Direct comparative effectiveness and safety between non-vitamin K antagonist oral anticoagulants for stroke prevention in nonvalvular atrial fibrillation

a systematic review and meta-analysis of observational studies

Li, G.; Lip, G.Y.H.; Holbrook, A.; Chang, Y.; Larsen, T.B.; Sun, X.; Tang, J.; Mbuagbaw, L.; Witt, D.M.; Crowther, M.; Thabane, L.; Levine, M.A.H.

Published in:
European Journal of Epidemiology

DOI (link to publication from Publisher):
10.1007/s10654-018-0415-7

Publication date:
2019

Document Version
Accepted author manuscript, peer reviewed version

Direct comparative effectiveness and safety between non-vitamin K antagonist oral anticoagulants for stroke prevention in nonvalvular atrial fibrillation: A systematic review and meta-analysis of observational studies

Running title: Direct NOAC comparison

Guowei Li[1,2,3]*, PhD;
Gregory Y. H. Lip[4,5], MD;
Anne Holbrook[2,3,6], MD;
Yaping Chang[2], MSc;
Torben B. Larsen[5,7], MD, PhD;
Xin Sun[8], PhD;
Jie Tang[9], PhD;
Lawrence Mbuagbaw[2,3], MD, PhD;
Daniel M. Witt[10], PhD;
Mark Crowther[2,3,6], MD;
Lehana Thabane[2,3], PhD;
Mitchell A. H. Levine[2,3,6], MD

[1]Guangdong Second Provincial General Hospital, Guangzhou 510317, China
[2]Department of Health research methods, Evidence, and Impact, McMaster University, Hamilton, Canada
[3]St. Joseph's Healthcare Hamilton, Hamilton, Canada
[4]University of Birmingham Institute of Cardiovascular Sciences, City Hospital, Birmingham, UK
[5]Aalborg Thrombosis Research Unit, Department of Clinical Medicine, Aalborg University, Aalborg, Denmark
[6] Department of Medicine, McMaster University, Hamilton, ON, Canada
[7]Department of Cardiology, Cardiovascular Research Centre, Aalborg University Hospital, Aalborg, Denmark
[8]Chinese Evidence-based Medicine Centre, West China Hospital, Sichuan University, Chengdu, China
[9]School of Public Health, Guangzhou Medical University, Guangzhou, China
[10]Department of Pharmacotherapy, University of Utah College of Pharmacy, Salt Lake City, USA

***Corresponding Author:**

Guowei Li, PhD, MSc, MBBS

Guangdong Second Provincial General Hospital, 466 Newport Middle Road, Haizhu District, Guangzhou 510317, Guangdong Province, China

Telephone: 86-020-89168066;

Fax: 86-020-89168021

E-mail: lig.mileo@yahoo.com

Abstract

Background: The non-vitamin K antagonist oral anticoagulants (NOACs) have been increasingly prescribed in clinical practice for stroke prevention in patients with nonvalvular atrial fibrillation (AF). Direct comparisons between NOACs in trials are lacking, leaving an important clinical decision-making gap. We aimed to perform a systematic review and meta-analysis to summarize the evidence of observational studies for direct comparative effectiveness and safety amongst NOACs in patients with AF.

Methods: Conference proceedings and electronic databases including MEDLINE, CINAHL, EMBASE and PUBMED were systematically searched. We included observational studies directly comparing individual NOACs in patients with nonvalvular AF who were aged ≥18 years for stroke prevention. Primary outcome included effectiveness outcome (stroke or systemic embolism) and safety outcome (major bleeding). Data were extracted in duplicated by two reviewers independently. A random-effects meta-analysis was conducted to synthesize the data from included observational studies. We used the Grading of Recommendations Assessment, Development and Evaluation (GRADE) to rate the overall quality of evidence for each outcome.

Results: Fifteen studies were included for qualitative synthesis, twelve studies for meta-analyses. It was found that rivaroxaban and dabigatran were similar with regard to risk of stroke or systemic embolism (Hazard ratio [HR] = 1.00, 95% CI: 0.91 - 1.10; evidence quality: low), but rivaroxaban was associated with higher risk of major bleeding (HR = 1.39, 95% CI: 1.28 - 1.50; evidence quality: moderate). Compared with apixaban, a significantly higher risk of major bleeding was observed with rivaroxaban (HR = 1.71, 95% CI: 1.51 - 1.94; evidence quality: low). Apixaban was associated with lower risk of major bleeding, in comparison with dabigatran (HR = 0.80, 95% CI: 0.68 - 0.95; evidence quality: low). No differences in risk of stroke or systemic embolism was observed between rivaroxaban vs. apixaban, and apixaban vs. dabigatran.

Conclusions: In this study, apixaban was found to have the most favorable safety profile amongst the three NOACs. No significant difference was observed in risk of stroke or systemic embolism between the NOACs. Such findings may provide some decision-making support for physicians regarding their choices amongst NOACs in patients with AF.

Registration: PROSPERO (identifier: CRD42016052908)

Keywords: Non-vitamin K antagonist oral anticoagulant; Stroke; Major bleeding; Atrial Fibrillation; Direct comparison

Introduction:

Atrial fibrillation (AF) is a highly prevalent, age-related cardiac arrhythmia and independently increases the risk of stroke by five fold [1,2]. The use of antithrombotic prophylaxis remains the mainstay to prevent stroke in patients with nonvalvular AF [3,4]. The non-vitamin K antagonist oral anticoagulants (NOACs) have been increasingly prescribed in clinical practice, due to their advantages over warfarin such as the decreased need for monitoring, fewer food and drug interactions, and more predictable pharmacodynamic effect [5-7].

The efficacy and safety of NOACs compared with warfarin have been presented in respective multicenter Phase III randomized controlled trials (RCTs) [8-11]. However, no head-to-head comparison between NOACs is available from RCTs; therefore there is a lack of direct clinical outcome evidence to inform physicians and patients on the choice amongst NOACs. Some studies employed the RCT data to conduct indirect comparison analyses for relative effect estimates between NOACs by using the common comparator arm (warfarin) in all the trials [12-18]. Nevertheless, the utility and credibility of their results are limited given the difference in populations, outcomes, study methodology and designs, and time in therapeutic range in warfarin groups between the respective RCTs [19,20]. Besides, whether and how NOACs in real-world circumstances would show different effectiveness-safety profiles from those in the ideal RCT settings, and whether similar comparative effectiveness-safety profiles would be observed amongst NOACs, remains to be further explored. Observational studies provide a platform for direct comparative evaluation amongst NOACs in heterogeneous populations in real-world clinical practice, which could supply some evidence to physicians to aid in decision-making regarding their choices amongst NOACs.

In this study, we aimed to conduct a systematic review and meta-analysis to summarize the evidence of direct comparison from observational studies for the comparative effectiveness and safety between NOACs in patients with nonvalvular AF.

Methods

We conducted this study based on guidance from the Cochrane Handbook of Systematic Reviews and reported results according to PRISMA (the Preferred Reporting Items for Systematic Reviews and Meta-analyses) recommendations [21,22]. The study protocol was registered in PROSPERO (Prospective Register of Ongoing Systematic Reviews; identifier: CRD42016052908).

Search strategy

We searched the following electronic databases to identify eligible observational studies: MEDLINE, CINAHL and EMBASE from Jan 1st, 2009 to November 30th, 2016, because the first NOAC (dabigatran) in AF was reported for licensing in 2009 [10]. We also updated the PUBMED search from November 2016 up to August 3rd, 2017. We used descriptors including synonyms for observational studies, NOACs, stroke or bleeding, and atrial fibrillation in the search (detailed terms for search were presented in **Supplemental Table 1**). Reference lists of included studies and other review or editorial articles were also searched for relevant reports. No language restriction was used. Three conference proceedings were searched for unpublished and ongoing studies: American College of Cardiology (2009 - 2016), European Society of Cardiology (2009 - 2016), and International Society on Thrombosis and Haemostasis (2009 - 2016).

Study eligibility criteria

Case-control and cohort studies directly comparing individual NOACs in patients with nonvalvular AF who were aged ≥18 years for stroke prevention were eligible for inclusion. We focused on factor Xa inhibitors (rivaroxaban, edoxaban, and apixaban) and the direct thrombin inhibitor (dabigatran). Therefore the comparisons amongst NOACs included dabigatran vs rivaroxaban, dabigatran vs apixaban, dabigatran vs edoxaban, rivaroxaban vs apixaban, rivaroxaban vs edoxaban, and apixaban vs edoxaban.

If data from the same participants were published in multiple reports or at different time points, we chose the study with the largest sample size and longest follow-up. We excluded studies if their objectives were not comparative effectiveness or safety profiles of NOACs, or if they could not provide data on comparative effectiveness or safety amongst NOACs in patients with AF. Moreover, some studies may compare one NOAC (e.g., dabigatran) with a combination of the other NOACs (e.g., rivaroxaban and apixaban). These studies were not included if no data on direct comparison (e.g., dabigatran vs rivaroxaban, or dabigatran vs apixaban) could be isolated or extracted. Furthermore, we excluded studies comparing NOACs in patients for cardioversion or ablation of AF, because of their short-term treatment duration and follow-up.

Outcomes

In this study, the primary outcomes included the effectiveness outcome (a composite of stroke or systemic embolism) and the safety outcome (a composite of major bleeding). Given that the included studies may define primary outcomes differently, we adopted the definitions from the included individual studies and presented their definitions in **Table 1**. Our secondary outcomes were myocardial infarction (MI) and all-cause mortality.

Data extraction and individual study quality assessment

Two reviewers (G.L. and Y.C.) independently screened and chose eligible studies for inclusion. We used the Kappa statistic to quantify the agreement between the two reviewers [23]. Disagreement was addressed by discussion between the two reviewers, with a third arbiter available if no consensus could be reached. The two reviewers extracted data independently including information on study design, patient characteristics, anticoagulant information, outcome assessment, follow-up period, and comparative treatment effect estimates.

The Cochrane Collaboration ROBINS-I (Risk of Bias In Non-randomized Studies- of Interventions) evaluation tool was used to assess the individual observational study quality [24]. Each study was rated as either low, moderate, serious, or critical risk of bias, according to the domains of confounding, participant selection, intervention classification, departure from intended intervention, missing data, outcome measures, and selective reporting.

Statistical analyses

We performed a random-effects meta-analysis to synthesize the data by pooling the results of the cohort and case-control studies, respectively. We used the adjusted hazard ratios (HRs) for cohort studies and odds ratios (ORs) for case-control studies for meta-analyses. Treatment effect estimates

were reported with pooled HRs and ORs for cohort studies and case-control studies respectively, each with 95% confidence intervals (CIs).

Data on the composite outcomes may not be extracted in some studies, because they may only report individual components of the composite outcomes (e.g., they presented results for stroke and systemic embolism respectively). For these studies, we only pooled data on stroke for effectiveness outcome, and intracranial hemorrhage (ICH) for safety outcome respectively, to avoid duplicate counting of the same patients with multiple events in the meta-analyses [25]. Likewise, if multiple doses of a NOAC were studied and not combined, we included data only on the highest dose for meta-analysis.

Statistical heterogeneity for included studies was estimated using the I^2 statistic, in which a p-value of less than 0.1 or an I^2 of over 50% indicated significant heterogeneity [21]. To explain heterogeneity in primary outcomes, for each comparison amongst NOACs, we conducted the following three predefined subgroup analyses by: 1) individual component of composite outcomes (i.e, stroke and systemic embolism for effectiveness outcome, and ICH and major gastrointestinal (GI) bleeding for safety outcome, respectively), 2) lengths of follow-up (where the median follow-up was used to categorize studies as having long- or short follow-up), and 3) different $CHADS_2$ (> 2 vs. ≤ 2) or CHA_2DS_2-VASc (> 3 vs. ≤ 3) scores, and HAS-BLED scores (>2 vs. ≤2). Three sensitivity analyses were performed to evaluate the robustness of our main results by: 1) employing a fixed-effects model for the meta-analysis, 2) only including low-risk-of-bias studies for analysis, and 3) only pooling data on standard doses of NOACs (150 mg b.i.d. for dabigatran, 20 mg o.d. for rivaroxaban, 5 mg b.i.d. for apixaban, and 60 mg o.d. for edoxaban).

Assessment of publication bias and quality of a body of evidence across included studies

We used the Begg's rank correlation and Egger's regression tests for primary outcomes to evaluate potential publication bias statistically [21]. Funnel plots were also constructed for visual inspection of asymmetry. The quality of a body of evidence for this study was rated using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) tool [26]. The quality of evidence across included studies could be categorized as very low, low, moderate, or high, based on the judgement about the study design, directness of evidence, precision of results, inconsistency of results or unexplained heterogeneity, and publication bias [26].

Results

There were 1,449 records included for screening. After title and abstract screening and duplicate removal, we assessed a total of 92 full-text articles for eligibility with an inter-rater Kappa statistic of 0.80 (95% CI: 0.63 – 0.97) between the assessors. Fifteen studies (nine full texts [20,27-34] and six abstracts [35-40]) were eligible to be included for qualitative synthesis, among which there were twelve studies (seven full texts [20,28,30-34] and five abstracts [35,37-40]) included for quantitative synthesis (**Supplemental Figure 1**).

