doi:10.1136/openhrt-2024-002792

For 'Presented at statement' see end of article.

Received 11 June 2024
Accepted 2 July 2024

For numbered affiliations see end of article.

Correspondence to
Dr Wallis C Y Lau; wallis.lau@ucl.ac.uk

Dr Pajaree Mongkhon; pajaree.mo@up.ac.th

ABSTRACT

Aims Women with atrial fibrillation (AF) are under-represented in randomised controlled trials (RCTs) of direct oral anticoagulants (DOACs). This systematic review and meta-analysis of RCTs and observational studies examined sex-specific outcomes of DOACs in AF.

Methods PubMed, Embase, Web of Science and Cochrane Library were searched from January 2008 to November 2022. Sex-specific comparative outcomes of stroke/systemic embolism (SE), major bleeding, intracranial haemorrhage (ICH) and gastrointestinal bleeding (GIB) between oral anticoagulants were pooled using random effects models. P values for interaction were calculated to examine differences in results between sexes. RCTs and observational studies were meta-analysed separately.

Results 5 RCTs and 33 observational studies were included, totalling 1 085 931 women and 1 387 123 men. Meta-analyses showed that for both sexes, DOAC versus warfarin was generally associated with lower risk of stroke/SE, major bleeding and ICH; in DOAC–DOAC comparisons, rivaroxaban versus dabigatran had higher GIB risk. The only sex-specific difference observed was that when compared with warfarin, women had higher GIB risk with rivaroxaban (women: pooled risk ratio (pRR)=1.34, 95% CI=1.18 to 1.51; men: pRR=0.97, 95% CI=0.85 to 1.10; p value for interaction (p for interaction)<0.001) and possibly dabigatran (women: pRR=1.25, 95% CI=0.92 to 1.70; men: pRR=0.83, 95% CI=0.72 to 0.97; p-for-interaction=0.02). The sex difference in GIB remained for rivaroxaban when a Bonferroni-corrected significance level was used (α=0.003). No sex-specific GIB data for apixaban and edoxaban was available for the meta-analysis.

Conclusions For both sexes, DOACs generally demonstrated favourable effectiveness and safety over warfarin. However, observational data suggested that women may have higher GIB risk with rivaroxaban and possibly dabigatran than warfarin. Further studies are warranted to verify our findings and elucidate sex-specific GIB risk with apixaban and edoxaban, of which the data is currently lacking.

PROSPERO registration number CRD42022325027.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Randomised controlled trials (RCTs) in patients with atrial fibrillation have demonstrated direct oral anticoagulants (DOACs) to be at least as effective as warfarin in reducing stroke with lower overall bleeding risk. However, the RCTs under-represented women and were not designed to investigate sex-specific outcomes, obscuring potential sex-specific differences in the effects of DOACs.

WHAT THIS STUDY ADDS

⇒ This systematic review and meta-analysis of RCTs and observational studies found both sexes to generally demonstrate favourable safety and effectiveness with DOACs compared with warfarin, but observational data indicates that gastrointestinal bleeding (GIB) risk may be raised in women with rivaroxaban and dabigatran compared with warfarin.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Data from observational studies suggests that GIB risk may differ with the types of DOAC in women. Further research studies are warranted to verify our findings and elucidate sex-specific GIB risk with apixaban and edoxaban, of which the data is currently lacking.

INTRODUCTION

Atrial fibrillation (AF) is the most common sustained cardiac arrhythmia worldwide.[1] Women with AF have higher risks of stroke than men.[2] Sex is therefore considered a risk modifier for stroke in AF, informing the decision to include women in the CHA2DS2-VASc Score.[3 4] Higher stroke risk in women could reflect differing pathophysiological mechanisms for stroke[5] and sex-specific interactions with the pharmacodynamics and pharmacokinetics of cardiovascular drugs, particularly warfarin.[6] However, whether there are sex differences in the effects of direct oral

anticoagulants (DOACs), which are currently recommended for use over warfarin,[7 8] is unclear.[6]

In randomised controlled trials (RCTs), DOACs are at least as effective as warfarin in reducing stroke with lower overall bleeding risk.[9–12] However, as the RCTs were not designed to have adequate power to investigate sex-specific outcomes, important sex-based interactions with DOACs could have been undetected. Women have been under-represented in RCTs assessing DOACs, and the generalisability of RCT findings to real-world practice is limited by the strict eligibility criteria.[13] Although recent observational studies have contributed data on sex-specific DOAC effectiveness and safety, a comprehensive assessment of the sex-specific outcomes of DOACs from the available evidence is lacking.

This systematic review and meta-analysis aimed to summarise the published evidence from RCTs and observational studies to compare the sex-specific effectiveness and safety between DOACs and warfarin. We also examined if the outcomes vary between anticoagulant users from different geographical regions.

METHODS

This study was conducted in accordance with the 2020 Preferred Reporting Items for Systematic Reviews and Meta-analyses statement.[14] The protocol was registered in PROSPERO, the international prospective register of systematic reviews, (CRD42022325027).

To clarify, the terminology 'sex' is used and not 'gender'. When mentioning sex, we are referring to the biology of living things, that is, biological features, such as chromosomal genotypes and reproductive organs that distinguish men and women at birth.

Data sources and search strategy

A systematic literature search was conducted through PubMed, Embase, Web of Science and Cochrane Library for studies published from 1 January 2008, the year when the first DOAC (dabigatran) was marketed, to 23 November 2022. Full search strategies are available in online supplemental tables S1–S4.

Study selection

Three investigators (PM, ED and JDC) independently screened the titles and abstracts of all identified records and screened the full texts of the potentially relevant articles to assess their eligibility. The reference lists of the included studies, prior systematic reviews and introduction and discussion sections of retrieved studies were also reviewed to identify additional relevant studies. Disagreements were resolved by discussion or consultation with a fourth investigator (WL).

Eligibility criteria

Studies were included if they: (1) were RCTs or longitudinal observational studies; (2) were conducted in patients with AF who received oral anticoagulant treatment; and (3) compared stroke or systemic embolism (SE), or bleeding outcomes between any DOAC (dabigatran, rivaroxaban, apixaban and edoxaban) and warfarin or other vitamin K antagonists (VKAs) in men and women. The primary outcome was stroke/SE. The secondary outcomes were bleeding which included major bleeding, intracranial haemorrhage (ICH), gastrointestinal bleeding (GIB) and any bleeding. Studies which did not explicitly define their bleeding outcomes as major bleeding, ICH or GIB and included other bleeding events or a composite of bleeding outcomes were classified as any bleeding. Outcome definitions as reported by each included study can be found in online supplemental table S5.

Studies were excluded if they were: (1) reviews or systematic reviews, cross-sectional studies, case reports, conference abstracts, editorials or commentaries, (2) animal or in vitro studies, (3) not published in English or (4) did not report sex-specific outcomes.

Data extraction

Three investigators (PM, ZW and JDC) extracted the data independently using prespecified forms. We gathered data on (1) study characteristics; (2) patient characteristics; (3) specific intervention/exposure group (DOAC type and dosage) and control groups; and (4) outcomes of interest and follow-up. Studies with incomplete data were clarified by contacting the corresponding author where possible. When authors did not respond, we used information reported to calculate the required data or excluded the study from the meta-analyses.

Quality assessment

Three investigators (PM, ZW and JDC) independently appraised the quality of the included studies using the revised Cochrane Risk of Bias Tool for Randomised Trials (RoB V.2.0)[15] and the Newcastle-Ottawa Scale for observational studies (see online supplemental appendix 1 for full details).[16]

Statistical analyses

In the primary meta-analyses, we pooled the results of studies that reported outcomes for all DOAC users as a group. Prespecified subgroup analyses were performed on individual DOACs and geographical regions of the study populations where data permitted (Asia, Europe, North America). Subgroup analyses were only possible with observational data as the number of RCTs was too small. Post hoc analyses for DOAC head-to-head comparisons were performed.

RCTs and observational data were analysed separately. Valvular heart disease was analysed separately from patients without valvular disease. For observational studies, we extracted results which had the greatest adjustment for potential confounding factors. The results from all included studies were expressed as hazard ratios (HRs) or risk ratios (RRs). HRs were considered comparable to RRs.[17] The DerSimonian and Laird random effects model was used to estimate sex-specific pooled RRs (pRR) with 95% confidence intervals (CI) as the common effect

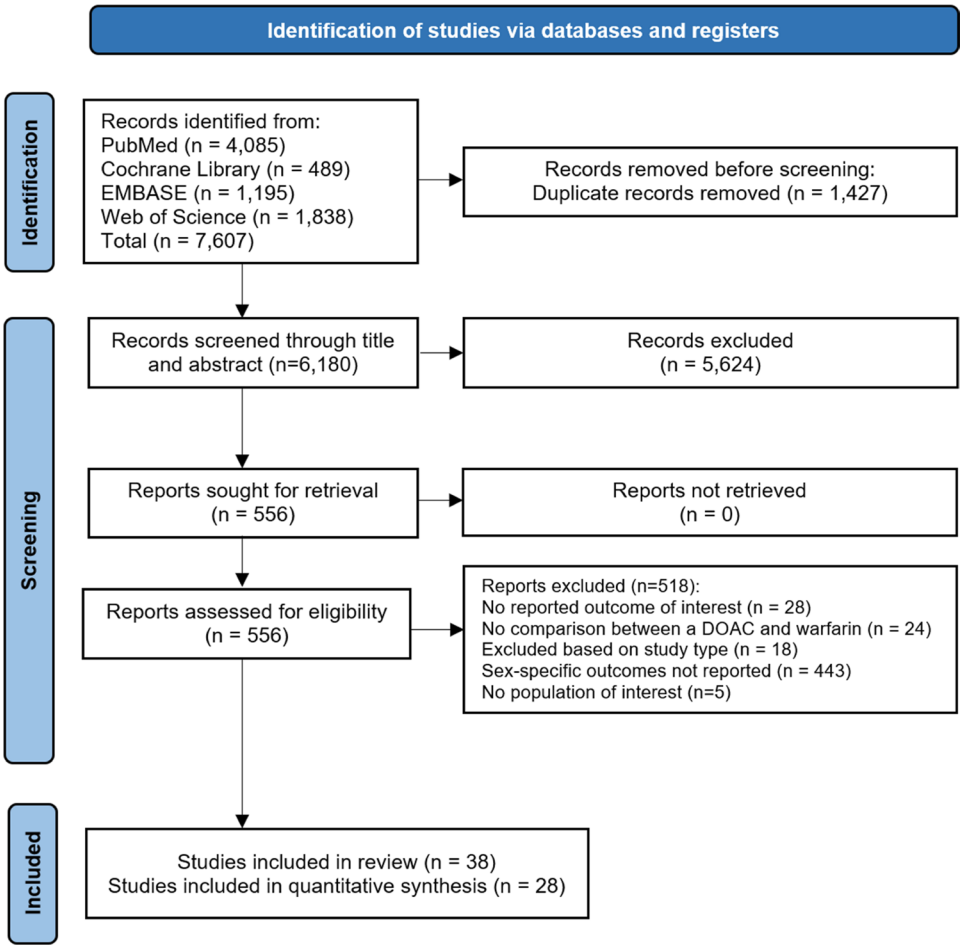


Figure 1 Study selection flowchart. DOAC, direct oral anticoagulant.

estimate. Heterogeneity between studies was investigated using I^2 with low (I^2<25%), moderate (I^2=25%–75%) and high (I^2>75%) thresholds. A p value for interaction (p-for-interaction) was calculated to assess differences in pRR between sexes and geographical regions. A p-for-interaction<0.1 indicated a statistically significant subgroup difference.[18] Post hoc, we applied Bonferroni corrected significance levels of 0.003 and 0.001 for the sex-specific oral anticoagulant and geographical region comparisons, respectively (online supplemental appendix 2). Studies ineligible for meta-analysis due to incomplete data or overlapping study populations were narratively reviewed (online supplemental appendix 3).

Analyses were conducted using R V.4.2.2. Risk of bias plot of RoB V.2.0 was created by robvis.[19]

RESULTS

Study selection and baseline characteristics

6180 unique records were identified, of which 38 studies met the inclusion criteria and were included in the systematic review and 28 were included in the meta-analyses (figure 1). 5 RCTs and 33 observational studies were included in the systematic review (online supplemental table S6). The RCTs were all multicentre and international studies. Four RCTs were large (n≥14 263) and conducted in patients with AF. One smaller-sized RCT (n=1426) was conducted in patients with AF after a successful transcatheter aortic valve replacement. Collectively, the RCTs had 45 713 men and 27 396 women.

All observational studies were cohort study designs using data from national administrative/clinical databases, medical institutions or stroke centres. 16 observational studies were conducted in North America, 11 in Asia and 6 in Europe. Most observational studies were conducted in an unselected AF population. Selected AF populations included patients with type 2 diabetes mellitus, chronic kidney disease, bioprosthetic heart valves, liver disease, patients aged≥80 years and patients with body mass index>30 kg/m^2.

Four RCTs were eligible for meta-analysis for stroke/SE (44 965 men and 26 718 women) and three for bleeding outcomes (33 451 men and 20 119 women). 19 observational studies including 8 02 483 men and 6 56 375 women and 24 observational studies including 1 076 058 men and 7 54 115 women were eligible for meta-analyses on stroke/SE and bleeding outcomes, respectively. Warfarin was considered the comparator group in the meta-analysis as only a minority of patients from two observational studies may have included VKAs other than warfarin.[20 21]

Quality assessment

Three of the five RCTs were judged as low risk of bias and two were rated as some concerns (online supplemental table S7 and online supplemental figure S1). For observational studies, 31 out of 33 received a good quality rating and two studies received a fair quality rating (online supplemental table S8 and online supplemental appendix 1).

Sex-specific outcomes for DOACs versus warfarin

Stroke/SE

Meta-analysis of four RCTs showed both sexes had a lower risk of stroke/SE using DOACs versus warfarin, with no evidence of sex-specific interaction (women: pRR=0.79, 95% CI=0.66 to 0.94, I^2=36%; men: pRR=0.84, 95% CI=0.75 to 0.93, I^2=0%; p-for-interaction=0.60). Results were similar for observational studies (women: pRR=0.75, 95% CI=0.58 to 0.97, I^2=85%; men: pRR=0.81, 95% CI=0.67 to 0.98, I^2=76%; p-for-interaction=0.64) (figure 2).

Major bleeding, GIB and ICH

Meta-analysis of three RCTs suggest DOACs have lower risk of major or non-major clinically relevant bleeding versus warfarin in women but not in men (women: pRR=0.75, 95% CI=0.59 to 0.94, I^2=77%; men: pRR=0.91, 95% CI=0.71 to 1.16, I^2=90%). There was no statistical difference between the sex-specific estimates (p-for-interaction=0.27). In the meta-analysis of observational studies, both sexes had lower risks of major bleeding with DOACs than warfarin (women: pRR=0.87, 95% CI=0.77 to 0.97, I^2=65%; men: pRR=0.79, 95% CI=0.68 to 0.91, I^2=81%; p-for-interaction=0.33) (figure 3).

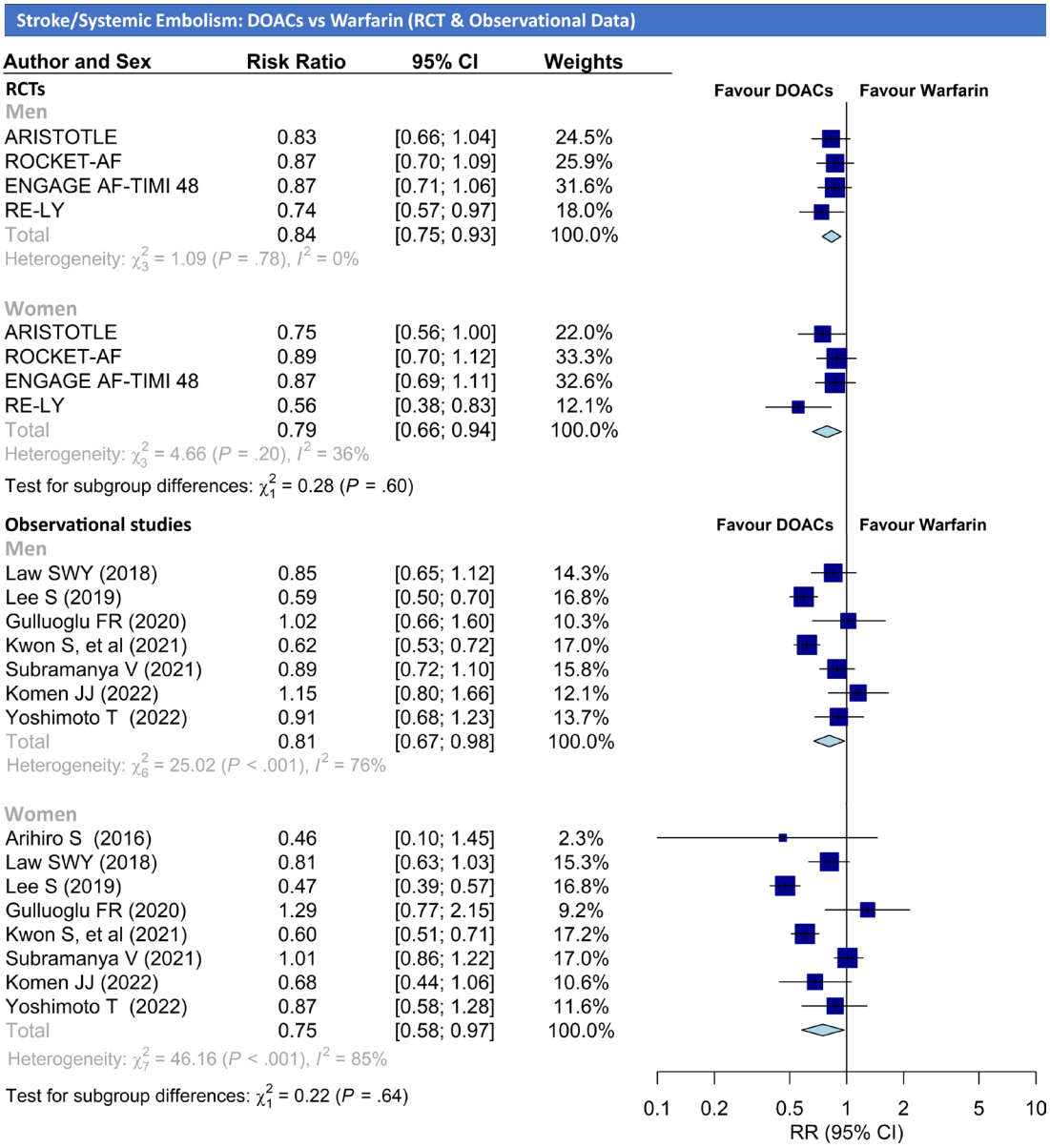
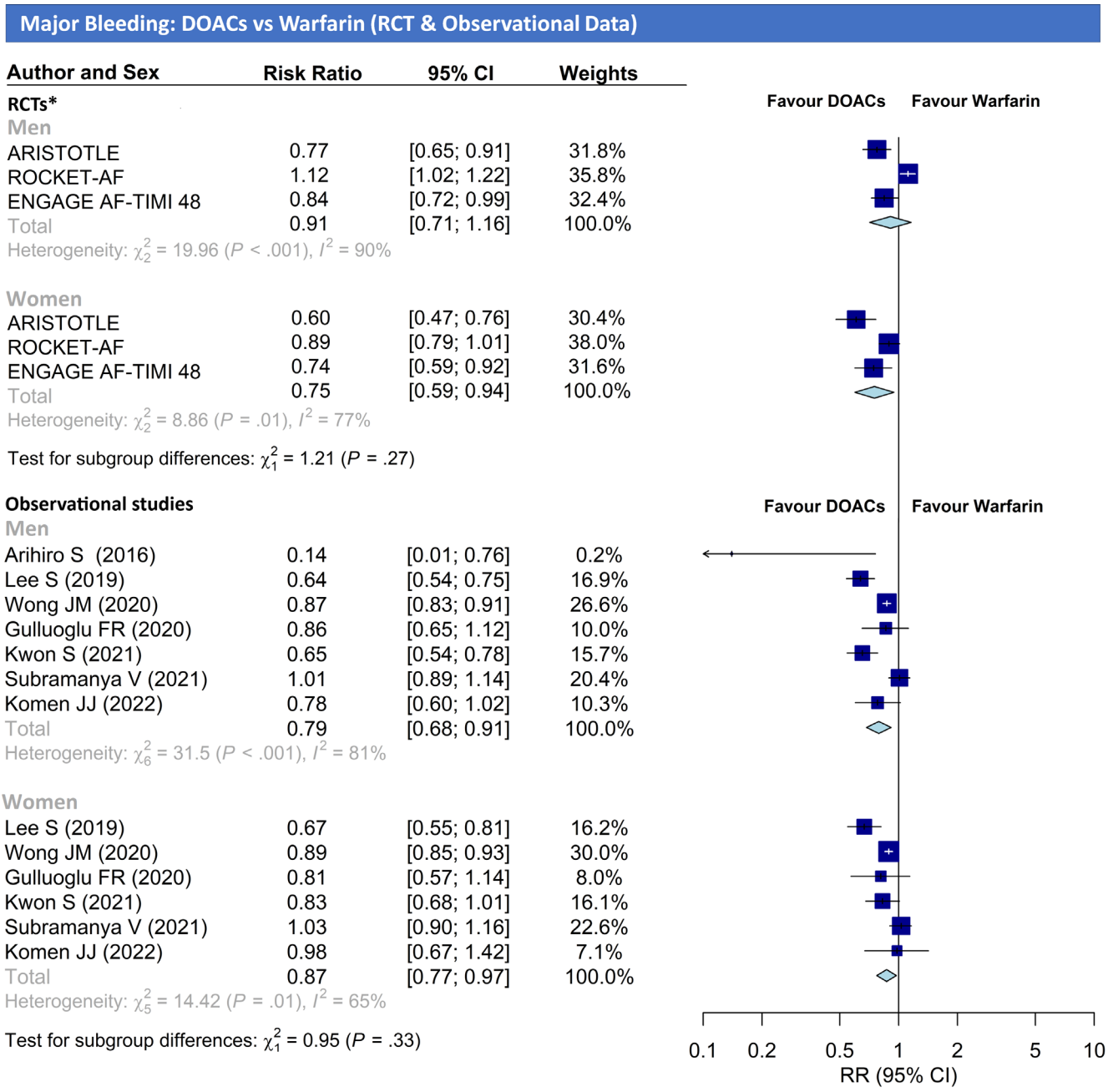


Figure 2 Forest plot of meta-analysis for stroke/systemic embolism with direct oral anticoagulants (DOACs) versus warfarin by sex. RCT, randomised controlled trial; RR, risk ratio.



