Strengths and weaknesses of ‘real-world’ studies involving non-vitamin K antagonist oral anticoagulants

A John Camm,1,2 Keith A A Fox3

ABSTRACT

Randomised controlled trials (RCTs) provide the reference standard for comparing the efficacy of one therapy or intervention with another. However, RCTs have restrictive inclusion and exclusion criteria; thus, they are not fully representative of an unselected real-world population. Real-world evidence (RWE) studies encompass a wide range of research methodologies and data sources and can be broadly categorised as non-interventional studies, patient registries, claims database studies, patient surveys and electronic health record studies. If appropriately designed, RWE studies include a patient population that is far more representative of unselected patient populations than those of RCTs, but they do not provide a robust basis for comparing treatment strategies. RWE studies can have very large sample sizes, can provide information on treatments in patient groups that are usually excluded from RCTs, are generally less expensive and quicker than RCTs, and can assess a broad range of outcomes. Limitations of RWE studies can include low internal validity, lack of quality control surrounding data collection and susceptibility to multiple sources of bias for comparing outcomes. RWE studies can complement the findings from RCTs by providing valuable information on treatment practices and patient characteristics among unselected patients. This information is necessary to guide treatment decisions and for reimbursement and payment decisions. RWE studies have been extensively applied in the postmarketing approval assessment of non-vitamin K antagonist oral anticoagulants since 2010. However, the benefits, costs, limitations and methodological challenges associated with the different types of RWE must be considered carefully when interpreting the findings.

INTRODUCTION

Randomised controlled trials (RCTs) are the gold standard for demonstrating the efficacy of a particular therapy or intervention.1–5 However, RCTs are, by design, limited to a subset of patients who are not fully representative of the unselected real-world population.12 Patients enrolled in RCTs tend to show higher treatment adherence than those in clinical practice, they are positively disposed to some aspect of the treatment (as shown by their agreement to participate in the study) and they have fewer comorbidities in order to limit the impact of competing risks.4 RCTs tend to exclude patients who are very young, very old or who have major comorbidities, thereby limiting the relevance of their findings to wider clinical practice.5 In addition, large-scale RCTs are expensive to conduct (~$50 million for a pivotal study) and have a long lead time from inception to completion.4

Real-world evidence (RWE) is an umbrella term that has been broadly defined as information derived from the analysis of data on health interventions used for healthcare decision-making, which has not been collected as part of an RCT but is typically from clinical practice.2 3 6 RWE is considered to better represent routine practice compared with the idealised conditions of an RCT.3 6 Data from RWE studies can complement findings from RCTs and, if appropriately designed, can provide valuable information about practice patterns and patient characteristics in a real-world setting.4 RWE studies are subject to some patient selection because such studies are unable to capture all types of patients, for example, patients who have not presented to a physician or are yet to be diagnosed. RWE studies provide a valuable reflection of the range and distribution of patients observed in the clinical practice setting.2

This review article discusses the utility, strengths and limitations of RWE studies, as well as potential developments to further bridge the gap between RCTs and daily clinical practice.

STRENGTHS AND LIMITATIONS OF RWE STUDIES

RWE studies include a wide range of research methodologies and data sources, although they can be broadly categorised into non-interventional studies, patient registries, claims database studies, patient surveys and electronic health record studies (table 1).2 5

They can also be categorised into prospective studies, which generally require primary data collection, or retrospective studies,
<table>
<thead>
<tr>
<th>Study type</th>
<th>Design</th>
<th>Conventional uses</th>
<th>Key characteristics compared with RCTs</th>
</tr>
</thead>
</table>
| Non-interventional studies, including pragmatic trials | Prospective or retrospective, non-interventional, observational Includes cohort, cross-sectional and case–control studies | ► Incidence, cause and prognosis (cohort)  
► Prevalence and cause (cross-sectional)  
► Predictors of outcome (case–control) | Advantages  
► Cohort: important in assessing risk factors where an RCT may be unethical; chronology enables clear distinction between cause and effect; several outcomes may be studied simultaneously; quick and inexpensive.  
► Cross-sectional: most efficient way to determine prevalence; quick and inexpensive; several outcomes may be studied simultaneously.  
► Case–control: case-efficient; useful for studying rare conditions or those with a long latency between exposure and disease; can assess many variables. |
| Patient registries                              | Prospective, observational cohort study        | ► Natural history of disease  
► Real-world safety and effectiveness  
► Prognosis and quality of life  
► Quality of care  
► Cost-effectiveness | Advantages  
► Larger and more diverse population; longer follow-up.  
► Few or no required study centre visits, evaluations or procedures.  
► Can identify the most cost-effective treatment approaches.  
► Data are captured in real time. |
| Administrative and claims database studies      | Retrospective                                   | ► Longitudinal and cross-sectional analyses of clinical and economic outcomes  
► Natural history of disease  
► Real-world safety and effectiveness | Advantages  
► Very large in size; can be used to identify rare events more easily, assess economic impact of various interventions and gain insight into associations between interventions and outcomes.  
► Quick and inexpensive.  
► Useful in assessing healthcare resource utilisation and costs. |
|                                                |                                               | Disadvantages  
► Lack of randomisation means that patient groups may not be comparable; problems of dose adjustment.  
► Data quality issues: potential for missing data; variance in the quality of data between registries, due to differences in audit and control measures.  
► Limits to the amount of data that can be collected.  
► Data may not be collected at fixed time intervals. |
|                                                |                                               | Disadvantages specific to claims database studies  
► Absence of a population denominator.  
► Patients can switch between insurance companies, which may limit duration of follow-up. |

Continued
Arrhythmias and sudden death

which use secondary data to look back in time (ie, data initially collected for other purposes; figure 1).