Table 1 shows characteristics of the fifteen included studies. Most studies (n = 14) were retrospective cohort designs using data from electronic health databases, while only one study was prospective cohort research [34]. Thirteen studies focused on NOAC-naive users. All the primary outcome measures were identified from ICD-9-CM or ICD-10-CM codes. All fifteen studies investigated

comparisons between rivaroxaban and dabigatran (number of patients: 337,661), nine studies [27,29,31,32,35,37-40] for rivaroxaban vs. apixaban (number of patients: 128,548), and nine studies [27,29,31,32,35,37-40] for apixaban vs. dabigatran (number of patients: 100,724). No studies provided data on edoxaban, reflecting its very recent approval. For rivaroxaban vs. dabigatran, the included studies were conducted in USA (n =10), China (n = 3), Denmark (n = 1) and Sweden (n = 1). Patients had a median age of approximately 72 years, with a median $CHADS_2$ score of 2 and a median HAS-BLED score of 2. The follow-up period varied from 110 days to 400 days. Studies with data on rivaroxaban vs. apixaban or apixaban vs. dabigatran were performed in USA (n = 8) and Sweden (n = 1). Patients' median age was 73 years, median $CHADS_2$ score 2 and median HAS-BLED score 2. Only one study provided data on follow-up period of approximately 160 days [31].

Among all the included studies, eight used multivariable survival regression, six propensity score method, and one multivariable logistic regression to quantify comparative evaluation amongst NOACs, respectively (**Table 1**). Study quality was evaluated for the nine full texts. Seven studies were rated as low-risk-of-bias for effectiveness and safety outcomes [20,28,29,31,32]. There was one study [30] rated as moderate-risk-of-bias for stoke or systemic embolism because it did not provide information on missing data and it measured transient ischemic attack as a component of effectiveness outcome. One study [27] was graded as moderate-risk-of-bias for safety outcomes because of the potential selective reporting and no information on missing data.

Rivaroxaban vs. dabigatran

Figure 1 and **Table 2** display results of comparative effectiveness and safety between rivaroxaban and dabigatran. Seven studies that provided data on HRs were synthesized, while the other study [29] that reported adjusted ORs was not included for meta-analyses. No significant difference in risk of stroke or systemic embolism was found between rivaroxaban and dabigatran (HR = 1.00, 95% CI: 0.91 – 1.10, p = 0.97; **Figure 1a**). There was marginally significant heterogeneity observed for risk of stroke or systemic embolism (I^2 = 44%, p-value = 0.1). Compared with dabigatran, rivaroxaban was significantly associated with increased risk of major bleeding (HR = 1.39, 95% CI: 1.28 - 1.50, p < 0.001; **Figure 1b**). Regarding secondary outcomes, no significant difference was found in risk of MI between rivaroxaban and dabigatran (HR = 0.87, 95% CI: 0.72 - 1.05, p = 0.15; **Supplemental Figure 2**), while a higher risk of all-cause death was found with rivaroxaban (HR = 1.28, 95% CI: 1.14 - 1.43, p < 0.001; **Supplemental Figure 3**). No statistically significant heterogeneity was found for risks of major bleeding, MI and death, with all the I^2 of < 50% and p-values of > 0.1.

Likewise, as shown in **Table 2**, rivaroxaban was non-significantly associated with risk of stroke, but significantly associated with increased risk of major GI bleeding, compared with dabigatran. However, no significant association was observed for ICH (p = 0.46). Similar results were found in the subgroup analysis by HAS-BLED score (p-values > 0.05 for subgroup differences; **Table 3**). Sensitivity analyses also yielded similar results to the main analyses (**Table 3**).

Rivaroxaban vs. apixaban

Compared with apixaban, no difference in risk of stroke or systemic embolism was found in rivaroxaban (HR = 1.09, 95% CI: 0.96 – 1.24, p = 0.19; **Figure 1c** and **Table 2**). However a significantly higher risk of major bleeding was observed in rivaroxaban (HR = 1.71, 95% CI: 1.51 –

1.94, p < 0.001; **Figure 1d** and **Table 2**) with significant heterogeneity found (I^2 = 56%, p = 0.04). No analyses for effectiveness outcomes or subgroup analyses were conducted due to insufficient studies or data available. Similar results were found in sensitivity analyses (**Table 3**).

Apixaban vs. dabigatran

In comparison with dabigatran, apixaban was not significantly associated with decreased risk of stroke or systemic embolism (HR = 0.94, 95% CI: 0.83 – 1.06, p = 0.32; **Figure 1e** and **Table 2**), but significantly associated with decreased risk of major bleeding (HR = 0.80, 95% CI: 0.68 – 0.95, p = 0.01; **Figure 1f** and **Table 2**). There was significant heterogeneity found for risk of major bleeding: I^2 = 61%, p = 0·03. No analyses for effectiveness outcomes or subgroup analyses were performed. Sensitivity analyses produced similar findings to the main analyses (**Table 3**).

Assessment of publication bias and quality of a body of evidence

There was no evidence of publication bias found in the comparison amongst NOACs, with all the p-values of > 0·05 for Begg's and Egger's tests (**Supplemental Figure 4-7**). The quality of a body of evidence across included studies was rated as low-quality for the effectiveness outcome of rivaroxaban vs. dabigatran and for the safety outcomes of rivaroxaban vs. apixaban or apixaban vs. dabigatran, due to the non-randomized design and unexplained heterogeneity. The evidence for the safety outcome of rivaroxaban vs. dabigatran was graded as moderate-quality because of the non-randomized design (**Supplemental Table 2**).

Discussion

In this systematic review and meta-analysis, we summarized the evidence from observational studies of direct comparative effectiveness and safety amongst NOACs in patients with AF. No significant differences in risk of stroke or systemic embolism were found between rivaroxaban vs. dabigatran, rivaroxaban vs. apixaban, or apixaban vs. dabigatran. Apixaban was found to have the most favorable safety profile amongst the three NOACs.

Apixaban was associated with a lower risk of major bleeding when compared with dabigatran or rivaroxaban (**Table 2**). This finding may provide some decision-making support for physicians regarding their choices amongst NOACs, especially when considering the equivalent effect of the NOACs on effectiveness outcomes. The evaluated risk of major bleeding and mortality in rivaroxaban compared with dabigatran or apixaban may reflect the true difference in safety outcomes between the three NOACs. The once-daily dosing of rivaroxaban and twice-daily administration of dabigatran might also explain the higher risk of major bleeding in rivaroxaban, given its higher peak in plasma concentrations than dabigatran [20]. However, the observed results (no difference in effectiveness, but better safety) between rivaroxaban and dabigatran or apixaban may also be partly due to selective prescribing. Patients in ROCKET-AF (Rivaroxaban Once Daily Oral Direct Factor Xa Inhibition Compared with Vitamin K Antagonism for Prevention of Stroke and Embolism Trial in Atrial Fibrillation) were older and frailer, required more orthopedic procedures, and had more baseline comorbidities than in RE-LY (Randomized Evaluation of Long-Term Anticoagulant Therapy) or ARISTOTLE (Apixaban for the Prevention of Stroke in Subjects With Atrial Fibrillation) [8-10], therefore physicians may prefer prescribing rivaroxaban to patients at higher risk of adverse health outcomes [30,32]. In addition, once-daily dosing may be preferred for patients on other multiple drugs

or with memory problems, to decrease pill burden. Although all the included studies used multivariable or propensity score adjustment to estimate the relative effect, the non-randomized design could not fully adjust for the effect of selective prescribing or prevent the potential residual confounding.

The numbers of studies included for quantitative syntheses of the primary outcomes were relatively small (**Table 2**). Three studies [27,29,36] investigating risk of major bleeding could not be used for meta-analyses, because two studies [27,36] did not provide data on the relative effect and the other study [29] only reported adjusted ORs (rather than HRs). Nevertheless, they consistently reported higher incidence rates of major bleeding during follow-up in rivaroxaban compared with dabigatran or apixaban. Likewise, due to insufficient studies available and suboptimal reporting, no analyses of using standard NOAC doses or subgroup analyses could be conducted for rivaroxaban vs apixaban or apixaban vs dabigatran. Therefore the significant heterogeneity could not be further explored, leading to the quality of a body of evidence being low (**Supplemental Table 2**).

Three studies summarizing the observational evidence of direct comparisons amongst NOACs have been published [41-43]. Although our findings were in general agreement with their results, the other studies have limitations, either only exploring the comparison between rivaroxaban and dabigatran [42,43], or only assessing the safety profiles amongst NOACs [41,43]. There are several studies using data from RCTs to indirectly compare efficacy and safety outcome between NOACs [12-18]. Our study found similar effectiveness but higher risk of major bleeding in rivaroxaban compared with dabigatran, which was not consistent with the indirect comparison studies that showed higher risk of stroke or systematic embolism in rivaroxaban but no difference in major bleeding [12,13,15,18]. Indirect comparison should be interpreted with caution, given that such comparison is essentially observational design across trials and may suffer apparent and latent biases including confounding [21,44]. Specifically, the difference in the three RCTs (ROCKET-AF, RE-LY, and ARISTOTLE) yielded the indirect comparison questionable and even misleading [19,45], because it remained unclear whether and to what extent the difference in risk of outcome could be attributed to the drug alone. Indirect comparison studies can be used to generate hypotheses that are further corroborated ideally in direct comparative RCTs [18]. Given that no such RCT is available currently or in the near future, findings from real-world studies with direct comparative assessment amongst NOACs may assist in decision-making in clinical practice. The large-scale direct design with multivariable or propensity score adjustment in the relatively homogeneous patients for each individual study may provide more credible evidence than indirect comparison, although an observational study is prone to biases due to its non-randomization and should be interpreted with caution. Moreover, four included studies evaluated the comparison between NOACs and warfarin and reported consistent findings with the respective trials [28,30,31,37]. This would also support the validity of the included observational studies and our current review. However, further large-scale, well-designed and transparently-reported observational studies or eventually head-to-head clinical trials are needed to update the evidence and inform decision-making, because of insufficient studies or data available in our study including limited evidence for subgroup evaluations and for risks of MI and death.

Strengths and limitations

This study is the first systematic review to summarize evidence from observational studies for direct comparison amongst NOACs, to our knowledge. An exhaustive and comprehensive search was conducted to obtain all relevant and most-updated studies. Study processes including screening, data extraction and analyses were performed in duplicate with a good level of agreement. Results from sensitivity analyses supported the robustness of findings from the main analyses.

Some limitations exist in our study. First, the non-randomized comparisons in observational studies may suffer from biases, which could impair the findings and thus weaken the strength of evidence. Secondly, due to limited studies or data, we could not further evaluate the comparative outcomes of interest amongst all NOACs, especially with no data on edoxaban available. Similarly, no analyses could be performed in subgroup populations including patients with or without renal dysfunction, with different sex, at low or high risk of stroke and/or major bleeding, with high or low drug adherence, with or without concomitant over-the-counter antiplatelets, and at different ages. Thirdly, the statistical methods used in the included studies including multivariable regression and propensity score methods were performed to estimate different relative treatment effects. For instance, the propensity score matching was used to estimate effects in the patients who received NOACs; the propensity score covariate adjustment was used for conditional effects within levels of the propensity scores; and the propensity score inverse probability of treatment weighting and the multivariable regression were used to estimate effects in all the patients with AF who were eligible for a NOAC [46]. However no analyses could be conducted to compare the different targeted effects due to the small number of included studies. Furthermore, all the included studies used ICD-9-CM or ICD-10-CM (International Classification of Diseases, Ninth/Tenth Revision, Clinical Modification) codes to identify outcomes and no chart reviews were performed to validate outcome measures, which was a common limitation of observational analyses based on electronic health databases. For example, it was reported that the outcome data (incidences of cardiovascular and bleeding events) identified from medical claims after MI were generally lower than from physician adjudication [47]. Therefore caution is needed when interpreting such observational studies that depend on the data from electronic health databases alone. Additionally, because the follow-up periods were relatively short ranging from 110 days to 400 days (**Table 1**), little was known about the long-term comparative effectiveness and safety between NOACs in patients with AF in the current study.

Conclusion

This systematic review and meta-analysis based on observational studies of direct comparative effectiveness and safety amongst NOACs in patients with AF found increased risk of major bleeding with rivaroxaban compared to dabigatran and apixaban. Apixaban was associated with lower risk of major bleeding than dabigatran. No significant difference was observed in risk of stroke or systemic embolism amongst the three NOACs. Such findings may provide some decision-making support for physicians regarding their choices amongst NOACs in patients with AF.

Author contributions

GL, GYHL, AH and MAHL: conceived and designed the study. GL, YC, LM and LT: acquired data, performed statistical analyses and interpretation, and drafted the manuscript. GYHL, AH, TBL, XS, JT, DMW, MC, and MAHL: provided professional and statistical support, and made several critical revisions to the manuscript. GYHL, AH, LT and MAHL: supervised the study. All authors read and approved the final manuscript. GL acts as the guarantor of this work.

Funding disclosure

This study received no grant from any funding agency in the public, commercial or not-for-profit sectors.

Declaration of interests

GYHL has served as a consultant for Bayer, BMS/Pfizer, Daiichi-Sankyo, Biotronik, Medtronic and Boehringer Ingelheim and has been on the speaker's bureau for Bayer, BMS/Pfizer, Boehringer Ingelheim, Daiichi-Sankyo, Medtronic. TBL has been on the speakers' bureaus for Bayer, BMS/Pfizer, Roche Diagnostics, and Boehringer- Ingelheim. MC has sat on advisory boards for Janssen, Leo Pharma, Portola, and AKP America; and he has received funding for presentations from Leo Pharma, Bayer, Celgene, Shire, and CSL Behring.