*Outcomes for ARISTOTLE and ENGAGE AF-TIMI 48 are major bleeding. ROCKET-AF include major and nonmajor clinically relevant bleeding.

Figure 3 Forest plot of meta-analysis for major bleeding with direct oral anticoagulants (DOACs) versus warfarin. Randomised controlled trial (RCT) data compares major or non-major clinically relevant bleeding of DOACs versus warfarin by sex. Observational data compares major bleeding of DOACs versus warfarin by sex. RR, risk ratio; Apixaban for Reduction in Stroke and Other Thromboembolic Events in Atrial Fibrillation, ARISTOTLE; Effective Anticoagulation with Factor Xa Next Generation in Atrial Fibrillation–Thrombolysis in Myocardial Infarction 48, ENGAGE AF-TIMI 48; Rivaroxaban Once Daily Oral Direct Factor Xa Inhibition Compared with Vitamin K Antagonism for Prevention of Stroke and Embolism Trial in Atrial Fibrillation, ROCKET-AF.

For GIB and ICH, sex-specific data were available only from observational studies. DOAC versus warfarin was associated with a lower risk of GIB in men but not women, with no evidence of sex-specific interaction (women: pRR=0.98, 95% CI=0.85 to 1.13, I^2=50%; men: pRR=0.86, 95% CI=0.75 to 0.99, I^2=56%; p-for-interaction=0.22) (figure 4). For ICH, a lower risk with DOACs was found in both sexes (women: pRR=0.56, 95% CI=0.42 to 0.74, I^2=63%; men: pRR=0.54, 95% CI=0.44 to 0.68, I^2=52%; p-for-interaction=0.86) (online supplemental figure S2).

Valvular heart disease

Two observational studies provided sex-specific data on the outcomes of DOACs as a group (dabigatran, rivaroxaban or apixaban) versus warfarin in patients with AF and bioprosthetic heart valves. Meta-analyses showed that

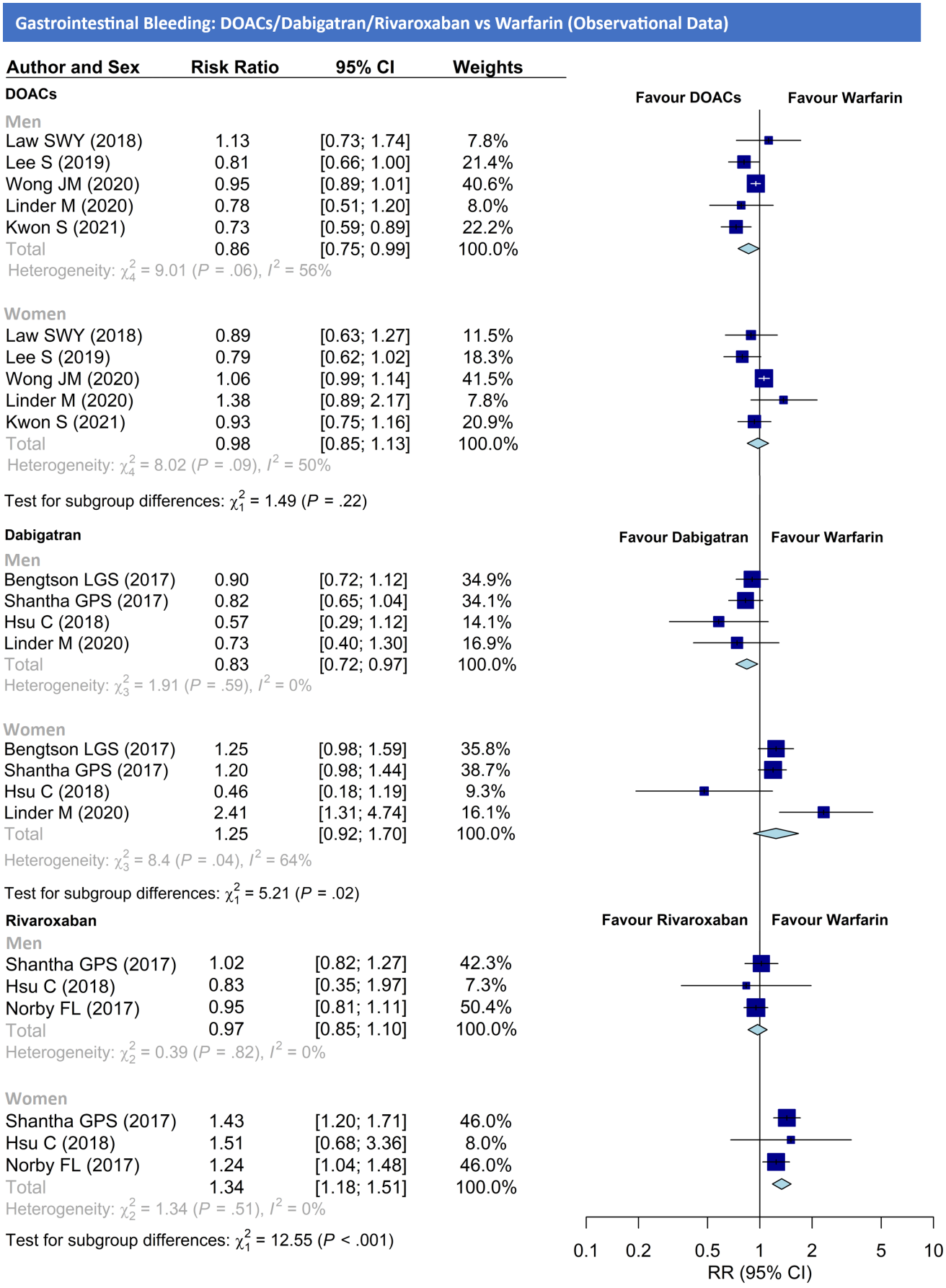


Figure 4 Forest plot of meta-analysis for observational studies comparing gastrointestinal bleeding of direct oral anticoagulants (DOACs), dabigatran and rivaroxaban versus warfarin by sex. RR, risk ratio.

in both sexes, there was no difference between DOACs and warfarin for stroke/SE (women: pRR=1.01, 95% CI=0.56 to 1.82, I^2=30%; men: pRR=1.38, 95% CI=0.88 to 2.15, I^2=0%; p-for-interaction=0.42) and major bleeding (women: pRR=0.66, 95% CI=0.37 to 1.18, I^2=0%; men: pRR=1.13, 95% CI=0.55 to 2.31, I^2=35%; p-for-interaction=0.25) (online supplemental figure S3).

One RCT assessed edoxaban against VKAs in patients with AF after a successful transcatheter aortic-valve replacement but did not report numeric estimates. Forest plots showed both sexes with higher incidence of major bleeding with edoxaban versus VKAs. No interaction tests were conducted but overlapping CIs suggest substantial sex difference is unlikely. Sex-specific data on stroke/SE, ICH or GIB were not reported.

Subgroup analyses

Individual DOACs

Stroke/SE

Meta-analysis showed dabigatran associated with lower risk of stroke/SE versus warfarin in both sexes, while rivaroxaban and apixaban had lower or similar risk. There was no indication of sex-specific interaction in each comparison (online supplemental figure S4). Only one study reported data for edoxaban finding no precise differences in stroke/SE versus warfarin in both sexes (online supplemental table S9).

Bleeding

Sex differences in the relative GIB risk versus warfarin were identified for dabigatran and rivaroxaban. Rivaroxaban versus warfarin was associated with a higher risk of GIB in women, but not men (women: pRR=1.34, 95% CI=1.18 to 1.51, I^2=0%; men: pRR=0.97, 95% CI=0.85 to 1.1, I^2=0%; p-for-interaction<0.001). For dabigatran, the point estimate for women suggests potentially higher risk of GIB versus warfarin but with 95% CI overlapping the null, whereas men had lower risk (women: pRR=1.25, 95% CI=0.92 to 1.70, I^2=64%; men: pRR=0.83, 95% CI=0.72 to 0.97, I^2=0%; p-for-interaction=0.02) (figure 4). Statistical evidence for GIB risk differences between sex remained only for rivaroxaban after Bonferroni correction (α=0.003). GIB data for apixaban was not available.

For major bleeding, ICH and any bleeding, there was no indication of sex-specific interactions in each DOAC comparison (online supplemental figures S2, S5 and S6). Meta-analysis for major bleeding showed both sexes using dabigatran or apixaban with lower associated risk versus warfarin. For rivaroxaban, major bleeding risk was comparable to warfarin in both sexes. For ICH, meta-analysis for dabigatran and rivaroxaban showed lower associated risk of ICH versus warfarin in both sexes. For any bleeding, both sexes with dabigatran and rivaroxaban were associated with lower or similar risk versus warfarin. ICH and any bleeding data for apixaban was unavailable.

Data for edoxaban was provided by one study. Both sexes with edoxaban had lower associated risk of major bleeding versus warfarin. For GIB and ICH, point estimates for both sexes suggested lower risk versus warfarin, but estimates were imprecise (online supplemental table S9).

Analysis by geographical regions

With each DOAC, Asians had lower stroke/SE and major bleeding risk versus warfarin and exhibited lower RRs for stroke/SE compared with other regions (online supplemental figure S7 and online supplemental table S10). For major bleeding, rivaroxaban versus warfarin was associated with lower risk among Asians, but similar or raised risk in other regions, whereas DOACs as a group among men were associated with greater reductions in major bleeding for Asians. Apixaban and dabigatran had lower or comparable major bleeding risk versus warfarin in all regions and sexes. GIB risk in men was lower or similar across regions with each DOAC versus warfarin comparison. Among women, GIB risk was lower or similar in Asians using DOACs, but comparable or raised in Europeans and North Americans. For ICH, DOACs as a group and dabigatran were associated with a lower risk versus warfarin for both sexes except for Europe which showed no precise difference in ICH risk. Some statistically significant differences in stroke/SE and major bleeding between regions remained after Bonferroni correction (α=0.001), but not for GIB and ICH.

Head-to-head DOAC comparisons

Meta-analysis of three observational studies found similar risk of stroke/SE between rivaroxaban and dabigatran. Two of these studies provided data for GIB and ICH. Meta-analysis showed both sexes with increased risk of GIB with rivaroxaban versus dabigatran, and for ICH, point estimates suggest increased risk with rivaroxaban in both sexes, although estimates were imprecise (online supplemental figure S8). Meta-analysis for major bleeding was not possible due to overlapping populations, but individual estimates from two studies showed both sexes with rivaroxaban associated with a higher risk (online supplemental table S11). Apixaban was compared with rivaroxaban and dabigatran in one study, reporting lower risk of stroke/SE and major bleeding with apixaban than dabigatran and rivaroxaban in both sexes (online supplemental table S11).

Narrative review

The excluded data and narrative summaries were generally consistent with meta-analyses for stroke/SE and major bleeding, with no noticeable differences between sexes across DOACs. One study reported data showing lower GIB with DOACs versus warfarin in both sexes among Asians, consistent with the geographical analysis (online supplemental table S9). One study that was narratively reviewed reported raised GIB in women with dabigatran versus warfarin consistent with our results (online supplemental table S12).

DISCUSSION

Key findings

This systematic review and meta-analysis compared the sex-specific effectiveness and safety of DOACs to warfarin.

Our study identified sex-specific interactions for GIB, with observational data suggesting women may have potentially higher risk of GIB with rivaroxaban and dabigatran compared with warfarin, which were not observed in men. The sex-specific interaction for GIB with rivaroxaban was observed even after Bonferroni correction. No other sex-specific interaction was found, with DOACs generally being associated with lower risk of stroke/SE, major bleeding and ICH compared with warfarin in both sexes. To our knowledge, this is the first and most comprehensive systematic review and meta-analysis to investigate the effectiveness and safety of DOACs in AF by sex, with the inclusion of representative real-world data outside RCT settings.

Comparison to other studies

A previous systematic review and meta-analysis reported sex-specific estimates of GIB risk with DOACs, using observational and RCT data published until October 2018.[22] The study found women to have raised GIB risk with DOACs as a group versus warfarin but not men.[22] However, the study neither evaluated sex-specific GIB risk by individual DOACs nor primarily intended to investigate sex-specific outcomes, and no sex-specific interaction tests were reported. Using updated data up to November 2022, our study identified that the raised relative GIB risk against warfarin in women may apply to rivaroxaban and possibly dabigatran. It is unclear why women may experience raised GIB. The pharmacokinetics of drugs frequently differ between sexes due to differences in body size, fat content, gastrointestinal physiology and renal functions. This can influence the processing, absorption and excretion of drugs, potentially altering drug safety and explaining the raised GIB in women.[23] Supporting this, women patients treated with DOACs have been observed to have higher rates of GIB compared with men,[24] although this is based on limited research and more studies are required to investigate differences in GIB risk between the sexes.

A meta-analysis[25] of four landmark RCTs in patients with AF found reduced risk of stroke/SE and major bleeding with DOACs and no evidence of sex-specific interaction. Our results for DOACs, which contribute further by including observational studies, are consistent with those reported results. Additionally, our subgroup meta-analysis using real-world data for dabigatran, apixaban and rivaroxaban demonstrated lower stroke/SE risk versus warfarin in both sexes, generally aligning with the landmark RCTs.[9 11 12] For edoxaban, the one available observational study[26] showed consistency with the landmark RCT,[27] reporting similar stroke/SE risk versus warfarin in both sexes. For ICH, RCTs have established reduced risk of ICH with DOACs versus warfarin,[9 11 12 27] but to our knowledge, there are no published sex-specific assessments. Our findings are consistent with a reduced risk of ICH for both sexes. This is important given the uncertainty of managing patients with AF and ICH.[28]

In our geographical analysis, reduced stroke/SE risk with DOACs versus warfarin was consistently observed in Asians. Furthermore, our findings suggest Asians with DOACs may experience improved risk reductions in bleeding compared with other regions. These results agree with a post hoc meta-analysis of RCTs[29] showing DOACs versus VKAs to reduce stroke/SE and major bleeding more in Asians relative to non-Asians. Asians are known to have enhanced pharmacokinetic and pharmacodynamic profiles with antithrombotic agents and greater natural tendency of bleeding compared with Caucasians.[30] Thus, Asians often have lower target international normalised ratio levels with warfarin which could increase thromboembolism risk, and therefore may experience greater reduction of stroke/SE with DOACs.[30–32] Additionally, Asians are prone to excessive bleeding with warfarin possibly due to their lower body weight and genetic susceptibility to overanticoagulation with warfarin.[30 33] Asians could therefore benefit more from DOACs regarding major bleeding risk.[34]

A systematic review and meta-analysis[35] comparing rivaroxaban and dabigatran showed similar stroke/SE risk, but increased GIB with rivaroxaban. Our post hoc meta-analysis of observational data agrees with these findings, and further demonstrates this for both men and women seperately. Given the post hoc nature of the analyses and that these are based on solely observational data, we emphasise that these results should be interpreted carefully.

Implications for clinical practice

For both sexes, our results demonstrate DOACs generally exhibiting improved effectiveness and safety versus warfarin in terms of reducing stroke and major bleeding risk. This reaffirms the use of DOACs in both sexes with AF, concurring with the current guidelines recommending DOACs over warfarin.[7 8] However, with observational data, our study identified sex-specific differences for GIB. Specifically, GIB risk may be raised with rivaroxaban and possibly dabigatran in women but not men, and other recent evidence is indicative of higher risk of GIB with DOACs than warfarin in women.[22 24] Further research should verify the sex-specific difference in GIB as this result was generated using pooled data from a small number of observational studies subject to potential confounding bias. In addition, GIB data for apixaban and edoxaban was not available and is urgently needed to better understand if sex-specific differences in GIB exist and whether there are preferable DOAC choices in women.[36] Thus, we call for future studies to report sex-specific data when examining outcomes of DOACs to elucidate these research gaps. Furthermore, other approaches to reduce GIB can be considered in patients with higher risk, such as the use of gastroprotective agents.[37]

Strengths and limitations

To our knowledge, this is the first and most comprehensive systematic review and meta-analysis comparing sex-specific effectiveness and safety of the DOACs versus warfarin using both RCTs and more representative real-world data. We conducted analyses on individual DOACs, across geographical settings and post hoc head-to-head DOAC comparisons. We summarised all the best available evidence on several important outcomes directly relevant to clinical practice, enabling useful interpretations which improve therapeutic decision-making and inform avenues for future research.

This study has limitations. There was limited literature assessing sex-specific outcomes and most studies were not designed for sex-specific analyses, reducing statistical power. Furthermore, subgroup analyses of individual DOACs contained a small number of studies. Statistical assessment of publication bias was not conducted due to limited studies in each meta-analysis (n<10). There was substantial heterogeneity between studies, likely representing the variation of individual DOACs in the pooled DOAC groups and differences in DOAC dosages, outcome definitions, study populations and durations of follow-up. Furthermore, observational data are limited by residual confounding, although adequate methods to account for confounding were adopted by included studies, and meta-analyses of RCTs were mostly consistent with observational data. Finally, the generalisability of these findings to younger patients is not possible, as the mean age of patients in most studies was >65 years.

Directions for future research

Studies are required to verify our findings on sex-specific GIB risk discrepancies. Our subgroup analysis contained a small number of observational studies, and the mechanistic reasons for sex-specific differences in GIB risk need exploration. Additionally, sex-specific GIB data for apixaban and edoxaban is needed. Furthermore, future research is needed to investigate effective approaches to reduce GIB risk, such as the use of gastroprotective agents, in women and high-risk patients with AF using DOACs.[37] Age-specific interactions with DOACs also need investigation as age may modify the risk of GIB in women.[38]

CONCLUSION

Among patients with AF, both sexes demonstrated generally favourable effectiveness and safety with DOACs compared with warfarin, supporting the preference of DOACs over warfarin in both sexes. However, meta-analysis of observational data suggests that GIB risk may be raised in women with AF using rivaroxaban and possibly dabigatran when compared with warfarin. Further studies are required to verify this finding and elucidate sex-specific GIB risk with apixaban and edoxaban, of which data is currently lacking.

Author affiliations
[1]Research Department of Practice and Policy, UCL School of Pharmacy, London, UK
[2]Centre for Medicines Optimisation Research and Education, University College London Hospitals NHS Foundation Trust, London, UK
[3]Department of Non-Communicable Disease Epidemiology, Faculty of Epidemiology & Population Health, London School of Hygiene and Tropical Medicines, London, UK
[4]School of Pharmacy, Institute of Clinical Sciences, College of Medical & Dental Sciences, University of Birmingham, Birmingham, UK
[5]Laboratory of Data Discovery for Health (D24H), Hong Kong Science Park, Hong Kong, Hong Kong
[6]Centre for Safe Medication Practice and Research, Department of Pharmacology and Pharmacy, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, Hong Kong
[7]Liverpool Centre for Cardiovascular Science, University of Liverpool, Liverpool John Moores University and Liverpool Heart & Chest Hospital, Liverpool, UK
[8]Danish Center for Health Services Research, Department of Clinical Medicine, Aalborg University, Aalborg, Denmark
[9]Institute of Cardiovascular Sciences, University College London and University College London British Heart Foundation Research Accelerator, London, UK
[10]Department of Medicine, School of Clinical Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, Hong Kong
[11]Aston School of Pharmacy, Aston University, Birmingham, UK
[12]Unit of Excellence on Cardiovascular Archive Research and Clinical Epidemiology, School of Pharmaceutical Sciences, University of Phayao, Phayao, Thailand
[13]Pharmacoepidemiology, Social and Administrative Pharmacy (P-SAP) Research Unit, Division of Social and Administrative Pharmacy (SAP), Department of Pharmaceutical Care, School of Pharmaceutical Sciences, University of Phayao, Phayao, Thailand

Presented at
An earlier version of this work was presented at the 39th International Conference on Pharmacoepidemiology and Therapeutic Risk Management (ICPE), Halifax, Canada, 25-27 August 2023. The citation is: Chobanov J, Mongkhon P, Wang Z et al. Sex-specific comparative effectiveness and safety of direct oral anticoagulants versus warfarin in patients with atrial fibrillation–A systematic review and meta-analysis [Abstract #855], the 39th International Conference on Pharmacoepidemiology and Therapeutic Risk Management (ICPE), Halifax, Canada, 25–27 August 2023. Pharmacoepidemiology and Drug Safety 2023;32 32 Suppl 1:3–612.

Contributors WL, PM and JDC were involved in the study concept and design. PM and JDC drafted the manuscript with input from all authors. All authors were involved in the acquisition, analysis or interpretation of data and critical revision of the manuscript for important intellectual content and have read and approved the final manuscript. WL, PM and JDC are the guarantors of this work.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests ZW's current post is partly supported by the AIR@InnoHK administered by Innovation and Technology Commission. KKCM was supported by the CW Maplethorpe Fellowship and received research funding from the National Institute of Health Research, UK, European Commission Horizon 2020 framework and Hong Kong Research Grants Council and personal fee from IQVIA, outside the scope of the submitted work. GYHL is consultant and speaker for BMS/Pfizer, Boehringer Ingelheim, Anthos and Daiichi-Sankyo, no fees was received personally. ICKW has received research funding from the National Institute of Health Research, UK, European Commission Horizon 2020 framework, the Hong Kong Research Grants Council, the Hong Kong Health and Medical Research Fund, Amgen, Janssen, Novartis and GSK for medication safety research, outside the scope of the submitted work. He also received funding from Bristol-Myers Squibb, Pfizer and Bayer on DOACs research but it is not associated with the current study. His current post is partly funded by the AIR@InnoHK administered by Innovation and Technology Commission. WL reported research funding from the AIR@InnoHK administered by Innovation and Technology Commission, outside the scope of the submitted work. There are no other relationships or activities to disclose.

Patient and public involvement statement Patients and/or the public were not involved in the design, conduct, reporting or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval Not applicable. The study is a systematic review and meta-analysis of existing published data.

Provenance and peer review Not commissioned; externally peer reviewed.

