

Comparative Efficacy and Safety of Direct Oral Anticoagulants in Atrial Fibrillation and Heart Failure: A Bayesian Network Meta-Analysis of Reconstructed IPD and RMST

I Komang Raditya Putra Pratama

pratama.2302511213@student.unud.ac.id

Faculty of Medicine, Udayana University

Ni Wayan Pradnyadhari Kusumaputri

Faculty of Medicine, Udayana University

I Nyoman Wiryawan

Doctor of Medicine Program, Faculty of Medicine, Udayana University

Research Article

Keywords: Atrial Fibrillation, Heart Failure, Direct Oral Anticoagulants, Individual Patient Data Reconstruction, Restricted Mean Survival Time

Posted Date: June 14th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10002222/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Purpose

To determine the optimal direct oral anticoagulant (DOAC) for patients with atrial fibrillation (AF) and heart failure (HF) by evaluating absolute comparative efficacy and safety using reconstructed individual patient data (IPD) and restricted mean survival time (RMST).

Methods

HF subgroup data from four landmark DOAC trials (N = 25,762) were extracted. Pseudo-IPD were algorithmically reconstructed from Kaplan-Meier curves. We applied a Bayesian network meta-analysis to estimate hazard ratios and SUCRA probabilities, followed by survival modeling to calculate absolute event-free days gained via RMST up to a 2.25-year truncation time.

Results

For stroke or systemic embolism, apixaban 5mg BID and dabigatran 150mg BID ranked highest (SUCRA: 70.3% and 69.6%). RMST analysis demonstrated dabigatran 150mg and apixaban 5mg significantly extended stroke-free survival by + 5.74 days (95% CI: 2.18–9.31) and + 3.23 days (95% CI: 0.92–5.55) versus vitamin K antagonists (VKAs), respectively. For major bleeding, apixaban 5mg ranked highest (SUCRA: 73.4%), preserving an additional + 8.32 bleeding-free days (95% CI: 5.14–11.50) over VKAs, followed by edoxaban 60mg (+ 4.47 days; 95% CI: 0.55–8.39). Efficacy and safety profiles were temporally stable and consistent across both reduced and preserved ejection fraction phenotypes ($P > 0.05$).

Conclusion

Translating relative hazards into absolute temporal metrics establishes that apixaban 5mg BID provides the optimal net clinical benefit in AF with HF, uniquely maximizing both stroke-free and bleeding-free survival across all ejection fraction phenotypes.

INTRODUCTION

Heart failure (HF) and atrial fibrillation (AF) are intertwined clinical epidemics. They share underlying risk factors and worsen each other's disease trajectory. [1] When AF develops in a patient with pre-existing HF, the risk of thromboembolism spikes. This happens because the HF syndrome creates an ideal environment for thrombosis through left ventricular stasis, endothelial injury, and a baseline hypercoagulable state.[2]

Direct oral anticoagulants (DOACs) have largely replaced vitamin K antagonists (VKAs) for stroke prevention in routine AF management.[3] Yet, applying these agents to the HF population is not straightforward. Major clinical guidelines currently group all DOACs as a single therapeutic class, offering little guidance on which specific agent fits best when HF is present.[4] Biologically, VKA efficacy relies heavily on time-in-therapeutic range, which frequently drops when hepatic congestion or polypharmacy, such as amiodarone use interferes. DOACs avoid some of these liver-related issues, but their heavy dependence on renal clearance makes them vulnerable to the fluctuating kidney function typically seen in cardiorenal syndrome. Because of these distinct pharmacokinetic hurdles, it remains unclear if one DOAC outperforms another, or VKAs, particularly when evaluating HF with reduced versus preserved ejection fraction, specifically HF_rEF and HF_pEF, separately.[5]

A few network meta-analyses (NMAs) have investigated this issue. However, most of them have a major limitation, namely that they rely entirely on aggregate data and assume a proportional hazard (PH) model.[6–8] Relying on conventional hazard ratios (HRs) means assuming the treatment effect stays constant over time. In a chronic condition like HF, where renal reserve and cardiac geometry progressively change, antithrombotic efficacy and bleeding risks do not remain static. On top of that, relative risks mean very little at the bedside because they cannot tell a clinician how many actual days of event-free survival a patient might gain.[9]

We designed this study to bypass those methodological barriers. Using an algorithm-based approach, we reconstructed individual patient data (IPD) from the landmark DOAC trials and integrated it into a Bayesian NMA framework.[10] This allowed us to drop the PH assumption entirely. Instead, we calculated the restricted mean survival time (RMST) to translate relative risks into absolute event-free days for both efficacy, specifically stroke or systemic embolism, and safety, specifically major bleeding and intracranial hemorrhage. We also applied piecewise Cox regression to observe how treatment effects shifted over time, stratifying all outcomes by HF_rEF and HF_pEF. The main goal was to move away from class-wide assumptions and map out an individualized DOAC strategy for patients battling both AF and HF.

METHODS

Search Strategy and Selection Criteria

This systematic review and network meta-analysis adhered strictly to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) extension for Network Meta-Analyses guidelines.[11] The study protocol was prospectively registered in the International Prospective Register of Systematic Reviews (PROSPERO: CRD420261373577).

A comprehensive, systematic literature search was executed across three major biomedical databases: PubMed/MEDLINE, Embase, and the Cochrane Central Register of Controlled Trials (CENTRAL), capturing records from database inception through to the present. The search syntax was constructed using a combination of Medical Subject Headings (MeSH) and free-text keywords, linked by Boolean operators. Core search domains encompassed “atrial fibrillation,” “heart failure,” “direct oral anticoagulants,” “novel oral anticoagulants,” “apixaban,” “dabigatran,” “edoxaban,” “rivaroxaban,” and “vitamin K antagonists.” To ensure maximum sensitivity, no language or geographic restrictions were imposed. Additionally, manual forward and backward citation tracking of pertinent review articles, clinical guidelines, and previously published meta-analyses was performed to capture any residual eligible literature.