All other authors have reported that they have no relationships relevant to the contents of this paper to disclose.

Acknowledgements

None declared.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

The data appeared in this study are already publicly available in the literature.

Abbreviations:

NOACs: non-vitamin K antagonist oral anticoagulants
AF: atrial fibrillation
GRADE: Grading of Recommendations Assessment, Development and Evaluation
HR: hazard ratio
CI: confidence interval
RCT: randomized controlled trial
MI: myocardial infarction

References:

1. Rockson SG, Albers GW (2004) Comparing the guidelines: anticoagulation therapy to optimize stroke prevention in patients with atrial fibrillation. J Am Coll Cardiol 43: 929-935.
2. Chugh SS, Havmoeller R, Narayanan K, Singh D, Rienstra M, et al. (2014) Worldwide epidemiology of atrial fibrillation: a Global Burden of Disease 2010 Study. Circulation 129: 837-847.
3. Camm AJ, Lip GY, De Caterina R, Savelieva I, Atar D, et al. (2012) 2012 focused update of the ESC Guidelines for the management of atrial fibrillation: an update of the 2010 ESC Guidelines for the management of atrial fibrillation. Developed with the special contribution of the European Heart Rhythm Association. Eur Heart J 33: 2719-2747.
4. Freedman B, Potpara TS, Lip GY (2016) Stroke prevention in atrial fibrillation. Lancet 388: 806-817.
5. Olesen JB, Sorensen R, Hansen ML, Lamberts M, Weeke P, et al. (2015) Non-vitamin K antagonist oral anticoagulation agents in anticoagulant naive atrial fibrillation patients: Danish nationwide descriptive data 2011-2013. Europace 17: 187-193.
6. Dentali F, Riva N, Crowther M, Turpie AG, Lip GY, et al. (2012) Efficacy and safety of the novel oral anticoagulants in atrial fibrillation: a systematic review and meta-analysis of the literature. Circulation 126: 2381-2391.
7. Weitz JI, Semchuk W, Turpie AG, Fisher WD, Kong C, et al. (2015) Trends in Prescribing Oral Anticoagulants in Canada, 2008-2014. Clin Ther 37: 2506-2514 e2504.
8. Granger CB, Alexander JH, McMurray JJ, Lopes RD, Hylek EM, et al. (2011) Apixaban versus warfarin in patients with atrial fibrillation. New England Journal of Medicine 365: 981-992.
9. Patel MR, Mahaffey KW, Garg J, Pan G, Singer DE, et al. (2011) Rivaroxaban versus warfarin in nonvalvular atrial fibrillation. New England Journal of Medicine 365: 883-891.
10. Connolly SJ, Ezekowitz MD, Yusuf S, Eikelboom J, Oldgren J, et al. (2009) Dabigatran versus warfarin in patients with atrial fibrillation. N Engl J Med 361: 1139-1151.
11. Giugliano RP, Ruff CT, Braunwald E, Murphy SA, Wiviott SD, et al. (2013) Edoxaban versus warfarin in patients with atrial fibrillation. N Engl J Med 369: 2093-2104.
12. Schneeweiss S, Gagne JJ, Patrick AR, Choudhry NK, Avorn J (2012) Comparative efficacy and safety of new oral anticoagulants in patients with atrial fibrillation. Circ Cardiovasc Qual Outcomes 5: 480-486.
13. Baker WL, Phung OJ (2012) Systematic review and adjusted indirect comparison meta-analysis of oral anticoagulants in atrial fibrillation. Circ Cardiovasc Qual Outcomes 5: 711-719.
14. Fernandez MM, Wang J, Ye X, Kwong WJ, Sherif B, et al. (2015) Systematic review and network meta-analysis of the relative efficacy and safety of edoxaban versus other nonvitamin K antagonist oral anticoagulants among patients with nonvalvular atrial fibrillation and CHADS2 score 2. SAGE Open Med 3: 2050312115613350.
15. Lip GY, Larsen TB, Skjoth F, Rasmussen LH (2012) Indirect comparisons of new oral anticoagulant drugs for efficacy and safety when used for stroke prevention in atrial fibrillation. J Am Coll Cardiol 60: 738-746.
16. Skjoth F, Larsen TB, Rasmussen LH, Lip GY (2014) Efficacy and safety of edoxaban in comparison with dabigatran, rivaroxaban and apixaban for stroke prevention in atrial fibrillation. An indirect comparison analysis. Thromb Haemost 111: 981-988.
17. Rasmussen LH, Larsen TB, Graungaard T, Skjoth F, Lip GY (2012) Primary and secondary prevention with new oral anticoagulant drugs for stroke prevention in atrial fibrillation: indirect comparison analysis. BMJ 345: e7097.
18. Mantha S, Ansell J (2012) An indirect comparison of dabigatran, rivaroxaban and apixaban for atrial

fibrillation. Thromb Haemost 108: 476-484.
19. Cannon CP, Kohli P (2012) Danger ahead: watch out for indirect comparisons! J Am Coll Cardiol 60: 747-748.
20. Graham DJ, Reichman ME, Wernecke M, Hsueh YH, Izem R, et al. (2016) Stroke, Bleeding, and Mortality Risks in Elderly Medicare Beneficiaries Treated With Dabigatran or Rivaroxaban for Nonvalvular Atrial Fibrillation. JAMA Intern Med 176: 1662-1671.
21. Hannink G, Gooszen HG, Rovers MM (2013) Comparison of registered and published primary outcomes in randomized clinical trials of surgical interventions. Ann Surg 257: 818-823.
22. Moher D, Liberati A, Tetzlaff J, Altman DG (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 339: b2535.
23. Viera AJ, Garrett JM (2005) Understanding interobserver agreement: the kappa statistic. Fam Med 37: 360-363.
24. Sterne JA, Hernan MA, Reeves BC, Savovic J, Berkman ND, et al. (2016) ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 355: i4919.
25. Li G, Holbrook A, Jin Y, Zhang Y, Levine MA, et al. (2016) Comparison of treatment effect estimates of non-vitamin K antagonist oral anticoagulants versus warfarin between observational studies using propensity score methods and randomized controlled trials. Eur J Epidemiol 31: 541-561.
26. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, et al. (2011) GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 64: 383-394.
27. Al-Khalili F, Lindstrom C, Benson L (2016) The safety and persistence of non-vitamin-K-antagonist oral anticoagulants in atrial fibrillation patients treated in a well structured atrial fibrillation clinic. Current Medical Research and Opinion 32: 779-785.
28. Chan YH, Kuo CT, Yeh YH, Chang SH, Wu LS, et al. (2016) Thromboembolic, Bleeding, and Mortality Risks of Rivaroxaban and Dabigatran in Asians with Nonvalvular Atrial Fibrillation. Journal of the American College of Cardiology 68: 1389-1401.
29. Deitelzweig S, Bruno A, Trocio J, Tate N, Gupta K, et al. (2016) An early evaluation of bleeding-related hospital readmissions among hospitalized patients with nonvalvular atrial fibrillation treated with direct oral anticoagulants. Current Medical Research & Opinion 32: 573-582.
30. Gorst-Rasmussen A, Lip GYH, Bjerregaard Larsen T (2016) Rivaroxaban versus warfarin and dabigatran in atrial fibrillation: comparative effectiveness and safety in Danish routine care. Pharmacoepidemiology and Drug Safety 25: 1236-1244.
31. Lip GYH, Keshishian A, Kamble S, Pan X, Mardekian J, et al. (2016) Real-world comparison of major bleeding risk among non-valvular atrial fibrillation patients initiated on apixaban, dabigatran, rivaroxaban, or warfarin: A propensity score matched analysis. Thrombosis and Haemostasis 116: 975-986.
32. Noseworthy PA, Yao X, Abraham NS, Sangaralingham LR, McBane RD, et al. (2016) Direct Comparison of Dabigatran, Rivaroxaban, and Apixaban for Effectiveness and Safety in Nonvalvular Atrial Fibrillation. Chest 150: 1302-1312.
33. Hernandez I, Zhang Y (2017) Comparing Stroke and Bleeding with Rivaroxaban and Dabigatran in Atrial Fibrillation: Analysis of the US Medicare Part D Data. Am J Cardiovasc Drugs 17: 37-47.
34. Li WH, Huang D, Chiang CE, Lau CP, Tse HF, et al. (2017) Efficacy and safety of dabigatran, rivaroxaban, and warfarin for stroke prevention in Chinese patients with atrial fibrillation: the Hong Kong Atrial Fibrillation Project. Clin Cardiol 40: 222-229.
35. Deitelzweig (a) S, Bruno A, Gupta K, Trocio J, Tate N (2015) Comparison of all-cause and bleeding-related

hospitalizations among non-valvular atrial fibrillation patients receiving oral anticoagulants. Circulation Conference: American Heart Association's 132.

36. Lai YH, Huang CH, Cheng CL, Yang YHK (2014) Comparative safety of new oral anticoagulants in nonvalvular atrial fibrillation-a single medical center experience. Pharmacoepidemiology and Drug Safety 23: 460.
37. Lin I, Masseria C, Mardekian J, Frean M, Phatak H, et al. (2015) Real-world bleeding risk among non-valvular atrial fibrillation (NVAF) patients prescribed apixaban, dabigatran, rivaroxaban and warfarin: Analysis of electronic health records. European Heart Journal 36: 1084.
38. Amin A, Keshishian A, Xie L, Baser O, Price K (2015) Comparison of Major-bleeding Risk and Health Care Costs Among Treatment-naïve Non-valvular Atrial Fibrillation Patients Initiating Apixaban, Dabigatran, Rivaroxaban or Warfarin. Circulation 132(Suppl 3): A19672.
39. Adeboyeje G, Sylwestrzak G, White J, Rosenberg A, Abarca J, et al. (2016) Comparative Effectiveness and Safety of Anticoagulant Therapy With Warfarin, Dabigatran, Apixaban, or Rivaroxaban in Patients With Nonvalvular Atrial Fibrillation. Circulation: Cardiovascular Quality and Outcomes 9(Suppl 2): A2.
40. Deitelzweig (b) S, Bruno A, Tate N, Ogbonnaya A, Shah M, et al. (2015) Major bleeding, hospitalisation rates and healthcare costs among non-valvular atrial fibrillation patients naive to oral anticoagulation and newly treated with novel oral anticoagulants. Eur Heart J 36: 338.
41. Deitelzweig S, Farmer C, Luo X, Vo L, Li X, et al. (2017) Risk of major bleeding in patients with non-valvular atrial fibrillation treated with oral anticoagulants: a systematic review of real-world observational studies. Current Medical Research and Opinion: 1-12.
42. Bai Y, Deng H, Shantsila A, Lip GY (2017) Rivaroxaban Versus Dabigatran or Warfarin in Real-World Studies of Stroke Prevention in Atrial Fibrillation: Systematic Review and Meta-Analysis. Stroke 48: 970-976.
43. Bundhun PK, Soogund MZ, Teeluck AR, Pursun M, Bhurtu A, et al. (2017) Bleeding outcomes associated with rivaroxaban and dabigatran in patients treated for atrial fibrillation: a systematic review and meta-analysis. BMC Cardiovasc Disord 17: 15.
44. Song F, Altman DG, Glenny AM, Deeks JJ (2003) Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. BMJ 326: 472.
45. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, et al. (2011) GRADE guidelines: 8. Rating the quality of evidence--indirectness. J Clin Epidemiol 64: 1303-1310.
46. Austin PC (2011) An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res 46: 399-424.
47. Guimaraes PO, Krishnamoorthy A, Kaltenbach LA, Anstrom KJ, Effron MB, et al. (2017) Accuracy of Medical Claims for Identifying Cardiovascular and Bleeding Events After Myocardial Infarction : A Secondary Analysis of the TRANSLATE-ACS Study. JAMA Cardiol 2: 750-757.