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

ORCID iDs
Jan D Chobanov http://orcid.org/0009-0002-5904-1897
Wallis C Y Lau http://orcid.org/0000-0003-2320-0470

REFERENCES

1 Benjamin EJ, Muntner P, Alonso A, *et al*. Heart disease and stroke statistics—2019 update: a report from the American heart association. *Circulation* 2019;139:e56–528.
2 Wang TJ, Massaro JM, Levy D, *et al*. A risk score for predicting stroke or death in individuals with new-onset atrial fibrillation in the community the Framingham heart study. *JAMA* 2003;290:1049–56.
3 Nielsen PB, Overvad TF. Female sex as a risk modifier for stroke risk in atrial fibrillation: using Cha2Ds2-Vasc versus Cha2Ds2-VA for stroke risk stratification in atrial fibrillation: a note of caution. *Thromb Haemost* 2020;120:894–8.
4 Nielsen PB, Skjøth F, Overvad TF, *et al*. Female sex is a risk modifier rather than a risk factor for stroke in atrial fibrillation: should we use a Cha2Ds2-VA score rather than Cha2Ds2-Vasc? *Circulation* 2018;137:832–40.
5 Schnabel RB, Benjamin EJ. Sex and stroke risk in atrial fibrillation – more work to be done. *JACC Clin Electrophysiol* 2018;4:615–7.
6 Rosano GMC, Lewis B, Agewall S, *et al*. Gender differences in the effect of cardiovascular drugs: a position document of the working group on pharmacology and drug therapy of the ESC. *Eur Heart J* 2015;36:2677–80.
7 January CT, Wann LS, Calkins H, *et al*. AHA/ACC/HRS focused update of the 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation. *J Am Coll Cardiol* 2019;74:104–32.
8 Hindricks G, Potpara T, Dagres N, *et al*. ESC guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European association for cardio-thoracic surgery (EACTS): the task force for the diagnosis and management of atrial fibrillation of the European society of cardiology (ESC) developed with the special contribution of the European heart rhythm association (EHRA) of the ESC. *Eur Heart J* 2020;42:373–498.
9 Connolly SJ, Ezekowitz MD, Yusuf S, *et al*. Dabigatran versus warfarin in patients with atrial fibrillation. *N Engl J Med* 2009;361:1139–51.
10 Giugliano RP, Ruff CT, Braunwald E, *et al*. Edoxaban versus warfarin in patients with atrial fibrillation. *N Engl J Med* 2013;369:2093–104.
11 Granger CB, Alexander JH, McMurray JJV, *et al*. Apixaban versus warfarin in patients with atrial fibrillation. *N Engl J Med* 2011;365:981–92.
12 Patel MR, Mahaffey KW, Garg J, *et al*. Rivaroxaban versus warfarin in nonvalvular atrial fibrillation. *N Engl J Med* 2011;365:883–91.
13 Hägg L, Johansson C, Jansson J-H, *et al*. External validity of the ARISTOTLE trial in real-life atrial fibrillation patients. *Cardiovasc Ther* 2014;32:214–8.
14 Page MJ, McKenzie JE, Bossuyt PM, *et al*. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:71.
15 Sterne JAC, Savović J, Page MJ, *et al*. Rob 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 2019;366.
16 Wells GA, Shea B, O'Connell D, *et al*. *The newcastle-ottawa scale (nos) for assessing the quality of nonrandomised studies in meta-analyses*.
17 Stare J, Maucort-Boulch D. Odds ratio, hazard ratio and relative risk. *Adv Meth Stat* 2016;13:59–67.
18 Richardson M, Garner P, Donegan S. Interpretation of subgroup analyses in systematic reviews: a tutorial. *Clin Epidemiol Glob Health* 2019;7:192–8.
19 McGuinness LA, Higgins JPT. Risk-of-bias visualization (robvis): an R package and shiny web App for visualizing risk-of-bias assessments. *Res Synth Methods* 2021;12:55–61.
20 Komen JJ, Pottegård A, Mantel-Teeuwisse AK, *et al*. Oral anticoagulants in patients with atrial fibrillation at low stroke risk: a multicentre observational study. *Eur Heart J* 2022;43:3528–38.
21 Balsam P, Lodziński P, Gawałko M, *et al*. Antithrombotic management and long-term outcomes of patients with atrial fibrillation. *J Clin Med* 2021;10.
22 Gu Z-C, Wei A-H, Zhang C, *et al*. Risk of major gastrointestinal bleeding with new vs conventional oral anticoagulants: a systematic review and meta-analysis. *Clin Gastroenterol Hepatol* 2020;18:792–9.
23 Madla CM, Gavins FKH, Merchant HA, *et al*. Let's talk about sex: differences in drug therapy in males and females. *Adv Drug Deliv Rev* 2021;175:113804.
24 Ferroni E, Denas G, Gennaro N, *et al*. Gender related differences in gastrointestinal bleeding with oral anticoagulation in atrial fibrillation. *J Cardiovasc Pharmacol Ther* 2022;27.
25 Ruff CT, Giugliano RP, Braunwald E, *et al*. Comparison of the efficacy and safety of new oral anticoagulants with warfarin in patients with atrial fibrillation: a meta-analysis of randomised trials. *Lancet* 2014;383:955–62.
26 Lee S-R, Choi E-K, Han K-D, *et al*. Edoxaban in Asian patients with atrial fibrillation: effectiveness and safety. *J Am Coll Cardiol* 2018;72:838–53.
27 Zelniker TA, Ardissino M, Andreotti F, *et al*. Comparison of the efficacy and safety outcomes of edoxaban in 8040 women versus 13 065 men with atrial fibrillation in the ENGAGE AF-TIMI 48 trial. *Circulation* 2021;143:673–84.
28 Ivany E, Lotto RR, Lip GYH, *et al*. Managing uncertainty: physicians' decision making for stroke prevention for patients with atrial fibrillation and intracerebral hemorrhage. *Thromb Haemost* 2022;122:1603–11.
29 Wang K-L, Lip GYH, Lin S-J, *et al*. Non–vitamin K antagonist oral anticoagulants for stroke prevention in Asian patients with nonvalvular atrial fibrillation: meta-analysis. *Stroke* 2015;46:2555–61.
30 Kim HK, Tantry US, Smith SC, *et al*. The East Asian paradox: an updated position statement on the challenges to the current Antithrombotic strategy in patients with cardiovascular disease. *Thromb Haemost* 2021;121:422–32.
31 Chao T-F, Guo Y. Should we adopt a standard International normalized ratio range of 2.0 to 3.0 for Asian patients with atrial fibrillation. *Thromb Haemost* 2020;120:366–8.
32 Pandey AK, Xu K, Zhang L, *et al*. Lower versus standard INR targets in atrial fibrillation: a systematic review and meta-analysis of randomized controlled trials. *Thromb Haemost* 2020;120:484–94.
33 Mega JL, Walker JR, Ruff CT, *et al*. Genetics and the clinical response to warfarin and edoxaban: findings from the randomised, double-blind ENGAGE AF-TIMI 48 trial. *Lancet* 2015;385:2280–7.
34 Gorog DA, Gue YX, Chao T-F, *et al*. Assessment and mitigation of bleeding risk in atrial fibrillation and venous thromboembolism: executive summary of a European and Asia-Pacific expert consensus paper. *Thromb Haemost* 2022;122:1625–52.
35 Bai Y, Deng H, Shantsila A, *et al*. Rivaroxaban versus dabigatran or warfarin in real-world studies of stroke prevention in atrial fibrillation. *Stroke* 2017;48:970–6.
36 Lau WCY, Torre CO, Man KKC, *et al*. Comparative effectiveness and safety between apixaban, dabigatran, edoxaban, and rivaroxaban among patients with atrial fibrillation: a multinational population-based cohort study. *Ann Intern Med* 2022;175:1515–24.
37 Steffel J, Verhamme P, Potpara TS, *et al*. The 2018 European heart rhythm association practical guide on the use of non-vitamin K antagonist oral anticoagulants in patients with atrial fibrillation. *Eur Heart J* 2018;39:1330–93.
38 Graham DJ, Reichman ME, Wernecke M, *et al*. Cardiovascular, bleeding, and mortality risks in elderly medicare patients treated with dabigatran or warfarin for nonvalvular atrial fibrillation. *Circulation* 2015;131:157–64.

SUPPLEMENTAL MATERIAL

Sex-Specific Comparative Outcomes between Oral Anticoagulants in Patients with Atrial Fibrillation-A Systematic Review and Meta-Analysis

Supplementary

Chobanov JD, *et al. Open Heart* 2024; 11:e002792. doi: 10.1136/openhrt-2024-002792

Supplementary Appendix-1: Quality assessment using the revised Cochrane Risk of Bias Tool for Randomised Trials and the Newcastle-Ottawa Scale

The revised Cochrane Risk of Bias Tool for Randomised Trials (RoB version 2.0)(1) was used to appraise the quality of the included randomised control trials. The tool assesses multiple sources of bias including: the study's randomisation process, bias due to deviations from intended interventions, missing outcome data, measurement of the outcome, and selection of the reported result. A risk of bias rating (high, low, or some concerns) was given for each randomised control included within the systematic review. A low overall risk indicated a high-quality study. To determine the quality of observational studies, the Newcastle-Ottawa Scale (NOS) was used(2). This scale gave a maximum 9-star rating by considering the risk of bias across three key domains: the selection and comparability of cohorts and the assessment of outcomes. A higher score and a well-distributed allocation of points across the domains indicated a better-quality study.

Two RCTs(3,4) were rated as having some concerns as participants were aware of their assigned intervention, which could lead to deviations from intended interventions. For observational studies, two studies received fair quality ratings. One study(5) received a fair quality rating as it did not specify how participants were censored, and it was conducted in only two hospitals, resulting in a lack of representativeness. The other study(6) received a fair quality rating as it did not specify how participants were censored, and did not mention the mean or median length of follow-up, making it unclear whether follow-up was long enough for outcomes to occur. The rest of the RCTs and observational studies were rated as having low risk of bias and good quality, respectively (Supplementary Tables S7 and S8).

Supplementary Appendix-2: Bonferroni Correction

Post-hoc, we applied a Bonferroni correction to mitigate the inflated type 1 error due to multiple testing. There were 30 p-for-interaction tests conducted comparing sex-specific outcomes with oral anticoagulants and 70 p-for-interaction tests were conducted comparing differences between geographical regions.

Taking the p-for-interaction<0.1 statistically significant level for detecting subgroup differences, we calculated Bonferroni corrected levels of 0.003 (0.1 / 30) and 0.001 (0.1 / 70) for the sex-specific oral anticoagulant and geographical regions comparisons respectively.

Supplementary Appendix-3: Data or studies excluded from meta-analysis

When study populations overlapped between observational studies, the effect estimates from the study with the largest sample size and most comprehensive adjustment for potential confounders were used in the meta-analyses. The other overlapping effect estimates were excluded. Other studies were ineligible for meta-analysis due to incomplete data for a meta-analysis (e.g., no numeric estimates were provided by the study).

A narrative review was conducted on the studies and data excluded from the meta-analyses.

Author	Reason for exclusion from meta-analysis	Sex-specific Outcome data excluded
Chang S (2019)	No sex-specific numeric estimates were reported	NA - no exact numeric estimates were reported
Seeger JD (2015)	No sex-specific numeric estimates were reported	NA - no exact numeric estimates were reported
Graham DJ (2015)	Sex-specific estimates were stratified by age incompatible with our other data	NA - Sex-specific estimates were stratified by age incompatible with our other data
Lee S (2018)	The sole observational study to report sex-specific data for edoxaban vs warfarin	NA – insufficient data for any meta-analysis for edoxaban
Mieghem NMV (2021)	No sex-specific numeric estimates were reported	NA - no exact numeric estimates were reported
Kwon S (2020)	Overlapping study population	DOAC vs warfarin data for stroke/SE, major bleeding, GIB, and ICH
Lip GYH (2017)	Overlapping study population	Dabigatran/rivaroxaban/apixaban vs warfarin data for stroke/SE
Bengtson LGS (2017)	Overlapping study population	Dabigatran vs warfarin data for stroke/SE
Shantha GPS (2017)	Overlapping study population	Dabigatran/rivaroxaban vs warfarin data for stroke/SE and major bleeding
Huybrechts KF (2019)	Overlapping study population	Dabigatran vs warfarin data for stroke/SE and major bleeding
Weir MR (2020)	Overlapping study population	Rivaroxaban vs warfarin data for stroke/SE and major bleeding
Norby FL (2017)	Overlapping study population	Rivaroxaban vs warfarin data for stroke/SE
Costa OS (2020)	Overlapping study population	Rivaroxaban vs warfarin data for stroke/SE and major bleeding
Baker WL (2019)	Overlapping study population	Rivaroxaban vs warfarin data for major bleeding

Abbreviations: DOACs=direct oral anticoagulant, SE = systemic embolism, GIB = gastrointestinal bleeding, ICH = intracranial haemorrhage, NA = Not available

Appendix References

1. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ (Clinical research ed).2019Aug28;366:l4898.

2. Wells G SB O'Connell D, Peterson J, Welch V, Losos M, Tugwell P. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses [Internet]. Available from: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp

3. Connolly SJ, Ezekowitz MD, Yusuf S, et al. Dabigatran versus Warfarin in Patients with Atrial Fibrillation. New England Journal of Medicine.2009Sep17;361(12):1139–51.

4. Van Mieghem NM, Unverdorben M, Hengstenberg C, et al. Edoxaban versus Vitamin K Antagonist for Atrial Fibrillation after TAVR. New England Journal of Medicine.2021Dec 2;385(23):2150–60.

5. Balsam P, Lodziński P, Gawałko M, et al. Antithrombotic Management and Long-Term Outcomes of Patients with Atrial Fibrillation. Insights from CRAFT Trial. Journal of clinical medicine.2021Apr19;10(8).

6. Subramanya V, Claxton JS, Lutsey PL, et al. Sex differences in treatment strategy and adverse outcomes among patients 75 and older with atrial fibrillation in the MarketScan database. BMC Cardiovascular Disorders.2021Dec16;21(1):598.

Table S1. The search strategies: PubMed database from 2018 to 23rd November 2022

Search number	Query	Results
#1	(((((((Male[MeSH Terms]) OR (Female[MeSH Terms])) OR (men)) OR (women)) OR (gender)) OR (sex)) OR (sex comparison*)) OR (sex difference*)	13,050,225
#2	("direct oral anticoagulant" OR non-vitamin K antagonist oral anticoagulant* OR "novel oral anticoagulant" OR "new oral anticoagulant" OR oral anticoagulant* OR DOAC* OR NOAC* OR TSOAC* OR Factor Xa inhibitors [MeSH terms] OR factor IIa inhibitor* OR direct thrombin inhibitor* OR Rivaroxaban [MeSH terms] OR Rivaroxaban OR Xarelto OR Apixaban OR Eliquis OR Dabigatran [MeSH terms] OR Dabigatran OR Pradaxa OR Edoxaban OR Lixiana OR Savaysa)	48,779
#3	(Warfarin [MeSH terms] OR warfarin OR vitamin k antagonist* OR Coumarins [MeSH terms] OR coumarin*)	81,616
#4	(Stroke [MeSH terms] OR ischaemic stroke OR ischemic stroke OR bleeding OR Hemorrhage [MeSH terms] OR haemorrhage OR hemorrhage)	5,579,981
#5	#1 AND #2 AND #3 AND #4	5,803
#6	Limit from 2008 – 2022	4,534
#7	Limit to human studies and English language	4,085

Table S2. The search strategies: EMBASE database from 2018 to 23rd November 2022

Search number	Query	Results
1	exp male/ or Male.mp.	10,776,274
2	exp female/ or Female.mp.	10,949,934
3	men.mp.	809,275
4	women.mp.	1,529,360
5	exp gender/ or gender.mp.	699,917
6	sex.mp. or exp sex/	1,463,259
7	sex comparison*.mp.	348
8	exp sex difference/ or sex difference*.mp.	421,201
9	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8	14,659,183
10	"direct oral anticoagulant".mp.	2,627
11	non-vitamin K antagonist oral anticoagulant*.mp.	1,966
12	"new oral anticoagulant".mp.	691
13	oral anticoagulant*.mp.	30,463
14	DOAC*.mp.	7,725
15	NOAC*.mp.	6,229
16	TSOAC*.mp.	104
17	Factor Xa inhibitors.mp. or exp blood clotting factor 10a inhibitor/	103,727
18	factor IIa inhibitor*.mp.	90
19	direct thrombin inhibitor*.mp.	4,138
20	Rivaroxaban.mp. or exp rivaroxaban/	25,754
21	Xarelto.mp.	1,415
22	Apixaban.mp. or exp apixaban/	19,062
23	Eliquis.mp.	828
24	exp dabigatran etexilate/ or exp dabigatran/ or Dabigatran.mp.	20,403
25	Pradaxa.mp.	1,210
26	Edoxaban.mp. or exp edoxaban/	7,310
27	Lixiana.mp.	144
28	Savaysa.mp.	163
29	10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28	127,580
30	Warfarin.mp. or exp warfarin/	105,253
31	exp warfarin/ or vitamin k antagonist*.mp.	109,227
32	Coumarins.mp. or exp coumarin derivative/	142,414
33	coumarin*.mp.	29,865
34	30 or 31 or 32 or 33	157,764
35	Stroke.mp.	525,270
36	ischaemic stroke.mp. or exp brain ischemia/	212,259
37	ischemic stroke.mp.	106,933
38	bleeding.mp. or exp bleeding/	1,128,275
39	Hemorrhage.mp.	460,625

Search number	Query	Results
40	haemorrhage.mp.	56,016
41	35 or 36 or 37 or 38 or 39 or 40	1,668,740
42	9 and 29 and 34 and 41	15,460
43	Limit 42 to (human and English language and year="2008-2022" and (article or article in press) and journal)	7,594
44	Limit 43 to clinical studies	1,195

Table S3. The search strategies: Web of Science database from 2018 to 23rd November 2022

Search number	Query	Results
#1	TS=(Male OR Female OR men OR women OR gender OR sex OR sex comparison* OR sex difference*)	3396895
#2	TS=("direct oral anticoagulant" OR non-vitamin K antagonist oral anticoagulant* OR "novel oral anticoagulant" OR "new oral anticoagulant" OR oral anticoagulant* OR DOAC* OR NOAC* OR TSOAC* OR Factor Xa inhibitors OR factor IIa inhibitor* OR direct thrombin inhibitor* OR Rivaroxaban OR Rivaroxaban OR Xarelto OR Apixaban OR Eliquis OR Dabigatran OR Dabigatran OR Pradaxa OR Edoxaban OR Lixiana OR Savaysa)	36259
#3	TS=(Warfarin OR vitamin k antagonist* OR Coumarins OR coumarin*)	54581
#4	TS=(Stroke OR ischaemic stroke OR ischemic stroke OR bleeding OR hemorrhage OR haemorrhage)	570625
#5	#1 AND #2 AND #3 AND #4	2066
#6	#1 AND #2 AND #3 AND #4 and English (Languages)	1995
#7	#1 AND #2 AND #3 AND #4 and English (Languages) and Article (Document Types)	1838

Table S4. The search strategies: Cochrane Library database from 2008 to 23rd November 2022

Search number	Query	Results
#1	MeSH descriptor: [Male] in all MeSH products	461950
#2	MeSH descriptor: [Female] in all MeSH products	486912
#3	men	94819
#4	women	184834
#5	MeSH descriptor: [Gender Identity] explode all trees	278
#6	sex	61972
#7	sex comparison*	11421
#8	sex difference*	27198
#9	#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8	706913
#10	"direct oral anticoagulant"	659
#11	non-vitamin K antagonist oral anticoagulant*	186
#12	"novel oral anticoagulant"	213
#13	"new oral anticoagulant"	620
#14	oral anticoagulant*	4029
#15	DOAC*	467
#16	NOAC*	676
#17	TSOAC*	9
#18	MeSH descriptor: [Factor Xa Inhibitors] explode all trees	641
#19	factor IIa inhibitor*	189
#20	direct thrombin inhibitor*	595
#21	MeSH descriptor: [Rivaroxaban] explode all trees	674
#22	Xarelto	135
#23	Apixaban	1186
#24	Eliquis	61
#25	MeSH descriptor: [Dabigatran] explode all trees	350
#26	Pradaxa	71
#27	Edoxaban	716
#28	Lixiana	34
#29	Savaysa	9
#30	#10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23 OR #24 OR #25 OR #26 OR #27 OR #28 OR #29	6670
#31	MeSH descriptor: [Warfarin] explode all trees	1798
#32	vitamin k antagonist*	1426
#33	MeSH descriptor: [Coumarins] explode all trees	2387
#34	coumarin*	407
#35	#31 OR #32 OR #33 OR #34	3777
#36	MeSH descriptor: [Stroke] explode all trees	11850
#37	ischaemic stroke	18861
#38	ischemic stroke	18861
#39	MeSH descriptor: [Hemorrhage] explode all trees	15786
#40	Hemorrhage	34695
#41	Haemorrhage	34700
#42	#36 OR #37 OR #38 OR #39 OR #40 OR #41	61280
#43	#9 AND #30 AND #35 AND #42	489
#44	#43 Limit to 2008- November 2022	489

Table S5: Definitions of stroke/systemic embolism or bleeding outcomes as reported by each study in the systematic review and meta-analysis

Author (Year)	Stroke/systemic embolism definition	Bleeding definition
Connolly SJ, et al. (2009): RE-LY[1]	Stroke was defined as the sudden onset of a focal neurologic deficit in a location consistent with the territory of a major cerebral artery and categorized as ischemic, haemorrhagic, or unspecified. Systemic embolism was defined as an acute vascular occlusion of an extremity or organ, documented by means of imaging, surgery, or autopsy.	Major bleeding was defined as a reduction in the haemoglobin level of at least 20 g/L, transfusion of at least 2 unit of blood, or symptomatic bleeding in a critical area or organ. Life-threatening bleeding was a subcategory of major bleeding.
Granger CB, et al. (2011): ARISTOTLE[2]	Stroke was defined as a focal neurologic deficit, from a nontraumatic cause, lasting at least 24 hours and was categorized as ischemic, haemorrhagic, or of uncertain type.	Primary safety outcome: the major bleeding was defined according to the International Society on Thrombosis and Haemostasis (ISTH) criteria, as clinically overt bleeding accompanied by a decrease in the haemoglobin level of at least 2 g/dL or transfusion of at least 2 units of packed red cells, occurring at a critical site, or resulting in death. Secondary safety outcome: a composite of major bleeding and clinically relevant nonmajor bleeding.
Patel MR, et al. (2011): ROCKET-AF[3]	Stroke was defined as a sudden focal neurologic deficit of presumed cerebrovascular aetiology that persisted beyond 24 hours. Systemic embolism was defined as abrupt vascular insufficiency associated with clinical or radiological evidence of arterial occlusion in the absence of another likely mechanism.	Principle safety endpoint: a composite of major and nonmajor clinically relevant bleeding. Major bleeding was defined as clinically overt bleeding associated with any of the following: fatal outcome, involvement of a critical anatomic site, fall in haemoglobin concentration ≥ 2g/dL, transfusion of ≥ 2 units of whole blood or packed red cells, or permanent disability. Non-major clinically relevant bleeding was defined as overt bleeding not meeting the criteria for major bleeding but requiring medical intervention, unscheduled contact with a physician, temporary interruption of study drug, pain, or impairment of daily activities.
Zelniker TA, et al. (2021): Secondary analysis of ENGAGE AF-TIMI 48[4]	A stroke is defined as an abrupt onset, over minutes to hours, of a focal neurologic deficit that is generally in the distribution of a single brain artery and that is not due to an identifiable nonvascular cause. The deficit must either be associated with symptoms lasting >24 hours or result in death within 24 hours of symptom onset. Systemic embolic event is defined as an abrupt episode of arterial insufficiency associated with clinical or radiologic evidence of arterial occlusion in the absence of other likely mechanisms	Major bleeding was defined according to the International Society on Thrombosis and Haemostasis (ISTH) criteria.
Mieghem NMV, et al. (2021): ENVISAGE-TAVI AF[5]	Stroke classified by VARC-2; Acute episode of a focal or global neurological deficit with at least one of the following: change in the level of consciousness, hemiplegia, hemiparesis, numbness, or sensory loss affecting one side of the body, dysphasia or aphasia, hemianopia, amaurosis fugax, or other neurological signs or symptoms consistent with stroke. Confirmation of the diagnosis by at least one of the following: neurologist or neurosurgical specialist; neuroimaging procedure (CT scan or brain MRI), but stroke may be diagnosed on clinical grounds alone. Classified into Ischemic, haemorrhagic, or undetermined.	The primary safety outcome was the incidence of major bleeding, designated according to ISTH definitions as clinically overt bleeding associated with a reduced haemoglobin level, blood transfusion, symptomatic bleeding at a critical site, or death
Graham DJ, et al. (2014)[6]	The effectiveness outcome was ischemic stroke defined by ICD-9-CM codes.	Safety outcomes were major bleeding with a specific focus on intracranial or gastrointestinal bleeding. Major bleeding was defined as a fatal bleeding event, a hospitalized bleeding event requiring transfusion, or hospitalization with hemorrhage into a critical site. The outcomes were defined according to ICD-9-CM codes.