Study eligibility was governed by the Population, Intervention, Comparison, Outcomes, and Study Design (PICOS) framework. Inclusion criteria dictated the selection of phase III randomized controlled trials (RCTs), including their pre-specified HF substudies, that enrolled adult patients (≥ 18 years) diagnosed with non-valvular AF and concomitant HF. The intervention arm must involve treatment with a DOAC, specifically apixaban, dabigatran, edoxaban, or rivaroxaban across standard or dose-reduced regimens, compared directly against dose-adjusted VKAs, with target international normalized ratio [INR] 2.0–3.0. Eligible records were required to report time-to-event outcomes (hazard ratios or verifiable Kaplan-Meier curves) for at least one primary efficacy endpoint (stroke or systemic embolism) or primary safety endpoint (International Society on Thrombosis and Haemostasis [ISTH]-defined major bleeding). To rigorously preserve network transitivity and clinical homogeneity, observational cohorts, cross-sectional analyses, non-randomized registries, and phase II dose-escalation trials were strictly excluded.

Data Extraction and Quality Assessment

Two investigators, IKRPP and NWPK, independently conducted the literature screening process via a sequential two-stage approach, starting with initial evaluation of titles and abstracts, followed by rigorous full-text appraisal of potentially eligible articles. Any inter-reviewer discrepancies regarding study eligibility were resolved through discussion to reach a consensus, or via adjudication by a third, senior cardiovascular methodologist.

Data extraction was performed independently by two reviewers, IKRPP and NWPK, utilizing pre-piloted, standardized electronic extraction forms. Extracted variables encompassed study design characteristics, baseline demographic profiles, mean follow-up duration, and the specific diagnostic criteria utilized for HF, including the baseline distribution of left ventricular ejection fraction [LVEF] and New York Heart Association [NYHA] functional classes. For survival outcomes, we systematically extracted trial-level hazard ratios (HRs) with their corresponding 95% confidence intervals (CIs). In instances where trial outcomes were stratigraphically reported based on LVEF phenotypes (HF with reduced ejection fraction [HFrEF] versus HF with preserved ejection fraction [HFpEF]), these distinct subgroup estimates were retrieved to facilitate predefined categorical subgroup and interaction analyses.

The methodological rigor and potential for bias within the included trials were meticulously appraised utilizing the revised Cochrane risk-of-bias tool for randomized trials (RoB 2).^[12] Each trial was evaluated independently across five critical domains: bias arising from the randomization process, deviations from intended interventions, missing outcome data, measurement of the outcome, and selection of the reported result. The overall risk of bias for each RCT was subsequently classified as “low risk,” “some concerns,” or “high risk.” To determine the ultimate reliability of the analytical outputs, the overall certainty of evidence for each network estimate was appraised using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) framework adapted specifically for NMAs, encompassing evaluations of study limitations, indirectness, inconsistency, imprecision, and publication bias.^[13]

Algorithm-based IPD Reconstruction

To overcome the limitations of traditional AD meta-analyses, we reconstructed pseudo-IPD for time-to-event outcomes (stroke/systemic embolism and major bleeding) utilizing the established algorithm by Guyot et al.^[10] High-resolution digital extraction of the published Kaplan-Meier curves from the HF substudies of the ARISTOTLE, ENGAGE AF-TIMI 48, RE-LY, and ROCKET-AF trials was performed using WebPlotDigitizer (version 4.6). The extracted spatial coordinates representing survival probabilities and time intervals were merged with the corresponding at-risk tables. For trials reporting cumulative hazard rates (e.g., RE-LY), spatial coordinates were mathematically transformed into survival probabilities ($S(t) = e^{-H(t)}$) prior to algorithmic processing. These processed coordinates were then solved through a customized iterative root-finding algorithm in R (version 4.3.0, *IPDfromKM* package) to deduce the underlying event times and censoring statuses, thereby yielding a reconstructed pseudo-IPD dataset of marginal survival times and censoring events. It must be noted that while this algorithm accurately approximates time-to-event trajectories at the cohort level, it does not retrieve patient-level baseline covariates (e.g., individual CHADS2 scores or continuous LVEF).

Statistical Analysis

The augmented NMA was executed in a two-stage hierarchical framework. First, aggregate outcomes were analyzed within a Bayesian framework to estimate pooled hazard ratios (HRs) and 95% credible intervals (CrIs). To determine the most parsimonious and statistically appropriate approach, both fixed-effects and random-effects models were initially fitted. Final model selection was rigorously guided by the Deviance Information Criterion (DIC) and the assessment of between-study heterogeneity (evaluated via the variance parameter τ and the I^2 statistic). Formal model comparison indicated a comparable fit between the two approaches ($\Delta DIC < 3$ across all clinical outcomes); however, to conservatively account for potential clinical heterogeneity across the landmark trials (such as varying baseline stroke risks and HF definitions), a random-effects model was consistently utilized for all primary analyses.

Models were fitted via Markov Chain Monte Carlo (MCMC) simulations using non-informative vague priors, running across four parallel chains. Each chain consisted of 100,000 iterations following a 20,000-iteration burn-in period to ensure optimal parameter space exploration. Model convergence was rigorously verified through visual inspection of trace plots and the Brooks-Gelman-Rubin diagnostic statistic (potential scale reduction factor, $\hat{R}$, with values strictly maintained at ≤ 1.05 indicating adequate convergence). Treatment rankings were probabilistically determined via the surface under the cumulative ranking curve (SUCRA). Crucially, because the evidence base formed a star network with VKA serving as the sole common comparator, closed-loop inconsistency

testing (e.g., node-splitting) was mathematically unfeasible. Consequently, the validity of all indirect DOAC comparisons relies strictly on the assumption of clinical and methodological transitivity across the trials.