Potential strengths of RWE studies

RWE studies aim to include patient populations that are far more representative of the unselected population than those of RCTs; they can have very large sample sizes and can provide information on treatment practices in specific populations that are usually excluded from RCTs (eg, elderly patients or those with renal impairment).\(^5\)

This enhances the external validity (ie, the generalisability) of their findings compared with RCTs, which have restricted inclusion characteristics.\(^5\) RWE studies are less expensive and faster to complete than RCTs, and they can assess treatment patterns across a much broader range of outcomes. As well as broadly assessing safety and clinical treatment patterns, and effectiveness outcomes (in studies comparing new treatments with standard of care), RWE studies can include assessments of treatment adherence, treatment persistence, epidemiology, burden of

<table>
<thead>
<tr>
<th>Study type</th>
<th>Design</th>
<th>Conventional uses</th>
<th>Key characteristics compared with RCTs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electronic health record</td>
<td>Retrospective, observational, medical record</td>
<td>Clinical treatments, procedures and</td>
<td>Advantages</td>
</tr>
<tr>
<td>surveys</td>
<td></td>
<td>outcomes</td>
<td>► Capture real-time clinical treatment, outcomes, techniques and procedures.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>► Can study rare conditions or those with a long latency between exposure and disease.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>► Quick and inexpensive.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Disadvantages</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>► Requires sophisticated data management and statistical tools.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>► Data quality issues: missing data, recording/coding errors, interpretive or recall biases.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>► Lack of randomisation means patient groups may not be comparable.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>► Typically limited to a small number of study centres.</td>
</tr>
<tr>
<td>Patient surveys</td>
<td>Online, interview or paper-based questionnaire</td>
<td>Health status and well-being, patient preferences, healthcare resource utilisation, treatment patterns and expenditure</td>
<td>Advantages</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>► Methodologically rigorous in their collection of data.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>► Can provide information on the generalisability of treatments, their impact, healthcare utilisation and costs.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Disadvantages</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>► Lack of relevant data on specific treatments/products.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>► Recall and subjectivity bias.</td>
</tr>
</tbody>
</table>

RCT, randomised controlled trial; RWE, real-world evidence.

Figure 1 Illustration to highlight the design and analysis time frame (relative to the study start or index date) of different types of real-world evidence studies. Arrows depict prospective or retrospective studies of various durations.
disease, prescribing patterns, health resource utilisation and cost-effectiveness.\(^5\) They also have the capacity to assess overall outcomes in various patient and risk groups and, because they may be conducted in an unselected population over a long time frame, can provide insights into toxicity, long-term safety and rare adverse events.\(^5\)

Potential limitations of RWE studies

Traditionally, methodological and study design issues, such as the risk of confounding and bias, have prevented non-randomised observational studies from being considered useful for the assessment of new medical treatments or interventions. Even a well-designed, robust, non-randomised study would be considered to be lower grade evidence compared with a poorly designed RCT.\(^7\) Currently, the value of RWE is much better understood, with RWE studies now forming a key part of the postauthorisation evaluation of an approved drug, at the request of regulatory authorities.\(^7,8\) However, possible limitations should still be considered.\(^7\) The key limitation of RWE studies is the lack of randomisation. Although this contributes to the high external validity of the data, it reduces the internal validity of the data (ie, the extent to which any differences between the intervention and control groups can be attributed to the intervention itself, as opposed to other factors).\(^5,9\)

In addition, without a controlled clinical study environment, there may be little or no control over the quality of the data collection in RWE studies, although measures can be taken to optimise this.\(^5\) For example, prospective patient registries may have quality-control measures and auditing procedures to improve data integrity, including the use of real-time electronic data capture, personnel training, independent adjudication of outcomes and the use of data or coding dictionaries to ensure standardised outcome definitions.\(^2,10\) Retrospective comparisons are prone to multiple biases, including sampling bias, recall bias, confounding by indication and changes in practice and/or disease biology.\(^5,11\) Statistical approaches have been developed that aim to adjust for multiple types of bias (eg, propensity score adjustment, covariate adjustment and others).\(^2\) However, serious confounding may remain due to the impact of unmeasured confounders and unmeasured risks on treatment decisions. Therefore, retrospective study data do not meet the reliability and accuracy afforded by the methodological rigour of RCTs.\(^5\)

The methods employed in developing treatment guidelines do not rank observational data as a reliable basis for comparing treatments or strategies.\(^2,13\)

Applications of RWE studies

Given their ability to evaluate clinical practice patterns and monitor long-term and infrequent safety events, RWE studies have been evaluated in postmarketing approval assessments of the non-vitamin K antagonist oral anticoagulants (NOACs). Four NOACs (apixaban, dabigatran, edoxaban and rivaroxaban) are currently available for the prevention of stroke in patients with atrial fibrillation (AF) and the treatment of venous thromboembolism (VTE), but it is recommended that they are used with caution in patients with conditions where there is an increased risk of bleeding.\(^11,14-17\) Therefore, it is essential that factors such as dose selection and treatment adherence are monitored outside of a controlled clinical environment to see if they differ from pivotal RCTs\(^8-21\) and to determine their impact on clinical outcomes such as major bleeding, stroke or recurrent VTE. For this reason, many RWE studies have been, and are being, conducted to evaluate NOAC treatments; these will be used as examples to illustrate the value and limitations of the information provided by the different RWE study types.

