ORIGINAL REPORT

Rivaroxaban was found to be noninferior to warfarin in routine clinical care: A retrospective noninferiority cohort replication study

Turki A. Althunian[1] | Anthonius de Boer[1,2] | Rolf H. H. Groenwold[1,3] |
Katrien O. Rengerink[3] | Patrick C. Souverein[1] | Olaf H. Klungel[1,3]

[1]Division of Pharmacoepidemiology and
Clinical Pharmacology, Utrecht Institute for
Pharmaceutical Sciences, Utrecht University,
Utrecht, The Netherlands

[2]Chairman of the Dutch Medicines Evaluation
Board, Dutch Medicines Evaluation Board,
Utrecht, The Netherlands

[3]Julius Center for Health Sciences and Primary
Care, University Medical Center Utrecht,
Utrecht, The Netherlands

Correspondence
Olaf H. Klungel, Division of
Pharmacoepidemiology and Clinical
Pharmacology, Utrecht University, P.O. Box
80082, 3408 TB Utrecht, The Netherlands.
Email: o.h.klungel@uu.nl

Funding information
The Saudi Food and Drug Authority, Grant/
Award Number: None

Abstract

Purpose: To compare the effectiveness and safety of a drug in daily practice with the outcomes of a target non-inferiority trial by rigorously mimickingin an observational study the trial's design features.

Methods: This cohort study was conducted using the British Clinical Practice Research Datalink (CPRD) to emulate the 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) trial. Patients with atrial fibrillation who were newly prescribed (>=12 months of no use) either rivaroxaban or warfarinfrom October 2008 to December 2017 were included. Non-inferiority of rivaroxaban to warfarin in the prevention of stroke or systemic embolism was assessed in different analysis populations (intention-to-treat [ITT], per-protocol [PP], and as-treated populations) using a hazardratio (HR) of 1.46 as the non-inferiority margin. Major bleeding (safety outcome) was also assessed and compared to that of the target trial. All outcomes were analyzed using Cox-proportional hazard analyses.

Results: We included 25,473 incident users of rivaroxaban (n=4,008) or warfarin (n=21,465). Similar to the trial, non-inferiority in the primary out come was demonstrated in all three analysis populations: HR=1.04 (95%CI 0.84 to 1.30) (ITT), HR=0.98 (95%CI 0.70 to 1.38) (PP), and HR=1.11 (95%CI 0.86 to 1.42) (as-treated). Risk of major bleeding was also similar to the target trial.

Conclusion: The results of this study provide supportive evidence to the effectiveness of rivaroxaban and adds knowledge on the usefulness of emulating a non-inferiority trial to assess drug effectiveness.

KEYWORDS

effectivness, electronic health care records, noninferiority, observational studies, pharmacoepidemiology, real-world evidence, methodology

1 | INTRODUCTION

It is well accepted that running a randomized double blind controlled trial (RCT) is the best approach to evaluate drug effects in an unbiased way.[1-3] An important question is whether the observed drug effects from RCTs can be extrapolated to routine daily practice.[4,5] Effects might be different because patients who are treated in clinical practice (the "real world") may not be eligible for the RCTs, due to strict in and exclusion criteria. For example, they may receive more concomitant drugs and have more comorbidities than patients who are enrolled in trials. Also, treatment adherence may be lower in daily practice compared to the adherence in trials.[6-10]

Evaluating drug effectiveness and safety in daily practice is hampered by the observational design of studies vulnerable to information, selection, and confounding bias.[4-10]

Many observational studies have been performed with the objective to try to find similar drug effects as have been observed in RCTs. Some studies found similar effects, others did not.[6-17]

An important step to try to evaluate whether drug effects from RCTs are similar in daily practice is to try to mimic the target RCT as much as possible.[11-17] Obvious measures such as randomization and blinding of patients and health care practitioners for the studied drugs are impossible in observational studies. However, by applying the same selection criteria in the observational study as were used in the target trial, by making the same comparison between drug dosages, and by considering similar follow-up time, obvious causes that might lead to different drug effects can be prevented.[6-17]

Such observational studies mimicking target RCTs have been performed. Interestingly in almost all cases such comparisons were done for superiority trials and only once for a noninferiority trial.[11-17] This last comparison was hampered by the absence of noninferiority comparison between study arms in the observational study.[17]

Therefore, the aim of our study was to compare the effectiveness and safety of a drug in daily practice with RCT outcomes by rigorously mimicking in an observational study the design features of the selected target noninferiority RCT.

2 | METHODS

2.1 | Sources of data

This retrospective cohort study was conducted using the Clinical Practice Research Datalink (CPRD) to mimic the design of the 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) noninferiority trial.[18,19] In the ROCKET AF efficacy of rivaroxaban, an oral factor Xa inhibitor, in preventing stroke or systemic embolism (primary efficacy outcome) in patients with atrial fibrillation (AF) was evaluated in almost 14 000 patients.[18] Noninferiority of rivaroxaban to warfarin was demonstrated in the primary outcome: hazard ratio (HR) of 0.79 (95% confidence interval [CI] 0.66-0.96).