Second, the reconstructed IPD dataset was subjected to advanced survival modeling. To quantify treatment effects in an absolute temporal scale without relying on the proportional hazards (PH) assumption, we calculated the restricted mean survival time (RMST), defined as the area under the survival curve from randomization up to a clinically relevant truncation time (τ) of 2.25 years, representing the maximal shared follow-up duration across the trials. Between-group RMST differences were evaluated to determine the absolute event-free time gained. Additionally, a landmark analysis incorporating piecewise Cox proportional hazards regression was conducted to assess the temporal consistency of treatment effects. The follow-up period was dichotomized at the 1-year threshold to evaluate early (≤ 1 year) versus late (> 1 year) treatment interactions. To preclude confounding by baseline risk heterogeneities across different trial populations (Simpson’s paradox), all Cox models were robustly stratified by trial origin.

To rigorously assess the robustness of our primary findings, a pre-specified sensitivity analysis was performed. This analysis was designed to evaluate whether the pooled NMA estimates were disproportionately driven by trials with methodological limitations or highly distinct baseline populations. Specifically, we re-executed the NMA models after selectively excluding studies judged to have a ‘Moderate’ or ‘High’ risk of bias (e.g., the RE-LY trial, attributed to its PROBE [Prospective Randomized Open, Blinded End-point] design) and studies enrolling a demonstrably distinct, higher-risk cohort (i.e., the ROCKET-AF trial). Comparing the primary estimates with this robustness set ensures that our central conclusions remain stable and are not unduly influenced by trial-specific biases or patient-level baseline disparities. Finally, the overall certainty of evidence was appraised employing the GRADE framework customized for NMAs. All statistical analyses were programmed and executed in R (packages: *gemtc*, *rjags*, *survival*, *survRM2*, *multinma*).

RESULTS

Study Selection, Characteristics, and Network Geometry

The systematic literature search and screening process are summarized in Fig. 1A. A total of 4 landmark phase III randomized controlled trials (RCTs) and their respective HF substudies—ARISTOTLE, ENGAGE AF-TIMI 48, RE-LY, and ROCKET-AF—met the eligibility criteria and were included in the network meta-analysis.[14–17] These trials yielded a cumulative cohort of 25,762 patients with concomitant AF and HF.

The baseline characteristics of the included HF populations were well-balanced across treatment arms and are detailed in Table 1. The mean age of the participants ranged from 70.0 to 72.0 years, with a predominant male distribution (ranging from 60% to 70%). The clinical profile reflected a high-risk population, with mean CHADS₂ or CHA₂DS₂-VASc scores ranging from 3.3 to 4.8. HF phenotypes varied across trials: the ROCKET-AF HF cohort was exclusively composed of patients with a left ventricular ejection fraction (LVEF) $\leq 40\%$, whereas the other trials included a broader spectrum of HFrEF and HFpEF phenotypes, with LVEF $\leq 40\%$ prevalence ranging from 44% to 48.5%. The median follow-up duration across the trials spanned from 1.8 to 2.8 years.

Table 1
Baseline characteristics of HF populations in the included RCT

Trial	Treatment Arm	N (HF Cohort)	Mean Age (Years)	Male (%)	LVEF $\leq 40\%$ (%)	NYHA III-IV (%)	Mean CHADS ₂ / CHA ₂ DS ₂ -VASc	Median Follow-up (Years)
ARISTOTLE [14]	Apixaban 5 mg vs. VKA	6.422	70.0	70.4	47.0	31.0	3.3	1.8
ENGAGE AF[15]	Edoxaban 60 mg vs. VKA	5.403	71.0	60.5	48.5	26.0	3.3	2.8
RE-LY[16]	Dabigatran* vs. VKA	4.904	72.0	64.0	44.0	30.0	3.3	2.0
ROCKET-AF [17]	Rivaroxaban 20 mg vs. VKA	9.033	72.0	60.0	100.0	20.0	4.8	1.9
*Dabigatran evaluated at 110 mg and 150 mg BID doses.								

The network geometry for the primary outcomes is illustrated in Fig. 1B. The network architecture was centered on VKAs as the common comparator, forming a star-shaped configuration with five distinct DOAC regimens: apixaban (5 mg BID), dabigatran (110 mg and 150 mg BID), edoxaban (60 mg QD), and rivaroxaban (20 mg QD). All comparisons were anchored by robust direct evidence from large-scale HF substudies, ensuring high transitivity across the network.

Reconstructed IPD Validation

The accuracy of the pseudo-IPD reconstruction was rigorously validated by comparing the algorithmically derived survival curves against the original published Kaplan-Meier plots for stroke/systemic embolism and major bleeding. Visual inspection demonstrated near-identical overlay, with the reconstructed curves capturing all major inflection points and censoring patterns of the original trial data. Statistical validation further confirmed the high fidelity of the reconstruction; the total number of events and patients at risk at each landmark interval in the pseudo-IPD dataset deviated by less than 1% from the values reported in the primary publications. This high degree of concordance provided a robust foundation for the subsequent RMST and piecewise Cox analyses.

Primary Outcomes

Stroke/SE

In the aggregate NMA, DOACs demonstrated comparable or superior efficacy to VKA for the prevention of stroke or systemic embolism. Apixaban 5 mg BID ranked highest in the probabilistic hierarchy (SUCRA: 70.3%), followed by dabigatran 150 mg BID (SUCRA: 69.6%). However, indirect comparisons between individual DOACs yielded credible intervals crossing 1.00, indicating no definitive statistical superiority of one DOAC over another within the network.

Translating these relative estimates into absolute temporal gains, the reconstructed IPD RMST analysis at the 2.25-year truncation point revealed significant gains in stroke-free survival for the top-ranking agents. Dabigatran 150 mg BID provided the highest absolute benefit, extending stroke-free survival by 5.74 days (95% CI: 2.18 to 9.31; $P = 0.0016$) compared to VKA. Apixaban 5 mg BID followed with a significant gain of 3.23 days (95% CI: 0.92 to 5.55; $P = 0.0062$). In contrast, while edoxaban 60 mg QD and rivaroxaban 20 mg QD showed positive trends (gains of 2.36 days and 2.07 days, respectively), these differences did not reach statistical significance in the absolute time domain over the 2.25-year follow-up period ($P > 0.05$).

Major Bleeding

The safety profile, defined by the incidence of ISTH major bleeding, revealed distinct heterogeneity among the antithrombotic regimens. Apixaban 5 mg BID demonstrated the most robust safety profile, ranking highest in the network (SUCRA: 73.4%).