Tables and figure legends:

Table 1. Patients' characteristics of included studies

Table 2. Results of comparative effectiveness and safety amongst NOACs in patients with AF

Table 3. Results of subgroup and sensitivity analyses for comparison between NOACs

Figure 1. Results of direct comparisons amongst NOACs: a - results for risk of stroke or systemic embolism comparing rivaroxaban with dabigatran; b - results for risk of major bleeding comparing rivaroxaban with dabigatran; c - results for risk of stroke or systemic embolism comparing rivaroxaban with apixaban; d - results for risk of major bleeding comparing rivaroxaban with apixaban; e - results for risk of stroke or systemic embolism comparing apixaban with dabigatran; f - results for risk of major bleeding comparing apixaban with dabigatran

Supplemental Table 1. Ovid search terms modified for MEDLINE, EMBASE and CINAHL (from Jan 1st, 2009 to Nov 30th, 2016)

Supplemental Figure 1. Study flow diagram showing the study selection process

Supplemental Figure 2. Relationship between rivaroxaban and risk of myocardial infarction compared with dabigatran

Supplemental Figure 3. Relationship between rivaroxaban and risk of all-cause death compared with dabigatran

Supplemental Figure 4. Funnel plot for stroke or systemic embolism in the comparison between rivaroxaban and dabigatran

Supplemental Figure 5. Funnel plot for major bleeding in the comparison between rivaroxaban and dabigatran

Supplemental Figure 6. Funnel plot for major bleeding in the comparison between rivaroxaban and apixaban

Supplemental Figure 7. Funnel plot for major bleeding in the comparison between apixaban and dabigatran

Supplemental Table 2. Summary of findings for direct comparative effectiveness and safety between NOACs in patients with atrial fibrillation

Table 1. Patients' characteristics of included studies

First author, publication year	Country	Study design	Data source	Study period	Comparison	Population characteristics				
						Name of NOAC (dose)	All new NOAC-users?	Sample size (% for females)	Age: years	$CHADS_2$ score
Full texts (n =9)										
Al-Khalili, 2016[27]	Sweden	Retrospective cohort	Stockholm Heart Center (a cardiology outpatient clinic)	2011 Dec - 2015 Jan	Rivaroxaban vs Apixaban; Apixaban vs Dabigatran; Rivaroxaban vs Dabigatran	Rivaroxaban	Yes	282 (50%)	Mean 73	Median 3[3]
						Apixaban		251 (49%)	Mean 73	Median 3[3]
						Dabigatran		233 (49%)	Mean 72	Median 3[3]
Chan, 2016[28]	China	Retrospective cohort	Taiwan National Health Insurance Research Database	2013 Feb - 2013 Dec	Rivaroxaban vs dabigatran	Rivaroxaban (10, 15 and 20 mg once daily)	No; some patients had experience with ≥ 1 of study drugs	3,916 (46%)	Mean 76	Mean 4.12[3]
						Dabigatran (110 and 150 mg twice daily)		5,921 (42%)	Mean 75	Mean 4.08[3]
Deitelzweig, 2016* (a)[29]	USA	Retrospective cohort	Premier Hospital Database	2012 Jan - 2014 Mar	Rivaroxaban vs Apixaban; Apixaban vs Dabigatran	Rivaroxaban	Unknown; all patients received a NOAC on their first hospitalization due to AF (index hospitalization)	37,754 (49%)	Mean 72.3	Mean 2.04
						Apixaban		4,138 (51%)	Mean 73.6	Mean 2.19
						Dabigatran		32,838 (46%)	Mean 71.9	Mean 2.09

Deitelzweig, 2016* (b)[29]	USA	Retrospective cohort	Cerner Health Facts Hospital Database	2012 Jan - 2014 Aug	Rivaroxaban vs Apixaban; Apixaban vs Dabigatran	Rivaroxaban	Unknown (same as above in Deitelzweig, 2016 (a))	6,635 (48%)	Mean 72.1	Mean 2.06
						Apixaban		1,813 (51%)	Mean 74.9	Mean 2.35
						Dabigatran		5,753 (45%)	Mean 72.4	Mean 2.15
Graham, 2016[20]	USA	Retrospective cohort	Medicare databases	2011 Nov - 2014 Jun	Rivaroxaban vs dabigatran	Rivaroxaban (20 mg once daily)	Yes	66,651 (47%)	N/A (all patients were ≥ 65 years)	Median 2
						Dabigatran (150 mg twice daily)		52,240 (47%)		Median 2
Gorst-Rasmussen, 2016[30]	Denmark	Retrospective cohort	Danish National Prescription Registry, Danish National Patient Register, and Danish Civil Registration System	2011 Feb - 2014 Jul	Rivaroxaban vs dabigatran	Rivaroxaban (20mg once daily)	Yes	1,629 (49%)	Mean 72.8	Mean 1.5
						Dabigatran (150mg twice daily)		5,320 (37%)	Mean 66.0	Mean 1.0
Hernandez, 2017	USA	Retrospective cohort	Medicare Part D data from the Centers for Medicare and Medicaid Services	2010 - 2013	Rivaroxaban vs dabigatran	Rivaroxaban (20 mg once daily)	Yes	5,799 (54%)	Mean 75.4	Mean 3.29
						Dabigatran (150mg twice daily)		7,322 (50%)	Mean 75.6	Mean 3.28
Li, 2017	China	Prospective cohort	Hospital-based AF registry in Queen Mary Hospital, Hong Kong	2008 Jan - 2014 Dec	Rivaroxaban vs dabigatran	Rivaroxaban	Yes	669 (40%)	Mean 73.3	Mean 2.5
						Dabigatran		467 (47%)	Mean 71.9	Mean 2.2
Lip, 2016[31]	USA	Retrospective cohort	Truven MarketScan®	2013 Jan - 2014 Dec	Rivaroxaban vs dabigatran	Rivaroxaban	Yes	4,657 (36%)	Mean 66.3	Mean 1.6
						Dabigatran		4,657 (35%)	Mean 66.5	Mean 1.6

			Commercial and Medicare supplemental US claims database		Rivaroxaban vs apixaban	Rivaroxaban	Yes	7,399 (39%)	Mean 68.3	Mean 1.7
						Apixaban		7,399 (39%)	Mean 68.4	Mean 1.8
					Apixaban vs dabigatran	Apixaban	Yes	4,407 (36%)	Mean 67.0	Mean 1.6
						Dabigatran		4,407 (36%)	Mean 66.9	Mean 1.7
Noseworthy, 2016[32]	USA	Retrospective cohort	Optum Labs Data Warehouse	2010 Oct - 2015 Feb	Rivaroxaban vs dabigatran	Rivaroxaban	Yes	15,787 (40%)	Median 70	Median 4[3]
						Dabigatran		15,787 (41%)	Median 71	Median 4[3]
					Rivaroxaban vs apixaban	Rivaroxaban	Yes	6,565 (46%)	Median 73	Median 4[3]
						Apixaban		6,565 (46%)	Median 73	Median 4[3]
					Apixaban vs dabigatran	Apixaban	Yes	6,542 (46%)	Median 73	Median 4[3]
						Dabigatran		6,542 (46%)	Median 73	Median 4[3]
Abstracts (n =6)										
Adeboyeje, 2016	USA	Retrospective cohort	A US commercial claims database	2010 Nov - 2015 Feb	Rivaroxaban vs Dabigatran; Dabigatran vs Apixaban; Rivaroxaban vs Apixban	Rivaroxaban	Yes	8,398	Mean 67	N/A
						Apixaban		3,689	Mean 69	
						Dabigatran		8,539	Mean 66	
Amin, 2015	USA	Retrospective cohort	Optum Research Database	2013 Jan - 2014 Dec	Dabigatran vs Apixaban; Rivaroxaban vs Apixban	Rivaroxaban	Yes	8,740	N/A	Mean 4.0[3]
						Apixaban		3,762		Mean 4.2[3]
						Dabigatran		2,677		Mean 4.0[3]
						Apixaban		833		
						Dabigatran		2,150		
Deitelzweig, 2015** (a)[33]	USA	Retrospective cohort	Humana Medicare Advantage Database	2009 Jul – 2014 Sep	Dabigatran vs Apixaban; Rivaroxaban vs Apixban	Rivaroxaban	Yes	7,667	N/A	N/A
						Apixaban		2,028	Mean 75.5	
						Dabigatran		5,644	N/A	
Deitelzweig,	USA	Retrospectiv	PharMetrics Plus	2012 Jan -	Dabigatran vs	Rivaroxaban	Yes	6,167	Mean 63.4	Mean 1.8

2015** (b)		e cohort	data.	2014 Jan	Apixaban; Rivaroxaban vs Apixban	Apixaban		833		
						Dabigatran		2,150		
Lai, 2014[34]	China	Retrospective cohort	A Taiwan Medical Center	2013 Mar – 2013 Oct	Rivaroxaban vs Dabigatran	Rivaroxaban (15mg once daily)	Yes	57 (47%)	Mean 74.8	Mean 3.0
						Dabigatran (110mg twice daily)		56 (43%)	Mean 77.1	Mean 3.1
Lin, 2015[35]	USA	Retrospective cohort	Humedica Electronic Health Record Database	2013 Jan – 2014 Jun	Dabigatran vs Apixaban; Rivaroxaban vs Apixban	Rivaroxaban	Yes	6,407	N/A	N/A
						Apixaban		2,038		
						Dabigatran		2,440		

* This study was presented by study (a) and (b) because it reported separate findings for the two databases and no combined data could be extracted.

** These two studies were attached with (a) and (b) to avoid confusion because they were from the same first author and published in the same year

[1] Outcomes were identified from ICD-9-CM or ICD-10-CM codes

[2] Modified HAS-BLED score ranged from 0 to 8 because labile international normalized ratio was not applicable to NOAC users

[3] Data were for CHA2DS2-VASc score;

[4] Data were for Charlson comorbidity index;

[5] Data were for aspirin use only;

[6] Data were for antiplatelet or NSAID use.

Table 1. (continued)

First author, publication year	Population characteristics				Outcome measures[1]		Follow-up period	Primary statistical analysis used
	HAS-BLED score[2]	Baseline Charlson-Deyo comorbidity index	Baseline renal dysfunction: %	Baseline non-study antiplatelet use: %	Effectiveness endpoint definition	Safety endpoint definition		
Full texts								
Al-Khalili, 2016[27]	N/A	N/A	10%	29%	N/A	Major bleeding defined according to the ISTH (International Society of Thrombosis and Hemostasis) criteria	Median 432 days	Multivariable Cox survival regression
			12%	40%			Median 348 days	
			10%	36%			Median 367 days	
Chan, 2016[28]	Mean 3.11	N/A	22%	41%	Hospital discharge diagnoses for ischemic stroke and systemic embolism	Major bleeding required hospitalization, including ICH, GI bleeding and other critical site bleeding	N/A	Multivariable Cox survival regression
	Mean 3.12		22%	45%				
Deitelzweig, 2016 (a) [29]	Mean 2.35	Mean 2.09[4]	N/A	N/A	N/A	All major bleeding resulted in a hospital readmission within one month of the index hospitalization including ICH, GI bleeding, and from other sites	N/A	Multivariable logistic regression
	Mean 2.56	Mean 2.35[4]						
	Mean 2.33	Mean 2.12[4]						
Deitelzweig, 2016 (b) [29]	Mean 2.31	Mean 2.39[4]	N/A	N/A	N/A	Major bleeding (same as above in Deitelzweig, 2016 (a))	N/A	Multivariable logistic regression
	Mean 2.50	Mean 2.71[4]						
	Mean 2.37	Mean 2.47[4]						
Graham, 2016[20]	Median 2	N/A	11%	15%	Thromboembolic stroke	ICH; Major extracranial bleeding defined as a fatal bleeding event, a hospitalized bleeding requiring transfusion, or hospitalization with hemorrhage into an extracranial critical site	Mean 111 days	Propensity score method (inverse probability of treatment weighting)
	Median 2		13%	13%			Mean 108 days	
Gorst-Rasmus	Mean 2.3	N/A	1.5%	44.0%[5]	Ischemic stroke or	Major bleeding including ICH, GI	Median 1.08 years	Propensity score

sen, 2016[30]	Mean 1.9	N/A	1.1%	36.1%[5]	systemic embolism or transient ischemic attack	bleeding and bleeding from other sites	Median 1.08 years	method (covariate adjustment)
Hernandez, 2017	N/A	N/A	28.6%	6.1%	Inpatient, emergency room, or outpatient claim for ischemic stroke, systemic embolism, transient ischemic attack, or pulmonary embolism	Major bleeding including ICH, hemoperitoneum, and inpatient or emergency room stays for GI, hematuria, or not otherwise specified hemorrhage	Mean 251 days	Propensity score method (inverse probability of treatment weighting)
	N/A	N/A	26.3%	7.1%			Mean 385 days	
Li, 2017	Mean 2.0	N/A	0.4%	N/A	Ischemic stroke that led to hospital admission	ICH that led to hospital admission	Mean 651 days	Multivariable Cox survival regression
	Mean 2.0	N/A	0.9%	N/A				
Lip, 2016[31]	Mean 1.9	Mean 1.6	7.2%	N/A	N/A	Major bleeding defined as bleeding requiring hospitalisation during the period of drug use or within 30 days after the last days of supply of the treatment prescription. Using hospital claims	Mean 173 days	Propensity score method (matching)
	Mean 2.0	Mean 1.6	7.2%				Mean 177 days	
	Mean 2.1	Mean 1.7	7.9%				Mean 182 days	
	Mean 2.2	Mean 1.8	8.5%				Mean 148 days	
	Mean 2.0	Mean 1.6	6.6%				Mean 146 days	
	Mean 2.0	Mean 1.6	7.4%				Mean 179 days	
Noseworthy, 2016[32]	Median 2	Median 2	13.3%	10.8%[6]	Inpatient admission for stroke (ischemic and hemorrhagic stroke) and systemic embolism	Inpatient admission for major bleeding including ICH, GI bleeding and major bleeding from other sites	N/A	Propensity score method (matching)
	Median 2	Median 2	13.7%	11.1%[6]				
	Median 2	Median 2	19.0%	11.7%[6]				
	Median 2	Median 2	19.1%	12.3%[6]				
	Median 2	Median 2	18.8%	12.2%[6]				
	Median 2	Median 2	18.3%	11.9%[6]				
Abstracts								