Author (Year)	Stroke/systemic embolism definition	Bleeding definition
Tsadok MA, et al. (2015)[7]	Hospital admission or emergency room visit for stroke or TIA. Stroke was defined as ischemic cerebrovascular disease, with the inclusion of TIA and retinal infarct. The outcomes were identified by ICD-9 and ICD-10.	Hospital admission or emergency room visit for any bleeding event which were a composite of intracranial hemorrhage, gastrointestinal hemorrhage, and other hemorrhage (the composite outcome was classified as any bleeding in our study). The outcomes were identified by ICD-9 and ICD-10.
Lauffenburger JC, et al. (2015)[8]	Clinical effectiveness was defined as a composite of the occurrence of ischemic stroke, TIA, and other thromboembolic events. The outcomes were measured by using validated ICD-9 coding algorithms.	Harm was defined as a composite of intracranial hemorrhage or haemorrhagic stroke, gastrointestinal hemorrhage, or other bleeding (the composite outcome was classified as any bleeding in our study). The outcomes were measured by using validated ICD-9 coding algorithms.
Seeger JD, et al. (2015)[9]	Hospitalization for haemorrhagic or ischemic stroke. The outcomes were identified by ICD-9.	Hospitalization for major bleeding including intracranial and extracranial bleeding. The outcomes were identified by ICD-9.
Arihiro S, et al. [SAMURAI-NVAF] (2016)[10]	The effectiveness outcome was stroke or systemic embolism within three months.	The safety outcome was major bleeding within three months defined according to the ISTH including fatal bleeding, symptomatic bleeding in a critical area or organ, or bleeding causing a fall in haemoglobin level of 2.0 g/dL or more or leading to transfusion of two or more units of whole or red blood cells.
Norby FL, et al. (2017)[11]	Ischemic stroke was defined based on the presence of ICD-9-CM codes 434.xx (occlusion of cerebral arteries) and 436.xx (acute but ill-defined cerebrovascular disease) as the primary discharge diagnosis in any inpatient claim following the index date.	Intracranial bleeding was defined based on the presence of ICD-9-CM codes 430 (subarachnoid haemorrhage) and 431 (intracerebral haemorrhage) as the primary discharge diagnosis. GI bleeding was defined by presence of bleeding-related ICD-9-CM codes in inpatient claims as primary and secondary diagnoses, presence of transfusion codes, and presence/absence of trauma codes.
Lip GYH, et al. (2017)[12]	Efficiency endpoint was the occurrence of ischemic stroke or systemic embolism which was extracted from hospital discharge code (ICD-10 codes).	The safety endpoint was bleeding events which were a composite of intracranial, gastrointestinal, traumatic intracranial, and clinically relevant nonmajor bleeding (the composite outcome was classified as any bleeding in our study).
Bengtson LGS, et al. (2017)[13]	Effectiveness outcome was ischemic stroke which was defined according to ICD-9-CM discharge diagnosis codes.	Safety outcomes were intracranial bleeding or gastrointestinal bleeding. The outcomes were defined according to ICD-9-CM discharge diagnosis codes.
Shantha GPS, et al. (2017)[14]	Inpatient admission for acute ischemic stroke which was defined based on the primary ICD-9-CM diagnosis.	Inpatient admission for acute major bleeding which was defined based on the primary ICD-9-CM diagnosis. The secondary outcomes were subdivision of major bleeding, defined as intracranial hemorrhage, gastrointestinal hemorrhage (GIH), and other major non-GIH.
Hsu C, et al. (2018)[15]	The outcome assessed was ischemic stroke or thromboembolism which was defined as a hospitalization for cerebral infarction, unspecified cerebral infarction, arterial embolism and thrombosis, other transient cerebral ischemia attacks, and related symptoms, or unspecified transient ischemic attack. The outcomes were identified by using ICD9.	A composite safety endpoint which included intracranial hemorrhage, gastrointestinal hemorrhage, and haematuria (the composite outcome was classified as any bleeding in our study). The outcomes were identified by using ICD9.
Lee S, et al. (2018)[16]	The outcome assessed was ischemic stroke which was defined by ICD-10-CM and diagnostic definition.	The safety outcomes were intracranial hemorrhage, hospitalization for GI bleeding, and hospitalization for major bleeding (ICH and GI bleeding), which were defined by ICD-10-CM code and diagnostic definition.

Author (Year)	Stroke/systemic embolism definition	Bleeding definition
Law SWY, et al. (2018)[17]	The effectiveness outcome was the composite of ischemic stroke or systemic embolism which was identified using ICD-9-CM.	The bleeding outcomes included intracranial hemorrhage and gastrointestinal bleeding which were identified using ICD-9-CM.
Lip GYH, et al. (2018)[18]	The effectiveness outcome was a hospitalization for ischemic stroke, haemorrhagic stroke, and systemic embolism which were identified by ICD-9 diagnosis codes.	The safety outcome was a hospitalization for major bleeding including gastrointestinal bleeding, intracranial hemorrhage, and major bleeding at other key sites. The outcomes were identified by ICD-9 diagnosis and procedure codes.
Baker WL, et al. (2019)[19]	The primary endpoints were major adverse cardiac events (MACE) which was defined as ischemic stroke or myocardial infarction (≥ 1 ICD diagnosis in the primary position during a hospitalization or emergency department visit).	The safety endpoint was major bleeding, defined using the validated Cunningham algorithm for detection of bleeding-related hospitalizations.
Chang S, et al. (2019)[20]	Admission with a primary diagnosis of ischemic stroke, transient ischemic attack, or systemic thromboembolism. The outcomes were identified by using ICD-9 and ICD-10.	Admission with a primary diagnosis of major bleeding of gastrointestinal, intracranial, or urogenital tract. The outcomes were identified by using ICD-9 and ICD-10.
Lee S, et al. (2019)[21]	The effectiveness outcome was ischemic stroke which was defined according to ICD-10-CM code and diagnostic definition.	The safety outcomes were intracranial hemorrhage, a hospitalization for gastrointestinal bleeding, and hospitalization for major bleeding. The outcomes were identified using ICD-10-CM code and diagnostic definition.
Huybrechts KF, et al. (2019)[22]	The effectiveness outcomes were hospitalization for stroke including haemorrhagic, ischemic, or stroke of uncertain classification. The outcomes were identified using ICD-9 diagnosis codes.	The safety outcome was major bleeding including i) major intracranial and extracranial bleeding and ii) major upper or lower GI bleeding. The outcomes were identified using ICD-9 diagnosis codes.
Kwon S, et al. (2020)[23]	The effectiveness outcome was ischemic stroke which was identified by ICD-10-CM code and diagnostic definition.	The safety outcomes were intracranial hemorrhage, gastrointestinal bleeding, and major bleeding. The outcomes were identified by ICD-10-CM code and diagnostic definition.
Weir MR, et al. (2020)[24]	The effectiveness outcome was a hospitalization for ischemic stroke or systemic embolism in the 2 years post-index. The outcomes were defined according to ICD-9-CM and ICD-10-CM diagnosis codes.	The safety outcome was a hospitalization for major bleeding in the 2 years post-index. The outcomes were defined according to ICD-9-CM and ICD-10-CM diagnosis codes.
Wong JM, et al. (2020)[25]	NA	Major bleeding outcomes were categorized as i) intracranial bleeding, ii) gastrointestinal bleeding, iii) other bleeding, and iv) any bleeding. The outcomes were defined using ICD-9 codes using the linked CMS database.

Author (Year)	Stroke/systemic embolism definition	Bleeding definition
Bang OY, et al. (2020)[26]	The effectiveness outcome was stroke (ischemic and haemorrhagic stroke) and systemic embolism. The outcomes were identified from the diagnosis codes using hospitalization and brain CT/MRI records.	The safety outcome was major bleeding including intracranial hemorrhage, gastrointestinal and other bleeding. The outcomes were identified from the related diagnosis codes using hospitalization records.
Costa OS, et al. (2020)[27]	The effectiveness outcome was the stroke and systemic embolism defined by an appropriate inpatient discharge ICD-9 diagnosis code in the primary coding position.	The safety outcome was major bleeding using the validated Cunningham algorithm.
Linder M, et al. (2020)[28]	NA	The safety outcomes were gastrointestinal bleeding, intracranial bleeding, and any bleeding (a composite of other bleeding, gastrointestinal bleeding, and intracranial bleeding).
Duan L, et al. (2021)[29]	The effectiveness outcome was a composite of ischemic stroke, transient ischemic attack (TIA) or systemic embolism. The outcomes were identified using ICD-9 and ICD-10 codes in the primary discharge diagnosis position for inpatient hospitalizations.	The safety outcome was a composite of major bleeding including gastrointestinal bleeding, intracranial hemorrhage, and bleeding from other sites. The outcomes were identified using ICD-9 and ICD-10 codes in the primary discharge diagnosis position for inpatient hospitalizations.
Gulluoglu FR, et al. (2020)[30]	The effectiveness outcome was a composite of ischemic stroke, unspecified stroke, and haemorrhagic stroke. The outcomes were defined using the UK Read code system.	The safety outcome was major bleeding, defined as a composite of intracranial bleeding, gastrointestinal bleeding, and bleeding on other clinically relevant sites. The outcomes were defined using the UK Read code system.
Balsam P, et al. (2021)[31]	The effectiveness endpoint was thromboembolic events/ischemic events which were defined by ICD-10 codes. Ischemic events consist of diagnosis codes for ischemic stroke, transient ischemic attack, and peripheral thromboembolism.	The safety endpoint was haemorrhagic events which were defined by ICD-10 codes. Haemorrhagic events were a composite of gastrointestinal, intracranial, and other locations of bleeding related codes (the composite outcome was classified as any bleeding in our study).
Kwon S, et al. (2021)[32]	The effectiveness outcome was ischemic stroke defined by ICD-10-CM codes and diagnostic definitions.	Safety outcomes included intercranial haemorrhage, gastrointestinal bleeding, and major bleeding. Major bleeding was defined as a composite of intercranial haemorrhage or major gastrointestinal bleeding requiring hospitalisation. The outcomes were defined using ICD-10-CM codes and diagnostic definitions.
Coleman CI, et al. (2021)[33]	NA	Outcome was major or clinically non-major bleeding requiring hospitalisation (classified as any bleeding in our study). Major bleeding component was intended to approximate the International Society of Thrombosis and Haemostasis (ISTH). ICD-9/10-CM diagnosis codes, CPT-4, HCPCS, ICD-9/10-PCS procedure codes or laboratory, vital signs, and other patient observation results were used.

Author (Year)	Stroke/systemic embolism definition	Bleeding definition
Halvorsen S, et al. (2022)[34]	The effectiveness endpoint was a composite of any stroke (ischemic or haemorrhagic) or systemic embolism requiring an acute hospitalization with an overnight stay. The endpoint was identified using ICD-10 codes.	The safety endpoint was major bleeding defined as any bleeding (intracranial, gastrointestinal, or other) requiring acute hospitalization with an overnight stay. The endpoint was identified using ICD-10 codes.
Subramanya V, et al. (2021)[35]	Defined as initial hospitalisation for an ischemic stroke after a diagnosis of AF. The endpoint was identified using ICD-9-CM codes and applying validated algorithms.	The safety endpoint initial hospitalisation for a major bleeding episode after a diagnosis of AF. The endpoint was identified using ICD-9-CM codes and applying validated algorithms.
Komen JJ, et al. (2022)[36]	The effectiveness outcome included ischemic or unspecified stroke. This were identified using ICD-10 codes	The safety outcome was any major bleeding. These were identified using ICD-10 codes.
Moon I, et al. (2022)[37]	The outcome was a composite of hospitalization under the diagnosis of ischemic stroke or systemic embolism after three weeks of anticoagulation. The outcome was defined by the ICD-10-CM codes.	Bleeding outcome included major bleeding (intracranial hemorrhage and gastrointestinal bleeding) which was defined by the ICD-10-CM codes.
Yoshimoto T, et al. (2022)[38]	The endpoint was the incidence of stroke/systemic embolism ascertained by collecting at patient's most recent outpatient visit after the onset of the end point or, in some cases, by hospital visit records or reports from other medical institutions .	Major bleeding was defined according to the International Society on Thrombosis and Haemostasis Statement.

Table S6. Summary characteristics of included studies

Author (Year)	Study design	Country/ Region	Data source	Population	Men/Women (n)	Men/Women (n) after matching or weighting	Age (years)	Mean $CHADS_2$ / CHA_2DS_2-VASc Score	Intervention (compared with warfarin)	Duration of follow-up (years)	% INR in the therapeutic range (2.0-3.0)
Connolly SJ, et al. (2009) [1]: RE-LY	RCT	Multicentre (44 countries)	951 clinical centres (44 countries)	AF with at least one risk factor for stroke	11,514/6,599	N/A	Mean=71	Mean $CHADS_2$=2.1	Dabigatran 110 mg BID, 150 mg BID	Median=2.0	Mean=64%
Granger CB, et al. (2011) [2]: ARISTOTLE	RCT	Multicentre (39 countries)	Multicentre (39 countries)	AF with at least one risk factor for stroke	11,785/6,416	N/A	Median (IQR)=70 (63-76)	Mean $CHADS_2$ (SD)=2.1 (1.1)	Apixaban 5 mg BID, 2.5 mg BID	Median=1.8	Median =66.0%
Patel MR, et al. (2011) [3]: ROCKET-AF	RCT	Multicentre (45 countries)	Multicentre (45 countries)	AF at moderate to high risk for stroke	8,601/5,663	N/A	Median (IQR)=73 (65-78)	Mean $CHADS_2$=3.5	Rivaroxaban 20 mg QD, 15 mg QD	Median=1.9	Median=58%
Zelniker TA, et al. (2021) [4]	RCT (Secondary analysis from ENGAGE AF-TIMI 48)	Multicentre (46 countries)	Multicentre (46 countries)	AF at moderate to high risk for stroke	13,065/8,040	N/A	Median (IQR): men=71 (63-77), women=74 (67-79)	Mean $CHADS_2$ (SD): men=2.8 (1.0), women 2.9 (1.0)	Edoxaban 60 mg QD (high dose), 30 mg QD (low dose)	Median=2.8	Mean=64.9±18.7%
Mieghem NMV, et al. (2021) [5]: ENVISAGE-TAVI AF	RCT	Multicentre (14 countries)	Multicentre (14 countries)	AF after successful transcatheter aortic-valve replacement	748/678	N/A	Mean (SD): 82.1 (5.4)	Median CHA2DS2-VASc (IQR): 4 (4-5) Mean CHA2DS2-VASc (SD): 4.5 (1.4)	Edoxaban 60 mg QD (high dose), 30 mg QD (low dose)	Median = 1.5	Mean = 63.5% Median = 68.2%

Author (Year)	Study design	Country/ Region	Data source	Population	Men/Women (n)	Men/Women (n) after matching or weighting	Age (years)	Mean $CHADS_2$ / CHA_2DS_2-VASc Score	Intervention (compared with warfarin)	Duration of follow-up (years)	% INR in the therapeutic range (2.0-3.0)
Graham DJ, et al. (2015) [6]	Retrospective cohort	USA	Patients enrolled in Medicare	NVAF	NR	65,190/69,223 (Table 1)	NR	NR	Dabigatran	NR	NR
Tsadok MA, et al. (2015) [7]	Retrospective cohort	Canada	Administrative data	AF	31,324/31,786 (Table 1)	No matching/weighting performed; 31,324/31,786 (Table 1)	Mean (SD): men=76.3 (9.3), women=80.3 (8.8)	Mean CHA_2DS_2-VASc Score (SD): men=2.6 (1.4), women=3.9 (1.3)	Dabigatran 110 BID, 150 mg BID	Median=1.3	NR
Lauffenburger JC, et al. (2015) [8]	Retrospective cohort	USA	the Truven Health MarketScan Commercial Claims and Encounters and Medicare supplement databases	NVAF	42,334/22,601 (In-text)	38,925/26,010 (Table 6)	Mean (SD)=69.9 (12.4)	Mean CHA_2DS_2-VASc Score: Dabigatran=2.3 (1.6), Warfarin=2.9 (1.7)	Dabigatran	Mean=0.98 (0.6)	NR
Seeger JD, et al. (2015) [9]	Retrospective cohort	USA	Two commercial health insurance databases (MarketScan, Truven and Clinformatics, Optum)	NVAF with CHA_2DS_2-VASc score ≥ 1	36,789/22,874 (Table 1A)	24,440/13,938 (Figure 3)	Mean=68	Mean CHA_2DS_2-VASc Score: MarketScan dabigatran =3.06 (1.6), warfarin=3 (1.5); Clinformatics dabigatran=2.81 (1.6), warfarin=2.78 (1.61)	Dabigatran	Dabigatran: mean=0.42, warfarin: mean=0.34	NR
Arihiro S, et al. [SAMURAI-NVAF] (2016) [10]	Prospective cohort	Japan	18 stroke centers balanced regionally across Japan	NVAF after onset of ischemic stroke or TIA	645/492 (Table 1)	No matching/weighting performed; 645/492 (Table 1)	Mean (SD)=77 (10)	Median (IQR) CHA_2DS_2-VASc Score: DOACs=5 (4-6), warfarin=6 (5-6)	DOACs (dabigatran, rivaroxaban, apixaban)	3 months	NR
Norby FL, et al. (2017) [11]	Retrospective cohort	USA	the Truven Health MarketScan Commercial Claims and Encounters and Medicare supplement and Coordination of Benefits databases	NVAF	87093/54007 (Table S1)	69504/42401 (Table 1)	Mean = 70	Mean CHA2DS2-VASc Score: Rivaroxaban = 3 (1.9) Warfarin = 3.2 (2)	Rivaroxaban	Mean = 12 months	NR

Author (Year)	Study design	Country/ Region	Data source	Population	Men/Women (n)	Men/Women (n) after matching or weighting	Age (years)	Mean $CHADS_2$ / CHA_2DS_2-VASc Score	Intervention (compared with warfarin)	Duration of follow-up (years)	% INR in the therapeutic range (2.0-3.0)
Lip GYH, et al. (2017) [12]	Retrospective cohort	Denmark	Danish registries	AF with 1 low-risk, nonsex-related stroke risk factor	8,869/5,151	NR	Median (IQR)=66.5 (61.1-70.4)	NR	Standard dose of DOACs (dabigatran 150 mg BID, rivaroxaban 20 mg QD, apixaban 5 mg BID)	Mean (SD)=2.6 (1.6)	NR
Bengtson LGS, et al. (2017) [13]	Retrospective cohort	USA	the US MarketScan databases	NVAF	NR	36,937/22,611 (Table 1; dabigatran & warfarin new-users)	Warfarin: mean=70.8 (12.1), dabigatran: mean=68.5 (12.3), rivaroxaban: mean=70.4 (12)	$CHADS_2$ score: warfarin =2.2 (1.5), dabigatran =2.0 (1.4), rivaroxaban =2.4 (1.5)	Dabigatran or rivaroxaban	Dabigatran vs warfarin: median=1.25, rivaroxaban vs warfarin: median=0.67	NR
Shantha GPS, et al. (2017) [14]	Retrospective cohort	USA	the Centers for Medicare and Medicaid Services	AF aged ≥ 66 years	66, 234/81,137 (Supplement Table 1)	22,854/33,093 (Table 1)	Mean (SD); men: dabigatran=75.9 (6.1), rivaroxaban=75.1 (6.2), warfarin=74.8 (6.1); women: dabigatran=76.8, rivaroxaban=76.8, warfarin=76.8	Mean CHA_2DS_2-VASc Score: men=3.8, women=4.8	Rivaroxaban 20 mg QD, dabigatran 150 mg BID	Median=1.17	NR
Hsu C, et al. (2018) [15]	Retrospective cohort	Taiwan	Nationwide P4P diabetes care program which was implemented by Taiwan's National Health Insurance (NHI) Administration	NVAF with T2DM	1289/1252 (Table 1)	614/597 (Table 1)	Mean (SD): Dabigatran= 75.1 (9.1), warfarin1=73.9 (8.7), rivaroxaban= 75.2 (8.7), warfarin2= 74.4 (8.2)	Reported in %, CHA_2DS_2-VASc Score of 5: 22.3% in dabigatran, 20.0% in warfarin1, 19.3% in rivaroxaban, 20.6% in warfarin2	Dabigatran, rivaroxaban	Median (IQR)=1.7 (0.5-3.7)	NR
Lee S, et al. (2018) [16]	Retrospective cohort	Korea	the national health claims database established by the National Health Insurance Service (NHIS) of Korea	NVAF	21,656/14,109 (Table 1)	9136/7108 (Table 1)	Mean=70	Mean CHA_2DS_2-VASc score=3.2	Edoxaban 60 mg QD and 30 mg QD	Median (IQR)=0.3 (0.1-0.5)	NR
Law SWY, et al. (2018) [17]	Retrospective cohort	Hong Kong	the Clinical Data Analysis and Reporting System (CDARS)	NVAF	7,900/7,392 (Online Table 2)	4,972/4,834 (Table 1)	Mean (SD): men=71.7 (10.8), women=75.8 (10.1)	Mean (SD): CHA_2DS_2VASc: men=2.96 (1.68), women=4.34 (1.79)	DOACs (dabigatran, rivaroxaban, and apixaban)	Mean (SD): men=1.23 (1.33), women=1.29 (1.4)	Mean TTR (SD): men=45.1 (29.1), women=46.0 (29.0)