The RMST analysis further underscored the superior safety of apixaban in terms of absolute bleeding-free survival. Over 2.25 years of follow-up, apixaban 5 mg BID significantly preserved an additional 8.32 days of bleeding-free survival compared to VKA (95% CI: 5.14 to 11.50; $P < 0.0001$). This represented the most substantial absolute safety benefit across the entire network. Edoxaban 60 mg QD also demonstrated a significant safety gain of 4.47 days (95% CI: 0.55 to 8.39; $P = 0.025$). In the aggregate NMA, apixaban 5 mg BID also demonstrated the most robust relative safety profile against VKA (HR 0.76, 95% CrI 0.52 to 1.13). Conversely, the absolute safety profile of other DOACs in the HF cohort was comparable to VKA, without significant differences in RMST.

Table 2 Unified network meta-analysis league table for primary outcomes						
Apixaban 5mg	Apixaban 5mg Apix_5	Dabigatran 110mg 1.30 (0.69–2.47)	Dabigatran 150mg 0.98 (0.51–1.85)	Edoxaban 60mg 1.15 (0.65–2.03)	Rivaroxaban 20mg 1.06 (0.60–1.88)	VKA 1.31 (0.87–1.98)
Dabigatran 110mg	0.92 (0.51–1.64)	Dabi_110	0.75 (0.37–1.54)	0.87 (0.47–1.65)	0.81 (0.43–1.55)	1.01 (0.61–1.66)
Dabigatran 150mg	0.96 (0.54–1.73)	1.05 (0.57–1.92)	Dabi_150	1.16 (0.62–2.23)	1.08 (0.57–2.07)	1.33 (0.79–2.21)
Edoxaban 60mg	0.93 (0.54–1.60)	1.02 (0.57–1.79)	0.97 (0.54–1.69)	Edox_60	0.92 (0.53–1.64)	1.15 (0.77–1.69)
Rivaroxaban 20mg	0.73 (0.42–1.26)	0.80 (0.45–1.40)	0.76 (0.43–1.33)	0.78 (0.47–1.33)	Riva_20	1.24 (0.82–1.87)
VKA	0.76 (0.52–1.13)	0.83 (0.54–1.27)	0.79 (0.52–1.20)	0.82 (0.57–1.19)	1.04 (0.72–1.50)	VKA
Notes: Estimates are Hazard Ratios (95% Credible Intervals) for Column vs Row. The upper-right triangle displays the comparative efficacy (Stroke/Systemic Embolism), whereas the lower-left triangle presents the comparative safety (Major Bleeding). An HR < 1.00 favors the column-defining treatment.						

Table 3 Absolute treatment effects: gain in event-free days over 2.25 years of follow-up (RMST analysis).			
Outcome / Comparison (vs. VKA)	Est. Gain in Days	95% Confidence Interval	P-value
Stroke or Systemic Embolism			
Apixaban 5 mg BID	+ 3.23	0.92 to 5.55	0.006
Dabigatran 150 mg BID	+ 5.74	2.18 to 9.31	0.001
Edoxaban 60 mg QD	+ 2.36	-0.94 to 5.66	0.161
Rivaroxaban 20 mg QD	+ 2.07	-0.77 to 4.91	0.153
Major Bleeding			
Apixaban 5 mg BID	+ 8.32	5.14 to 11.50	< 0.001
Edoxaban 60 mg QD	+ 4.47	0.55 to 8.39	0.025
All estimates are derived from reconstructed individual patient data (IPD) with a truncation time (τ) of 2.25 years (821 days).			

Secondary Outcome

Intracerebral Hemorrhage

Incidence of ICH was evaluated as a secondary safety endpoint. While point estimates for all DOACs suggested a lower relative hazard compared to VKA, the 95% credible intervals were wide and uniformly crossed unity (e.g., dabigatran 110 mg HR 0.28, 95% CrI 0.05–1.67). This imprecision is likely driven by the inherent sparsity of ICH events within the isolated HF substudies. Consequently, definitive claims of statistical superiority among the antithrombotic regimens could not be established. Although the SUCRA probabilistic hierarchy positioned dabigatran 110 mg (SUCRA: 70.3%) and 150 mg (61.9%) as the most favorable options and VKA as the highest-risk agent (SUCRA: 9.2%), these rank probabilities must be interpreted with caution given the lack of statistically significant credible intervals.

Validation of Proportional Hazards and Temporal Stability

A critical limitation of traditional network meta-analysis is the reliance on the constant hazard ratio assumption, which may obscure time-varying treatment effects, particularly in chronic conditions like HF. To mitigate this, we conducted a landmark analysis using piecewise Cox proportional hazards regression on the reconstructed individual patient data, dichotomizing the follow-up period at the 1-year threshold (early [< 1 year] versus late [> 1 year] hazard).

The interaction analysis revealed that the antithrombotic efficacy and bleeding hazard profiles of all evaluated DOACs remained remarkably stable over time. Specifically, the interaction terms between treatment assignments and the late period (> 1 year) were not statistically significant for any agent (e.g., apixaban, dabigatran, edoxaban, and rivaroxaban all demonstrated 95% confidence intervals for the interaction exponentiated coefficients crossing 1.00). This temporal stability statistically confirms that the protective benefits and safety profiles of the highest-ranking DOACs do not wane during long-term maintenance therapy in the HF cohort.

Subgroup Analysis

HFpEF vs HFrEF

To determine whether the treatment effects of DOACs were modified by the baseline left ventricular ejection fraction, we conducted a subgroup network meta-analysis dichotomized into HF with reduced ejection fraction (HFrEF) and preserved ejection fraction (HFpEF). For the primary efficacy outcome (stroke or systemic embolism), the protective benefits of DOACs compared with VKA were remarkably consistent across both phenotypes. In the HFpEF cohort, apixaban demonstrated an HR of 0.75 (95% CrI 0.45–1.26), while edoxaban yielded an HR of 0.86 (95% CrI 0.56–1.33). Parallel efficacy was observed in the HFrEF cohort (apixaban HR 0.76, 95% CrI 0.48–1.20; edoxaban HR 0.88, 95% CrI 0.59–1.32).