Adeboyeje, 2016	N/A	N/A	N/A	N/A	Effectiveness outcome defined as a composite of thromboembolic event or stroke	All Major bleeding required hospitalization	N/A	Propensity score method (inverse probability of treatment weighting)
Amin, 2015	N/A	N/A	N/A	N/A	N/A	Major bleeding events identified by the Cunningham algorithm plus additional major bleeding sites	N/A	Multivariable Cox survival regression
Deitelzweig, 2015**(a)[33]	N/A	N/A	N/A	N/A	N/A	All bleeding events required a hospitalization	N/A	Multivariable Cox survival regression
Deitelzweig, 2015**(b)	N/A	N/A	N/A	N/A	N/A	Major bleeding not specified	N/A	Multivariable Cox survival regression
Lai, 2014[34]	N/A	N/A	N/A	N/A	N/A	Major bleeding including cerebral, respiratory, gastrointestinal and urinary hemorrhage	Median 136 days Median 177 days	Multivariable Cox survival regression
Lin, 2015[35]	N/A	N/A	N/A	N/A	N/A	Major bleeding not specified	N/A	Multivariable Cox survival regression

Table 2. Results of comparative effectiveness and safety amongst NOACs in patients with AF

Outcome	Rivaroxaban vs Dabigatran		Rivaroxaban vs Apixaban		Apixaban vs Dabigatran	
	Number of studies/patients	Pooled HR (95% CI), p-value	Number of studies/patients	Pooled HR (95% CI), p-value	Number of studies/patients	Pooled HR (95% CI), p-value
Stroke or SE	7/198,445	1.00 (0.91 – 1.10), 0.97	2/25,217	1.09 (0.96 – 1.24), 0.19[2]	2/25,312	0.94 (0.83 – 1.06), 0.32[2]
Stroke	4/164,722	1.02 (0.89 – 1.16), 0.75[3]	1/13,130	-[1]	1/13,084	-[1]
SE	1/13,121	-[1]	0/0	-[1]	0/0	-[1]
Major bleeding	7/206,623	1.39 (1.28 – 1.51), < 0·001	7/77,657	1.71 (1.51 – 1.94), < 0·001	6/43,470	0.80 (0.68 – 0.95), 0·01
ICH	4/173,423	1.19 (0.75 – 1.88), 0.46	1/13,130	-[1]	1/13,084	-[1]
Major GI bleeding	3/141,849	1.26 (1.18 – 1.36), < 0·001	0/0	-[1]	0/0	-[1]
Myocardial infarction	2/128,728[2]	0.87 (0.72 – 1.05), 0.15	0/0	-[1]	0/0	-[1]
All-cause death	4/148,798	1.28 (1.14 – 1.43), <0.001	0/0	-[1]	0/0	-[1]

[1] No meta-analysis conducted due to insufficient studies or data available;

[2] Fixed-effects model was used due to only two studies included for analysis;

[3] Data were for ischemic stroke

Table 3. Results of subgroup and sensitivity analyses for comparison between NOACs

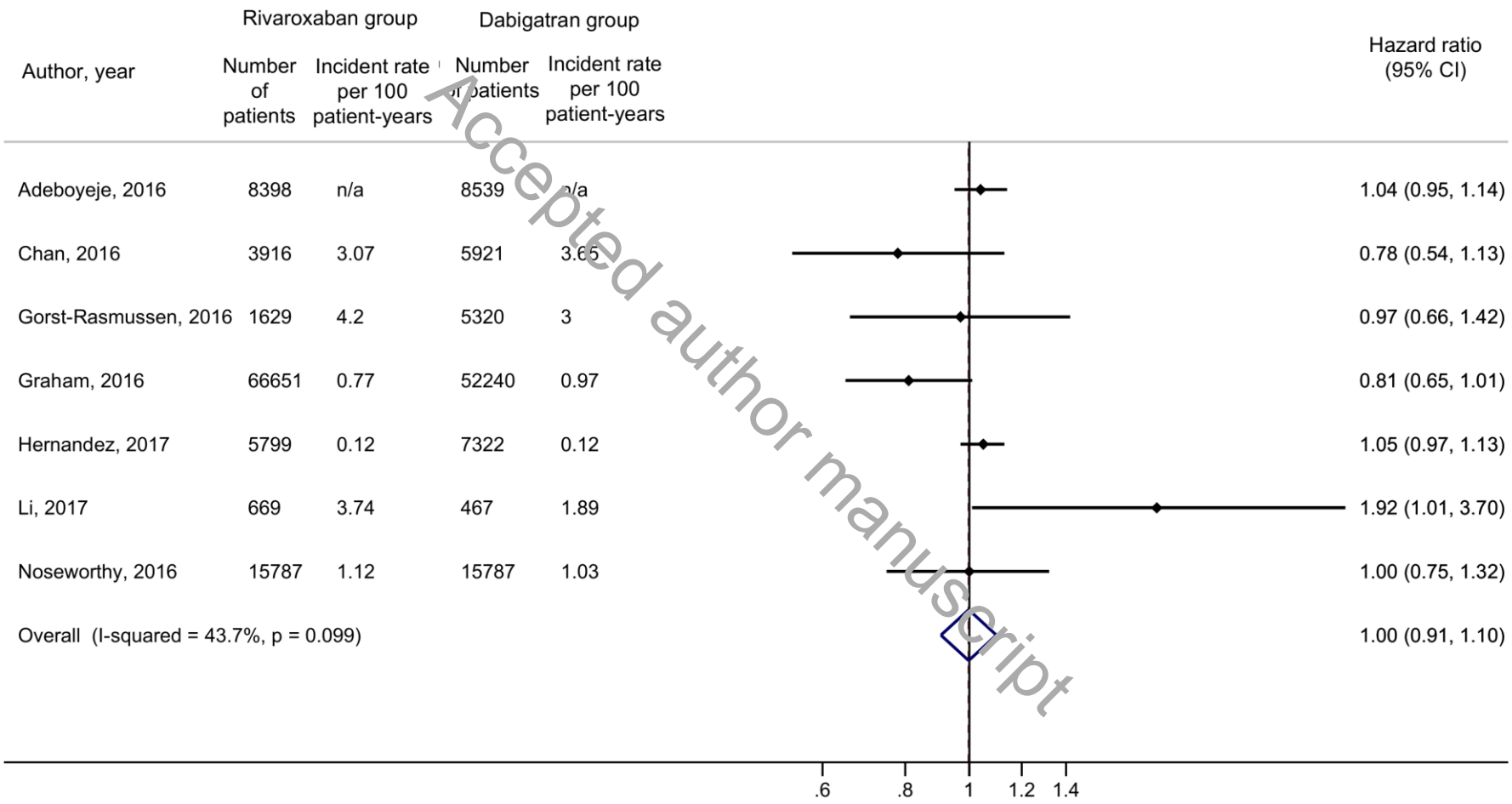
Analysis	Stroke or SE*	Major bleeding		
	Rivaroxaban vs Dabigatran	Rivaroxaban vs Dabigatran	Rivaroxaban vs Apixaban	Apixaban vs Dabigatran
Subgroup analysis[1]				
HAS-BLED score				
≤ 2	0.78 (0.60 – 1.02), 0.052[2]	1.41 (1.13 – 1.76), 0.002[2]	-[3]	-[3]
> 2	1.01 (0.80 – 1.27), 0.95[2]	1.35 (1.11 – 1.63), 0.001[2]	-[3]	-[3]
Sensitivity analysis				
Employing fixed-effects model	1.03 (0.97 – 1.08), 0.38	1.40 (1.31 – 1.49), < 0.001	1.65 (1.52 – 1.79), < 0.001	0.84 (0.76 – 0.92), <0.001
Only including low-risk-of-bias studies	1.00 (0.90 – 1.11), 0.95	1.35 (1.24 – 1.46), < 0.001	2.11 (1.70 – 2.63), < 0.001	0.69 (0.48 – 0.98), 0.039
Using data on standard doses of NOACs	0.96 (0.80 – 1.15), 0.63	1.45 (1.30 – 1.61), < 0.001	-[3]	-[3]

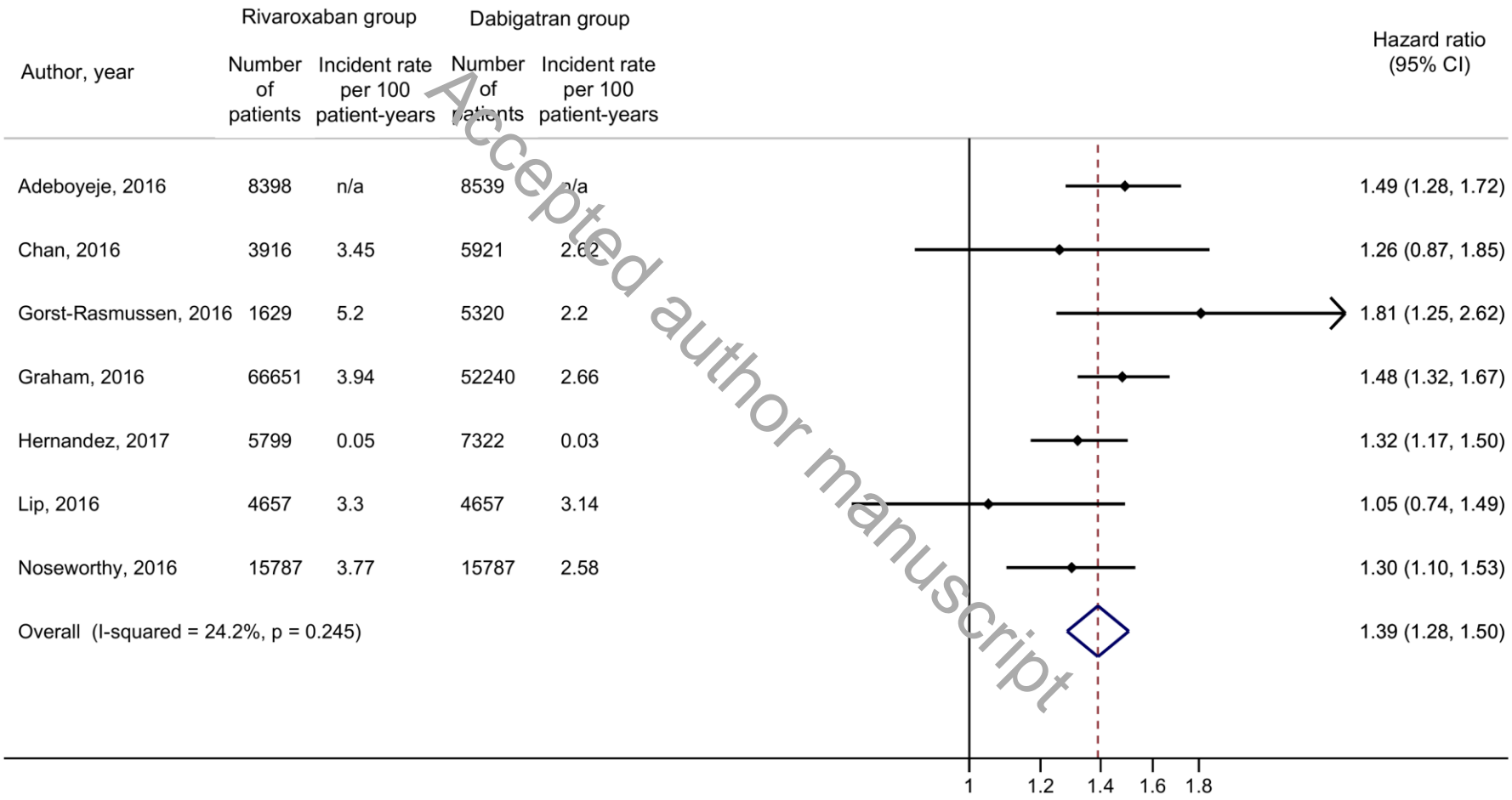
* No syntheses conducted for stroke or SE regarding Rivaroxaban vs Apixaban or Apixaban vs Dabigatran due to only 2 studies included and no data available for subgroup analyses

