

Interpreting the quality of health care database studies on the comparative effectiveness of oral anticoagulants in routine care

Sebastian Schneeweiss
Krista F Huybrechts
Joshua J Gagne

Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

Background: Dabigatran, an oral direct thrombin inhibitor, has now been available for 2 years in the US for the prevention of stroke in patients with nonvalvular atrial fibrillation, and direct Xa inhibitors are also starting to enter the market. Studies examining the effects of new oral anticoagulants in health care databases are beginning to emerge. The purpose of this study was to describe the validity of early published observational studies on the comparative safety and effectiveness of new oral anticoagulants in patients with atrial fibrillation.

Methods: We identified published nonrandomized post-marketing studies (articles or conference abstracts or posters) and critically appraised their internal validity, with a particular focus on their ability to control confounding and other biases.

Results: Two full-length journal articles, three conference posters, two conference presentation abstracts, and a US Food and Drug Administration analysis form the basis of the early comparative effectiveness and safety experience with new oral anticoagulants. Some published studies exhibit substantial biases and have insufficient precision for several important endpoints. Several studies suffer from biases arising from comparing ongoing users of the older drug, warfarin, who seem to tolerate it, to initiators of the new treatment who may have switched from warfarin or have had no prior experience with anticoagulants. Analyses tended to not adjust or not adjust adequately for confounding, and unsound propensity score application was also observed. Several studies introduced selection bias by excluding patients who died during follow-up and by restricting the study population to those with continuous database enrollment following cohort entry. We describe how these deficiencies can be avoided when studying new drugs.

Conclusion: The first published post-marketing observational studies may not be sufficient for decision-makers to assess fully the comparative effectiveness and safety of new oral anticoagulants. These studies have methodologic challenges that can be avoided by using sound pharmacoepidemiologic design and analysis strategies.

Keywords: anticoagulation, dabigatran, rivaroxaban, warfarin, confounding, epidemiology, claims data, new user design, propensity score

Introduction

Optimal clinical practice and coverage decision-making require assessment of the effectiveness and safety of new medications relative to existing treatments as soon as these new drugs reach the market and as they are used in routine care.¹ In a premarket-randomized controlled trial, the direct thrombin inhibitor dabigatran, the first new oral anticoagulant to reach the market, was more efficacious than warfarin in reducing the risk of stroke when given at a dose of 150 mg twice daily to patients with nonvalvular atrial fibrillation (hazard ratio 0.66; 95% confidence interval 0.53–0.82).² There was no substantial difference in major bleeding rates comparing dabigatran 150 mg

Correspondence: Sebastian Schneeweiss
Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, 1 Brigham Circle, Suite 3030, Boston, MA 02120, USA
Tel +1 617 278 0930
Fax +1 617 232 8602
Email schneeweiss@post.harvard.edu

with warfarin (hazard ratio 0.93; 95% confidence interval 0.81–1.07). The contraindications, narrow therapeutic index and need for therapeutic monitoring, and risk of drug–drug interactions associated with warfarin make new oral anticoagulants appealing.³ However, it is not known how the effectiveness and safety of these agents compare in routine care settings.

Differences in baseline risks, concomitant drug use, and adherence patterns among patients treated in actual practice compared with those in randomized trials in controlled research environments may hinder the generalizability of results from randomized controlled trials to routine care settings. Therefore, comparative safety and effectiveness information using large health care databases is an important complement to large randomized controlled trials to understand better the benefits and risks of new oral anticoagulants as they diffuse into the market. These ubiquitous and inexpensive data are collected in the routine provision of care and can provide the first insights into the use and outcomes associated with use of new drugs. There are many challenges to assessing the comparative effectiveness and safety of new drugs validly as they enter the market.⁴ For example, patients can differ with respect to baseline outcome risks and prior treatment patterns, and sparse data can hinder investigators' ability to address these potential sources of bias. As the first database study results become publicly available, it is imperative to review the methodologic quality of the studies from the perspective of sound epidemiologic principles that can reduce bias and that should be considered in executing such studies.

We sought to identify and evaluate the methodologic rigor of early observational studies of the comparative effectiveness and safety of new oral anticoagulants for the prevention of stroke in patients with atrial fibrillation.

Materials and methods

Search strategy

We searched PubMed and Web of Science in April 2013 for nonrandomized comparative studies of new oral anticoagulants in patients with atrial fibrillation for the prevention of stroke. Databases were searched from October 2010, ie, the date on which the first new oral anticoagulant was introduced, using the keywords (“dabigatran” or “rivaroxaban”) and (“atrial fibrillation”) and by limiting searches to journal articles. In addition, we searched selected cardiovascular and outcomes research conference abstract volumes (ie, American Heart Association, American College of Cardiology, European Society of Cardiology, International

Society of Pharmacoeconomics and Outcomes Research, and Academy Health). We also searched the reference list of each identified article and abstract to identify additional potentially relevant studies. The focus of this paper is on reviewing the methodologic rigor of early studies that health care providers and other decision-makers may encounter and use, rather than performing a comprehensive systematic review.