Consistent safety profiles were observed across phenotypes for the primary safety outcome. Interaction testing across these treatment effects revealed no significant effect modification by ejection fraction status (all P-interaction > 0.05). These findings suggest that the comparative efficacy and safety profiles of the evaluated DOACs against VKA remain stable regardless of the underlying HFrEF or HFpEF phenotype.

Network Inconsistency, Sensitivity Analysis & Publication Bias

Because the network was star-shaped without direct head-to-head DOAC trials, closed-loop inconsistency could not be statistically evaluated. The validity of the network therefore relies on the assumption of clinical transitivity, which warrants cautious interpretation given the distinct baseline risk profile observed in the ROCKET-AF cohort.

To ensure the stability of the primary estimates, a pre-specified sensitivity analysis was performed. We restricted the network strictly to low risk-of-bias trials and excluded the ROCKET-AF study to account for baseline clinical heterogeneity. This exclusion did not meaningfully alter the point estimates or the overall treatment hierarchy. For instance, apixaban maintained a consistent effect size against VKA for stroke or systemic embolism (HR 0.76, 95% CrI 0.50–1.14). Overall, the sensitivity analysis yielded results consistent with the primary network, confirming that our estimates were robust and not disproportionately driven by trial-specific biases or outlier populations (**Figure S6**). Furthermore, the model fit remained highly stable; Deviance Information Criterion (DIC) values showed minimal variation between the fixed-effects and random-effects approaches across the stroke/systemic embolism, major bleeding, and intracranial hemorrhage networks.

Finally, formal evaluations of publication bias using comparison-adjusted funnel plots or Egger's regression tests were omitted. Per established methodological guidelines, tests for funnel plot asymmetry are strictly underpowered and prone to spurious findings when a synthesized network comprises fewer than 10 primary studies.

Certainty of Evidence (GRADE Assessment)

The certainty of evidence for each outcome was systematically evaluated using the GRADE framework adapted for network meta-analysis. For the primary efficacy outcome (stroke or systemic embolism), the certainty of evidence for all DOACs versus VKA—including apixaban—was graded as moderate. Although point estimates consistently favored DOAC therapy, a serious concern for imprecision was identified because the 95% network credible intervals (CrIs) crossed the unity threshold (1.00), necessitating a single-level downgrade.

Regarding safety outcomes, the certainty of evidence for major bleeding compared with VKA was similarly graded as moderate across all evaluated DOACs (apixaban, edoxaban, dabigatran, and rivaroxaban). This downgrade was primarily driven by imprecision, as the wide network CrIs encompassed the null effect (1.00), despite favorable clinical trends. Likewise, the certainty of evidence for intracranial hemorrhage was graded as moderate for all agents due to wide credible intervals and resulting statistical imprecision within the network framework.

Overall, while the individual landmark trials exhibited high internal validity, the GRADE assessment reflects moderate certainty for the final network estimates. This highlights that sample size limitations within the specific HF subgroups inherently reduce the statistical precision of the relative treatment effects, though the direction of benefit remains highly consistent with primary trial findings.

DISCUSSION

In this network meta-analysis incorporating reconstructed individual patient data from 25,762 patients with AF and concomitant HF, DOACs demonstrated comparable or superior efficacy to VKAs. Traditional aggregate data analysis yielded wide credible intervals for indirect DOAC comparisons, limiting the assertion of definitive relative superiority among the individual agents. However, translating relative hazards into absolute time metrics via restricted mean survival time (RMST) revealed that apixaban and dabigatran 150 mg provided significant absolute gains in stroke-free survival. Apixaban further yielded the most substantial gain in major bleeding-free days. Piecewise Cox regression confirmed these treatment profiles remained temporally stable beyond one year, and subgroup analyses demonstrated structural consistency across both HFrEF and HFpEF phenotypes.

These findings contextualize and significantly expand upon recent aggregate-data meta-analyses. Prior network syntheses, such as those by Wang et al. and others, have generally evaluated HF patients as a monolithic cohort, concluding a class-wide DOAC superiority over VKA but frequently failing to definitively delineate an individual agent hierarchy. Our results diverge from these previous iterations through two methodological refinements. First, prior analyses likely masked agent-specific nuances, particularly regarding hemorrhagic risk, due to the mathematical constraints of conventional hazard ratios, which inherently obscure temporal variations in drug efficacy and safety.[6–8] By utilizing an algorithm-based IPD reconstruction,[10] we bypassed the proportional hazards assumption and calculated the RMST, a mathematically robust metric providing a more tangible measure of clinical benefit.[9, 18] Second, previous studies suffered from the dilution effect of mixing divergent HF pathophysiologies. By strictly stratifying HFrEF and HFpEF, our analysis demonstrates that the assumption of intra-class DOAC equivalence is flawed. We establish that apixaban's safety profile is not a statistical artifact of cohort pooling but a consistent physiological advantage across the HF continuum.

The coexistence of AF and HF presents intersecting prothrombotic and hemorrhagic vulnerabilities.[2] Systemic endothelial dysfunction and left ventricular stasis necessitate potent stroke prophylaxis, yet HF-associated renal congestion, altered hepatic metabolism, and polypharmacy simultaneously elevate bleeding risks. Our absolute time-to-event data indicate that apixaban optimizes this narrow therapeutic index. Patients with HFpEF, in particular, are typically older and carry a higher burden of profibrotic, microvascular, and renal comorbidities, classically rendering them highly susceptible to major hemorrhagic events.[19] The demonstration that apixaban maintains its bleeding-free survival advantage independent of baseline left ventricular ejection fraction addresses a persistent clinical equipoise, providing a high-certainty rationale for its utilization as a unified first-line agent. While the absolute RMST gains may appear numerically modest (e.g., ~ 3 to 8 event-free days over 2.25 years), their clinical implications in the HF population are substantial. In this demographic, marked by renal fragility and high baseline morbidity, even marginal absolute reductions in major bleeding or stroke translate directly to the prevention of hospitalization and cascading secondary complications.