> **KEY POINTS**
>
> - The ROCKET AF trial was emulated in an observational study using electronic health care records.
> - Noninferiority assessment has not been performed in previous emulation studies.
> - Noninferiority of rivaroxaban to warfarin in the prevention of stroke or systemic embolism in patients with atrial fibrillation was confirmed in this study.
> - Other findings were not replicated in this study which could be attributed to several reasons related to precision, (unmeasured) effect modification and/or confounding.
> - Further assessment is needed of observational studies that emulate a target trial to build more body of evidence towards the design challenges and the usefulness of these studies in effectiveness research.

CPRD covers de-identified data of over 10 million patients currently registered in more than 600 primary care practices in the United Kingdom.[19,20] The CPRD is one of the largest primary care databases in the world with longitudinal medical records, and CPRD patients are considerably representative of the UK population with respect to sex, age, and [19,20] CPRD has an acceptable level of data completeness and validity as concluded in previous validation studies.[21-24]

By linkage to Hospital Episode Statistics (HES) Admitted Care, data on study outcomes and admission were used. An Independent Scientific Advisory Committee (ISAC) protocol was submitted and approval of the study was granted.[25]

2.2 | Study cohort

Patients who were diagnosed with nonvalvular AF and were prescribed either rivaroxaban or warfarin from September 30, 2008 (the date of granting marketing authorization of rivaroxaban by the EMA) were included and followed to the end of the study period (December 31, 2017).[26] All patients were incident users (ie, those who did not use an anticoagulation therapy 12 months before the start of rivaroxaban or warfarin). The inclusion criteria were similar to those of the ROCKET AF trial which included patients with nonvalvular AF at moderate-to-high risk for stroke. This level of risk was indicated by a history of stroke, transient ischemic attack, or systemic emolism or at least two of the following risk factors: heart failure; hypertension; age ≥75 years, type I or II diabetes mellitus.[18] The following exclusion criteria could not be applied in this study: the use of the combined treatment of aspirin and thienopyridines within 5 days before cohort entry date (defined in section 2.3), the use of intravenous antiplatelets

within 5 days before the cohort entry date, and the use of fibrinolytics within 5 days before the cohort entry date.

2.3 | Medication exposure and analysis populations

The cohort entry date was defined as the date of the first prescription (as order, not dispensing) of either rivaroxaban (test drug) or warfarin (active comparator). In the follow-up period, current use of rivaroxaban, warfarin, or other anticoagulants/antiplatelet drug windows were created first for each patient based on prescription refills and permissible gaps using the British National Formulary codes, then periods of nonexposure were filled in between afterwards. A 90-day period was chosen as the permissible gap since some patients in the British primary care are given a prescription to cover a 90-day period. Similar to the target trial, the analysis population was split into three populations:

- Per-protocol (PP) population: this population included patients who remained adherent to the first allocated treatment, that is, did not exceed the permissible gap without a prescription of the first allocated treatment. Patients were censored at the end of the study period, when switching/stopping the allocated treatment, when lost to follow-up, or upon death.
- Intention-to-treat (ITT) population: this population included all patients, regardless of their adherence to their first treatment allocation. Patients were censored at the end of the study period, when lost to follow-up, or upon death.
- As-treated (time-varying) population: this population was constructed to take into account the time varying nature of the exposure to both study drugs in the routine clinical care in addition to time-varying confounders (section 2.5). Exposure and nonexposure windows in this population were additionally split into smaller windows to allow for time-varying confounders. Patients were censored at the end of the study period, when lost to follow-up, or upon death.

2.4 | Study outcomes

As in the target trial, the primary outcome of this study was the composite endpoint of stroke or systemic embolism (defined as arterial embolism and thrombosis). The primary outcome was assessed in the three analysis populations. The secondary outcomes were the composite endpoint of stroke, systemic embolism or vascular death (efficacy) and major bleeding (safety), both assessed in the ITT population (similar to the target trial). Vascular death was defined in the target trial as any death attributed to vascular causes (eg, due to myocardial infarction, stroke, heart failure, etc). Major bleeding was defined in the target trial as postoperative bleeding occurring after the first postoperative study dose; fatal bleeding; bleeding at a critical site or in a critical organ; overt bleeding warranting treatment cessation; bleeding

leading to re-operation; clinically overt bleeding associated with a fall in hemoglobin of 2 g/dL or more or leading to a transfusion of two or more units of blood. In this study, major bleeding was defined as any bleeding at a critical site or organ as follows: gastrointestinal, intracranial, intraspinal, intraocular, pericardial, intra-articular, intramuscular with compartment syndrome, or retroperitoneal. Table S1 shows the ICD-10 (International Statistical Classification of Diseases and Related Health Problems 10th Revision) codes that were utilized to identify the primary and secondary outcomes. Diagnostic accuracy of major bleeding and death ICD-10 algorithms were assessed in a couple of validation studies with mixed results.[27,28] Studies that assessed the validity of ICD-10 algorithms of other outcomes were not identified in the published literature.