NON-INTERVENTIONAL STUDIES

Non-interventional studies are defined as those where treatment is prescribed in accordance with the terms of the marketing authorisation, and where the assignment of the patient to a particular therapy is entirely at the physician’s discretion, rather than decided in advance by the study protocol.\(^22,23\) As such, treatment selection and management is separate from the decision to include the patient in the study, with no diagnostic or monitoring procedures performed above those deemed necessary by the attending physician for the conduct of the study.\(^22,25\)

In some instances, prospective, multicentre, non-interventional studies conducted in routine clinical practice may be referred to as pragmatic clinical trials.\(^24,25\)

With the exception of registries (detailed in the next section), non-interventional studies generally fall into three categories: cohort, cross-sectional and case–control studies.\(^11\)

Cohort studies

Cohort studies can be either prospective or retrospective, and typically assess the incidence, cause and prognosis of certain conditions/outcomes.\(^11\) In prospective cohort studies, patients are monitored until a relevant outcome occurs (eg, a major bleeding event, stroke or recurrent VTE for studies of NOACs).\(^26,27\) with data routinely collected on potential risk factors for the outcome.\(^11\) In retrospective cohort studies, the methodology is the same but the data have already been collected for a separate purpose and a post-hoc analysis is carried out.\(^11\)

Advantages of cohort studies include the ability to assess a broad range of risk factors,\(^26\) including those where an RCT may be unethical (such as the impact of smoking or exposure to toxins), and the fact that several outcomes can be monitored simultaneously.\(^11\) The chronology of the study also enables a clear distinction between cause and effect (unlike cross-sectional studies), although this also means that loss to follow-up can significantly affect the outcomes, and studying rare outcomes can be inefficient.\(^11,26\) Retrospective cohort studies can be conducted quickly and inexpensively compared with RCTs because the data have already been collected. However, retrospective cohort studies can suffer from limited or missing
data, less rigour in data collection (no prestudy analysis plan or protocol) and recall bias, all of which are distinct disadvantages.11

The first international, prospective, multicentre, non-interventional study of a NOAC for stroke prevention in patients with AF was the single-arm XANTUS (Xarelto for Prevention of Stroke in Patients with Atrial Fibrillation) study of rivaroxaban. This was then followed by the XALIA (Xarelto for Long-term and Initial Anticoagulation in venous thromboembolism) study of rivaroxaban versus standard anticoagulation in patients with VTE.26 27 These studies, both examples of pragmatic trials, supported the results of RCTs, showing low incidences of major bleeding (in both studies) and low rates of stroke (XANTUS) and recurrent VTE (XALIA) in patients receiving rivaroxaban in a real-world setting.26 27 The problem with selection bias (eg, the idea that agreement to participate in the survey might select a group that responds in a specific way) was noted in both studies, and efforts to minimise this using a single cohort design (XANTUS) or propensity score matching (XALIA) were employed.28 In both XANTUS and XALIA, outcomes were centrally adjudicated by investigators blinded to treatment in an attempt to reduce reporting bias.26 27 However, missing data and the effects of unmeasured confounding factors could not be addressed.26 27

Cross-sectional studies
Cross-sectional studies involve the assessment of a single group of patients at a single point in time, at which treatment and outcomes are determined simultaneously.11 29 They are typically used to assess prevalence and infer the cause of conditions/outcomes.11 29 Cross-sectional studies can be conducted relatively quickly and inexpensively compared with RCTs, and can assess multiple outcomes simultaneously. They are, therefore, the most efficient way to determine the prevalence of a condition.11 29 However, because the data are collected at a single time point, it is difficult to clearly distinguish cause and effect—patients who develop an outcome but die before the end of the study are not captured, and they are susceptible to selection bias. This method is also inefficient for studying rare conditions because, even in large sample sizes, there may be few or no patients with the disease.11 29-31 Additionally, cross-sectional studies are often completed using questionnaires, which have inherent problems, including low response rates and susceptibility to various sources of bias (see later section on patient surveys).2 11 32

In the case of the NOACs, cross-sectional studies have provided insights into the prevalence of underdosing, how to achieve more appropriate dose selection,30 how closely the pivotal RCTs resemble a real-world population33 and how successfully NOACs have been adopted into prescribing practices over time.31

Case–control studies
Unlike cohort and cross-sectional studies, case–control studies are usually conducted retrospectively.11 Patients who have experienced the outcome of interest are matched with a control group who have not experienced this outcome, and exposure to treatment or other factors are assessed from medical history to determine causality.11 29 Because the patient population is selected based on the outcome, case–control studies are especially useful in studying rare conditions or those with a long latency between exposure and disease, although the study is limited to investigating a single outcome.11 They can also consider many variables simultaneously, providing a case-efficient way of identifying potential predictors of specific outcomes.11 However, case–control studies are often conducted by an interview, which makes them susceptible to sampling bias (eg, recruitment of cases or controls from single sources), observational bias (eg, interviewers assessing cases differently from controls) and recall bias (eg, cases being more likely to recall past exposure than controls because they have considered the cause of the condition), as well as unmeasured confounders.11 29 34 Because of these limitations, the results of case–control studies are most appropriately used for hypothesis generation,11 34 35 and have primarily been used to assess whether NOAC treatment is a risk factor for rare outcomes or safety outcomes in specific patient populations.34-36

PATIENT REGISTRIES
Patient registries are non-interventional, observational cohort studies. They are typically prospective studies that involve standardised, ongoing data collection in a real-world setting to fulfil a specific predefined purpose, where management of treatment and care is determined by the patient and caregiver rather than the registry protocol.16 Ideally, the collection of registry data is uniform for all patients, including the type, method and frequency of collection. Information is then collated in a central registry database for analysis.