These findings must be interpreted within the strict context of several methodological limitations. Most critically, the primary data originate entirely from pre-specified subgroups of global AF trials rather than dedicated, prospectively designed HF trials.[14–17] Because these parent studies were not statistically powered to evaluate efficacy or safety endpoints specifically within the HF strata, much less within the highly specific HFrEF and HFpEF phenotypes, several of our indirect comparisons are undeniably underpowered. The wide credible intervals observed in specific treatment nodes reflect a high susceptibility to Type II errors. Consequently, the lack of a dedicated randomized controlled trial for DOACs in HF dictates that our comparative estimates remain hypothesis-generating rather than definitive.

Second, the network geometry is exclusively star-shaped; the absolute lack of direct, head-to-head DOAC trials necessitates a mathematical reliance on transitivity through a common VKA node, an assumption that cannot be statistically verified via closed-loop inconsistency testing. Finally, although the IPD reconstruction algorithm accurately approximates marginal survival and censoring, it fundamentally cannot retrieve patient-level baseline covariates.[10] Continuous multivariable adjustment for precise age, fluctuating creatinine clearance rates, or specific CHA₂DS₂-VASC scores was therefore unfeasible, leaving residual unmeasured confounding unchecked.

CONCLUSION

By translating relative risks into absolute event-free survival, this individual patient data network meta-analysis demonstrates that DOACs provide clinically meaningful advantages over VKAs in patients with AF and HF. While both apixaban and dabigatran 150 mg maximize stroke-free survival, apixaban uniquely delivers the most favorable safety profile by significantly avoiding major bleeding. Importantly, these treatment hierarchies remain consistent across both HFrEF and HFpEF phenotypes. Although indirect comparisons between individual DOACs did not establish definitive statistical superiority, the overall efficacy and safety profile of apixaban suggests that it may represent a preferred first-line anticoagulant option for patients with concomitant AF and HF, particularly in this high-risk population regardless of their baseline ventricular function. Further dedicated head-to-head randomized trials are warranted to validate these findings.

Declarations

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Author Contributions

All authors contributed to the study conception, design, and statistical analysis. Literature screening was independently conducted by two investigators, I Komang Raditya Putra Pratama and Ni Wayan Pradnyadhari Kusumaputri, via a sequential two-stage approach, starting with the initial evaluation of titles and abstracts, followed by a rigorous full-text appraisal of potentially eligible articles. Any inter-reviewer discrepancies regarding study eligibility were resolved through discussion to reach a consensus, or via adjudication by a third, senior cardiovascular methodologist. Data extraction was performed independently by two reviewers, I Komang Raditya Putra Pratama and Ni Wayan Pradnyadhari Kusumaputri, utilizing pre-piloted, standardized electronic extraction forms. Any disagreements during the appraisal process were resolved through consensus, with a third reviewer, I Nyoman Wiryawan, serving as an arbitrator when necessary. The first draft of the manuscript was written by I Komang Raditya Putra Pratama, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Ethics Approval

Not applicable. This study is a systematic review and network meta-analysis based entirely on previously published clinical trials and algorithmically reconstructed aggregate data. It did not involve any direct interventions with human participants or animals; therefore, institutional review board approval was not required.

Consent to Participate

Not applicable. This study utilized data extracted from previously published literature, and no primary individual patient data were collected directly from participants.

Consent to Publish

Not applicable. This manuscript does not contain any individual person's clinical data, identifiable details, images, or videos.

References

1. Carlisle MA, Finkel KV, Saleh M, Sellers MB, et al. Heart Failure and Atrial Fibrillation, Like Fire and Fury. *JACC Heart Fail.* 2019;7(6):447–56.
2. Zannad F, Rossignol P, Iraqi W. Heart failure is an antithrombotic and anticoagulant state. *Circulation.* 2014;129(16):1618–9.
3. Hindricks G, Potpara T, Dagres N, et al. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS). *Eur Heart J.* 2021;42(5):373–498.
4. McDonagh TA, Metra M, Adamo M, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur Heart J.* 2021;42(36):3599–726.
5. Savarese G, Giugliano RP, Rosano GM, et al. Efficacy and Safety of Novel Oral Anticoagulants in Patients With Atrial Fibrillation and Heart Failure: A Meta-Analysis. *JACC Heart Fail.* 2016;4(11):870–80.
6. Caldeira D, Alves M, Costa J, Pinto FJ, Ferreira JJ. Non-vitamin K antagonist oral anticoagulants in patients with atrial fibrillation and heart failure: a systematic review and meta-analysis. *ESC Heart Fail.* 2015;2(3):114–22.
7. Chen C, Cao Y, Zheng Y, et al. Effect and Safety of Direct Oral Anticoagulants vs. Warfarin in Patients with Atrial Fibrillation and Heart Failure: A Systematic Review and Meta-Analysis. *Front Cardiovasc Med.* 2021;8:626685.
8. Wei Y, Royston P. Evaluating the assumption of proportional hazards in survival analysis. *BMC Med Res Methodol.* 2017;17(1):1–14.
9. Uno H, Claggett B, Tian L, et al. Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. *J Clin Oncol.* 2014;32(22):2380–5.
10. Guyot P, Ades AE, Ouwens MJ, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. *BMC Med Res Methodol.* 2012;12:9.
11. Hutton B, Salanti G, Caldwell DM, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Ann Intern Med.* 2015;162(11):777–84.
12. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ.* 2019;366:l4898.
13. Brignardello-Petersen R, Bonner A, Alexander PE, et al. Advances in the GRADE approach to rate the certainty in estimates from a network meta-analysis. *J Clin Epidemiol.* 2018;93:36–44.
14. McMurray JJ, Ezekowitz JA, Lewis BS, et al. Left ventricular systolic dysfunction, heart failure, and the risk of stroke and systemic embolism in patients with atrial fibrillation: insights from the ARISTOTLE trial. *Circ Heart Fail.* 2013;6(3):451–60.
15. Magnani G, Giugliano RP, Ruff CT, et al. Efficacy and safety of edoxaban compared with warfarin in patients with atrial fibrillation and heart failure: insights from ENGAGE AF-TIMI 48. *Eur Heart J.* 2016;37(32):2528–36.
16. Ferreira J, Ezekowitz JA, Connolly SJ, et al. Dabigatran compared with warfarin in patients with atrial fibrillation and symptomatic heart failure: a subgroup analysis of the RE-LY trial. *Eur J Heart Fail.* 2013;15(9):1053–61.
17. van Diepen S, Hellkamp AS, Patel MR, et al. Efficacy and safety of rivaroxaban in patients with heart failure and intact or reduced ejection fraction: insights from the ROCKET AF trial. *Eur J Heart Fail.* 2013;15(12):1408–15.
18. Royston P, Parmar MK. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Med Res Methodol.* 2013;13:152.
19. Redfield MM, Borlaug BA. Heart Failure with Preserved Ejection Fraction: A Review. *JAMA.* 2023;329(10):827–38.