2.5 | Confounding adjustment

The analysis of the primary and secondary efficacy outcomes was adjusted for the following potentially confounding variables: age, sex, blood pressure (systolic and diastolic), body mass index (BMI), smoking status, alcohol consumption, CHA_2DS_2-VASc score,[29] coexisting conditions (previous stroke, systemic embolism, transient ischaemic attack, congestive heart failure, hypertension, diabetes mellitus, previous myocardial infarction, peripheral vascular disease, chronic obstructive pulmonary disease), kidney functions (kidney functions in CPRD are recorded under one of four categories, in the target trial, kidney functions were shown with the absolute values of creatinine clearance), previous use (12 months prior the index date) of warfarin/oral factor Xa inhibitor/aspirin, statins, calcium channel blockers, angiotensin converting enzyme inhibitors, angiotensin II blockers, diuretics, beta-blockers, centrally acting antihypertensive drugs, alpha-adrenoceptor blocking antihypertensive drugs, antipsychotics, selective serotonin reuptake inhibitors, nonsteroidal anti-inflammatory drugs, antiarrhythmic drugs, nitrates, and antidiabetics.

In the addition to the previously mentioned confounders (except for CHA_2DS_2-VASc score), the following confounders were adjusted for in the analysis of the risk of major bleeding: gastritis, oesophagitis, history of bleeding, liver functions, and the use of proton pump inhibitors/histamine 2 receptor antagonists. Potential confounders were selected, because they were reported in the literature as risk factors for the outcomes and we assumed that none of the selected variables would classify as an instrumental variable.

In the analysis of the primary outcome in the ITT and PP populations, and the secondary efficacy and safety outcomes in the former (ie, time-fixed analysis), all confounders were assessed at cohort entry (the use of confounding prescription drugs was assessed 6 months prior the index date except for the use of proton pump inhibitors/histamine 2 receptor antagonists which was assessed 3 months before the index date). In the time-varying analysis of the primary outcome in the as-treated population, we distinguished between different types of confounders: confounders that are time-varying and possibly have a direct (ie, instantaneous) effect on the outcome (type 1); confounders that are time-varying, yet

have a chronic character where the effect on the outcome may be delayed (type 2); confounders that are time-varying, yet can only change in one direction (type 3); and confounders that are not time-varying (type 4). The adjustment of these confounders were as follows:

- Type 1: each time the status of these confounders changes, the information in the datasets was updated. Hence, confounder information was assessed at the moment the exposure changes (start/stop/switch) as well as when then confounder status changes (start/stop/switch).
- Type 2: the exact timing of changes in the confounder status is probably less important. Therefore, information about the confounder status was updated at the moment (a) the exposure changes (start/stop/switch), and (b) if exposure remained constant every 6 months.
- Type 3: these concern chronic conditions that, once diagnosed, do not disappear anymore. Furthermore, the exact timing of changes in the confounder status is probably less important. Therefore, information about the confounder status was updated at the moment of diagnosis and the status remained the same afterwards.
- Type 4: these confounders are measured at cohort entry.

2.6 | Statistical analysis

The primary hypothesis of the study was that rivaroxaban is noninferior to warfarin in the reduction of the risk of stroke or systemic embolism (the primary outcome). This hypothesis was tested in the three analysis populations using Cox proportional hazards regression models. Because of the relatively large sample size, no differences were expected between confounding adjustment and other methods to correct for confounding (eg, propensity score matching).[30-32] The HRs of the primary outcome in these populations were estimated after adjusting for all confounders. The proportional hazards assumption was checked using scaled Schoenfeld residuals (graphical and statistical test diagnostics), and the right form of the continuous confounders was checked using Martingale residuals. Noninferiority of rivaroxaban to warfarin was considered to be demonstrated if the 95% CI of each HR lies entirely below the noninferiority margin (HR: 1.46). This margin was used in the ROCKET AF trial, which was defined based on the limit of the 95% CI of the risk ratio of warfarin vs placebo in the reduction of stroke, or systemic embolism that was pooled from six placebo-controlled trials in patients with AF (ie, noninferiority was analyzed using the fixed-margin method).[18,33,34] To assess the primary outcome, a minimum of 2300 patients in each group will provide a power of at least 80% to demonstraste noninferiority of rivaroxaban to warfarin. These calculations were based on the noninferiority margin HR 1.46, one-sided significance level of 0.025, a mean follow-up time of 2 years, and an event rate of the primary outcome of 2.3% per patient-years in both study groups with an anticipated HR of 1. The HRs of the secondary outcomes were also

estimated using Cox proportional-hazard regression models after adjusting for all confounders. The HRs of the primary and secondary outcomes were compared to those of the ROCKET AF trial.