Study evaluation

We extracted data from the methods sections of the identified studies. In particular, we abstracted information on the study inclusion and exclusion criteria; the study's ability to balance patient characteristics between comparison groups; the potential for time-related biases, including immortal time bias,⁵ or violations of principles of temporality, including adjustment for intermediate factors during follow-up;⁶ and the likelihood of exposure and outcome misclassification.⁷ The authors used their pharmacoepidemiologic expertise to interpret these methodologic aspects of each study and summarized them in a nonsystematic way. The focus of this paper is on methodologic considerations; we did not intend to synthesize the findings of the individual studies in a meta-analysis. Each study is discussed with an attempt to identify generalizable conclusions for conducting future health care database evaluations of new drugs.

Results

We identified three poster presentations and two abstracts from podium presentations at scientific conferences, two full-length articles, and a report from the US Food and Drug Administration (FDA) using data from the Mini-Sentinel project (Table 1). All are cohort studies based on administrative health care utilization data, although some studies do not specify the particular data source used⁸ or the source population.^{8,9}

Conference presentations

The poster presentations suffered from major avoidable methodologic shortcomings that compromised their validity. The two studies presented at the podium appear to have used more methodologically sound approaches.

Selection bias

In three studies, the cohort was restricted to patients continuously enrolled throughout the study period.^{8,10,11} Such a requirement restricts the study population to those patients who do not die during study follow-up. This causes selection

Table I Summary of study characteristics

Study	Source	Time period	Exposure groups	Follow-up	Outcomes
Conference posters					
Nanchanatt et al ¹⁰	Horizon Blue Cross Blue Shield of New Jersey	January 2010 to December 2012	Dabigatran (n = 4,316) Rivaroxaban (n = 1,276)	Not specified	Gastrointestinal bleed; stroke; intracranial hemorrhage; adherence
Trask et al ⁹	Medicaid	February 2011 to September 2012	Dabigatran (n = 42)	One year	Systemic embolic events; adverse events; hospitalizations; adherence
Washington et al ⁸	Health plan (not further specified)	November 2011 to October 2012	Dabigatran (n = 6,766) Rivaroxaban (n = 1,870) Warfarin (n = 54,494)	Not specified	Stroke or systemic embolism; pulmonary embolism; gastrointestinal bleed; intracranial hemorrhage; cost
Conference presentations					
Thelus et al ¹⁶	Military health care system	November 2010 to May 2012	Dabigatran (n = 14,297) Warfarin (n = 33,548)	One year	Gastrointestinal bleed; myocardial infarction; ischemic stroke; intracranial hemorrhage
Charland et al ¹¹	Claims database (not specified)	January 2010 to February 2011	Dabigatran (n = 1,090) Warfarin (n = 7,072)	4 months	Stroke or systemic embolism; bleeding hospitalizations
Full-length articles					
Larsen et al ¹⁹	Danish nationwide registries	August 2011 to December 2012	Dabigatran (n = 4,978) Warfarin (n = 8,936)	10.5 months (median)	Primary: stroke, systemic embolism, or intracranial bleeds Secondary: death, gastrointestinal bleed, traumatic intracranial bleed or major bleed Other: myocardial infarction, pulmonary embolism, hospitalization for any cause
Sorensen et al ¹⁸	Danish nationwide registries	August 2011 to December 2011	Dabigatran (n = 2,726) Vitamin K antagonists (n = 49,640)	≤4 months	Thromboembolic events; bleeds
Mini-Sentinel					
Mini-Sentinel ²⁵	Mini-Sentinel Distributed Database (specific databases not specified)	October 2010 to December 2011	Dabigatran (up to n = 24,916 depending on analysis) Warfarin (up to n = 120,977 depending on analysis)	Up to a mean of 109 days depending on analysis	Gastrointestinal bleed; intracerebral hemorrhage

bias if a treatment is associated with an increased or decreased risk of fatal events, since those who have a fatal drug-induced event are excluded from the analysis.¹²

Confounding bias

None of the poster presentations made adjustments for baseline differences in patient characteristics (eg, age, sex, prior events, cardiovascular risk factors, anticoagulant treatment history), making confounding bias likely. This potential threat to study validity is highest in studies that mixed prevalent and new users.