Advantages of registries over RCTs include the capacity to enrol a much larger and more diverse patient population with the potential for a longer follow-up period. This provides data that are more reflective of a real-world population and enables the study of longer term outcomes, including the identification of more infrequent safety outcomes.2 Registry studies also involve few or no required visits, evaluations or procedures at specialist centres because the data are collected by the attending physician as part of daily practice. Not only does this make registries potentially less expensive than RCTs, but it also means that treatment patterns reflect the daily clinical decision-making that is most relevant to healthcare providers and payers, and can help identify the most cost-effective treatment approaches.2 In addition, registries are sometimes linked with other databases, which can enable the assessment of separate outcomes such as healthcare utilisation and mortality.10

These factors make registries particularly useful for assessing (1) the natural history of a disease; (2) real-world

safety and effectiveness; (3) prognosis and quality of life; (4) quality of care; and (5) cost-effectiveness of treatment strategies.10

As with all observational studies, a disadvantage inherent to patient registries is that the lack of randomisation means there is no guarantee that the patient groups are similar (because of factors such as dose adjustments), which can reduce the internal validity of the findings. Importantly, retrospective registries are also subject to survivor bias because they exclude deaths prior to the registry start date, and time from diagnosis to enrolment is longer for the patients included compared with prospective studies, meaning these registries will not encompass high-risk periods after disease onset.2 Because the data in registries are collected from a wide variety of different centres, sometimes across different countries, there can also be considerable variation in the auditing and control measures employed and the quality of the data collected.2 In addition, variations between countries and centres regarding the requirement and processes for obtaining informed patient consent can also lead to authorisation or selection bias.37 38 Furthermore, there are limits to the amount of data that can be obtained in a routine clinical practice visit, and the visits may not be scheduled at regularly timed intervals, hindering comparisons between patient groups. Despite this, the advancement of modern technologies such as mobile phone apps and cloud computing may make it far easier to collect data in a more uniform fashion and in real time, thus enabling more rigorous monitoring of data collection.2

Data on NOAC use in patients with AF or VTE are provided by several large registries, including the regional Dresden NOAC Registry, the national ORBIT AF I/II (Outcomes Registry for Better Informed Treatment of Atrial Fibrillation) registries and the multinational GARFIELD-AF (Global Anticoagulant Registry in the FIELD-Atrial Fibrillation), GARFIELD-VTE (Global Anticoagulant Registry in the FIELD-Venous Thromboembolism) and GLORIA-AF (Global Registry on Long-Term Oral Antithrombotic Treatment in Patients with Atrial Fibrillation) registries (table 2).

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Overview of key prospective registries for stroke prevention in patients with AF and treatment of VTE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Registry</td>
<td>Design</td>
</tr>
<tr>
<td>GARFIELD-AF39</td>
<td>International, multicentre registry; target n=55 000 patients at &gt;1000 centres in 50 countries.</td>
</tr>
<tr>
<td>ORBIT AF I70</td>
<td>Multicentre outpatient registry in the USA of over 10 000 patients.</td>
</tr>
<tr>
<td>ORBIT AF II43 February 2013–ongoing</td>
<td>Multicentre outpatient registry targeting 15 000 patients in 300 US practices with newly diagnosed AF and/or with AF who recently transitioned to NOAC treatment.</td>
</tr>
<tr>
<td>GLORIA-AF71</td>
<td>International, multicentre registry; target n=56 000 patients at ≤2200 sites in &lt;50 countries.</td>
</tr>
<tr>
<td>PREFER in AF registry72</td>
<td>Multicentre registry in 7 European countries; n=7243 patients enrolled from 461 centres.</td>
</tr>
<tr>
<td>Dresden NOAC Registry79</td>
<td>Registry of patients treated with NOACs in private practices and hospitals in the Saxony region of Germany.</td>
</tr>
<tr>
<td>RIETE 2001–ongoing73</td>
<td>International, multicentre registry of patients with VTE.</td>
</tr>
<tr>
<td>GARFIELD-VTE March 2014–ongoing74</td>
<td>International multicentre registry; target n=10 000 patients from ~500 sites in 28 countries.</td>
</tr>
<tr>
<td>PREFER in VTE registry76</td>
<td>Multicentre registry in 7 European countries; n=3545 patients enrolled from 381 centres.</td>
</tr>
<tr>
<td>SWIVTER44</td>
<td>Swiss multicentre registry.</td>
</tr>
</tbody>
</table>
Results from these registries have provided insights into the management of NOAC treatment in clinical practice, including switching from warfarin to NOACs, the identification of high-risk patient populations and the high levels of inappropriate NOAC dosing that occur in clinical practice. However, reported limitations to the interpretation of these findings include missing or incomplete data, selection bias and residual/unmeasured bias due to a lack of randomisation.