The success of replicating the noninferiority findings of the target trial was assessed using the regulatory agreement approach that was described by Franklin et al paper.[35] Agreement is achieved if the 95% CIs of the primary outcome comparison in the three analyses were below the noninferiority margin. However, and given the anticipated smaller number of rivaroxaban users in CPRD, the 95% CIs of our study might be wider compared with those of the target trial and even if noninferiority was demonstrated, a complete statistical replication might not be achieved (ie, superiority of rivaroxaban vs warfarin in the PP and as-treated populations that was demonstrated in the target trial). Therefore, an additional metric, the estimate agreement approach, was used which requires the effect estimate of each analysis population in our study to be within the 95% CI of the corresponding estimate in the target trial.[35] The probability of estimate agreement is a function of the ratio between the variance of the estimate in our study to that of the target trial assuming no bias in the former estimate. If equality is achieved (ie, ratio = 1), the probability of containing the effect estimate for each analysis population in our study within the 95% CI of its corresponding effect estimate in the target trial is 83%. The higher the ratio, the higher the probability of achieving agreement.[35]

A proportion of values were missing for some confounders (BMI = 71.3%, systolic and diastolic blood pressure = 19.1%, kidney functions = 34.0%, smoking status = 39%). Therefore, sensitivity analyses were conducted to assess the impact of missingness on all study results (these variables were excluded in main Cox regression models). The missing values of the confounders were assumed to be missing at random and were multiply imputed by means of multiple imputation by chained equations to create five imputed datasets: the continuous confounders were imputed using predictive mean matching, the binary confounders were imputed using logistic regression, and the categorical confounders were imputed using polytomous regression. The estimated coefficients and variances from the imputed datasets were pooled using Rubin's rule.[36,37]

All efficacy outcomes were planned to be compared in patients who met the selection criteria of the target trial vs those who would have been excluded (mainly due to safety reasons). The exclusion criteria of the target trial were applied to determine patients who are excluded. However, and due to the missing kidney functions in one-third of the included patients in this study, this comparison was performed in the sensitivity analyses after imputing missing values. A variable that classified those who met and those who did not meet the exclusion criteria was included in the Cox regression models of the sensitivity analyses, and an interaction term with the treatment variable was added to assess the effect of rivaroxaban vs warfarin in those two patient categories.

All statistical analyses were conducted using RStudio Version 1.1.383. Multiple imputation (sensitivity analysis) was performed using the "mice" package in RStudio.[38]

TABLE 1 Baseline characteristics of the intention-to-treat population

Baseline characteristic	Rivaroxaban N = 4008	Warfarin N = 21 465	SMD[a]
Age			
Mean	75.1	74.4	0.07
SD[b]	11.2	10.3	
Gender no. (%)			
Male	2200 (54.9)	12 107 (56.4)	0.03
Female	1808 (45.1)	9358 (43.6)	
BMI	29.0	29.1	0.03
Mean	6.1	6.3	
SD			
Systolic BP			
Mean	132.4	132.8	0.02
SD	17.5	17.7	
Diastolic BP			
Mean	76.6	77.1	0.04
SD	11.0	11.2	
Patients met exclusion criteria; no (%)	143 (3.6)	1317 (6.1)	0.12
Previous use of anticoagulants (vitamin K or DOAC)/aspirin	2296 (57.3)	14 392 (67.0)	0.20
History of stroke, transient ischemic attack, or systemic embolism; no (%)	337 (8.4)	1884 (8.8)	0.01
CHA$_2$DS$_2$-VASc risk score; no (%)			
0	165 (4.1)	924 (4.3)	
1	434 (10.8)	2043 (9.5)	
2	710 (17.7)	3792 (17.7)	0.05
3	1009 (25.2)	5240 (24.4)	
4	952 (23.8)	5307 (24.7)	
5	480 (12.0)	2675 (12.5)	
≥6	258 (6.4)	1484 (6.9)	
Hypertension; no (%)	2477 (61.8)	13 583 (63.3)	0.03
Ischemic heart disease; no (%)	818 (20.4)	5224 (24.3)	0.09
Peripheral arterial disease; no (%)	151 (3.8)	907 (4.2)	0.02
Deep vein thrombosis/pulmonary embolism; no (%)	98 (2.4)	752 (3.5)	0.06
Congestive heart failure; no (%)	417 (10.4)	2786 (13.0)	0.08
Major bleeding; no (%)	688 (17.2)	3253 (15.2)	0.05
Diabetes mellitus; no (%)	741 (18.5)	3756 (17.5)	0.03
Chronic obstructive pulmonary disease; no (%)	434 (10.8)	2224 (10.4)	0.02
Kidney functions (reduced creatinine clearance); no (%)			
Normal	2029 (77.2)	10 866 (76.6)	
Mildly impaired (80-50 mL/min)	395 (15.0)	1924 (13.6)	0.09
Moderately impaired (50-30 mL/min)	183 (7.0)	1182 (8.3)	
Severely impaired (<30 mL/min)	22 (0.8)	212 (1.5)	
Smoking history; no (%)			
Never smokers	1452 (36.4)	8050 (37.6)	
Current smokers	305 (11.1)	2185 (10.2)	0.03
Former smokers	2093 (52.5)	11 152 (52.1)	

(Continues)

TABLE 1 (Continued)