Prevalent versus new users

The studies presented as posters included both prevalent and incident users, which likely worsens the confounding bias and may affect measures of adherence. Long-time warfarin users who enter a study cohort are less likely to switch treatment

if they have been managed successfully on warfarin in the past. Such a “survivor cohort” will more likely select patients who do well on their current treatment (Figure 1). Users of newly marketed agents are more likely to be starting that treatment, with a higher probability of treatment switches and treatment failure due to a lack of tolerability shortly after starting the new treatment. As discussed below, this can be avoided by limiting the study to new users.^{13–15} Both studies available as presentation abstracts focused exclusively on new users.^{11,16}

Differential follow-up

A study comparing dabigatran with rivaroxaban considered outcomes that occurred between the first filling of an anti-coagulant and the end of follow-up (late 2012), and reported them as risks or cumulative incidences (ie, number of events/number of subjects).¹⁰ However, dabigatran was approved by

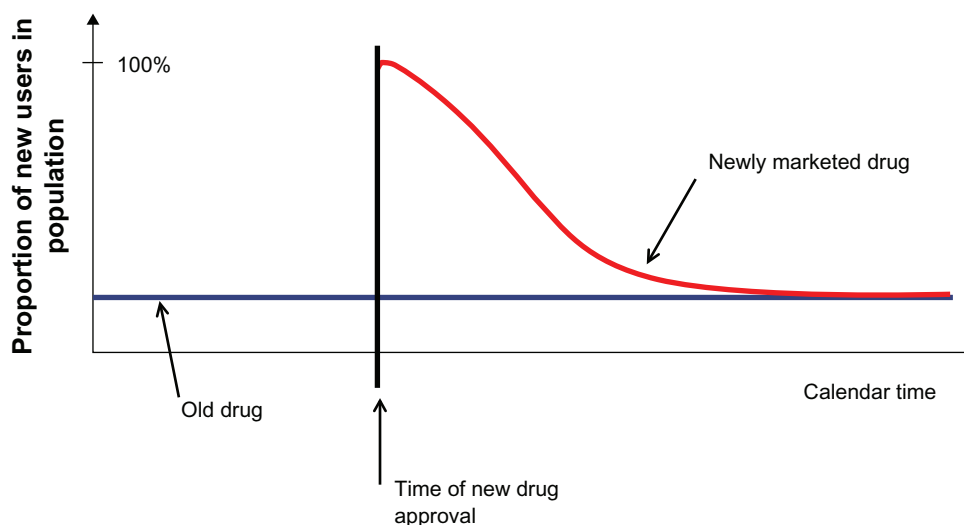


Figure 1 Proportion of new users to ongoing users of an older drug and a new drug shortly after introduction of the new drug.*

Notes: *Comparing all comers at the time of the new drug entering the market means comparing largely prevalent users of the old drug (“survivors” who tolerate the drug) against new users of the new drug who have never had a chance to experience the new drug. Comparing new users of either medication will reduce the resulting bias.

the FDA in October 2010, whereas rivaroxaban was approved in November 2012. With a 2010 study start, follow-up for dabigatran users will on average be longer than for rivaroxaban users, which will allow more total cases to accrue with the treatment that has been available for longer, even if they occur at the same rate in both groups. This can be avoided by reporting incidence rates and rate ratios rather than risks and risk ratios.¹⁷ The earlier availability and longer follow-up time of dabigatran also results in an increased opportunity for treatment switches compared with more recently introduced treatments, which may further bias an analysis that simply carries forward the first exposure without considering differential follow-up time.¹⁰

Mixed indication

One study identified subjects by medication use and included all indications, ie, atrial fibrillation as well as prophylaxis of venous thromboembolism.⁸ Depending on the distribution of indications, results may vary and make interpretation and comparison with other studies difficult.

Full-length articles

We identified two published cohort studies based on longitudinal, linked Danish registry data.^{18,19} Both used clearly defined populations and focused on incident events considering person–time denominators. The study by Sorensen et al analyzed a brief period of data during the very early marketing phase of dabigatran between August 22, 2011 and December 31, 2011.¹⁸ The study illustrates the pitfalls of comparing ongoing users of an older drug with new users

of or switchers to a new drug in nonrandomized studies. Of warfarin-exposed patients included in the analysis, 92% had used warfarin prior to the start of follow-up and did not see a reason to change their treatment (prevalent users). However, among dabigatran users, only 31% had previously used warfarin while 69% started an oral anticoagulant for the first time. Consistent with the bias expected from such an approach, the results showed a substantial increase in the risk of embolic events among those previously treated with warfarin as compared with those who were naïve to anticoagulants. The authors acknowledged that multiple unrecorded and therefore unadjusted patient risk factors may have triggered switching from warfarin to dabigatran,¹ and more importantly, the absence of complications promotes continued warfarin use. The authors conducted a subgroup analysis in which they limited the study population to new users of either dabigatran or warfarin. As expected, the associations changed substantially in the direction suggested by randomized trial evidence. The problem of comparing prevalent users with switchers without randomization is well described, for example, in the rheumatology literature (ongoing methotrexate users versus those switching to biologic disease-modifying antirheumatic drugs).²⁰ Such studies are well intentioned and try to answer a clinically relevant question, but randomization is usually necessary to obtain valid estimates when focusing on patients who are not treatment-naïve. It is widely acknowledged that comparing new users of one agent with new users of the comparator agent can substantially reduce such bias in nonrandomized studies.^{7,13,14} Finally, the follow-up time in this study was

short (up to 4 months), making it difficult to study ischemic events that are less frequent and may be affected by anticoagulation treatment only after several months.