Long-term clinical outcomes in the majority of the above registries are currently being acquired because the studies are still ongoing, but initial results support the use of NOACs in patients with AF and patients treated for VTE. For example, in SWIVTER (SWiss Venous ThromboEmbolism Registry), a study of patients treated for VTE, the risk of recurrent VTE was similar in the rivaroxaban treatment arm compared with the conventional anticoagulation arm (1.2% vs 2.1%, P=0.29) for the propensity score-adjusted population; the risk of major bleeding was also similar (0.5% vs 0.5%, respectively, P=1.00). Patients with AF receiving rivaroxaban for stroke prevention in the Dresden NOAC Registry were also analysed (n=1204). The combined endpoint including stroke, transient ischaemic attack and systemic embolism occurred at a rate of 2.03 per 100 patient-years in the intention-to-treat analysis (95% CI 1.5 to 2.7). Major bleeding in the on-treatment group occurred at a rate of 3.0 per 100 patient-years. Event outcomes in the on-treatment group were higher in patients receiving rivaroxaban 15 mg once daily compared with rivaroxaban 20 mg once daily (stroke/transient ischaemic attack/systemic embolism 2.7 vs 1.25 per 100 patient-years and major bleeding 4.5 vs 2.4 per 100 patient-years, respectively). Similar results were observed in analyses performed in patients with AF receiving apixaban and dabigatran for stroke prevention in the Dresden NOAC Registry.

**ADMINISTRATIVE AND HEALTHCARE CLAIMS DATABASE STUDIES**

These studies involve retrospective (or sometimes real-time) analysis of data from administrative and healthcare claims databases containing treatment information and clinical information, such as diagnosis codes and hospital admissions/discharge dates. As such, these databases are particularly suited to longitudinal and cross-sectional analyses of healthcare utilisation and costs at the patient, group or population level, although insights can also be gained into associations between interventions and outcomes using the (although sometimes limited) clinical information.

The key advantages of administrative and healthcare claims database studies are that they can be performed relatively quickly and inexpensively compared with RCTs, involve a very large established patient cohort and can have a long follow-up period. This enables the identification of rare events, the determination of longer term outcomes and insight into the economic impact of interventions. In some instances, database information can also be linked with clinical data, such as patient-reported outcomes, laboratory assessments, medical records and physician surveys. This practice is most common in countries such as Denmark and Sweden, where every member of the populace has a unique personal identification number, enabling lifelong follow-up and linkage between databases at the individual level. The information contained within these Scandinavian administrative databases has been used to assess the comparative effectiveness and safety of the NOACs for stroke prevention in patients with AF; regression or propensity-based adjustment was used to try to reduce bias from confounding variables inherent in a non-interventional observational analysis. Similar approaches have been used in numerous studies using data from healthcare claims databases, particularly to assess the risk of bleeding events in NOAC-treated patients and adherence/persistence to treatment (table 3).

Administrative and claims databases, however, lack key information about the choice of therapy. This choice can be influenced by clinical risk factors and physician and patient preferences which are not recorded in such databases. In addition, databases lack clarity in terms of insights into dosing. All these factors make inferences from administrative and claims databases, particularly direct drug comparisons, unreliable in relation to the clinical effectiveness of specific treatments. Data quality is also one of the major disadvantages of these database studies. Clinical data may not only be limited or missing (eg, health outcomes, health status, symptoms and patient characteristics, such as creatinine clearance levels, international normalised ratio values or time in therapeutic range), but the accuracy of reporting is variable (both between different countries and between centres within the same country) and coding errors are common. With respect to NOAC studies, there are many different codes and categories for bleeding events, further increasing the chances of coding errors (eg, gastrointestinal bleeding is often listed as ‘bleeding of unknown source’), and diagnosis codes (eg, International Classification of Diseases, Ninth or Tenth Revision) may not always differentiate between indications. Therefore, caution should always be exercised when interpreting treatment effectiveness and safety based on claims database analyses.

Even if the integrity of the data can be confirmed, there are additional issues inherent to claims database analyses, including limited validation, lack of a population denominator and lack of distinction between costs and charges; follow-up duration may be limited by patients switching between insurance plans. However, selection bias is perhaps the most common and challenging methodological issue affecting both administrative and claims database analyses because treatment selection, clinical outcomes and economic outcomes could be influenced by factors that are not recorded in the database (eg, baseline health status, symptomatology and comorbidities).
### Table 3  Examples of healthcare administrative and claims databases and associated NOAC studies