Baseline characteristic	Rivaroxaban N = 4008	Warfarin N = 21 465	SMD[a]
Liver functions (elevated liver enzymes); no (%)[c]			
Normal liver functions	3949 (98.5)	21 236 (98.9)	
Mildly elevated (<3 ULN[d])	57 (1.4)	216 (1.0)	0.04
Cancer. no (%)	1698 (42.4)	8121(37.8)	0.09
GERD/Gastritis; no (%)	993 (24.8)	4836 (22.5)	0.05
Anemia; no (%)	58 (1.4)	266 (1.2)	0.02
Aspirin; no (%)	1057 (26.4)	7569 (35.3)	0.19
Antiarrhythmic agents; no (%)	644 (16.1)	4876 (22.7)	0.17
NSAIDs; no (%)	191 (4.8)	1558 (7.3)	0.11
Antiplatelets; no (%)	1874 (46.8)	12 670 (59.0)	0.3
SSRIs. no (%)	313 (7.8)	1477 (6.9)	0.04
Antidiabetics; no (%)	524 (13.1)	2749 (12.8)	0.01
Statins; no (%)	1925 (48.0)	10 594 (49.4)	0.03
Calcium channel blockers; no (%)	1391 (34.7)	7776 (36.2)	0.03
ACEIs/A$_2$RBs; no (%)	1911 (47.7)	11 480 (53.5)	0.12
Diuretics; no (%)	1337 (33.4)	9173 (42.7)	0.19
Beta blockers; no (%)	1744 (43.5)	10 914 (50.8)	0.15
Centrally-acting antihypertensive agents; no (%)	35 (0.9)	210 (1.0)	0.01
Alpha blockers; no (%)	541 (13.5)	3072 (14.3)	0.02
Antipsychotics; no (%)	112 (2.8)	649 (3.0)	0.01
Nitrates; no (%)	158 (8.7)	1034 (11.4)	0.09
Proton-pump inhibitors/Histamine-2 receptor antagonists; no (%)	1490 (37.2)	7489 (34.9)	0.05

[a]Standardized mean difference.
[b]Standard deviation.
[c]Data on patients with severely elevated liver functions (>3 ULN) are not provided in this table since CPRD policy precludes us from revealing data about small proportions for confidentiality purposes.
[d]Upper limit of normal.

TABLE 2 Results of the analysis of the primary outcome in the observational study compared with the ROCKET AF trial

Arms	Observational study			ROCKET AF trial		
	ITT	PP	As-treated	ITT	PP	As-treated
Rivaroxaban						
Number of events	92	40	72	269	188	189
Event rate: number/100 person-years	0.90	1.01	1.19	2.10	1.70	1.70
Warfarin						
Number of events	987	254	462	306	241	243
Event rate: number/100 person-years	0.81	0.98	1.01	2.40	2.20	1.01
Hazard ratio (95% CI) Original analysis	1.04 (0.84-1.30)	0.98 (0.70-1.38)	1.11 (0.86-1.42)	0.88 (0.75-1.03)	0.79 (0.66-0.96)	0.79 (0.65-0.95)

TABLE 3 Results of the analysis of the secondary outcomes (efficacy and safety) in the observational study compared with the ROCKET AF trial

	Observational study		ROCKET AF trial	
Outcome	Riivaroxaban	Warfarin	Rivaroxaban	Warfarin
Composite endpoint of stroke, systemic embolism, or vascular death				
Number of events	274	3123	346	410
Event rate: number/100 person-years	5.03	4.76	3.11	3.63
Hazard ratio (95% CI)	1.18 (1.03-1.34)	Reference	0.86 (0.74-0.99)	Reference
Major bleeding				
Number of events	294	2849	395	386
Event rate: number/100 person-years	5.68	4.72	3.60	3.40
Hazard ratio (95% CI)	1.07 (0.95-1.21)	Reference	1.04 (0.90-1.20)	Reference

3 | RESULTS

3.1 | Primary analyses

We included 25 473 nonvalvular AF patients who were incident users of either rivaroxaban or warfarin (Table 1). The median follow-up time of rivaroxaban was lower compared with warfarin (1.14 vs 2.61 years, respectively). These follow-up times are different in comparison with those of the target trial (1.9 years in both groups). The median follow-up time in the warfarin group in this study was higher as expected due to longer study period. On the other hand, the median follow-up time in the rivaroxaban was smaller compared with that of the target trial which might be attributed to the late adoption of the rivaroxaban treatment in British practice. The baseline characteristics and the differences in the distribution of all confounders are provided in Table 1.

The unadjusted incidence rates of the primary outcome in the three analysis populations are shown in Table 2 which were lower compared with those of the target trial. Similar to the target trial, noninferiority of rivaroxaban to warfarin was demonstrated, based on the regulatory agreement approach, in all three analyses: HR = 1.04 (95% CI 0.84-1.30) in ITT population, HR = 0.98 (95% CI 0.70-1.38) in the PP population, and HR = 1.11 (0.86-1.42) in the as-treated population. However, the superiority findings in the PP and as-treated analyses of the target trial were not replicated, and the the effect estimates for the ITT (HR = 1.04) and PP (HR = 0.98) populations in our study were outside the 95% CI of the corresponding effect estimates in the target trial (ITT: 0.75-1.03, PP: 0.66-0.96). The probabilities of estimate agreement were 60%, 40%, and 67% in the ITT, PP, and as-treated populations; respectively.