Author (Year)	Study design	Country/ Region	Data source	Population	Men/Women (n)	Men/Women (n) after matching or weighting	Age (years)	Mean $CHADS_2$ / CHA_2DS_2-VASc Score	Intervention (compared with warfarin)	Duration of follow-up (years)	% INR in the therapeutic range (2.0-3.0)
Lip GYH, et al. (2018) [18]	Retrospective cohort	USA	the US Centers for Medicare and Medicaid Services Medicare data and 4 commercial claims databases in the United States	NVAF	251,072/ 215,963 (Supplemental Table 2)	278,546/247,524 (Table 1)	Mean (SD); Apixaban vs Warfarin: 76.1 (9.8) vs 76.0 (9.7); Dabigatran vs Warfarin: 73.2 (10.3) vs 73.3 (10.3); Rivaroxaban vs Warfarin: 75.6 (9.5) vs 75.7 (9.6)	Mean (SD): CHA2DS2VASc score; Apixaban vs Warfarin: 3.9 (1.7) vs 3.9 (1.6); Dabigatran vs Warfarin: 3.5 (1.7) vs 3.5 (1.7); Rivaroxaban vs Warfarin: 3.8 (1.6) vs 3.8 (1.6)	DOACs (dabigatran, rivaroxaban, and apixaban)	Varied across treatment groups ranging from 123-159 days	NR
Baker WL, et al. (2019) [19]	Retrospective cohort	USA	US Truven MarketScan data	NVAF with T2DM	NR	15667/8979 (Table 1)	Median (IRQ)=70 (62-79)	Median (IQR) CHA_2DS_2-VASc Score=4 (3-5)	Rivaroxaban	Median (IQR)=1.4 (0.6-2.7)	NR
Chang S, et al. (2019) [20]	Retrospective cohort	Taiwan	Chang Gung Memorial Hospital system (multicentre)	AF with CKD stage 4 or 5	355/445 (Table 1)	No matching/weighting performed; 355/445 (Table 1)	47.9% in DOACs aged between 75-84, 42.5% in warfarin aged between 75-84	CHA_2DS_2-VASc score: DOACs=4.7 (1.5), warfarin=4.6 (1.7)	DOACs	Mean=3.2	NR
Lee S, et al. (2019) [21]	Retrospective cohort	Korea	the Korean National Health Insurance Service database	NVAF with liver disease	22288/15 065 (Table 1)	22198/15155 (Table 1)	Mean (SD): DOACs=69 (9.6), warfarin=69.2 (10.5)	CHA_2DS_2-VASc score: DOACs=3.5 (1.6), warfarin=3.5 (1.9)	DOACs	Mean=1.2	NR
Huybrechts KF, et al. (2019) [22]	Retrospective cohort	USA	the IBM MarketScan Commercial Claims and Encounters database and Medicare Supplement and the Optum Research Database	NVAF	85,403/5 1,925 (Table 1)	37,677/21,219 (Table 1)	Mean (SD): dabigatran=67.84 (11.87), warfarin=67.65 (12.07)	CHA_2DS_2-VASc score: dabigatran=3.05 (1.6), warfarin=3.01 (1.56)	Dabigatran	NR	NR
Kwon S, et al. (2020) [23]	Retrospective cohort	Korea	the Korean Health Insurance Review and Assessment (HIRA) service	NVAF with very old aged (≥80 years)	9662/14, 997 (Table 1)	9650/15,009 (Table 1)	Mean (SD): DOACs=84 (3.7), warfarin=84.1 (3.6)	CHA_2DS_2-VASc score: DOACs=4.7 (1.2), warfarin=4.7 (1.2)	DOACs (dabigatran, rivaroxaban, apixaban, edoxaban)	NR	NR

Author (Year)	Study design	Country/ Region	Data source	Population	Men/Women (n)	Men/Women (n) after matching or weighting	Age (years)	Mean $CHADS_2$ / CHA_2DS_2-VASc Score	Intervention (compared with warfarin)	Duration of follow-up (years)	% INR in the therapeutic range (2.0-3.0)
Weir MR, et al. (2020) [24]	Retrospective cohort	USA	the Optum® Deidentified Electronic Health Record Database	NVAF with CKD stage 4 or 5	5583/5393 (Supplemental Material 13)	944/618 (Supplemental Material 13)	Mean (SD): rivaroxaban=79.9 (8.2), warfarin=79.9 (8.2)	CHA_2DS_2-VASc score: rivaroxaban=4.5 (1.5), warfarin=4.5 (1.5)	Rivaroxaban	Mean: rivaroxaban= 1.07, warfarin=1.01	NR
Wong JM, et al. (2020) [25]	Retrospective cohort	USA	American College of Cardiology's National Cardiovascular Disease Registry's Practice Innovation and Clinical Excellence (PINNACLE) registry	AF or flutter	153,103/115,917 (Table 1)	No matching/weighting performed; 153,103/115,917 (Table 1)	Mean (SD): rivaroxaban=75.6 (7.3), dabigatran=75.5 (7.3), apixaban=76.5 (7.4), warfarin=77.3 (7.5)	CHA_2DS_2-VASc score: rivaroxaban=4.1 (1.4), dabigatran=4.2 (1.4), apixaban=4.2 (1.4), warfarin=4.5 (1.4)	DOACs (rivaroxaban, dabigatran, and apixaban)	Median=1.4 (0.6)	NR
Bang OY, et al. (2020) [26]	Retrospective cohort	Korea	the Korean Health Insurance Review and Assessment Service (HIRA) database	NVAF	26,907/21,482 (Table 1)	After weighting % women in each group: apixaban 43.93%, dabigatran=41.01%, rivaroxaban=43.15% (Table 1)	After weighting mean age in each group: apixaban=71.64, dabigatran=70.68, rivaroxaban=71.83	After weighting mean CHA_2DS_2-VASc score in each group: apixaban 4.52, dabigatran=4.4, rivaroxaban=4.45	DOACs (rivaroxaban, dabigatran, and apixaban)	Median=149, 171, 175, and 105 days in apixaban, dabigatran, rivaroxaban, and warfarin, respectively	NR
Costa OS, et al. (2020) [27]	Retrospective cohort	USA	US Optum de-Identified Electronic Health Record data	NVAF with BMI > 30 kg/m^2	NR	42,842/28,384 (Table 1)	After matching median age (IQR) in rivaroxaban=67 (60-75), warfarin=69 (61-75)	After matching median (IQR) CHA_2DS_2-VASc score in rivaroxaban=3 (2-4), warfarin=3 (2-4)	Rivaroxaban	Median (IQR)=2.6 (1.2-4.1)	NR
Linder M, et al. (2020) [28]	Retrospective cohort	Sweden	The Swedish patient register (NPR) and The Swedish prescribed drug register (PDR)	NVAF	NR	% women among comparisons; DTI (46.9%) vs VKA (47.2%), XAB (46.8%) vs DTI (47.1%), XAB (49.7%) vs VKA (49.7%) (Table 1)	Mean age (SD) among comparisons: DTI=72.8 (7.7) vs VKA=73.1(7.8), XAB=73.0 (7.7) vs DTI=72.7 (7.7), XAB=75.6 (8.5) vs VKA=75.6 (8.3)	CHA_2DS_2-VASc score among comparisons: DTI=3.0 (1.5) vs VKA=3.0 (1.5), XAB=2.9 (1.5) vs DTI=2.9 (1.5), XAB=3.4 (1.6) vs VKA=3.4 (1.6)	DTI (dabigatran) and XAB (rivaroxaban and apixaban)	NR	NR

Author (Year)	Study design	Country/ Region	Data source	Population	Men/Women (n)	Men/Women (n) after matching or weighting	Age (years)	Mean $CHADS_2$ / CHA_2DS_2-VASc Score	Intervention (compared with warfarin)	Duration of follow-up (years)	% INR in the therapeutic range (2.0-3.0)
Duan L, et al. (2021) [29]	Retrospective cohort	USA	the Kaiser Permanente Southern California (KPSC) Health System	AF with bioprosthetic heart valves	1620/1052 (Table 1)	1622/1051 (Table 1)	56.3% in DOACs aged ≥75years, 56.5% in warfarin aged ≥75years	70% in both DOACs and warfarin had a CHA_2DS_2-VASc score ≥ 4	DOACs	Mean (SD)= 2.9 (2.2)	NR
Gulluoglu FR, et al. (2020) [30]	Retrospective cohort	UK	the Clinical Practice Research Datalink (CPRD)	AF with T2DM	5261/3294 (Table 1)	4201/2673 (Table C4)	Mean: DOACs=76.07, warfarin=75.05	CHA_2DS_2-VASc score: DOACs=4.18 (1.52), warfarin=4.12 (1.49)	DOACs	Mean follow-up for stroke outcomes: DOACs=1.3, warfarin=2.1, for bleeding outcomes: DOACs=1.3, warfarin= 2.03	NR
Balsam P, et al. (2021) [31]	Retrospective cohort	Poland	the MultiCenter expeRience in AFib patients Treated with oral anticoagulation (CRAFT)	AF	1,759/1,224 (Table 1)	% women among comparisons; Dabigatran (42%) vs VKA (41%), Rivaroxaban (47%) vs VKA (47%) (Table S3 & Table S4)	Median (IQR): rivaroxaban=74 (65-81), dabigatran=69 (62-78), VKA=68 (61-78)	Median CHA_2DS_2-VASc score (IQR): rivaroxaban=4 (3-5), dabigatran=3 (2-5), VKA=3 (2-5)	DOACs (rivaroxaban and dabigatran)	Mean=4.0	NR
Kwon S, et al. (2021) [32]	Retrospective cohort	Korea	The Korean Health Insurance Review and Assessment (HIRA) service database	AF with prior gastrointestinal bleeding	23,938/18,110 (Table 1)	23,845/18,118 (Table 1)	Mean (SD): Pooled DOAC = 71.8 (9.9) Warfarin = 71.8 (10.4)	Mean CHA2DS2-VASc score (SD): Pooled DOAC = 3.7 (1.5) Warfarin = 3.7 (1.6)	DOACs	Median (IQR) = 0.6 (0.2 - 1.7)	NR

Author (Year)	Study design	Country/ Region	Data source	Population	Men/Women (n)	Men/Women (n) after matching or weighting	Age (years)	Mean $CHADS_2$ / CHA_2DS_2-VASc Score	Intervention (compared with warfarin)	Duration of follow-up (years)	% INR in the therapeutic range (2.0-3.0)
Coleman CI, et al. (2021) [33]	Retrospective cohort	USA	US Optum de-Identified Electronic Health Record data	AF with T2DM	68,990/47,059 (Table 1)	69,049/47,000 (Table 1)	Mean age (SD): Rivaroxaban = 71 (10) Warfarin = 71 (10)	Mean CHA2DS2-VASc score (SD): Rivaroxaban = 4.3 (1.5) Warfarin = 4.3 (1.5)	Rivaroxaban	Mean = 2.9	Mean TTR = 47% Median TTR = 50%
Halvorsen S, et al. (2022) [34]	Retrospective cohort	Denmark, Norway, Sweden	the Danish National Patient Registry, Danish National Health Service Prescription Database, Danish Civil Registration System, Norwegian Patient Registry, Norwegian Prescription Database, National Population Register of Norway, Swedish National Patient Register, Swedish Prescribed Drug Register, Swedish Total Population Register, and Swedish Cause of Death Register	NVAF	124,713/94,832 (Table S3)	Apixaban (44%/56%) vs warfarin (43.9%/56.1%), dabigatran (39.7%/60.3%) vs warfarin (40%/60%), rivaroxaban (43.9%/56.1%) vs warfarin (43.8%/56.2%) (Table 1)	Median (IQR): apixaban=75.1 (67.8-82.8) vs warfarin=75.1 (67.8-82.8), dabigatran=71.7 (65.0-79.3) vs warfarin=71.7 (64.9-79.3), rivaroxaban=74.6 (67.4-82.3) vs warfarin=74.7 (67.4-82.4)	Mean: apixaban vs warfarin = 3.4, dabigatran vs warfarin = 2.9, rivaroxaban vs warfarin = 3.3	Apixaban, dabigatran, rivaroxaban	Median (IQR)=9.7 (3.9-21.5) months for stroke/SE and 9.6 (3.8-21.3) months for any bleeding	NR
Subramanya V, et al. (2021) [35]	Retrospective cohort	USA	the IBM MarketScan Medicare Supplemental and coordination of benefits database	AF aged (≥75 years)	62,898/58,186 (Table 3)	No matching/weighting performed; 62,898/58,186 (Table 3)	Mean (SD): men=82.5 (5.2) women=83.8 (5.6)	NR	DOACs	NR	NR
Komen JJ, et al. (2022) [36]	Retrospective cohort	Denmark, Norway, Scotland, and the Stockholm region in Sweden	four Western European databases, namely Denmark, Norway, Scotland, and the Stockholm region in Sweden.	NVAF at low risk (one non-sex-related CHA2DS2-VASc point)	20,693/12,433 (Table 1)	NR	Mean: DOACs=65.3, VKA=64.2	NR	DOACs (apixaban, dabigatran, edoxaban)	NR	NR

Author (Year)	Study design	Country/ Region	Data source	Population	Men/Women (n)	Men/Women (n) after matching or weighting	Age (years)	Mean $CHADS_2$ / CHA_2DS_2-VASc Score	Intervention (compared with warfarin)	Duration of follow-up (years)	% INR in the therapeutic range (2.0-3.0)
Moon I, et al. (2022) [37]	Retrospective cohort	Korea	The Korean National Health Insurance Service database (2013-2018)	AF with bioprosthetic heart valve	832/1010 (Table S4)	489/597 (Table 1)	Mean (SD): Warfarin = 78.9 (6.8) DOACs = 79 (7)	Mean (SD)=4.7 (1.4)	DOACs	Mean=1.2	NR
Yoshimoto T, et al. (2022) [38]	Prospective cohort	Japan	1273 medical institutions (2016-2018)	NVAF with aged ≥75 years (with previous ischemic stroke/TIA)	4583/2720 (Figure 3)	No matching/weighting performed; 4583/2720 (Figure 3)	Median (IQR)=81.0 (78.0-85.0)	Median (IQR): CHADS2=4(4-5), CHA2DS2-VASc=6 (5-7)	DOACs	Mean = 1.86 years	NR

Abbreviations: AF=atrial fibrillation; BMI=body mass index; CKD=chronic kidney disease; DOACs=direct oral anticoagulants; DTI=direct thrombin inhibitors; IQR=interquartile range; NR=not reported; NVAF=non-valvular atrial fibrillation; RCT=randomized controlled trial; SD=standard deviation; T2DM=type 2 diabetes mellitus; TIA=transient ischemic attack; VKA=vitamin K antagonist; XAB=direct factor Xa inhibitor (Xabans)

Table S7. Risk of bias of randomised controlled trials

Connolly SJ, et al. (2009)[1]**: RE-LY**

Bias domain	Risk-of-bias judgment	Comments
Bias arising from the randomization process	Low risk	Quote: All participants were randomly assigned to received one of two doses of dabigatran, or to receive warfarin, by means of a central, interactive, automated telephone system. Comment: Baseline characteristics were similar among three groups.
Bias due to deviations from intended interventions	Some concerns	A blinded manner was done in dabigatran groups and an open-label was used in warfarin group. However, the bias was reduced by the implementation of several validated procedures, including blinded evaluation of outcome events.
Bias due to missing outcome data	Low risk	Quote: complete follow-up was achieved in 99.9% of patients, with 20 patients lost to follow-up.
Bias in measurement of the outcome	Low risk	Comment: The methods used to measure the outcomes were appropriate. The ascertainment of the outcome was similar among groups. The blinded evaluation of outcome events was performed.
Bias in selection of the reported result	Low risk	There was no selection of the reported results.
Overall judgement	**Some concerns**	

Granger CB, et al. (2011)[2]**: ARISTOTLE**

Bias domain	Risk-of-bias judgment	Comments
Bias arising from the randomization process	Low risk	This study was a randomization-controlled trial. Two groups were well balanced with respect to baseline characteristics.
Bias due to deviations from intended interventions	Low risk	This study was a double-blind, double dummy design.
Bias due to missing outcome data	Low risk	Data for the outcomes was available nearly all and participants randomized (lost to follow-up in apixaban and warfarin group accounting for 0.4% each)
Bias in measurement of the outcome	Low risk	The methods used to measure the outcomes were appropriate. The ascertainment of the outcome was similar among groups. The blinded evaluation of outcome events was performed. Outcome definition was based on relevant criteria and clearly presented.
Bias in selection of the reported result	Low risk	There was no selection of the reported results.
Overall judgement	**Low risk**	

Patel MR, et al. (2011)[3]**: ROCKET-AF**

Bias domain	Risk-of-bias judgment	Comments
Bias arising from the randomization process	Low risk	This study was a randomization-controlled trial. Randomization was performed with the use of a central 24-hour, computerized, automated voice-response system.
Bias due to deviations from intended interventions	Low risk	This study was a double-blind, double dummy design.
Bias due to missing outcome data	Low risk	Data for the outcomes was available nearly all and participants randomized (99.3% in rivaroxaban, 99.4% in warfarin were available for outcome analyses).
Bias in measurement of the outcome	Low risk	The methods used to measure the outcomes were appropriate. The ascertainment of the outcome was similar among groups. Outcome definition was based on relevant criteria and clearly presented.
Bias in selection of the reported result	Low risk	There was no selection of the reported results.
Overall judgement	**Low risk**	

Zelniker TA, et al. (2021)[4]**: a secondary analysis from ENGAGE AF-TIMI 48**

Bias domain	Risk-of-bias judgment	Comments
Bias arising from the randomization process	Low risk	This study was a randomization-controlled trial. Randomization was performed with the use of a central, 24-hour, interactive, computerized response system.
Bias due to deviations from intended interventions	Low risk	This study was a double-blind, double-dummy design.
Bias due to missing outcome data	Low risk	Data for the outcomes was available nearly all and participants randomized (99.5% were available for outcome analyses).
Bias in measurement of the outcome	Low risk	The methods used to measure the outcomes were appropriate. The ascertainment of the outcome was similar among groups. Outcome definition was based on relevant criteria and clearly presented in the protocol.
Bias in selection of the reported result	Low risk	There was no selection of the reported results.
Overall judgement	**Low risk**	

Mieghem NMV, et al. (2021)[5]**: ENVISAGE-TAVI AF**

Bias domain	Risk-of-bias judgment	Comments
Bias arising from the randomization process	Low risk	This study was a randomization-controlled trial. Randomization was performed with the use of a central, 24-hour, interactive, computerized response system.
Bias due to deviations from intended interventions	Some concerns	This study was an open-label trial. However, blinded evaluation of endpoints was conducted to minimise potential biases.
Bias due to missing outcome data	Low risk	Data for the outcomes was available nearly all and participants randomized (only 1 (0.1%) patient was lost to follow-up in the edoxaban treatment arm
Bias in measurement of the outcome	Low risk	The methods used to measure the outcomes were appropriate. The ascertainment of the outcome was similar among groups. Outcome definition was based on relevant criteria and clearly presented in the protocol.
Bias in selection of the reported result	Low risk	There was no selection of the reported results.
Overall judgement	**Some concerns**	

Table S8. Quality assessment of observational studies using NOS

Author (year)	Selection				Comparability		Outcome			Score (9★) and quality of study
	Representativeness of the exposed cohort (DOAC group)	Selection of the non exposed cohort (Warfarin group)	Ascertainment of exposure	Demonstration that outcome of interest was not present at start of study	Study controls for age, gender, stroke risk and bleeding risk	Study controls for any additional factors	Assessment of outcome	Was follow-up long enough for outcomes to occur	Adequacy of follow up of cohorts	
Graham DJ, et al. (2015)[6]	★	★	★	★	★	★	★	0	★	(8★) Good quality
Tsadok MA, et al. (2015)[7]	★	★	★	★	★	★	★	★	★	(9★) Good quality
Lauffenburger JC, et al. (2015)[8]	★	★	★	★	★	★	★	★	★	(9★) Good quality
Seeger JD, et al. (2015)[9]	★	★	★	★	★	★	★	★	★	(9★) Good quality
Arihiro S, et al. (2016)[10]	0	★	★	★	★	★	★	★	★	(8★) Good quality
Norby FL, et al. (2017)[11]	★	★	★	★	★	★	★	★	★	(9★) Good quality
Lip GYH, et al. (2017)[12]	★	★	★	★	★	★	★	★	★	(9★) Good quality
Bengtson LGS, et al. (2017)[13]	★	★	★	★	★	★	★	★	★	(9★) Good quality
Shantha GPS, et al. (2017)[14]	★	★	★	★	★	★	★	★	★	(9★) Good quality
Hsu C, et al. (2018)[15]	★	★	★	★	★	★	★	★	★	(9★) Good quality
Lee S, et al. (2018)[16]	★	★	★	★	★	★	★	★	★	(9★) Good quality
Law SWY, et al. (2018)[17]	★	★	★	★	★	★	★	★	★	(9★) Good quality
Lip GYH, et al. (2018)[18]	★	★	★	★	★	★	★	★	★	(9★) Good quality
Baker WL, et al. (2019)[19]	★	★	★	★	★	★	★	★	★	(9★) Good quality
Chang S, et al. (2019)[20]	0	★	★	★	★	★	★	★	★	(8★) Good quality
Lee S, et al. (2019)[21]	★	★	★	★	★	★	★	★	★	(9★) Good quality
Huybrechts KF, et al. (2019)[22]	★	★	★	★	★	★	★	0	★	(8★) Good quality
Kwon S, et al. (2020)[23]	★	★	★	★	★	★	★	0	★	(8★) Good quality
Weir MR, et al. (2020)[24]	★	★	★	★	★	★	★	★	★	(9★) Good quality

Author (year)	Selection				Comparability		Outcome			Score (9★) and quality of study
	Representativeness of the exposed cohort (DOAC group)	Selection of the non exposed cohort (Warfarin group)	Ascertainment of exposure	Demonstration that outcome of interest was not present at start of study	Study controls for age, gender, stroke risk and bleeding risk	Study controls for any additional factors	Assessment of outcome	Was follow-up long enough for outcomes to occur	Adequacy of follow up of cohorts	
Wong JM, et al. (2020)[25]	★	★	★	★	0	★	★	★	★	(8★) Good quality
Bang OY, et al. (2020)[26]	★	★	★	★	★	★	★	★	★	(9★) Good quality
Costa OS, et al. (2020)[27]	★	★	★	★	★	★	★	★	★	(9★) Good quality
Linder M, et al. (2020)[28]	★	★	★	★	★	★	★	0	★	(8★) Good quality
Duan L, et al. (2020)[29]	★	★	★	★	★	★	★	★	★	(9★) Good quality
Gulluoglu FR, et al. (2020)[30]	★	★	★	★	★	★	★	★	★	(9★) Good quality
Balsam P, et al. (2021)[31]	0	★	★	★	★	★	★	★	0	(7★) Fair quality
Kwon S, et al. (2021)[32]	0	★	★	★	★	★	★	★	★	(8★) Good quality
Coleman CI, et al. (2021)[33]	★	★	★	★	★	★	★	★	★	(9★) Good quality
Halvorsen S, et al. (2022)[34]	★	★	★	★	★	★	★	★	★	(9★) Good quality
Subramanya V, et al. (2021)[35]	★	★	★	★	★	★	★	0	0	(7★) Fair quality
Komen JJ, et al. (2022)[36]	★	★	★	★	★	★	★	0	★	(8★) Good quality
Moon I, et al. (2022)[37]	★	★	★	★	★	★	★	★	★	(9★) Good quality
Yoshimoto T, et al. (2022)[38]	0	★	★	★	★	★	★	★	★	(8★) Good quality

Abbreviations: DOAC=direct oral anticoagulants; NOS= the Newcastle-Ottawa Scale

Table S9. Summary of data excluded from meta-analyses on the sex-specific effectiveness and safety of DOACs against warfarin by sex.