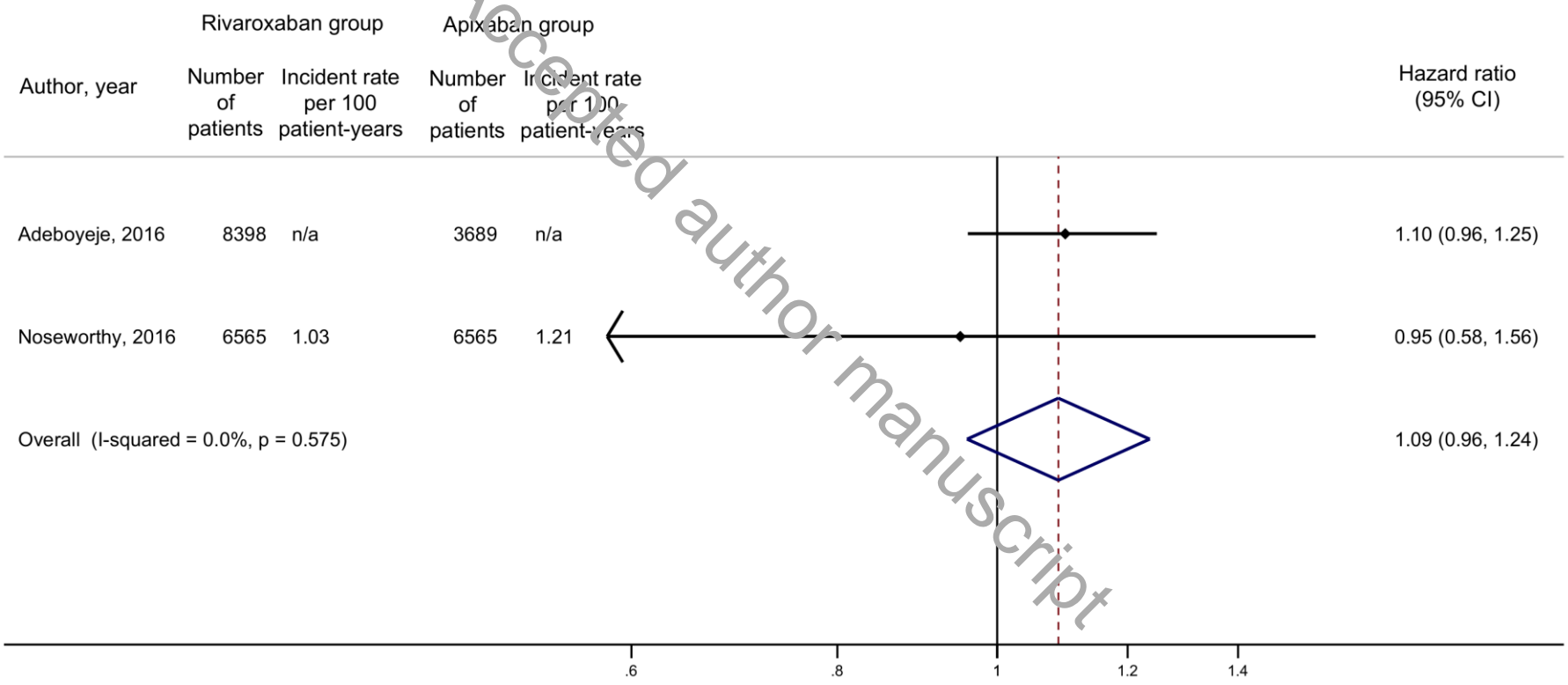
[1] No subgroup analyses performed by length of follow-up, $CHADS_2$ (> 2 vs. ≤ 2) score or CHA_2DS_2-VASc (> 3 vs. ≤ 3) score because of insufficient studies or data available;

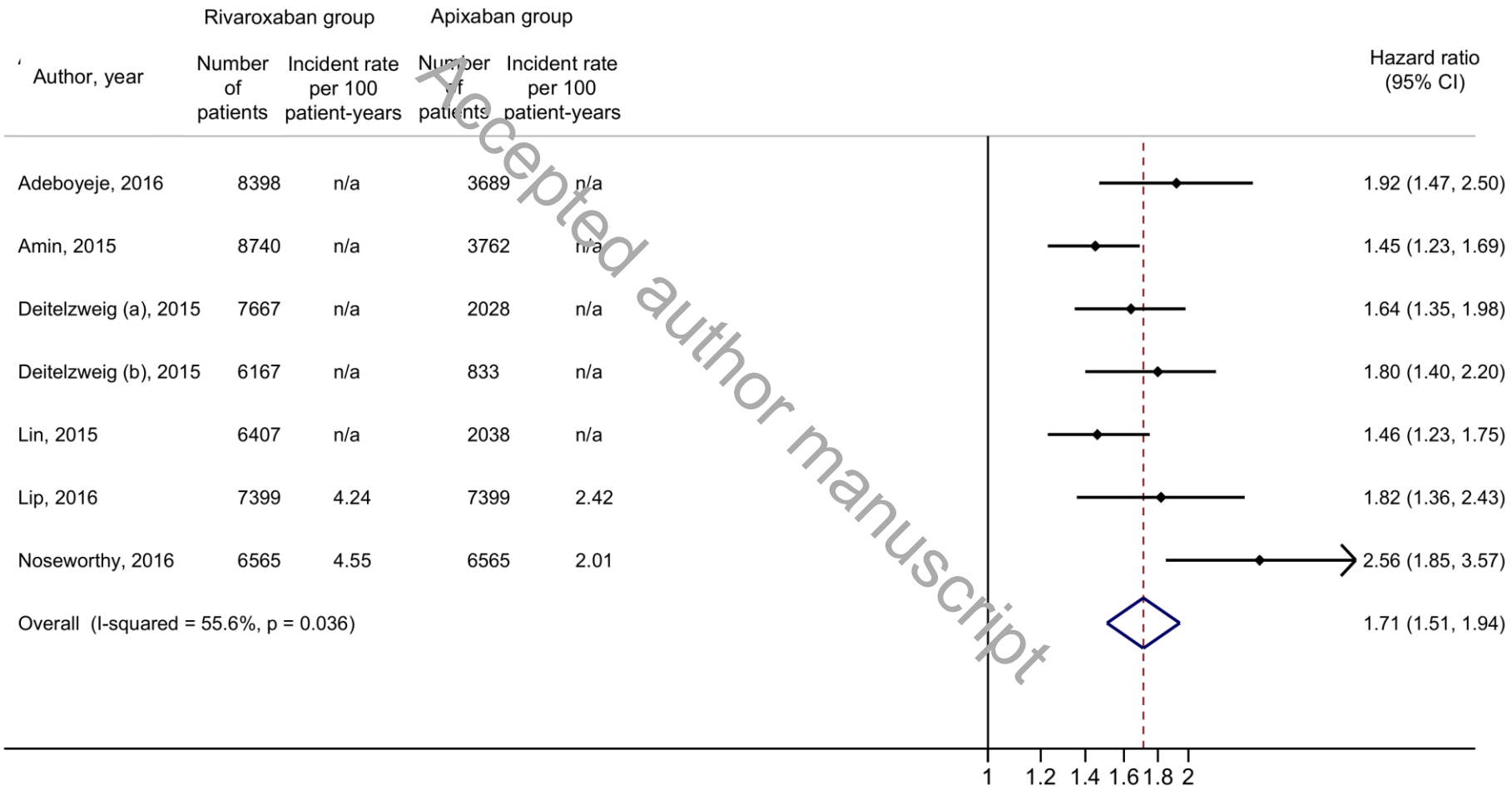
[2] Fixed-effects model was used due to only two studies included for analysis;

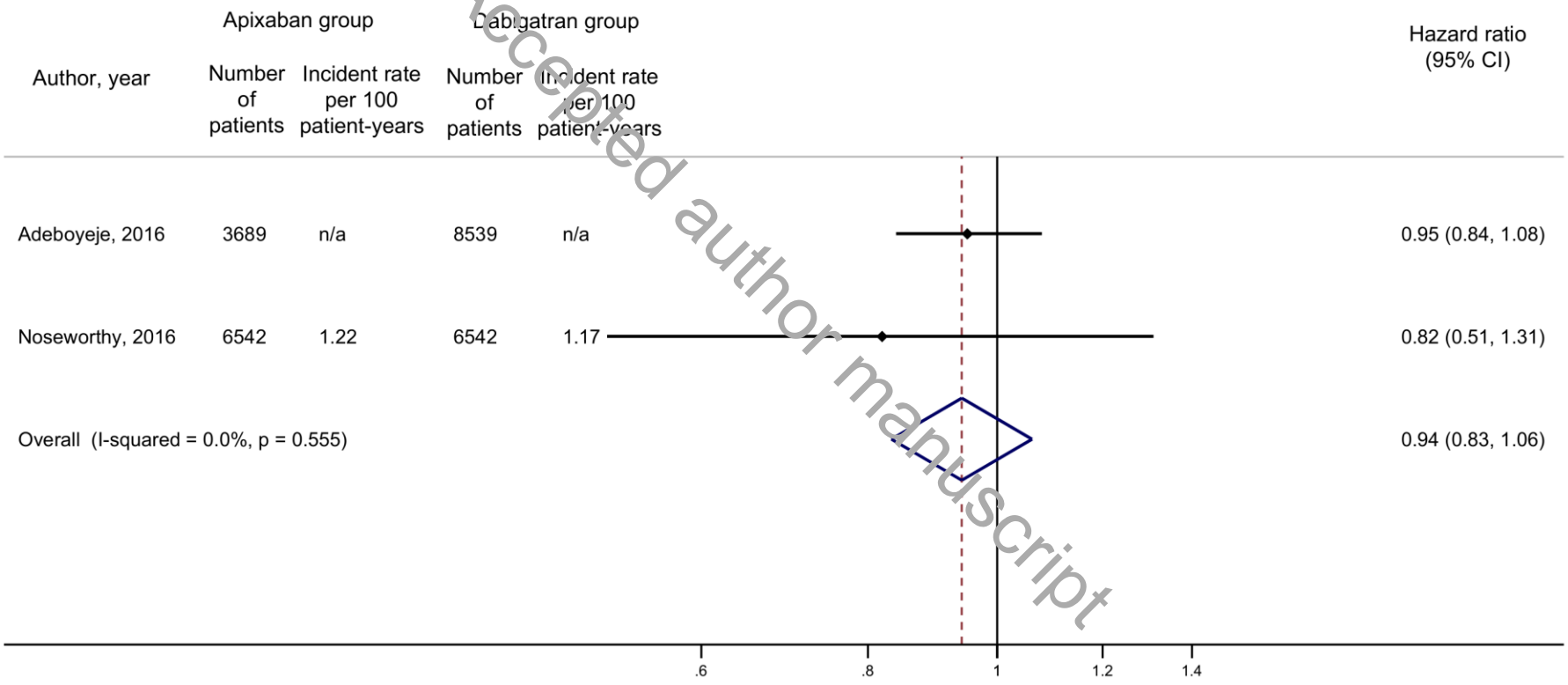
[3] No meta-analysis conducted due to insufficient studies or data available