<table>
<thead>
<tr>
<th>Claims database</th>
<th>Design</th>
<th>Study examples</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNOIDES multicentre database</td>
<td>USA and Canada; healthcare data from five Canadian provinces and the USA</td>
<td>Jun et al&lt;sup&gt;38&lt;/sup&gt; n=59 525</td>
<td>No increase in risk of major bleeding or all-cause mortality was associated with NOACs compared with warfarin in patients with venous thromboembolic events, within the first 90 days of treatment (major bleeding, HR=0.92, 95% CI 0.82 to 1.03; death, HR=0.99, 95% CI 0.84 to 1.16).&lt;sup&gt;46&lt;/sup&gt;</td>
</tr>
<tr>
<td>Medicare programme</td>
<td>USA; retrospective healthcare claims database.</td>
<td>Graham et al&lt;sup&gt;45&lt;/sup&gt; n=134 414; Graham et al&lt;sup&gt;67&lt;/sup&gt; n=118 681</td>
<td>Dabigatran was associated with a reduced risk of ischaemic stroke, ICH and death but was also associated with an increased risk of major GI bleeding, compared with warfarin in elderly patients with NVAF.&lt;sup&gt;54&lt;/sup&gt; Treatment with rivaroxaban was associated with a non-significant reduction in thromboembolic stroke compared with dabigatran. However, a significant increase in ICH and major bleeding rates compared with dabigatran was also noted.&lt;sup&gt;53&lt;/sup&gt;</td>
</tr>
<tr>
<td>US Truven Health MarketScan</td>
<td>USA; retrospective commercial claims and Medicare supplemental database.</td>
<td>Coleman et al&lt;sup&gt;67&lt;/sup&gt; n=22 822</td>
<td>Rivaroxaban was associated with a significantly lower combined rate of both stroke and ICH compared with warfarin (HR=0.61, 95% CI 0.45 to 0.92). This rate was lower but non-significant for apixaban compared with warfarin.&lt;sup&gt;45&lt;/sup&gt;</td>
</tr>
<tr>
<td>National database</td>
<td>Denmark; nationwide registry study linking data from across several Danish administrative registries.</td>
<td>Sindet-Pedersen et al&lt;sup&gt;77&lt;/sup&gt; n=12 318</td>
<td>Rivaroxaban was associated with a similar major bleeding risk and risk of recurrent VTE compared with VKA treatment (major bleeding 2.28% vs 2.10%; recurrent VTE 3.03% vs 3.13%, respectively).&lt;sup&gt;47&lt;/sup&gt;</td>
</tr>
<tr>
<td>1. Danish National Prescription Registry&lt;sup&gt;28&lt;/sup&gt; 1994</td>
<td>1. National database including purchase date, ATC code, package size and dose units.</td>
<td>Sørensen et al&lt;sup&gt;86&lt;/sup&gt;</td>
<td>In patients receiving dabigatran, an increased risk of thromboembolism (dabigatran 110 mg and 150 mg) and bleeding (dabigatran 110 mg only) was observed if patients had prior use of VKAs, compared with the VKA treatment arm.&lt;sup&gt;55&lt;/sup&gt;</td>
</tr>
<tr>
<td>2. Danish National Patient Register&lt;sup&gt;26&lt;/sup&gt; 1977</td>
<td>2. National database including hospital admission and discharge dates, discharge diagnoses (ICD).</td>
<td>Larsen et al&lt;sup&gt;90&lt;/sup&gt;</td>
<td>No significant difference was observed between NOACs and warfarin for rates of ischaemic stroke. For apixaban and dabigatran, the risk of death, any bleeding and major bleeding was significantly lower compared with warfarin. Rivaroxaban was associated with similar bleeding and death rates compared with warfarin.&lt;sup&gt;39&lt;/sup&gt;</td>
</tr>
<tr>
<td>3. Danish Civil Registration System&lt;sup&gt;75&lt;/sup&gt; 1968</td>
<td>3. National database including age, gender, date of birth, vital and emigration status.</td>
<td>Nielsen et al&lt;sup&gt;96&lt;/sup&gt;</td>
<td>When NOACs were used at a reduced dose, no significant difference was observed for the risk of ischaemic stroke/systemic embolism between NOACs and warfarin. Bleeding rates were significantly lower for dabigatran compared with warfarin and similar bleeding rates observed for rivaroxaban and apixaban compared with warfarin.&lt;sup&gt;49&lt;/sup&gt;</td>
</tr>
<tr>
<td>Vårdanalysdatabasen</td>
<td>Regional database of Stockholm area of Sweden, which includes data relating to hospitalisations/other healthcare consultations (including primary care and outpatient visits), diagnoses and prescription claims (derived from National Prescribed Drug Register).</td>
<td>Forslund et al&lt;sup&gt;57&lt;/sup&gt; n=22 198; Forslund et al&lt;sup&gt;62&lt;/sup&gt; n=13 390</td>
<td>Similar risks of TIA/ischaemic or unspecified stroke/death and severe bleeding were demonstrated for patients with AF treated with NOAC and warfarin (HR=0.94, 95% CI 0.85 to 1.05 and HR=1.02, 95% CI 0.98 to 1.19). Lower risks were associated with NOACs for intracranial bleeding (HR=0.72, 95% CI 0.53 to 0.97) and haemorrhagic stroke (HR=0.56, 95% CI 0.34 to 0.93) compared with warfarin, but a higher risk was observed with NOACs for GI bleeding (HR=1.28, 95% CI 1.04 to 1.59).&lt;sup&gt;57&lt;/sup&gt; For patients with NIAC-receiving OACs, the 1-year crude persistence rates were 85.9% with apixaban, 74.4% with dabigatran, 77.4% for rivaroxaban and 85.0% for warfarin.&lt;sup&gt;90&lt;/sup&gt;</td>
</tr>
<tr>
<td>Swedish Patient Register</td>
<td>National database including hospital admission and outpatient visits including diagnoses and procedures (ICD).</td>
<td>Friberg and Oldgren&lt;sup&gt;99&lt;/sup&gt; n=68 056</td>
<td>Risks for all-cause stroke and systemic embolism were similar with NOACs and warfarin (HR=1.04, 95% CI 0.91 to 1.19). Significantly lower risks were observed with NOACs for major bleeding (HR=0.85, 95% CI 0.76 to 0.96), ICH (HR=0.60, 95% CI 0.47 to 0.76) and all-cause mortality (HR=0.89, 95% CI 0.81 to 0.98) compared with warfarin. However, a higher risk was observed for GI bleeding in the NOAC group (HR=1.22, 95% CI 1.01 to 1.46).&lt;sup&gt;97&lt;/sup&gt;</td>
</tr>
<tr>
<td>Swedish Prescribed Drug Register</td>
<td>National database including details about every dispensed prescription.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swedish Cause of Death Register</td>
<td>National database including cause of death.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swedish socioeconomic longitudinal integration database for health insurance and labour market studies (USA) register</td>
<td>National database including demographic information (eg, educational level, occupation, income).</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AF, atrial fibrillation; ATC, Anatomical Therapeutic Chemical Classification; CNOIDES, Canadian Network for Observational Drug Effect Studies; GI, gastrointestinal; ICD, International Classification of Diseases; ICH, intracranial haemorrhage; NOAC, non–vitamin K antagonist oral anticoagulant; NVAF, non-valvular atrial fibrillation; OAC, oral anticoagulant; TIA, transient ischaemic attack; VKA, vitamin K antagonist; VTE, venous thromboembolism.