Opposite to the target trial, the incidence rates (unadjusted) of the composite endpoint of stroke, systemic embolism, or vascular death of rivaroxaban and warfarin were higher (Table 3). Addtionally, the hazard of this composite endpoint was higher in rivaroxaban: HR = 1.18 (95% CI 1.03-1.34). In the target trial, the risk was lower in the rivaroxaban group (HR = 0.86 [95% CI 0.74-0.99]). Although the incidence rates of major bleeding were higher in both groups in comparison with the rates in the target trial (Table 3), the HR and the 95% of rivaroxaban vs warfarin were similar to those of the target trial.

3.2 | Sensitivity analyses

The results of the sensitivity analysis of the primary outcome were consistent with those of the original analyses in those who did not meet the exclusion criteria of the target trial (ITT population: HR 1.03 [95% CI 0.82-1.30], PP population: 0.96 [95% CI 0.67-1.38], as-treated: HR 1.07 [95% CI 0.82-1.41]). On the other hand, noninferiority of rivaroxaban to warfarin in the primary outcome was not shown in those who met the exclusion criteria (ITT population: HR 1.20 [95% CI 0.52-2.75], PP population: 1.20 [95% CI 0.29-4.92], as-treated: HR 1.54 [95% CI 0.63-3.79]). This inconsistency was also observed in the sensitivity analysis of secondary efficacy outcome (the composite endpoint of stroke, systemic embolism, or vascular death). In those who did not meet the exclusion criteria, warfarin was superior to rivaroxaban in the prevention of this outcome (similar to the original analysis). In those who did meet the exclusion criteria, the result was different (HR 1.18 [95% CI 0.71-1.96]). However, the 95% CIs in the HRs of those who did not meet the eligibility criteria were very wide. This is reflective of the number of those who met the exclusion criteria compared with those who did not meet these criteria (1460 vs 15 750, respectively). Finally, the result of the sensitivity analysis of major bleeding was consistent with that of the original analysis (HR 1.04 [95% CI 0.91-1.18]).

4 | DISCUSSION

Noninferiority of rivaroxaban to warfarin in the prevention of stroke or systemic embolism was demonstrated in routine clinical care in three analysis populations (similar to the ROCKET AF trial). These findings are supportive for the effectiveness of rivaroxaban; however,

and as a result of the wider CIs in our study, the superiority findings in the PP and as-treated populations found in the target trial were not replicated. The latter is reflected in the probabilities of estimate agreement (in particularly that of the PP population comparison which was <50% of the equal variance probability [ie, 83%]). Additionally, a precise comparison of the results with a broader patient population was not possible due to the small number of patients who met the exclusion criteria of the target trial.

Similar to previous replication studies, we used an active comparator that has a similar probability of being prescribed for the indication to that of the test drug (both are first-line treatment with no preference for one over another according to the National Institute for Health and Care Excellence [NICE] guideline for the management of AF) and included incident users.[13-17,39] That helped in achieving an equal distribution of CHA2DS2-VASc between both groups. We also applied the selection criteria of the target trial and tried to assess the comparison between rivaroxaban and warfarin, with regard to the efficacy outcomes, in those who did not meet vs those who met the target trial's exclusion criteria. However, we found that most patients in our study were from the former category, which affected the precision of the analysis in patients who met the exclusion criteria (the 95% CIs of the HR of the comparison of rivaroxaban with warfarin in all efficacy outcomes were very wide among those who met the exclusion criteria). The results of analyzing the safety outcome were consistent with the target trial; however, the findings of analyzing the secondary efficacy outcome were opposite to those of the target trial (more stroke, systemic embolism, or vascular death in the rivaroxaban treatment group compared to warfarin). This could be attributed to the distribution of (unmeasured) effect modifiers that might be different to that in the target trial. It might also be attributed to unmeasured confounding.

The proportion of patients with previous or coexisting vascular conditions in our study is smaller compared with that of the target trial participants. The patients in our study have smaller percentage of previous strokes, systemic embolisms or transient ischemic attacks compared with patients in the target trial (8.6% vs 54.8%, respectively). Similarly, the percentage of patients with congestive heart failure in our study is very low (12.0% vs 62.5%). This might have contributed to having lower rates of the primary outcome in both groups of our study compared with those of the target trial.