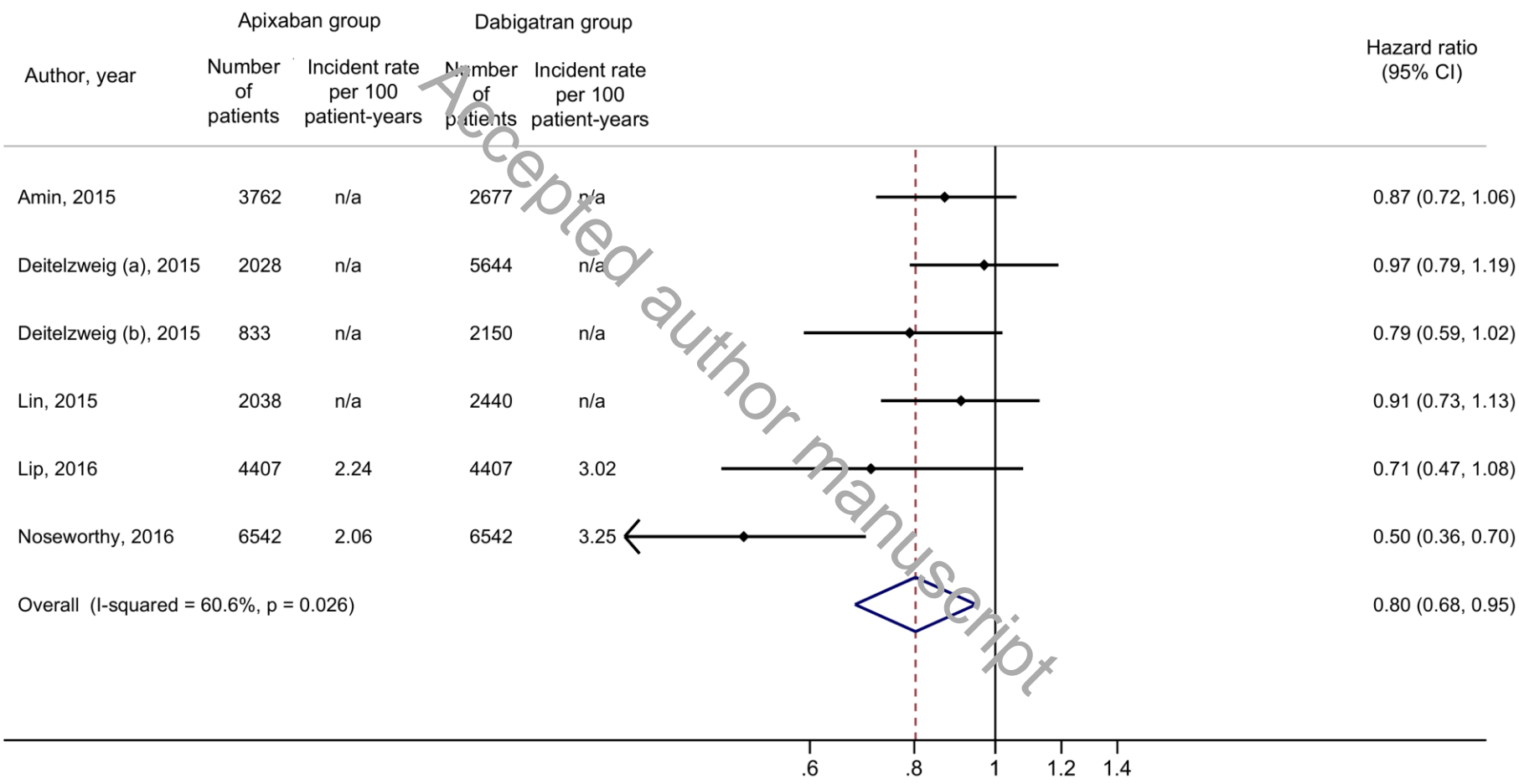






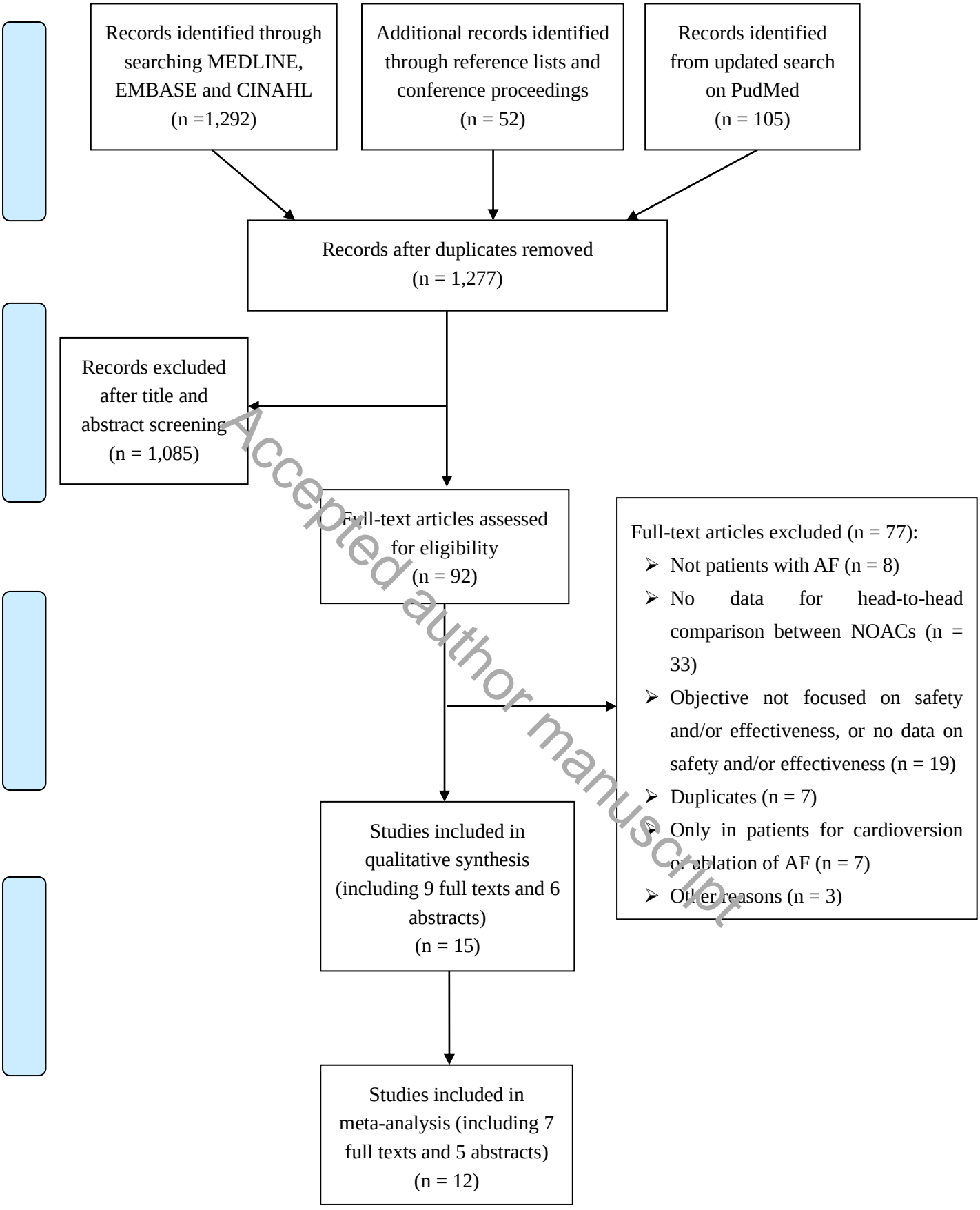




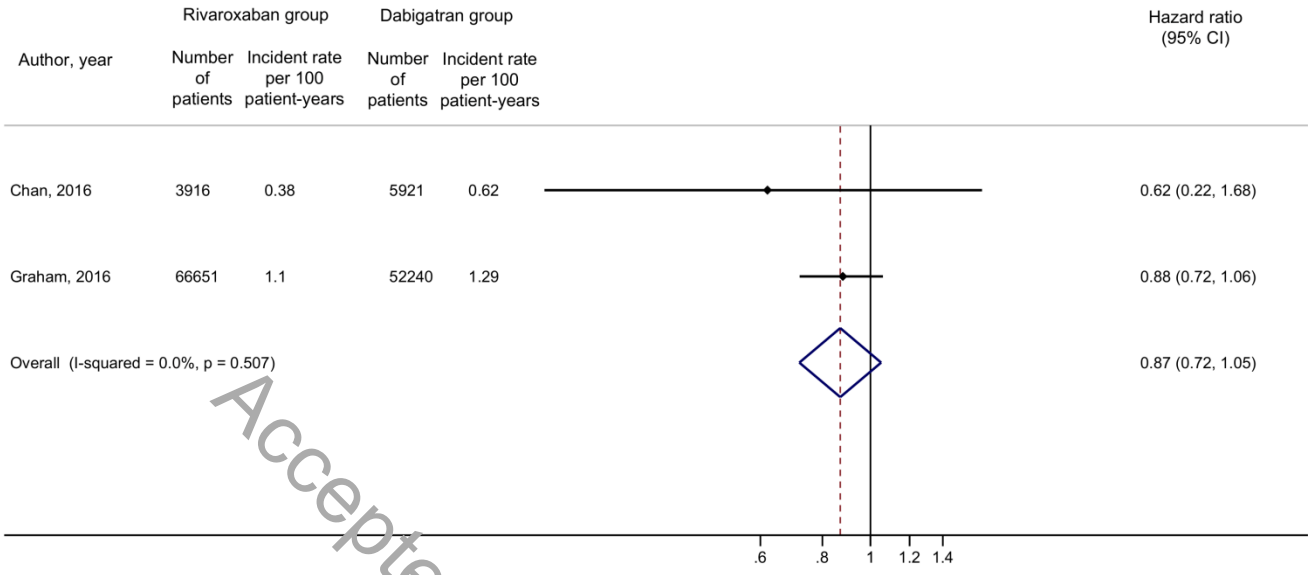


Supplemental Table 1. Ovid search terms modified for MEDLINE, EMBASE and CINAHL (from Jan 1st, 2009 to Nov 30th, 2016)

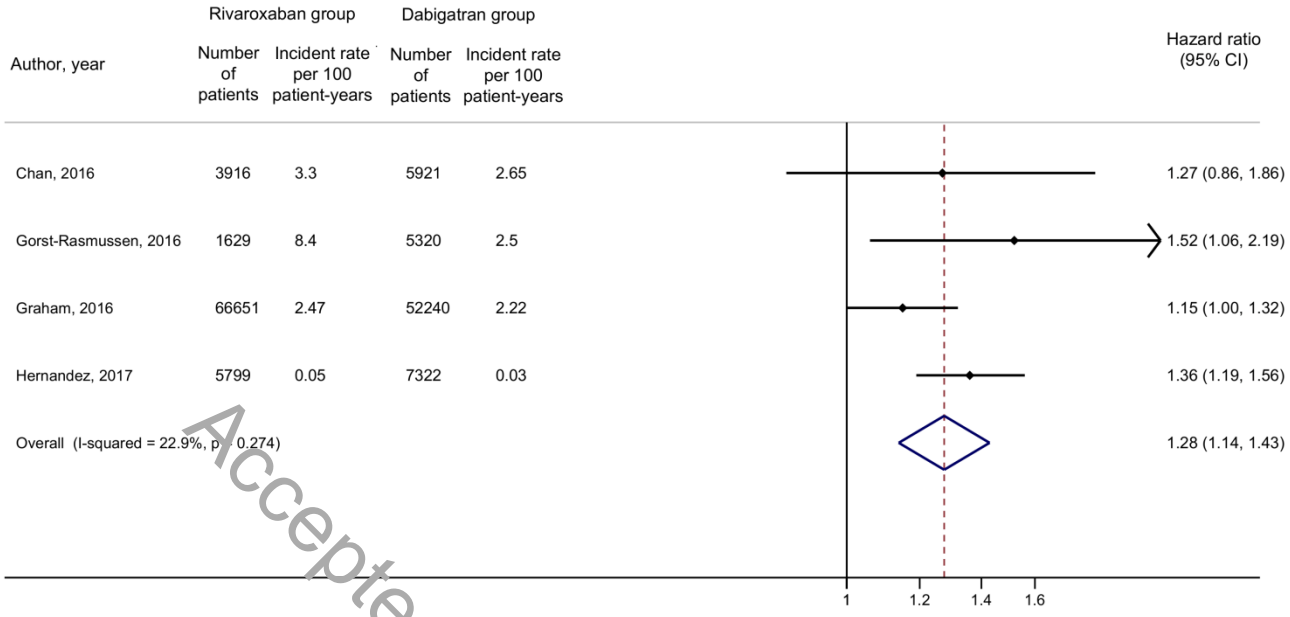
Search steps	Search terms
1	atrial fibrillation.mp. or heart atrium fibrillation/ [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
2	atrial flutter.mp. or heart atrium flutter/
3	1 or 2
4	(dabigatran or BIBR1048 or BIBR-1048 or "BIBR 1048").mp.
5	(rivaroxaban or "BAY 59 7939" or "BAY 59-7939" or "BAY 597939" or BAY59-7939 or BAY597939).mp.
6	(apixaban or BMS-562247 or BMS562247 or "BMS 562247").mp.
7	(edoxaban or DU-176b or DU176b or "DU 176b").mp.
8	("non-vitamin K antagonis$" or "non-vitamin K").tw.
9	4 or 5 or 6 or 7 or 8
10	("observational study" or "observational").mp.
11	exp cohort study
12	cohort.mp.
13	exp case-control study or case-control.mp.
14	10 or 11 or 12 or 13
15	3 and 9 and 14



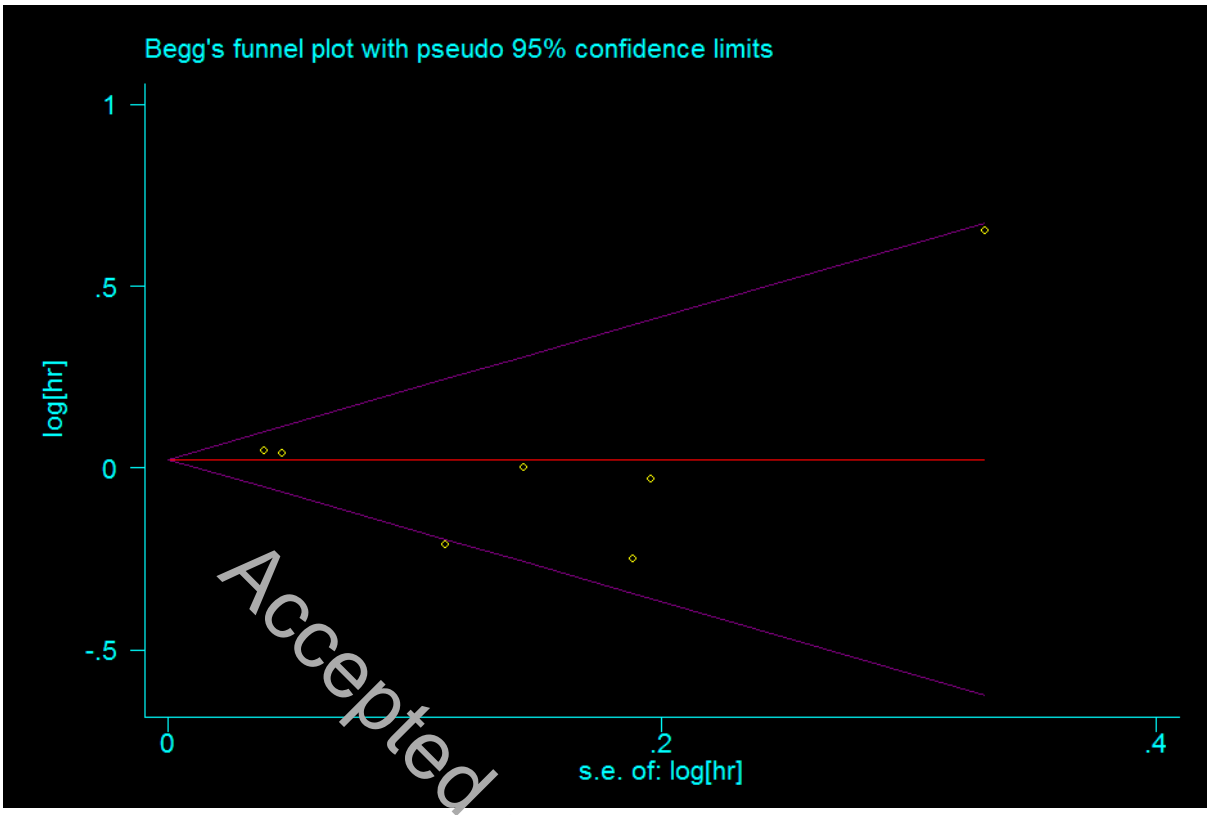
Supplemental Figure 1. Study flow diagram showing the study selection process