	Studies	RR (95% CI)	
		Men	Women
Outcome: Stroke/SE (reference group = warfarin)			
DOACs	Kwon S (2020)[23]	0.63 (0.47 - 0.84)	0.73 (0.58 - 0.92)
Dabigatran	Lip GYH (2017)[12]	0.9 (0.49 - 1.66)	0.72 (0.28 - 1.84)
	Bengtson LGS (2017)[13]	0.58 (0.43 - 0.79)	0.7 (0.51 - 0.96)
	Shantha GPS (2017)[14]	1.05 (0.75 - 1.47)	0.81 (0.62 - 1.04)
	Huybrechts KF (2019)[22]	0.71 (0.51 - 0.99)	0.82 (0.53 - 1.25)
Rivaroxaban	Lip GYH (2017)[12]	1.01 (0.43 - 2.37)	2.17 (0.89 - 5.31)
	Shantha GPS (2017)[14]	0.69 (0.48 - 0.99)	0.98 (0.76 - 1.25)
	Weir MR (2020)[24]	1.11 (0.35 - 3.53)	0.84 (0.34 - 2.07)
	Norby FL (2017)[11]	0.90 (0.70 – 1.17)	0.61 (0.46 – 0.81)
	Costa OS (2020)[27]	0.84 (0.71 – 1.00)	0.88 (0.74 - 1.05)
Apixaban	Lip GYH (2017)[12]	0.96 (0.39 - 2.36)	1.2 (0.4 - 3.59)
Edoxaban	Lee S (2018)[16]	0.72 (0.43 - 1.14)	0.7 (0.43 - 1.1)
Outcome: Major bleeding (reference group = warfarin)			
DOACs	Kwon S (2020)[23]	0.77 (0.57 - 1.03)	0.7 (0.55 - 0.89)
Dabigatran	Shantha GPS (2017)[14]	0.73 (0.59 - 0.9)	0.97 (0.82 - 1.14)
	Huybrechts KF (2019)[22]	0.69 (0.6 - 0.8)	0.76 (0.64 - 0.9)
Rivaroxaban	Shantha GPS (2017)[14]	0.91 (0.75 - 1.11)	1.2 (1.03 - 1.42)
	Baker WL (2019)[19]	0.94 (0.73 - 1.21)	1.05 (0.78 - 1.41)
	Weir MR (2020)[24]	0.82 (0.49 - 1.38)	1 (0.64 - 1.55)
	Costa OS (2020)[27]	0.78 (0.7 - 0.88)	0.84 (0.73 - 0.95)
Edoxaban	Lee S (2018)[16]	0.51 (0.29 - 0.84)	0.58 (0.31 - 1)
Outcome: GIB (reference group = warfarin)			
DOACs	Kwon S (2020)[23]	0.73 (0.53 - 1.02)	0.69 (0.53 - 0.91)
Edoxaban	Lee S (2018)[16]	0.66 (0.34 - 1.17)	0.55 (0.24 - 1.09)
Outcome: ICH (reference group = warfarin)			
DOACs	Kwon S (2020)[23]	0.95 (0.48 - 1.87)	0.68 (0.42 - 1.1)
Edoxaban	Lee S (2018)[16]	0.27 (0.06 - 0.72)	0.62 (0.21 - 1.44)

Abbreviations: RR=risk ratio, CI=confidence interval, DOACs=direct oral anticoagulant, SE = systemic embolism, GIB = gastrointestinal bleeding, ICH = intracranial haemorrhage.

Table S10. Summary of findings on effectiveness and safety of DOACs against warfarin across geographical settings and stratified by sex.

	Studies by Geographical regions	Study RR (95% CI)	
		Men	Women
Stroke/SE (reference group = warfarin)			
DOACs	**Asia**		
	Arihiro S (2016)[10]	NA	0.46 (0.1 - 1.45)
	Law SWY (2018)[17]	0.85 (0.65 - 1.12)	0.81 (0.63 - 1.03)
	Lee S (2019)[21]	0.59 (0.5 - 0.7)	0.47 (0.39 - 0.57)
	Kwon S (2021)[32]	0.62 (0.53 - 0.72)	0.6 (0.51 - 0.71)
	Yoshimoto T (2022)[38]	0.91 (0.68 - 1.23)	0.87 (0.58 - 1.28)
	Pooled RR	**0.71 (0.58 - 0.86)**	**0.64 (0.50 - 0.82)**
	Europe		
	Gulluoglu FR (2020)[30]	1.02 (0.66 - 1.6)	1.29 (0.77 - 2.15)
	Komen JJ (2022)[36]	1.15 (0.8 - 1.66)	0.68 (0.44 - 1.06)
	Pooled RR	**1.10 (0.83 - 1.45)**	**0.92 (0.49 - 1.73)**
	North America		
	Subramanya V (2021)[35]	0.89 (0.72 - 1.1)	1.01 (0.86 - 1.22)
	Asia vs Europe **Asia vs North America** **Europe vs North America**	**p-for-interaction = 0.01** **p-for-interaction = 0.11** **p-for-interaction = 0.25**	**p-for-interaction = 0.29** **p-for-interaction = 0.003** **p-for-interaction = 0.79**
Dabigatran	**Asia**		
	Bang OY (2020)[26]	0.63 (0.53 - 0.75)	0.57 (0.47 - 0.69)
	Europe		
	Balsam P (2021)[31]	1.36 (0.71 - 2.6)	0.46 (0.22 - 0.96)
	Halvorsen S (2022)[34]	0.89 (0.76 - 1.04)	0.9 (0.76 - 1.07)
	Pooled RR	**0.97 (0.69 - 1.36)**	**0.71 (0.38 - 1.34)**
	North America		
	Tsadok MA (2015)[7]	0.98 (0.78 - 1.23)	0.79 (0.56 - 1.04)
	Lauffenburger JC (2015)[8]	0.83 (0.74 - 0.93)	0.86 (0.76 - 0.98)
	Lip GYH (2018)[18]	0.85 (0.69 - 1.04)	0.79 (0.65 - 0.97)
	Pooled RR	**0.86 (0.78 - 0.94)**	**0.83 (0.75 - 0.92)**
	Asia vs Europe **Asia vs North America** **Europe vs North America**	**p-for-interaction = 0.02** **p-for-interaction = 0.002** **p-for-interaction = 0.48**	**p-for-interaction = 0.52** **p-for-interaction < 0.001** **p-for-interaction = 0.62**
Rivaroxaban	**Asia**		
	Bang OY (2020)[26]	0.65 (0.55 - 0.77)	0.59 (0.5 - 0.7)
	Europe		
	Balsam P (2021)[31]	0.93 (0.61 - 1.4)	1.32 (0.87 - 2)
	Halvorsen S (2022)[34]	0.98 (0.84 - 1.13)	1.07 (0.92 - 1.24)
	Pooled RR	**0.97 (0.85 - 1.12)**	**1.10 (0.95 - 1.26)**
	North America		
	Lip GYH (2018)[18]	0.8 (0.71 - 0.89)	0.78 (0.7 - 0.86)
	Asia vs Europe **Asia vs North America** **Europe vs North America**	**p-for-interaction < 0.001** **p-for-interaction = 0.04** **p-for-interaction = 0.03**	**p-for-interaction < 0.001** **p-for-interaction = 0.006** **p-for-interaction < 0.001**

	Studies by Geographical regions	Study RR (95% CI)	
		Men	Women
Apixaban	**Asia**		
	Bang OY (2020)[26]	0.66 (0.54 - 0.79)	0.58 (0.48 - 0.7)
	Europe		
	Halvorsen S (2022)[34]	1 (0.87 - 1.15)	0.9 (0.78 - 1.04)
	North America		
	Lip GYH (2018)[18]	0.74 (0.65 - 0.85)	0.56 (0.49 - 0.63)
	Asia vs Europe	**p-for-interaction < 0.001**	**p-for-interaction < 0.001**
	Asia vs North America	**p-for-interaction = 0.34**	**p-for-interaction = 0.76**
	Europe vs North America	**p-for-interaction = 0.002**	**p-for-interaction < 0.001**
Major bleeding (reference group = warfarin)			
DOACs	**Asia**		
	Arihiro S (2016)[10]	0.14 (0.01 - 0.76)	NA
	Lee S (2019)[21]	0.64 (0.54 - 0.75)	0.67 (0.55 - 0.81)
	Kwon S (2021)[32]	0.65 (0.55 - 0.78)	0.83 (0.68 - 1.01)
	Pooled RR	**0.64 (0.57 - 0.72)**	**0.74 (0.60 - 0.92)**
	Europe		
	Gulluoglu FR (2020)[30]	0.86 (0.65 - 1.12)	0.81 (0.57 - 1.14)
	Komen JJ (2022)[36]	0.78 (0.6 - 1.02)	0.98 (0.67 - 1.42)
	Pooled RR	**0.82 (0.68 - 0.99)**	**0.88 (0.69 - 1.14)**
	North America		
	Wong JM (2020)[25]	0.87 (0.83 - 0.91)	0.89 (0.85 - 0.93)
	Subramanya V (2021)[35]	1.01 (0.89 - 1.14)	1.03 (0.9 - 1.16)
	Pooled RR	**0.93 (0.80 - 1.07)**	**0.95 (0.82 - 1.09)**
	Asia vs Europe	**p-for-interaction = 0.03**	**p-for-interaction = 0.30**
	Asia vs North America	**p-for-interaction < 0.001**	**p-for-interaction = 0.06**
	Europe vs North America	**p-for-interaction = 0.31**	**p-for-interaction = 0.65**
Dabigatran	**Asia**		
	Bang OY (2020)[26]	0.7 (0.59 - 0.83)	0.66 (0.55 - 0.8)
	Europe		
	Halvorsen S (2022)[34]	0.85 (0.76 - 0.95)	0.95 (0.83 - 1.09)
	North America		
	Lip GYH (2018)[18]	0.65 (0.57 - 0.74)	0.78 (0.69 - 0.88)
	Asia vs Europe	**p-for-interaction = 0.06**	**p-for-interaction = 0.002**
	Asia vs North America	**p-for-interaction = 0.5**	**p-for-interaction = 0.14**
	Europe vs North America	**p-for-interaction = 0.002**	**p-for-interaction = 0.03**
Rivaroxaban	**Asia**		
	Bang OY (2020)[26]	0.81 (0.7 - 0.95)	0.65 (0.55 - 0.76)
	Europe		
	Halvorsen S (2022)[34]	1.21 (1.09 - 1.33)	1.1 (0.97 - 1.24)
	North America		
	Lip GYH (2018)[18]	0.98 (0.92 - 1.04)	1.14 (1.07 - 1.2)
	Asia vs Europe	**p-for-interaction < 0.001**	**p-for-interaction < 0.001**
	Asia vs North America	**p-for-interaction = 0.02**	**p-for-interaction < 0.001**
	Europe vs North America	**p-for-interaction < 0.001**	**p-for-interaction = 0.61**

	Studies by Geographical regions	Study RR (95% CI)	
		Men	Women
Apixaban	**Asia**		
	Bang OY (2020)[26]	0.65 (0.54 - 0.77)	0.52 (0.43 - 0.63)
	Europe		
	Halvorsen S (2022)[34]	0.78 (0.71 - 0.86)	0.66 (0.59 - 0.75)
	North America		
	Lip GYH (2018)[18]	0.56 (0.52 - 0.61)	0.63 (0.58 - 0.68)
	Asia vs Europe	**p-for-interaction = 0.08**	**p-for-interaction = 0.04**
	Asia vs North America	**p-for-interaction = 0.13**	**p-for-interaction = 0.07**
	Europe vs North America	**p-for-interaction < 0.001**	**p-for-interaction = 0.53**
GIB (reference group = warfarin)			
DOACs	**Asia**		
	Law SWY (2018)[17]	1.13 (0.73 - 1.74)	0.89 (0.63 - 1.27)
	Lee S (2019)[21]	0.81 (0.66 - 1)	0.79 (0.62 - 1.02)
	Kwon S (2021)[32]	0.73 (0.59 - 0.89)	0.93 (0.75 - 1.16)
	Pooled RR	**0.81 (0.67 - 0.98)**	**0.87 (0.75 - 1.01)**
	Europe		
	Linder M (2020)[28]	0.78 (0.51 - 1.2)	1.38 (0.89 - 2.17)
	North America		
	Wong JM (2020)[25]	0.95 (0.89 - 1.01)	1.06 (0.99 - 1.14)
	Asia vs Europe	**p-for-interaction = 0.86**	**p-for-interaction = 0.06**
	Asia vs North America	**p-for-interaction = 0.13**	**p-for-interaction = 0.02**
	Europe vs North America	**p-for-interaction = 0.37**	**p-for-interaction = 0.25**
Dabigatran	**Asia**		
	Hsu C (2018)[15]	0.57 (0.29 - 1.12)	0.46 (0.18 - 1.19)
	Europe		
	Linder M (2020)[28]	0.73 (0.4 - 1.3)	2.41 (1.31 - 4.74)
	North America		
	Bengtson LGS (2017)[13]	0.9 (0.72 - 1.12)	1.25 (0.98 - 1.59)
	Shantha GPS (2017)[14]	0.82 (0.65 - 1.04)	1.2 (0.98 - 1.44)
	Pooled RR	**0.86 (0.73 - 1.01)**	**1.22 (1.05 - 1.42)**
	Asia vs Europe	**p-for-interaction = 0.58**	**p-for-interaction = 0.005**
	Asia vs North America	**p-for-interaction = 0.24**	**p-for-interaction = 0.05**
	Europe vs North America	**p-for-interaction = 0.60**	**p-for-interaction = 0.04**
Rivaroxaban	**Asia**		
	Hsu C (2018)[15]	0.83 (0.35 - 1.97)	1.51 (0.68 - 3.36)
	North America		
	Norby FL (2017)[11]	0.95 (0.81 – 1.11)	1.24 (1.04 – 1.48)
	Shantha GPS (2017)[14]	1.02 (0.82 - 1.27)	1.43 (1.2 - 1.71)
	Pooled RR	**0.97 (0.86 – 1.11)**	**1.33 (1.16 – 1.53)**
	Asia vs North America	**p-for-interaction = 0.73**	**p-for-interaction = 0.76**

	Studies by Geographical regions	Study RR (95% CI)	
		Men	Women
ICH (reference group = warfarin)			
DOACs	**Asia**		
	Law SWY (2018)[17]	0.55 (0.27 - 1.1)	0.16 (0.06 - 0.4)
	Lee S (2019)[21]	0.45 (0.34 - 0.58)	0.54 (0.4 - 0.73)
	Kwon S (2021)[32]	0.44 (0.29 - 0.65)	0.52 (0.33 - 0.82)
	Pooled RR	**0.45 (0.37 - 0.56)**	**0.43 (0.26 - 0.70)**
	Europe		
	Linder M (2020)[28]	0.67 (0.38 - 1.16)	0.88 (0.51 - 1.54)
	North America		
	Wong JM (2020)[25]	0.65 (0.56 - 0.75)	0.64 (0.55 - 0.74)
	Asia vs Europe	**p-for-interaction = 0.20**	**p-for-interaction = 0.06**
	Asia vs North America	**p-for-interaction = 0.006**	**p-for-interaction = 0.13**
	Europe vs North America	**p-for-interaction = 0.92**	**p-for-interaction = 0.28**
Dabigatran	**Europe**		
	Linder M (2020)[28]	0.69 (0.33 - 1.44)	0.51 (0.19 - 1.3)
	North America		
	Bengtson LGS (2017)[13]	0.38 (0.18 - 0.8)	0.33 (0.12 - 0.89)
	Shantha GPS (2017)[14]	0.29 (0.12 - 0.75)	0.4 (0.22 - 0.71)
	Pooled RR	**0.34 (0.19 – 0.61)**	**0.38 (0.23 – 0.63)**
	Europe vs North America	**p-for-interaction = 0.14**	**p-for-interaction = 0.60**

Abbreviations: RR=risk ratio, CI=confidence interval, DOACs=direct oral anticoagulant, SE = systemic embolism, GIB = gastrointestinal bleeding, ICH = intracranial haemorrhage, NA=not applicable

Table S11. Summary of sex-specific effectiveness and safety for head-to-head DOAC comparisons by sex.

Target vs. Comparator	Studies	RR (95% CI)		P-for-interaction (men vs women)
		Men	Women	
Outcome: Stroke/SE				
Rivaroxaban vs dabigatran* (Meta-analysis)	Shantha GPS (2017)[14]	0.66 (0.45 - 0.96)	1.2 (0.93 - 1.61)	P = 0.46
	Norby FL (2017)[11]	1.02 (0.68 - 1.53)	0.57 (0.37 - 0.87)	
	Balsam P (2021)[31]	0.69 (0.35 - 1.37)	1.76 (0.82 - 3.79)	
	Pooled RR	**0.79 (0.59 - 1.06)**	**1.02 (0.56 - 1.84)**	
Rivaroxaban vs dabigatran*	Lip GYH (2018)[18]	0.85 (0.68 - 1.06)	0.96 (0.78 - 1.19)	P = 0.44
Apixaban vs dabigatran	Lip GYH (2018)[18]	0.81 (0.64 - 1.02)	0.62 (0.48 - 0.8)	P = 0.13
Apixaban vs rivaroxaban	Lip GYH (2018)[18]	0.92 (0.8 - 1.06)	0.72 (0.63 - 0.82)	P = 0.01
Outcome: Major bleeding				
Rivaroxaban vs dabigatran*	Shantha GPS (2017)[14]	1.24 (0.99 - 1.55)	1.27 (1.09 - 1.48)	P = 0.86
Rivaroxaban vs dabigatran*	Lip GYH (2018)[18]	1.41 (1.25 - 1.61)	1.39 (1.23 - 1.56)	P = 0.87
Apixaban vs dabigatran	Lip GYH (2018)[18]	0.81 (0.69 – 0.94)	0.75 (0.65 – 0.87)	P = 0.48
Apixaban vs rivaroxaban	Lip GYH (2018)[18]	0.57 (0.53 – 0.62)	0.54 (0.5 – 0.58)	P = 0.33
Outcome: GIB				
Rivaroxaban vs dabigatran (Meta-analysis)	Shantha GPS (2017)[14]	1.24 (0.98 - 1.59)	1.19 (1.01 - 1.44)	P = 0.69
	Norby FL (2017)[11]	1.30 (1.01 - 1.69)	1.26 (0.96 - 1.65)	
	Pooled RR	**1.27 (1.06 - 1.51)**	**1.21 (1.04 - 1.40)**	
Outcome: ICH				
Rivaroxaban vs dabigatran (Meta-analysis)	Shantha GPS (2017)[14]	2 (0.76 - 5.31)	1.65 (0.88 - 3.14)	P = 0.86
	Norby FL (2017)[11]	1.48 (0.57 - 3.82)	1.50 (0.67 - 3.36)	
	Pooled RR	**1.71 (0.87 - 3.38)**	**1.59 (0.97 - 2.62)**	

Abbreviations: RR=risk ratio, CI=confidence interval, DOAC=direct oral anticoagulant, SE = systemic embolism, GIB = gastrointestinal bleeding, ICH = intracranial haemorrhage.

*Overlapping study populations prevented pooling of all rivaroxaban vs dabigatran estimates for major bleeding and stroke/SE.

Table S12. Narrative summary of studies excluded from meta-analyses due to no exact sex-specific numeric estimates or incompatible data being provided.

Author	Study type	DOAC	Population	Reason for exclusion from meta-analysis	Narrative summary
Chang S (2019)[20]	Observational study	DOACs	AF with CKD stage 4 or 5	No sex-specific numeric estimates were reported	Forest plots compared DOACs and warfarin against no oral anticoagulation, finding DOACs and warfarin to raise major bleeding risk to a similar extent in men and women indicating no sex-specific differences.
Seeger JD (2015)[9]	Observational study	Dabigatran	NVAF with CHA2DS2-VASc score ≥ 1	No sex-specific numeric estimates were reported	Forest plots compared dabigatran and warfarin illustrating dabigatran having a lower risk of stroke/SE and major bleeding in both sexes, although confidence intervals for stroke/SE were imprecise but generally consistent with our meta-analyses.
Graham DJ (2015)[6]	Observational study	Dabigatran	NVAF	Sex-specific estimates were stratified by age incompatible with our other data	In this study, women aged 75+ had raised GIB risk with dabigatran, whereas for men, only the oldest age group (85+) experienced raised GIB with dabigatran, indicating that women may experience higher relative GIB risk with dabigatran as found in our meta-analyses. For stroke and ICH, dabigatran had either smaller or comparable risk to warfarin in both sexes, with no noticeable differences between sex.
Lee S (2018)[16]	Observational study	Edoxaban	NVAF	The sole observational study to report sex-specific data for edoxaban vs warfarin.	Numeric estimates showed no difference in stroke/SE risk between edoxaban vs warfarin in both sexes, although, point estimates were leaning towards lower stroke/SE risk. Edoxaban vs warfarin was associated with lower major bleeding risk in both sexes. For GIB and ICH, point estimates were suggestive of lower risk with edoxaban in both sexes, but estimates were imprecise.
Mieghem NMV (2021)[5]	RCT	Edoxaban	Patients with AF and a successful TAVR	No sex-specific numeric estimates were reported.	The RCT assessed edoxaban against VKAs in patients with AF after a successful TAVR. Both sexes had a higher incidence of major bleeding with edoxaban compared to VKAs. Although interaction tests were not conducted, the largely overlapping confidence intervals suggest substantial sex-difference is unlikely. Sex-specific data on stroke/SE, ICH or GIB were not presented by the study.