**ELECTRONIC HEALTH RECORD STUDIES**

Electronic health record studies are retrospective (or sometimes real-time), observational analyses of data sourced directly from medical records or charts. They are typically used to assess clinical treatments, procedures and outcomes. Like claims database studies, electronic health record studies can be performed relatively quickly and
inexpensively compared with RCTs, can involve a relatively large patient cohort and can have a longer follow-up. The study of rare conditions or those with a long latency between exposure and disease is also possible, as is the study of the real-world use of specific techniques, treatments and procedures. Although health record studies are usually limited to a small number of study centres, central databases such as the UK Clinical Practice Research Datalink are becoming more widely available with the proliferation of mobile technologies. These databases enable longitudinal, patient-level data collection from multiple sources and may contribute to more consistent recording and coding of information.

Although real-time clinical treatment and outcomes data are captured in electronic health record studies, there are similar issues to those of claims database studies regarding data integrity. Findings are heavily reliant on the accuracy of recorded information or the recall of individuals. Therefore, there is the potential for recording/coding and interpretive errors or bias. Transforming the data for research purposes requires sophisticated statistical tools, and making robust inferences on clinical effectiveness and safety based on chart review data remains a challenge.

Nevertheless, recent electronic health record studies of NOACs have supported the findings of the pivotal RCTs and provided insights into the safety of NOAC use and treatment persistence.

**PATIENT SURVEYS**

Patient and health surveys are designed to collect health status, well-being, healthcare resource utilisation, treatment patterns, education and treatment preference information from patients, healthcare providers or the general population. They can be conducted in a variety of ways, but typically involve either online, interview or paper-based questionnaires. Patient surveys usually include a rigorous methodology for the collection of data and can provide unique information on the generalisability of treatments, their impacts at a patient level and adherence in the real world. With regard to NOACs, patient surveys have proved particularly useful in identifying factors that influence patient–physician decision-making, the need for improved patient education and adherence (including physician adherence to appropriate monitoring), and both patient and physician treatment preferences.

Patient surveys, however, lack relevant data on specific treatments or products. Therefore, they cannot be used to evaluate the effectiveness or safety of treatments, other than by inference from patient or physician perceptions, and are susceptible to various forms of bias. These include recall bias, selection bias and subjectivity bias (eg, positive response bias, where a patient may respond in a way they believe to be appropriate rather than reporting their actual behaviour). Although efforts can be made to minimise these issues, such as validation of the questionnaire prior to the survey, question/response randomisation or validation of responses within the questionnaire itself, these biases should always be considered when interpreting the results of any survey.

**STUDY OVERVIEW AND COMPARISON WITH RCTS**

The completed RWE studies listed in this review show consistent results in terms of effectiveness and safety with other studies investigating NOACs, including the findings from phase III RCTs (table 4). However, it is important to note that no comparisons between RWE studies and RCTs or other types of studies can be made.

**PRACTICAL AND PROCEDURAL CONSIDERATIONS**

Overall, there are several key aspects to consider when evaluating and comparing data from RWE studies. Perhaps most important is whether the study design is appropriate for the study outcomes. As detailed in previous sections, different study designs lend themselves to different outcomes: RCTs are the gold standard for the evaluation of efficacy and for comparing treatments; claims database studies may be the most appropriate for healthcare utilisation; cross-sectional studies can be used to assess prevalence; case–control and electronic health record studies are useful for rare conditions; and patient registries can investigate the natural history of a condition. Real-world clinical safety and effectiveness can be assessed using various RWE study types (and there is an argument that these should be obtained to provide the most balanced overview for an intervention), but it appears that prospective, non-interventional studies and patient registries provide the most robust data on these outcomes.

It is also essential that the study population reflects the target population. For example, the interpretation of health claims database findings should consider the specificity of diagnosis codes and the breadth of insurance coverage.

It is vital to consider the inherent sources of bias in the study designs when drawing the appropriate conclusions from RWE studies and comparing the findings with previous studies. These are far more numerous and varied than those of RCTs, from the positive response bias in patient surveys to the selection bias and potential for data-mining in administrative and claims database studies. There are sophisticated statistical approaches that can adjust for bias, including logistic regression adjustment, inverse probability weighting, propensity (or other) matching, instrumental variable methods and panel data models. However, despite these approaches, RWE study data still do not meet the reliability and accuracy afforded by the methodological rigour of RCTs and do not provide a reliable way of comparing treatment strategies.