Our study included design features (eg, incident-user design, the use of active comparator, the use of a time-varying [as-treated] population, mimicking the eligibility criteria and outcome definitions of the target trial, etc) that aim at preventing confounding and immortal time bias, and direct the attribution of the differences in the findings between our study and the target trial to reasons other than having broader patients populations. However, the number of rivaroxaban users in our study was lower by almost 3000 patients compared with that of the target trial which may explain our failure to replicate the superiority findings in the PP and as-treated populations (it is important to notice that rivaroxaban was not claimed to be superior to warfarin since superiority was not achieved in the ITT population). Additionally, the target population of the trial is mostly treated in real-world settings (ie, the proportion of patients who did not meet the exclusion criteria was small). This affected the precision of the analysis in those who did not meet the exclusion criteria. Another limitation in our study was that the accuracy and completeness of recording stroke, systemic embolism, and vascular death in HES (using ICD-10) are not well assessed in the literature, and this may not provide us with high level of assurance against the risk of misclassification. Finally, for some confounders values were missing (proportion of missing values ranged from 0.39% to 71.3%); however, the results of the analyses using multiple imputed datasets were consistent with those of the analyses of the complete records.

In concordance with the ROCKET AF trial, noninferiority of rivaroxaban to warfarin in the prevention of stroke or systemic in patients with AF was demonstrated in routine clinical care using EHRs in a UK population. However, a future study that allows for larger number of rivaroxaban users and for a comparison of the results with different subgroups would provide more data on the effectiveness of rivaroxaban, the usefulness of EHRs in effectiveness research, and the applicability of noninferiority analysis in observational settings.

ETHICS STATEMENT

The study design was reviewed and approved by ISAC (ISAC protocol number 18_123).

CONFLICT OF INTEREST

All authors have completed and submitted the International Committee of Medical Journal Editors (ICMJE) Form for Disclosure of Potential Conflicts of Interest; and declare no support from any organization for the submitted work.

AUTHOR CONTRIBUTIONS

T.A., A.dB., P.S., K.R., R.G., and O.K. study concept and design. T.A., A.dB., P.S., K.R., R.G., and O.K. acquisition, analysis, or interpretation of data. T.A. drafting of the manuscript for important intellectual content. A.dB., P.S., K.R., R.G., and O.K. critical revision of the manuscript for important intellectual content. T.A. and R.G. statistical analysis. A.dB., P.S., K.R., R.G. and O.K. study supervision. All authors read and approved the final manuscript.

ORCID

Turki A. Althunian https://orcid.org/0000-0002-8030-7963

REFERENCES

1. Grobbee DE, Hoes AW. *Clinical Epidemiology: Principles, Methods, and Applications for Clinical Research*. 2nd ed. Massachusetts: Jones & Bartlett Learning; 2015.

2. Vandenbroucke JP. Observational research, randomised trials, and two views of medical science. *PLoS Med*. 2008;5(3):e67.

3. Schulz KF, Grimes DA. Allocation concealment in randomised trials: defending against deciphering. *Lancet*. 2002;359:614-618.

4. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. *J Chronic Dis*. 1967;20:637-648.

5. Eichler HG, Abadie E, Breckenridge A, et al. Bridging the efficacy–effectiveness gap: a regulator's perspective on addressing variability of drug response. *Nat Rev Drug Discov*. 2011;10(7):495-506.

6. McMahon AD. Approaches to combat with confounding by indication in observational studies of intended drug effects. *Pharmacoepidemiol Drug Saf*. 2003;12(7):551-558.

7. McMahon AD. Observation and experiment with the efficacy of drugs: a warning example from a cohort of nonsteroidal anti-inflammatory and ulcer-healing drug users. *Am J Epidemiol*. 2001;154(6):557-562.

8. Grobbee DE, Hoes AW. Confounding and indication for treatment in evaluation of drug treatment for hypertension. *BMJ*. 1997;315:1151-1154.

9. Rossouw JE, Anderson GL, Prentice RL, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's health initiative randomized controlled trial. *JAMA*. 2002;288:321-333.

10. Naudet F, Maria AS, Falissard B. Antidepressant response in major depressive disorder: a meta-regression comparison of randomized controlled trials and observational studies. *PLoS ONE*. 2011;6(6):e20811.

11. Curay JC, Sacks L, Woodcock J. Real-world evidence and real-world data for evaluating drug safety and effectiveness. *JAMA*. 2018;320(9):867-868.

12. Tannen RL, Weiner MG, Marcus SM. Simulation of the Syst-Eur randomized control trial using a primary care electronic medical record was feasible. *JCE*. 2006;59:254-264.

13. Tannen RL, Weiner MG, Xie D, Barnhart K. A simulation using data from a primary care practice database closely replicated the women's health initiative trial. *JCE*. 2007;60:686-695.

14. Weiner MG, Xie D, Tannen RL. Replication of the Scandinavian simvastatin survival study using a primary care medical record database prompted exploration of a new method to address unmeasured confounding. *Pharmacoepidemiol Drug Saf*. 2008;17(7):661-670.

15. Tannen RL, Weiner MG, Xie D. Use of primary care electronic medical record database in drug efficacy research on cardiovascular outcomes: comparison of database and randomised controlled trial findings. *BMJ*. 2009;338:b81.

16. Flory JH, Mushlin AI. Observational and clinical trial findings on the comparative effectiveness of diabetes drugs showed agreement. *J Clin Epidemiol*. 2015;68(2):200-210.