Figures

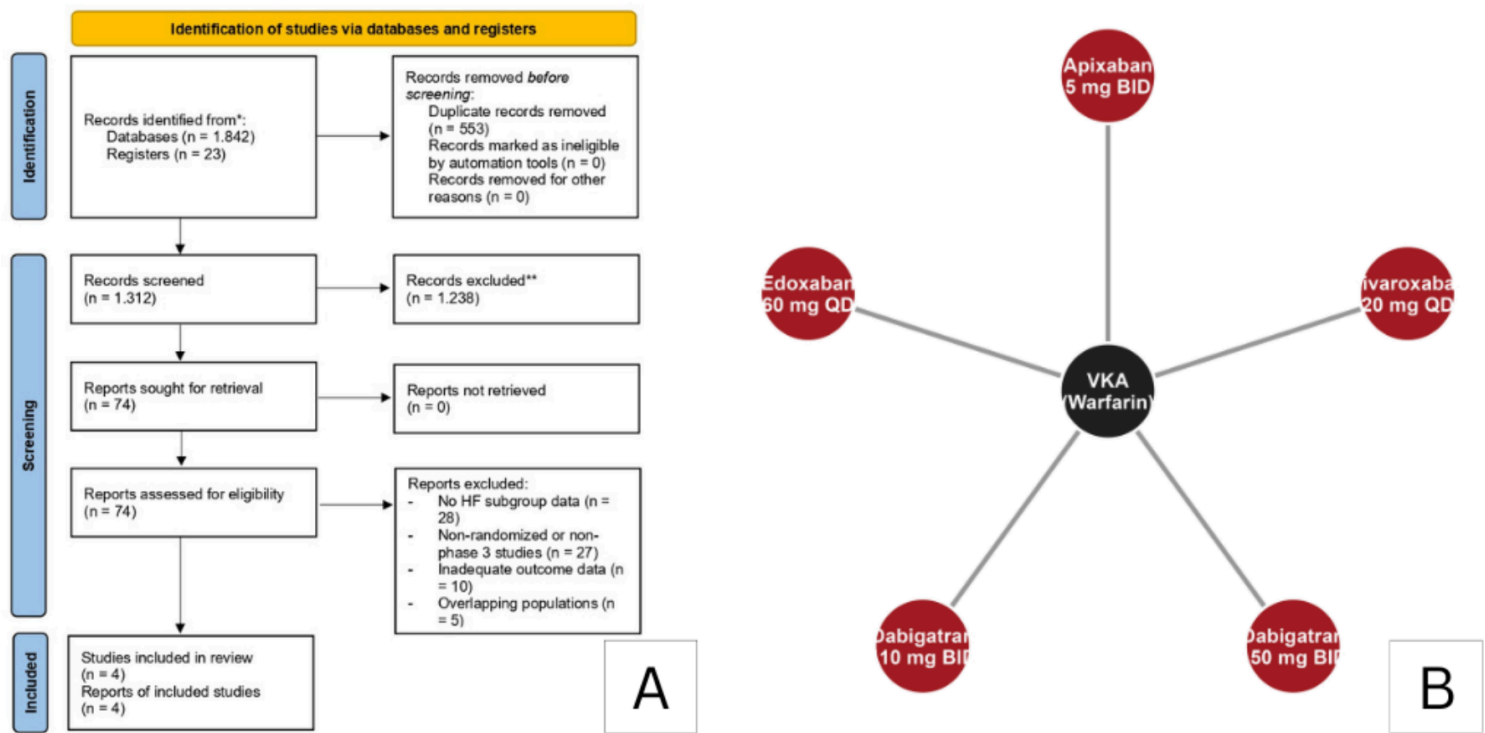


Figure 1
(A) PRISMA 2020 flow diagram of study selection; (B) Network geometry of treatment comparisons