Lastly, one of the most common issues with RWE studies is the quality and consistency of data collection. When comparing data between studies, the type of data...
Table 4  RWE study overview for use of NOACs in routine clinical practice (excluding registries and claims databases)

<table>
<thead>
<tr>
<th>Study</th>
<th>Design</th>
<th>Objective</th>
<th>Key findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>XANTUS</td>
<td>Prospective, observational, international study of patients treated with rivaroxaban for stroke prevention in patients with AF.</td>
<td>To investigate the safety and effectiveness of rivaroxaban in routine clinical use in the AF setting.</td>
<td>At the 1-year follow-up (n=6784), event rates of stroke and major bleeding in patients treated with rivaroxaban were low (event rates per 100 patient-years were 2.1 for treatment-emergent major bleeding, 1.9 for death and 0.7 for stroke).</td>
</tr>
<tr>
<td>XALIA</td>
<td>Prospective, non-interventional, international study of patients with DVT.</td>
<td>To assess the effectiveness and safety of rivaroxaban compared with standard anticoagulation therapy.</td>
<td>In the propensity score-adjusted population (n=4515), results confirmed that rivaroxaban was a well-tolerated and effective alternative to standard anticoagulation therapy. Rates of major bleeding and recurrent VTE were low in the rivaroxaban group compared with standard anticoagulation therapy (0.8% vs 2.1% (P=0.44) and 1.4% vs 2.3% (P=0.72), respectively).</td>
</tr>
<tr>
<td>NOAC-TURK study</td>
<td>Multicentre cross-sectional study in Turkey.</td>
<td>To evaluate the current patterns in NOAC treatment in Turkey, to include demographics and clinical outcomes.</td>
<td>A total of 2862 patients were included and the most frequent indication for NOACs was AF (83.3%). Bleeding events occurred in 7.6% of patients (1.1% major bleeding) and embolic events were observed in 1.3%. Rivaroxaban and dabigatran were preferred compared with apixaban, and 47.6% were receiving suboptimal doses of NOACs.</td>
</tr>
<tr>
<td>Tamayo et al</td>
<td>Nested case–control post-marketing safety surveillance study.</td>
<td>To assess major bleeding risk factors using data from an ongoing safety study of patients with AF treated with rivaroxaban.</td>
<td>The study included 542 cases and 2710 controls and the risk factors for major bleeding identified were generally consistent with those published previously. The strongest risk factors identified were increased age, anaemia, prior GI bleeding, heart failure and vascular disease (P&lt;0.0001, all factors).</td>
</tr>
<tr>
<td>European Heart Rhythm Association survey</td>
<td>Multinational self-assessment survey in patients with AF.</td>
<td>To assess patients’ attitudes, level of education and knowledge concerning OACs.</td>
<td>A total of 1147 responses were gathered from across Europe. OAC treatment was used by 77% of patients, with around one-third of these patients receiving NOACs. Compliance was identified as an issue in patients with AF, with 14.5% of patients temporarily discontinuing treatment and 26.5% missing at least one dose (self-reported). Further patient education is required to improve treatment adherence and knowledge of treatment monitoring.</td>
</tr>
<tr>
<td>Sauter et al</td>
<td>Paper-based pilot survey of Swiss practitioners on their knowledge of NOACs.</td>
<td>To assess the knowledge of Swiss general internal medicine physicians around follow-up, guidelines, dosing adjustments, complications and indications.</td>
<td>Of the 53 physicians who completed the survey, NOACs were seen to be well-accepted as the first-choice treatment in newly diagnosed patients. Two-thirds of the respondents adhered to clinical follow-up guidance and bleeding complication rates were low (1.9 events±2.87 in the previous 2 years), with the majority managed without admission to hospital.</td>
</tr>
</tbody>
</table>

AF, atrial fibrillation; DVT, deep vein thrombosis; GI, gastrointestinal; NOAC, non-vitamin K antagonist oral anticoagulant; OAC, oral anticoagulant; RWE, real-world evidence; VTE, venous thromboembolism; XALIA, Xarelto for Long-term and Initial Anticoagulation in venous thromboembolism; XANTUS, Xarelto for Prevention of Stroke in Patients with Atrial Fibrillation.

CONCLUSIONS

RWE studies can complement the findings from RCTs, provide valuable information on treatment practices and patient characteristics in a real-world setting, and are essential to the evidence base required for sound coverage and payment decisions. However, it is critical that the benefits, costs, limitations and methodological challenges associated with the different forms of RWE are carefully considered when interpreting the findings, as highlighted by guidance from regulatory and industry bodies. Moreover, despite the use of sophisticated statistical approaches to adjust for bias, there will always be residual confounding in comparative studies because of the absence of randomisation; therefore, results should be considered as hypothesis-generating and should not be viewed as a substitute for RCTs.

The registration of retrospective studies, in the same fashion as prospective trials, may help standardise data collection, interpretation and reporting, improve the consistency of methodological rigour across RWE studies, and help to bridge the gap between RCTs and clinical practice.

Acknowledgements  The authors would like to thank Stuart Wakelin for editorial assistance in the preparation of the manuscript.

Contributors  AJC and KAAF both contributed to the drafting of the manuscript.

Funding  Funding for editorial support was provided by Bayer AG.

Competing interests  AJC has received research grants and speaker’s honoraria, and participated in scientific advisory boards for Bayer, Daiichi Sankyo, Bristol-
REFERENCES