17. Fralick M, Kesselheim A, Avorn J, Schneeweiss S. Use of health care databases to support supplemental indications of approved medications. *JAMA Intern Med*. 2017;178(1):55-63.

18. Patel MR, Mahaffey WM, Garg J, et al. Rivaroxaban versus warfarin in nonvalvular atrial fibrillation. *N Engl J Med*. 2011;365:883-891.

19. Clinical Practice Research Datalink. Primary care data for public health research. https://www.cprd.com/primary-care. Accessed December 12, 2018.

20. Herrett E, Gallagher AM, Bhaskaran K, et al. Data resources profile: clinical practice research datalink (CPRD). *Int J Epidemiol*. 2015;44(3):827-836.

21. Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the general practice research database: a systematic review. *Br J Clin Pharmacol*. 2010;69:4-14.

22. Khan NF, Harrison SE, Rose PW. Validity of diagnostic coding within the general practice research database: a systematic review. *Br J Gen Pract*. 2010;60:e128.

23. Lewis J, Brensinger C. Agreement between GPRD smoking data: a survey of general practitioners and a population-based survey. *Pharmacoepidemiol Drug Saf*. 2004;13:437-441.

24. Van Staa TP, Abenhaim L, Cooper C, Zhang B, Leufkens HG. The use of a large pharmacoepidemiological database to study exposure to oral corticosteroids and risk of fractures: validation of study population and results. *Pharmacoepidemiol Drug Saf*. 2000;9:359-366.

25. Althunian TA, de Boer A, Groenwold RHH, et al. Replication of the rivaroxaban once daily oral direct factor Xa inhibition compared with vitamin K antagonism for prevention of stroke and embolism trial in atrial fibrillation (ROCKET AF) in a cohort study using electronic primary care data. *Clin Pract Res Datalink*. 2018; https://cprd.com/protocol/replication-rivaroxaban-once-daily-oral-direct-factor-xa-inhibition-compared-vitamin-k (Accessed December 12, 2018).

26. Committee for Medicinal Products for Human Use. European Medicines Agency (EMA). Summary of positive opinion for Xarelto. http://www.ema.europa.eu/docs/en_GB/document_library/Summary_of_opinion_-_Initial_authorisation/human/000944/WC500059301.pdf(Accessed October 15, 2017).

27. Ogre E, Botrel MA, Juchault C, Bouget J. Sensitivity and specificity of an algorithm based on medico-administrative data to identify hospitalized patients with major bleeding presenting to an emergency department. *BMC Med Res Methodol*. 2019;19:194.

28. Gallagher AM, Dedman D, Padmanabhan S, Leufkens HGM, de Vires F. The accuracy of date of death recording in the clinical practice research datalink GOLD database in England compared with the Office for National Statistics death registrations. *Pharmacoepidemiol Drug Saf*. 2019;28:563-569.

29. Lip GY, Nieuwlaat R, Pisters R, Lane DA, Crijns HJ. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the euro heart survey on atrial fibrillation. *Chest*. 2010;137:263-272.

30. Sturmer T, Rothman KJ, Glynn RJ. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. *J Clin Epidemiol*. 2006;59(5):437-447.

31. Toh S, Garcia Rodriguez LA, Hernan MA. Confounding adjustment via a semi-automated high-dimensional propensity score algorithm: an application to electronic medical records. *Pharmacoepidemiol Drug Saf*. 2019;20(8):849-857.

32. Elze MC, Gregson J, Baber U, et al. Comparison of propensity score methods and covariate adjustment: evaluation in 4 cardiovascular studies. *J Am Coll Cardiol*. 2017;69(3):345-357.

33. Althunian TA, de Boer A, Groenwold RHH, Klungel OH. Defining the non-inferiority margin and analyzing non-inferiority: an overview. *Br J Clin Pharmacol*. 2017;83:1636-1642.

34. Althunian TA, de Boer A, Klungel OH, Insani WN, Groenwold RHH. Methods of defining the non-inferiority margin in randomized, double-blind controlled trials: a systematic review. *Trials*. 2017;18(1):107.

35. Franklin JM, Pawar A, Martin D, et al. Nonrandomized real-world evidence to support regulatory decision making: process for a randomized trial replication project. *Clin Pharmacol Ther*. 2019;107(4):817–826. https://doi.org/10.1002/cpt.1633.

36. Leaf PJ, Frangakis C, Stuart EA, Azur MJ. Multiple imputation by chained equations: what is it and how does it work? *Int J Methods Psychiatr Res*. 2011;20(1):40-49.

37. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. New Jersey: Wiley; 2004.

38. van Buuren S. Multivariate imputation by chained equations: package 'mice' (version 2.46.0). https://cran.r-project.org/web/packages/mice/mice.pdf (Accessed January 18, 2018).

39. National Institute for Health and Care Excellence (NICE). Atrial fibrillation: management (clinical guideline). https://www.nice.org.uk/guidance/cg180/resources/atrial-fibrillation-management-pdf-35109805981381(Accessed October 15, 2017).