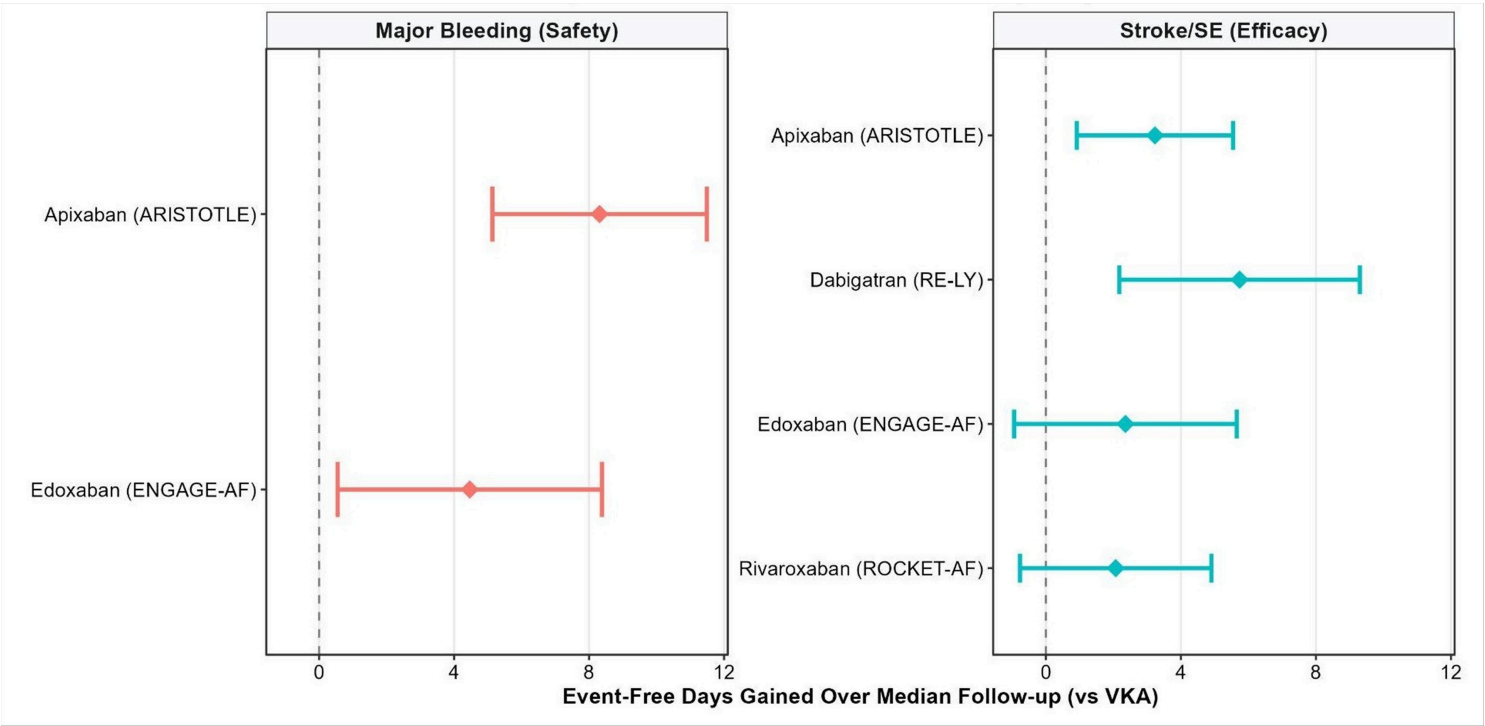


Figure 2
Restricted mean survival time (RMST) and Kaplan-Meier overlay

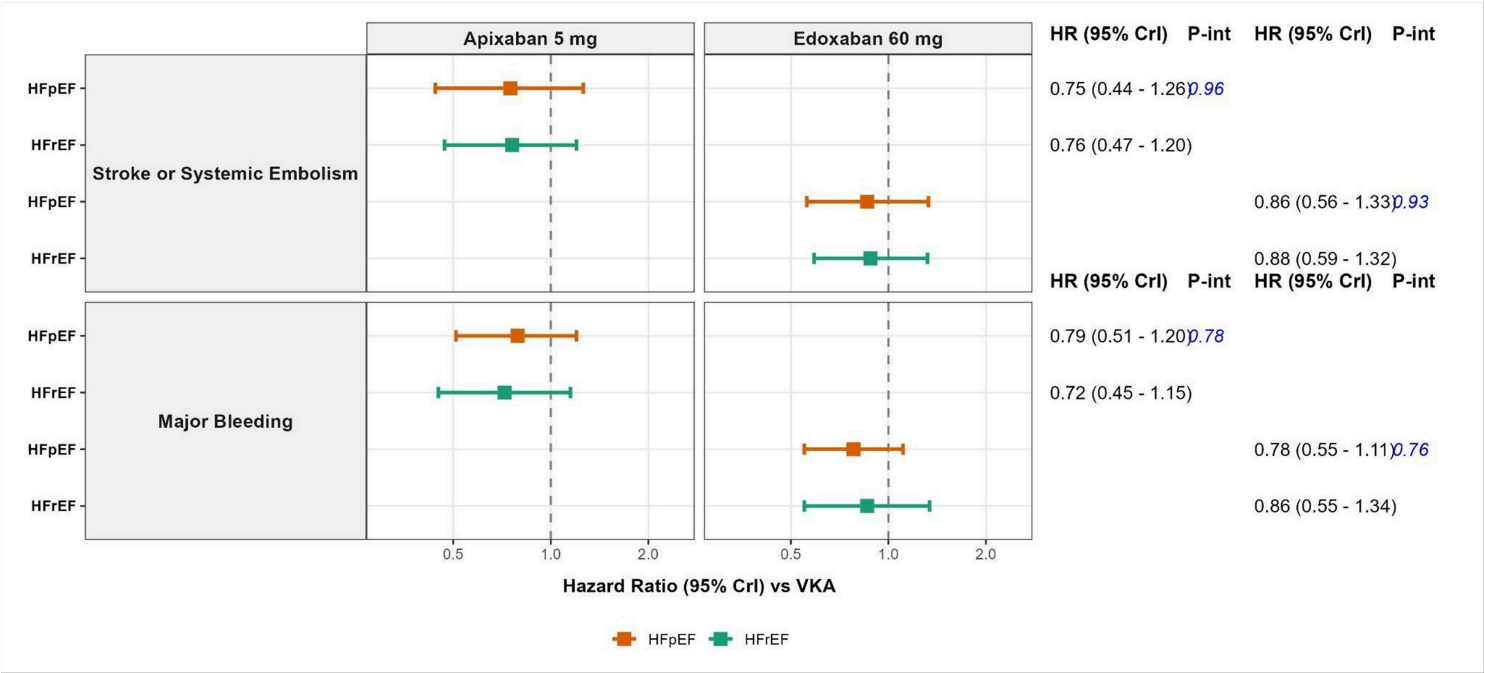


Figure 3
HFrEF vs HFpEF Subgroup Forest Plot

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterialComparativeEfficacyandSafetyofDirectOralAnticoagulantsinAtrialFibrillationandHeartFailureABayesianNetworkMetaAnalysisofRe](#)