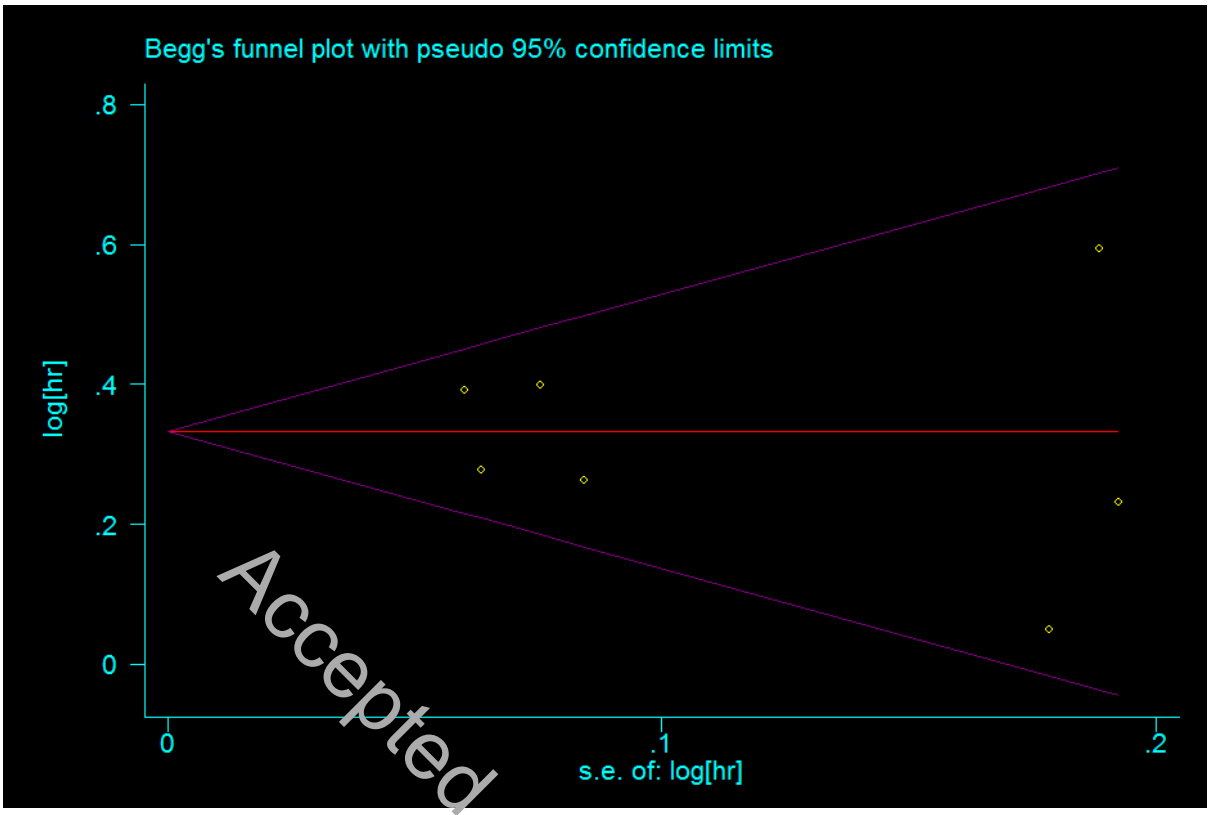
Supplemental Figure 2. Relationship between rivaroxaban and risk of myocardial infarction compared with dabigatran



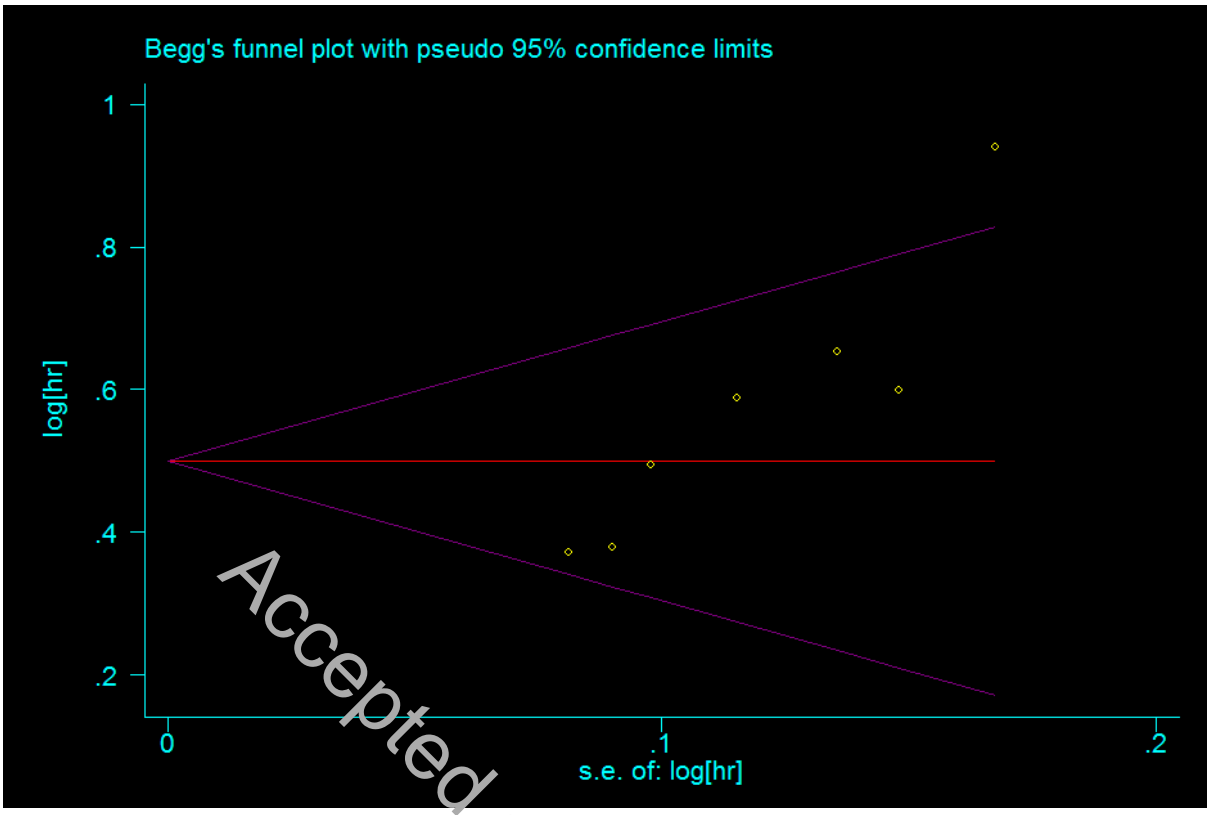
Supplemental Figure 3. Relationship between rivaroxaban and risk of all-cause death compared with dabigatran



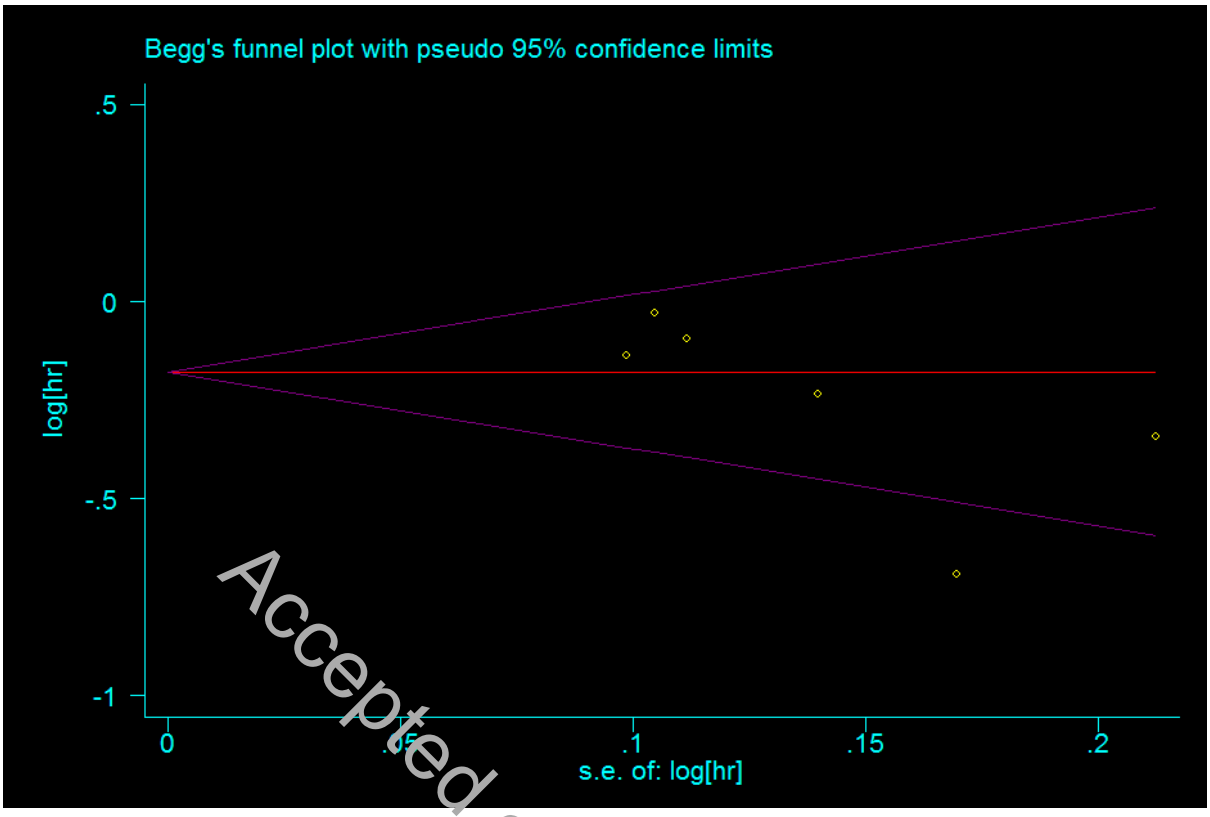
Supplemental Figure 4. Funnel plot for stroke or systemic embolism in the comparison between rivaroxaban and dabigatran



Supplemental Figure 5. Funnel plot for major bleeding in the comparison between rivaroxaban and dabigatran



Supplemental Figure 6. Funnel plot for major bleeding in the comparison between rivaroxaban and apixaban



Supplemental Figure 7. Funnel plot for major bleeding in the comparison between apixaban and dabigatran

Supplemental Table 2. Summary of findings for direct comparative effectiveness and safety between NOACs in patients with atrial fibrillation

Patient or population: Patients with nonvalvular atrial fibrillation (AF)
Settings: Multicenter, multinational data from observational studies
Intervention: NOAC **(**rivaroxaban, dabigatran, or apixaban)
Comparison: another NOAC as the reference

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	No of Participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk NOAC as the reference	**Corresponding risk** Another NOAC				
Results of rivaroxaban vs dabigatran						
Stroke or systemic embolism Follow-up: varied from 110 to 400 days	Incident rate ranging from 0.1% to 3.7% per 100 patient-years	Incident rate ranging from 0.1% to 4.2% per 100 patient-years	**1.00[1] (0.91 – 1.10)**	**198,445** (7 studies)	⊕⊕⊖⊖[2] **Low**	-
Major bleeding Follow-up: varied from 110 to 400 days	Incident rate ranging from 0.03% to 3.1% per 100 patient-years	Incident rate ranging from 0.05% to 5.2% per 100 patient-years	**1.39[1] (1.28 - 1.50)**	**206,623** (7 studies)	⊕⊕⊕⊖[3] **Moderate**	-
Results of rivaroxaban vs apixaban						
Stroke or systemic embolism Follow-up: not reported	Incident rate: 1.2% as reported in only one study	Incident rate: 1.0% as reported in only one study	**1.09[1] (0.96 – 1.24)**	**25,217** (2 studies)	**Undetermined[4]**	-
Major bleeding Follow-up: only one study provided data on follow-up period of approximately 160 days	Incident rate approximately 2% per 100 patient-years	Incident rate approximately 4-5% per 100 patient-years	**1.71[1] (1.51 – 1.94)**	**77,657** (7 studies)	⊕⊕⊖⊖[2] **Low**	-
Results of apixaban vs dabigatran						
Stroke or systemic embolism Follow-up	Incident rate: 1.2% as reported in only one study	Incident rate: 1.2% as reported in only one study	**0.94[1] (0.83 – 1.06)**	**25,312** (2 studies)	**Undetermined[4]**	-
Major bleeding Follow-up: only one study provided data on follow-up period of approximately 160	Incident rate approximately 3% per 100 patient-years	Incident rate approximately 2% per 100 patient-years	**0.77[1] (0.58 – 1.02)**	**43,470** (6 studies)	⊕⊕⊖⊖[2] **Low**	-

days					

*The basis for the assumed risk (e.g. the median control group risk across studies) is provided. The corresponding risk (and its 95% confidence interval) is based on the assumed risk in the comparison group and the relative effect of the intervention (and its 95% CI).

CI: Confidence interval; PS: propensity score; RCT: randomized controlled trial; NOACs: new oral anticoagulants;

GRADE Working Group grades of evidence

High quality: Further research is very unlikely to change our confidence in the estimate of effect.

Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.

Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.

Very low quality: We are very uncertain about the estimate.

[1] Expressed as hazard ratios (HRs)

[2] Due to the unexplained heterogeneity and non-randomized design in the included observational studies

[3] Due to the non-randomized design in the included observational studies

[4] Due to insufficient studies or data