Abbreviations: AF=atrial fibrillation; CKD=chronic kidney disease; DOACs=direct oral anticoagulants; GIB=gastrointestinal bleeding; ICH=intracranial haemorrhage; NVAF=non-valvular atrial fibrillation; RCT=randomised controlled trial; SE=systemic embolism; TAVR=transcatheter aortic valve replacement; VKA=vitamin K antagonist;

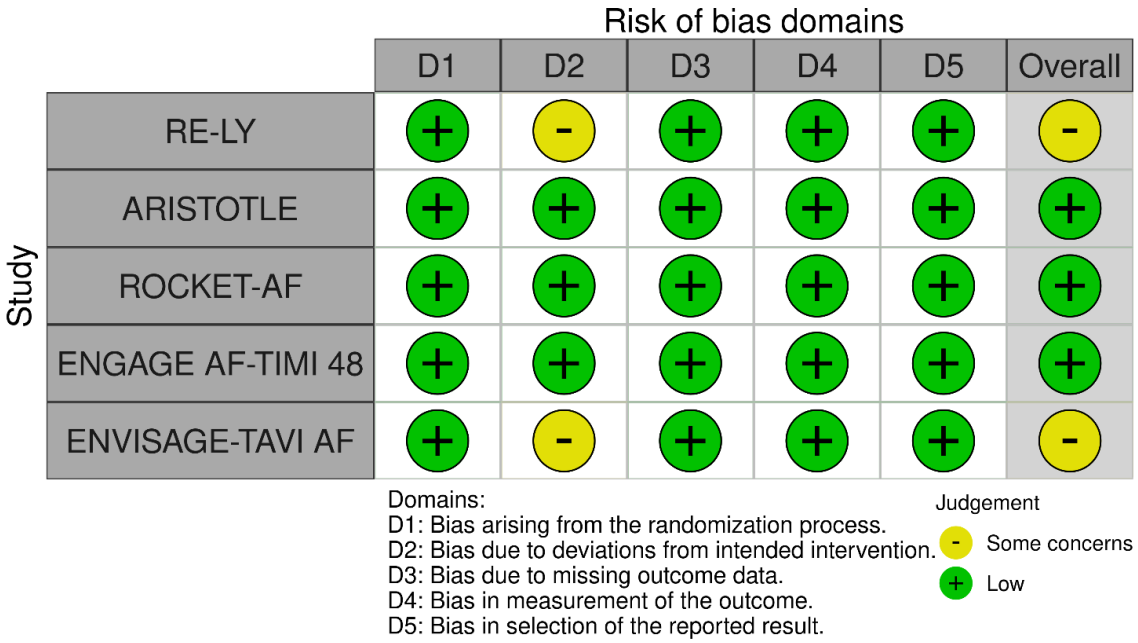


Figure S1. Visualisation for risk of bias assessment for randomised controlled trials

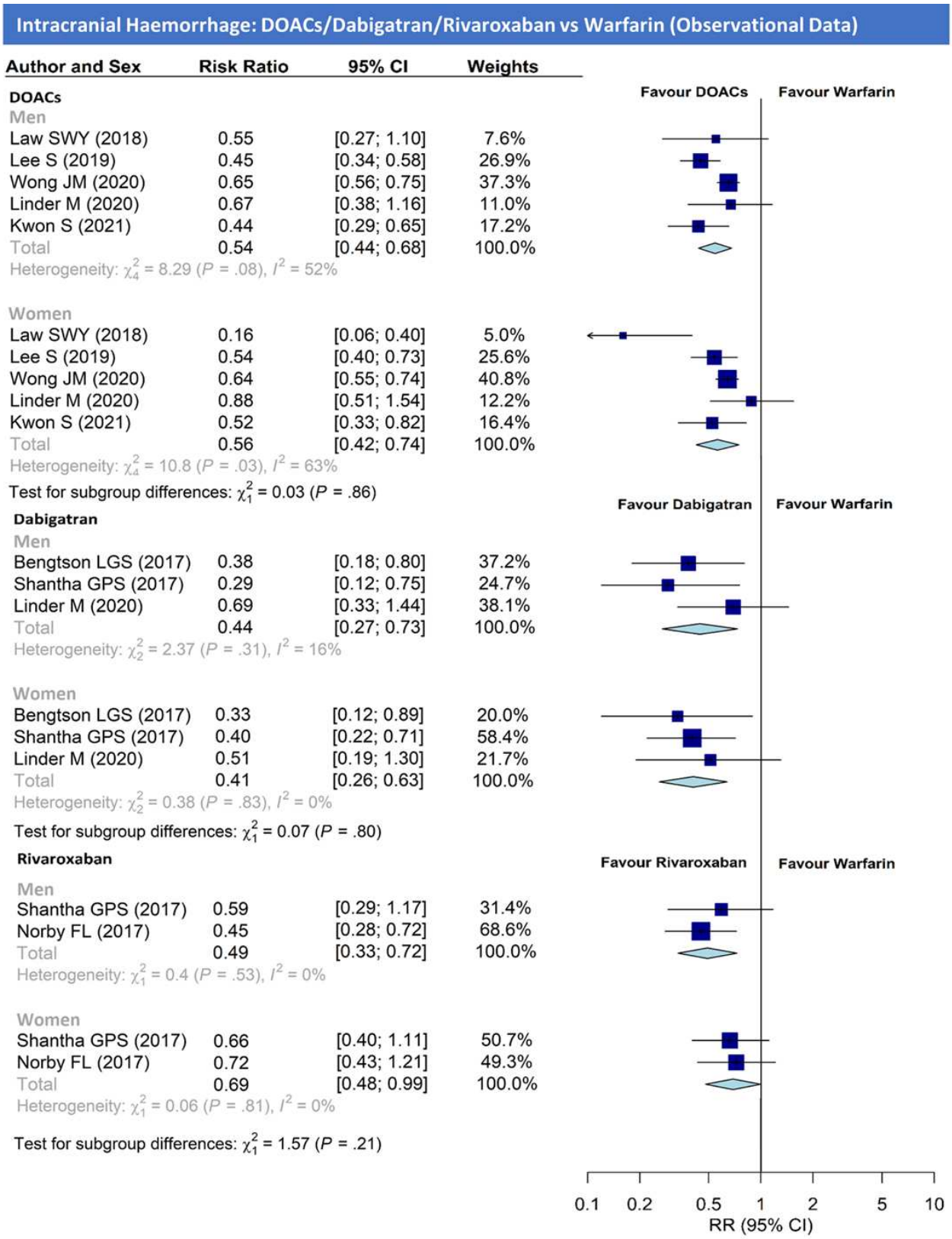


Figure S2. Forest plot of meta-analysis for observational studies comparing intracranial haemorrhage safety of DOACs as a group, dabigatran, and rivaroxaban vs warfarin by sex. CI, confidence interval; DOACs=direct oral anticoagulants

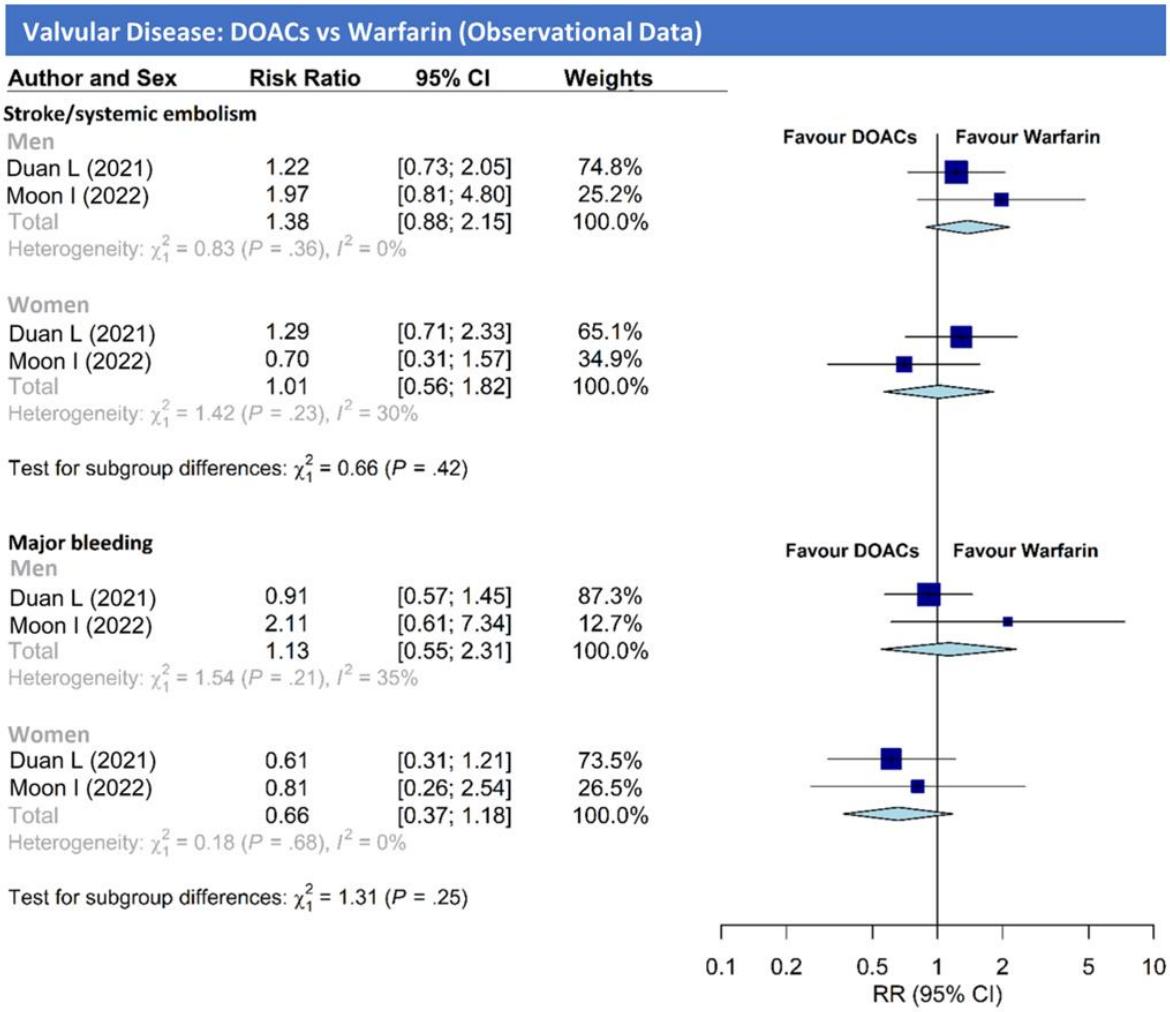
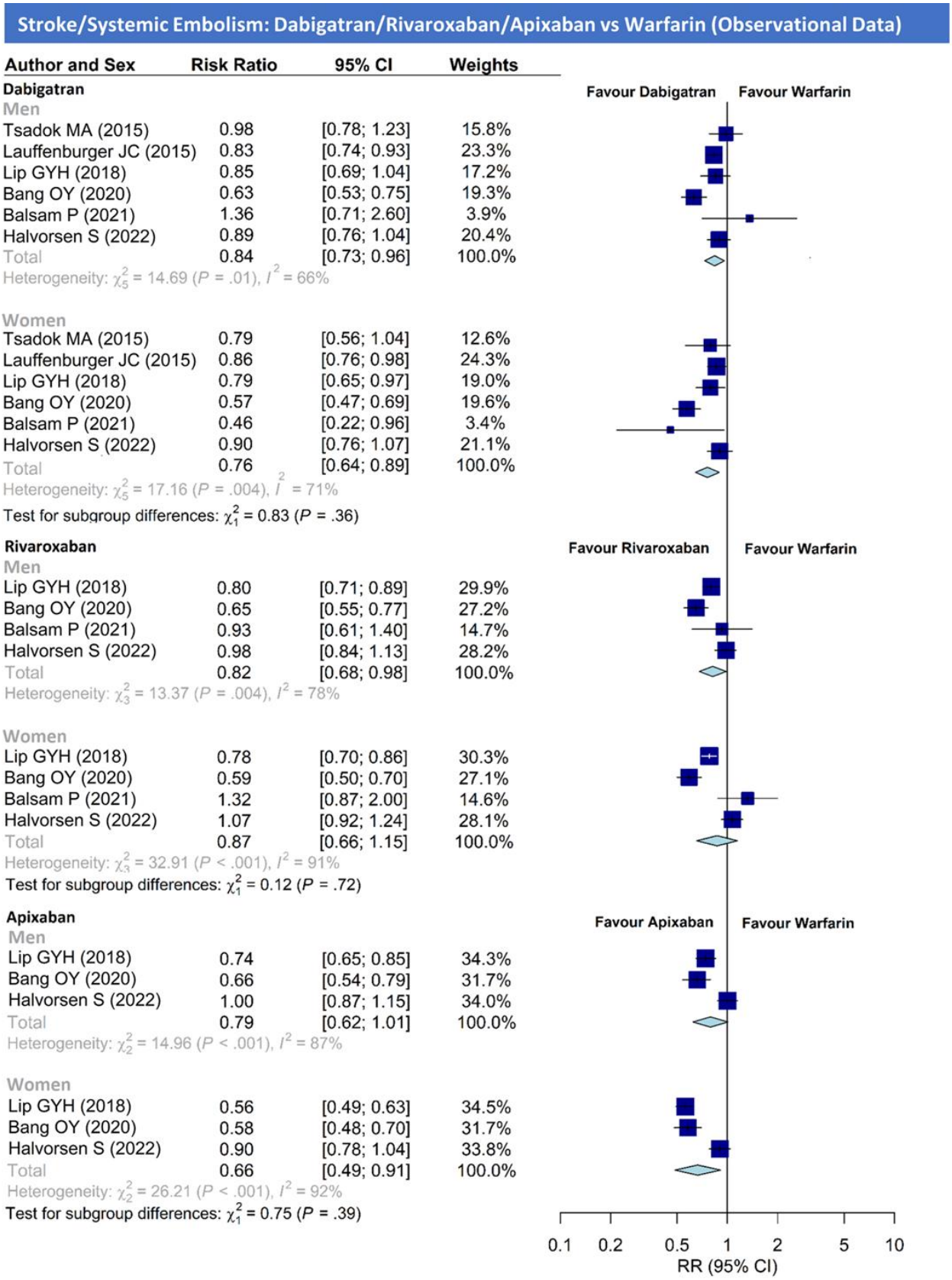


Figure S3. Forest plot of meta-analysis for observational studies comparing DOACs as a group vs warfarin on stroke/systemic embolism and major bleeding in patients with valvular heart disease by sex. CI = confidence interval; DOACs=direct oral anticoagulants



Author and Sex	Risk Ratio	95% CI	Weights
Dabigatran			
Men			
Tsadok MA (2015)	0.98	[0.78; 1.23]	15.8%
Lauffenburger JC (2015)	0.83	[0.74; 0.93]	23.3%
Lip GYH (2018)	0.85	[0.69; 1.04]	17.2%
Bang OY (2020)	0.63	[0.53; 0.75]	19.3%
Balsam P (2021)	1.36	[0.71; 2.60]	3.9%
Halvorsen S (2022)	0.89	[0.76; 1.04]	20.4%
Total	0.84	[0.73; 0.96]	100.0%
Heterogeneity: $\chi^2_5 = 14.69$ ($P = .01$), $I^2 = 66\%$			
Women			
Tsadok MA (2015)	0.79	[0.56; 1.04]	12.6%
Lauffenburger JC (2015)	0.86	[0.76; 0.98]	24.3%
Lip GYH (2018)	0.79	[0.65; 0.97]	19.0%
Bang OY (2020)	0.57	[0.47; 0.69]	19.6%
Balsam P (2021)	0.46	[0.22; 0.96]	3.4%
Halvorsen S (2022)	0.90	[0.76; 1.07]	21.1%
Total	0.76	[0.64; 0.89]	100.0%
Heterogeneity: $\chi^2_5 = 17.16$ ($P = .004$), $I^2 = 71\%$			
Test for subgroup differences: $\chi^2_1 = 0.83$ ($P = .36$)			
Rivaroxaban			
Men			
Lip GYH (2018)	0.80	[0.71; 0.89]	29.9%
Bang OY (2020)	0.65	[0.55; 0.77]	27.2%
Balsam P (2021)	0.93	[0.61; 1.40]	14.7%
Halvorsen S (2022)	0.98	[0.84; 1.13]	28.2%
Total	0.82	[0.68; 0.98]	100.0%
Heterogeneity: $\chi^2_3 = 13.37$ ($P = .004$), $I^2 = 78\%$			
Women			
Lip GYH (2018)	0.78	[0.70; 0.86]	30.3%
Bang OY (2020)	0.59	[0.50; 0.70]	27.1%
Balsam P (2021)	1.32	[0.87; 2.00]	14.6%
Halvorsen S (2022)	1.07	[0.92; 1.24]	28.1%
Total	0.87	[0.66; 1.15]	100.0%
Heterogeneity: $\chi^2_3 = 32.91$ ($P < .001$), $I^2 = 91\%$			
Test for subgroup differences: $\chi^2_1 = 0.12$ ($P = .72$)			
Apixaban			
Men			
Lip GYH (2018)	0.74	[0.65; 0.85]	34.3%
Bang OY (2020)	0.66	[0.54; 0.79]	31.7%
Halvorsen S (2022)	1.00	[0.87; 1.15]	34.0%
Total	0.79	[0.62; 1.01]	100.0%
Heterogeneity: $\chi^2_2 = 14.96$ ($P < .001$), $I^2 = 87\%$			
Women			
Lip GYH (2018)	0.56	[0.49; 0.63]	34.5%
Bang OY (2020)	0.58	[0.48; 0.70]	31.7%
Halvorsen S (2022)	0.90	[0.78; 1.04]	33.8%
Total	0.66	[0.49; 0.91]	100.0%
Heterogeneity: $\chi^2_2 = 26.21$ ($P < .001$), $I^2 = 92\%$			
Test for subgroup differences: $\chi^2_1 = 0.75$ ($P = .39$)			

Figure S4. Forest plot of meta-analysis for observational studies comparing the effectiveness of dabigatran, rivaroxaban and apixaban vs warfarin on stroke/systemic embolism by sex. CI = confidence interval; DOACs=direct oral anticoagulants

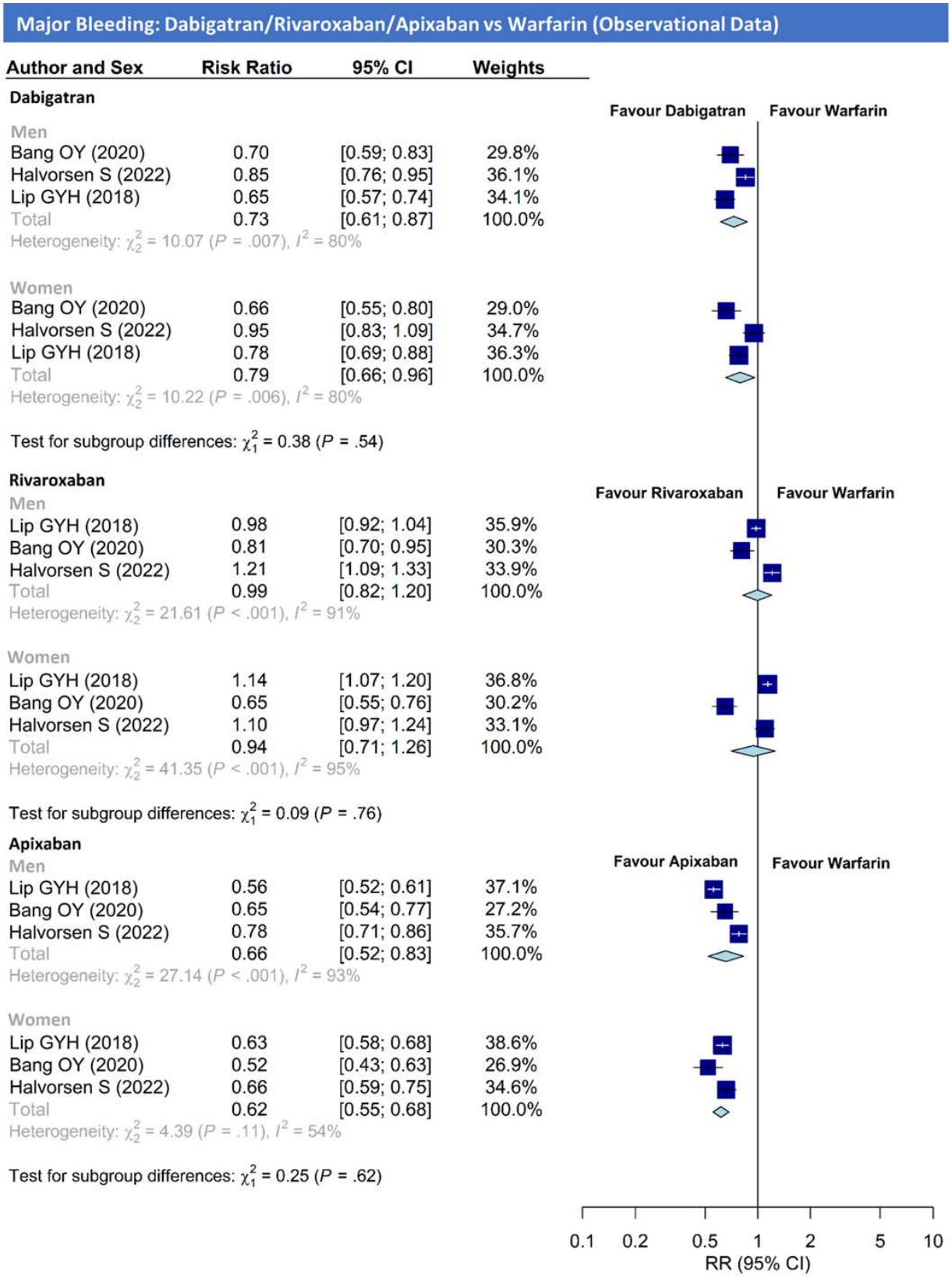


Figure S5. Forest plot of meta-analysis for observational studies comparing major bleeding safety of dabigatran, rivaroxaban and apixaban vs warfarin by sex. CI, confidence interval; DOACs=direct oral anticoagulants

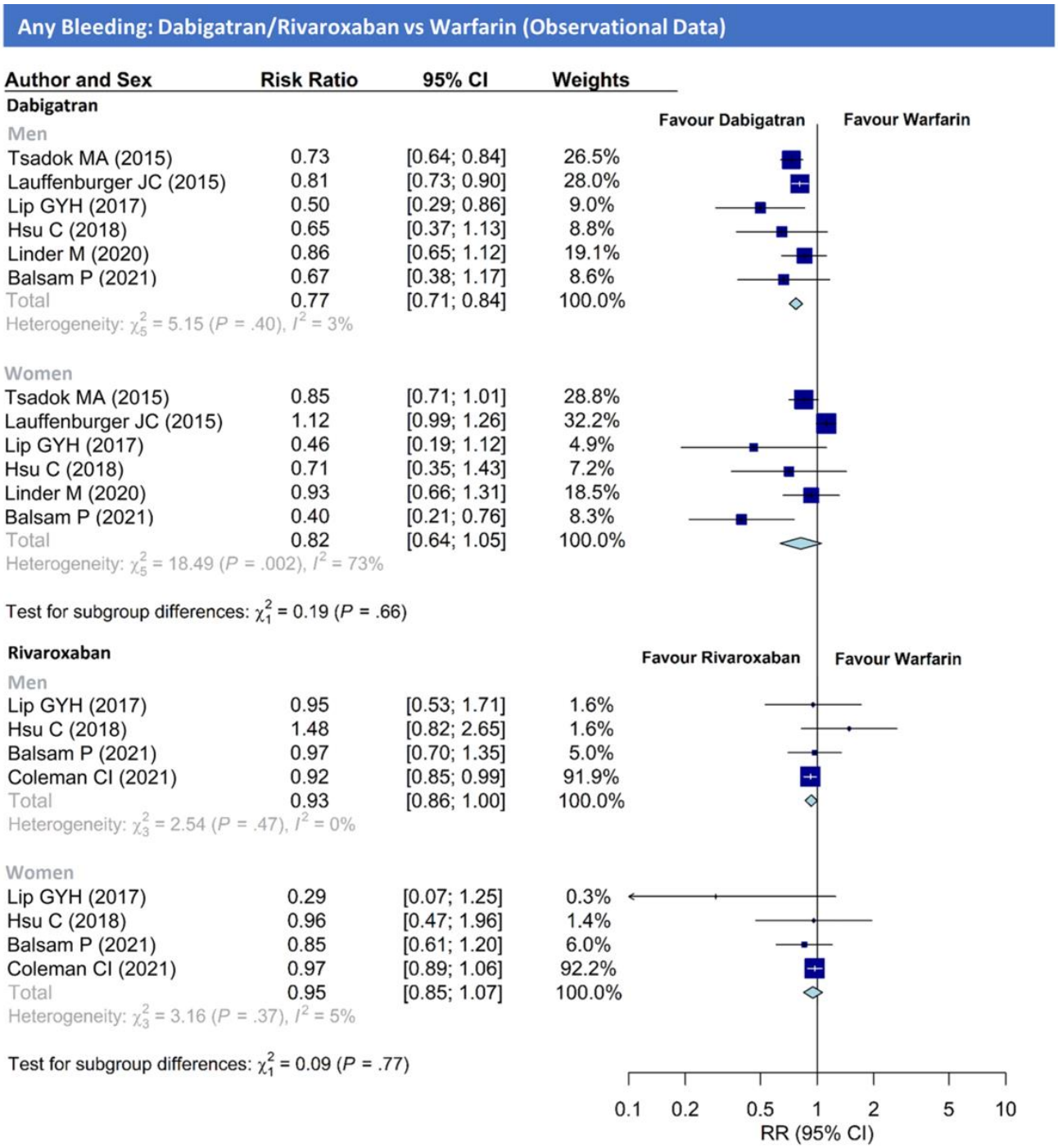


Figure S6. Forest plot of meta-analysis for observational studies comparing any bleeding safety of dabigatran and rivaroxaban vs warfarin by sex. CI, confidence interval; DOACs=direct oral anticoagulants

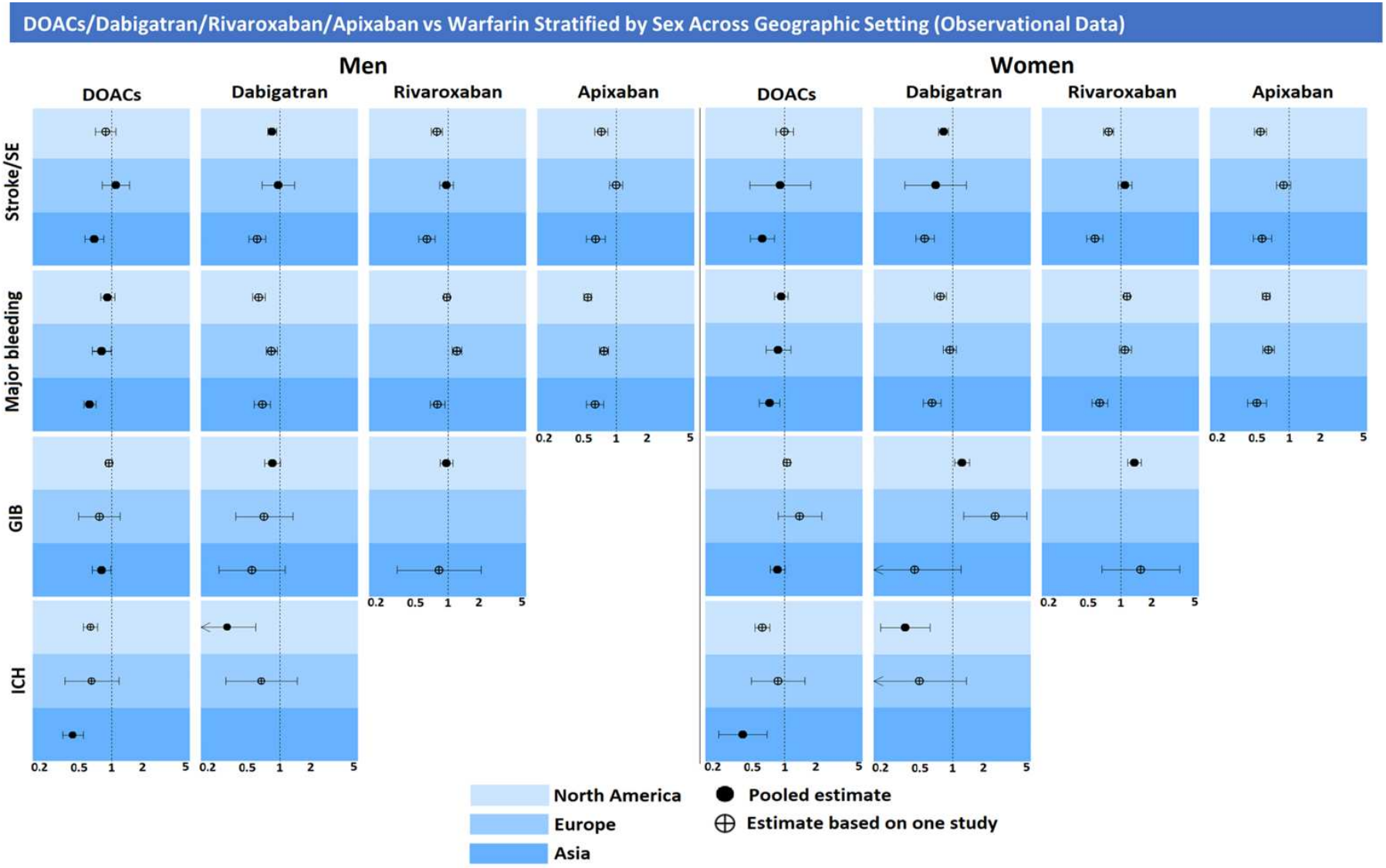


Figure S7. Graphical plots of the sex-specific effectiveness and safety of DOACs as a group, dabigatran, rivaroxaban and apixaban compared to warfarin across geographical settings. DOACs = direct oral anticoagulants; SE = systemic embolism; GIB = gastrointestinal bleeding; ICH = intracranial haemorrhage

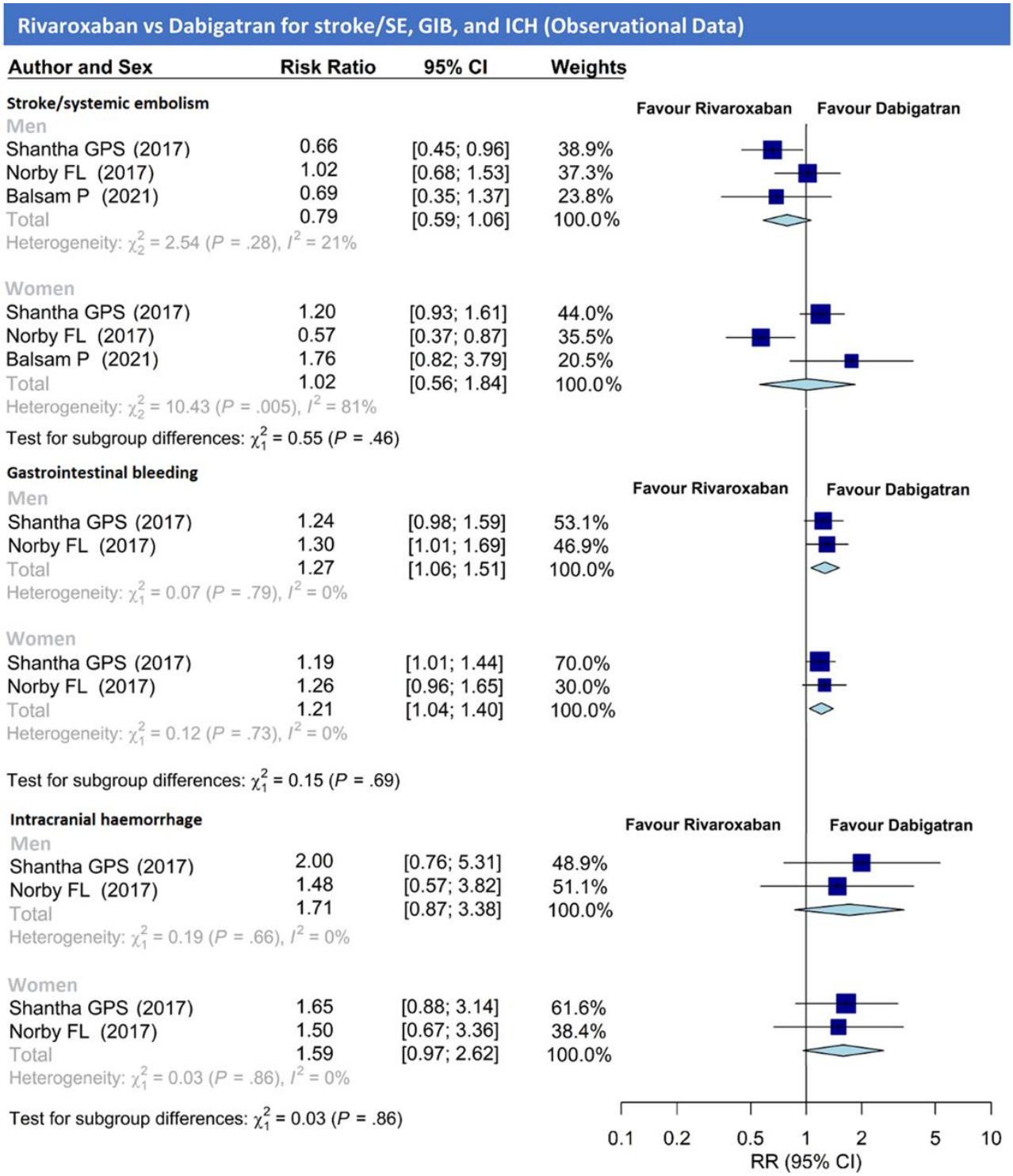


Figure S8. Forest plot of meta-analysis for observational studies comparing rivaroxaban against dabigatran on stroke/SE, GIB, and ICH by sex. CI, confidence interval; DOACs=direct oral anticoagulants; SE = systemic embolism; GIB = gastrointestinal bleeding; ICH = intracranial haemorrhage

References

1. Connolly SJ, Ezekowitz MD, Yusuf S *et al.* Dabigatran versus Warfarin in Patients with Atrial Fibrillation. *New England Journal of Medicine* 2009;**361**:1139–51.

2. Granger CB, Alexander JH, McMurray JJV *et al.* Apixaban versus Warfarin in Patients with Atrial Fibrillation. *New England Journal of Medicine* 2011;**365**:981–92.

3. Patel MR, Mahaffey KW, Garg J *et al.* Rivaroxaban versus warfarin in nonvalvular atrial fibrillation. *The New England journal of medicine* 2011;**365**:883–91.

4. Zelniker TA, Ardissino M, Andreotti F *et al.* Comparison of the Efficacy and Safety Outcomes of Edoxaban in 8040 Women Versus 13 065 Men With Atrial Fibrillation in the ENGAGE AF-TIMI 48 Trial. *Circulation* 2021;**143**:673–84.

5. Van Mieghem NM, Unverdorben M, Hengstenberg C *et al.* Edoxaban versus Vitamin K Antagonist for Atrial Fibrillation after TAVR. *New England Journal of Medicine* 2021;**385**:2150–60.

6. Graham DJ, Reichman ME, Wernecke M *et al.* Cardiovascular, bleeding, and mortality risks in elderly Medicare patients treated with dabigatran or warfarin for nonvalvular atrial fibrillation. *Circulation* 2015;**131**:157–64.

7. Tsadok MA, Jackevicius CA, Rahme E *et al.* Sex Differences in Dabigatran Use, Safety, And Effectiveness In a Population-Based Cohort of Patients With Atrial Fibrillation. *Circ-Cardiovasc Qual Outcomes* 2015;**8**:593–9.

8. Lauffenburger JC, Farley JF, Gehi AK *et al.* Effectiveness and safety of dabigatran and warfarin in real-world US patients with non-valvular atrial fibrillation: a retrospective cohort study. *Journal of the American Heart Association* 2015;**4**, DOI: 10.1161/jaha.115.001798.

9. Seeger JD, Bykov K, Bartels DB *et al.* Safety and effectiveness of dabigatran and warfarin in routine care of patients with atrial fibrillation. *Thrombosis and haemostasis* 2015;**114**:1277–89.

10. Arihiro S, Todo K, Koga M *et al.* Three-month risk-benefit profile of anticoagulation after stroke with atrial fibrillation: The SAMURAI-Nonvalvular Atrial Fibrillation (NVAF) study. *International journal of stroke : official journal of the International Stroke Society* 2016;**11**:565–74.

11. Norby FL, Bengtson LGS, Lutsey PL *et al.* Comparative effectiveness of rivaroxaban versus warfarin or dabigatran for the treatment of patients with non-valvular atrial fibrillation. *BMC Cardiovascular Disorders* 2017;**17**:238.

12. Lip GYH, Skjøth F, Nielsen PB *et al.* Effectiveness and Safety of Standard-Dose Nonvitamin K Antagonist Oral Anticoagulants and Warfarin Among Patients With Atrial Fibrillation With a Single Stroke Risk Factor: A Nationwide Cohort Study. *JAMA Cardiol* 2017;**2**:872–81.

13. Bengtson L, Lutsey P, Lin C *et al.* Comparative effectiveness of dabigatran and rivaroxaban versus warfarin for the treatment of non-valvular atrial fibrillation. *Journal of cardiology* 2017;**69**:868–76.

14. Palamaner Subash Shantha G, Bhave PD, Girotra S *et al.* Sex-Specific Comparative Effectiveness of Oral Anticoagulants in Elderly Patients With Newly Diagnosed Atrial Fibrillation. *Circulation Cardiovascular quality and outcomes* 2017;**10**, DOI: 10.1161/circoutcomes.116.003418.

15. Hsu CC, Hsu PF, Sung SH *et al.* Is There a Preferred Stroke Prevention Strategy for Diabetic Patients with Non-Valvular Atrial Fibrillation? Comparing Warfarin, Dabigatran and Rivaroxaban. *Thrombosis and haemostasis* 2018;**118**:72–81.

16. Lee S-R, Choi E-K, Han K-D *et al.* Edoxaban in Asian Patients With Atrial Fibrillation: Effectiveness and Safety. *Journal of the American College of Cardiology* 2018;**72**:838–53.

17. Law S, Lau WCY, Wong ICK *et al.* Sex-Based Differences in Outcomes of Oral Anticoagulation in Patients With Atrial Fibrillation. *Journal of the American College of Cardiology* 2018;**72**:271–82.

18. Lip GYH, Keshishian A, Li X *et al.* Effectiveness and Safety of Oral Anticoagulants Among Nonvalvular Atrial Fibrillation Patients: The ARISTOPHANES Study. *Stroke* 2018;**49**:2933–44.

19. Baker WL, Beyer-Westendorf J, Bunz TJ *et al.* Effectiveness and safety of rivaroxaban and warfarin for prevention of major adverse cardiovascular or limb events in patients with non-valvular atrial fibrillation and type 2 diabetes. *Diabetes, obesity & metabolism* 2019;**21**:2107–14.

20. Chang SH, Wu CV, Yeh YH *et al.* Efficacy and Safety of Oral Anticoagulants in Patients With Atrial Fibrillation and Stages 4 or 5 Chronic Kidney Disease. *The American journal of medicine* 2019;**132**:1335-1343.e6.

21. Lee SR, Lee HJ, Choi EK *et al.* Direct Oral Anticoagulants in Patients With Atrial Fibrillation and Liver Disease. *Journal of the American College of Cardiology* 2019;**73**:3295–308.

22. Huybrechts KF, Gopalakrishnan C, Bartels DB *et al.* Safety and Effectiveness of Dabigatran and Other Direct Oral Anticoagulants Compared With Warfarin in Patients With Atrial Fibrillation. *Clinical pharmacology and therapeutics* 2020;**107**:1405–19.

23. Kwon S, Lee SR, Choi EK *et al.* Non-vitamin K antagonist oral anticoagulants in very elderly east Asians with atrial fibrillation: A nationwide population-based study. *American heart journal* 2020;**229**:81–91.

24. Weir MR, Ashton V, Moore KT *et al.* Rivaroxaban versus warfarin in patients with nonvalvular atrial fibrillation and stage IV-V chronic kidney disease. *American heart journal* 2020;**223**:3–11.

25. Wong JM, Maddox TM, Kennedy K *et al.* Comparing Major Bleeding Risk in Outpatients With Atrial Fibrillation or Flutter by Oral Anticoagulant Type (from the National Cardiovascular Disease Registry's Practice Innovation and Clinical Excellence Registry). *The American journal of cardiology* 2020;**125**:1500–7.

26. Bang OY, On YK, Lee MY *et al.* The risk of stroke/systemic embolism and major bleeding in Asian patients with non-valvular atrial fibrillation treated with non-vitamin K oral anticoagulants compared to warfarin: Results from a real-world data analysis. *PloS one* 2020;**15**:e0242922.

27. Costa OS, Beyer-Westendorf J, Ashton V *et al.* Effectiveness and safety of rivaroxaban versus warfarin in obese nonvalvular atrial fibrillation patients: analysis of electronic health record data. *Current medical research and opinion* 2020;**36**:1081–8.

28. Linder M, Iliadou Nyman A, Kieler H *et al.* Assessing Safety of Direct Thrombin Inhibitors, Direct Factor Xa Inhibitors and Vitamin K Antagonists in Patients with Atrial Fibrillation: A Nation-Wide Propensity Score Matched Cohort from Sweden. *Clinical epidemiology* 2020;**12**:1029–38.

29. Duan L, Doctor J, Adams J *et al.* Comparison of Direct Oral Anticoagulants Versus Warfarin in Patients With Atrial Fibrillation and Bioprosthetic Heart Valves. *The American journal of cardiology* 2021;**146**:22–8.

30. Rustem Gulluoglu F, Souverein PC, van den Ham HA *et al.* Comparative effectiveness and safety of direct oral anticoagulants versus warfarin in UK patients with atrial fibrillation and type 2 diabetes: A retrospective cohort study. *Pharmacoepidemiology and drug safety* 2021;**30**:1293–320.

31. Balsam P, Lodziński P, Gawałko M *et al.* Antithrombotic Management and Long-Term Outcomes of Patients with Atrial Fibrillation. Insights from CRAFT Trial. *Journal of clinical medicine* 2021;**10**, DOI: 10.3390/jcm10081780.

32. Kwon S, Lee S-R, Choi E-K *et al.* Non-Vitamin K Antagonist Oral Anticoagulants in Patients With Atrial Fibrillation and Prior Gastrointestinal Bleeding. *Stroke* 2021;**52**:511–20.

33. Coleman CI, Costa OS, Brescia CW *et al.* Thromboembolism, bleeding and vascular death in nonvalvular atrial fibrillation patients with type 2 diabetes receiving rivaroxaban or warfarin. *Cardiovasc Diabetol* 2021;**20**:52.

34. Halvorsen S, Johnsen SP, Madsen M *et al.* Effectiveness and safety of non-vitamin K antagonist oral anticoagulants and warfarin in atrial fibrillation: a Scandinavian population-based cohort study. *European Heart Journal - Quality of Care and Clinical Outcomes* 2022;**8**:577–87.

35. Subramanya V, Claxton JS, Lutsey PL *et al.* Sex differences in treatment strategy and adverse outcomes among patients 75 and older with atrial fibrillation in the MarketScan database. *BMC Cardiovascular Disorders* 2021;**21**:598.

36. Komen JJ, Pottegård A, Mantel-Teeuwisse AK *et al.* Oral anticoagulants in patients with atrial fibrillation at low stroke risk: a multicentre observational study. *European Heart Journal* 2022;**43**:3528–38.

37. Moon I, Go T-H, Kim JY *et al.* Effectiveness and safety of non-vitamin K direct oral anticoagulants in atrial fibrillation patients with bioprosthetic valve. *PLOS ONE* 2022;**17**:e0268113.

38. Yoshimoto T, Toyoda K, Ihara M *et al.* Impact of Previous Stroke on Clinical Outcome in Elderly Patients With Nonvalvular Atrial Fibrillation: ANAFIE Registry. *Stroke* 2022;**53**:2549–58.

GRAPHICAL ABSTRACT: DOACs Vs Warfarin Stratified by Sex for Stroke/Systemic Embolism, Major Bleeding, Gastrointestinal Bleeding, and Intracranial Haemorrhage (Observational Data)