The second Danish study by Larsen et al suffers from an ill-implemented propensity score (PS) analysis.¹⁹ The goal of a PS analysis is to estimate the probability of choosing between treatment A and treatment B conditional on all observed patient characteristics and then to compare treatment groups by conditioning on this probability, ie, make them similar in terms of all predictors that influence treatment choice. In this study, two unconventional decisions were made. First, the study contrast was between historical warfarin users (2009–2010) and current dabigatran users (2010–2011), but the PS was estimated between current warfarin users (2010–2011) and current dabigatran users. The PS was thus estimated in a population different from that to which it was applied, which is inconsistent with the theory of PS analyses.²¹ Estimating a PS in one population and using it to control confounding in a separate population does not guarantee that, on average, variables used to estimate the PS will be balanced between the two treatment groups. Second, an important advantage of PS analyses is that they can easily enable investigators to avoid comparing patients who are dissimilar. This can be achieved by trimming on the PS, ie, by excluding patients in one group who have a PS value that is larger (or smaller) than that of any patients in the comparison group,²² or by matching within a defined caliper.²¹ If a patient cannot find a subject from the comparison group with a PS value within that caliper, then that patient will be excluded. However, in the analysis by Larsen et al, no caliper was defined and all patients were matched, suggesting that at least several patients were matched to those with quite different characteristics, leading to residual confounding within matched sets. After matching, substantial differences persisted in important patient characteristics, and the resulting adjusted point estimates were quite similar to the unadjusted estimates, suggesting that either there was no confounding in the first place (which is unlikely) or that confounding was not adequately adjusted (more likely).

Both Danish studies used only hospital discharge diagnoses to assess patient comorbidities because outpatient diagnoses are not available in the Danish registries. As a result, patients who had not been hospitalized did not contribute any diagnoses and likely appeared healthier, even though they might have had chronic conditions. If prior hospitalization is a determinant of whether patients get one drug over another, then this can lead to differential confounder misclassification and bias.

Analysis of Mini-Sentinel data by the FDA

Mini-Sentinel is a pilot program sponsored by the FDA and aims to develop an operational framework, data infrastructure, and methods to perform active monitoring of FDA-regulated medical products.²³ The Mini-Sentinel Distributed Database comprises more than a dozen electronic health care databases that cover a total of more than 120 million lives.²⁴ Each data partner converts its data into a common format against which standardized queries can be run.

The FDA used this system to analyze several scenarios and to estimate rates of gastrointestinal bleeding and intracerebral hemorrhage among initiators of dabigatran or warfarin.²⁵ Scenarios varied the drug of interest (dabigatran or warfarin), new user definition (whether patients had to be naïve to warfarin, dabigatran, or both), and duration of the washout period for defining new user status. The results for each scenario were presented as rates of the outcomes of interest during the first treatment episode. Follow-up times among new users were largely comparable between treatment groups. Rates were stratified by age groups, but no explicit direct comparisons between dabigatran and warfarin were made within strata. The results enable readers to compare event rates between initiators of each drug, either overall or within age strata. However, because these rates are not adjusted except for age stratification, comparisons may likely be confounded.²⁶

Precision of current studies

One study used a total of 42 Medicaid patients with a pharmacy claim for dabigatran to evaluate tolerability and treatment adherence in a real-world setting.⁹ The authors did not report a measure of precision, such as a 95% confidence interval, which would have been extremely wide and would have highlighted the substantial uncertainty of the results due to chance.

Even the larger studies from Denmark lacked substantial statistical power for certain outcomes. The Sorensen study included 1,114 patients exposed to dabigatran 150 mg and 49,640 exposed to warfarin, but only 765 and 4,237 patients were initiating treatment. Among anticoagulant-naïve patients, only 12 thromboembolic events occurred in initiators of dabigatran 150 mg. The Larsen study included 2,739 initiators of dabigatran 150 mg and 8,936 warfarin initiators in the main analysis. While the numbers of strokes and deaths permitted fairly precise effect estimation, the precision of effect estimates for other outcomes, including systemic embolism and intracranial bleeding, was limited. The studies using US databases had more subjects enrolled.

Still, wide confidence intervals and short follow-up may require combing several databases.

Discussion

The early evidence on the comparative safety and effectiveness of new oral anticoagulants includes several presentations at scientific conferences, two full-length journal articles, and a report using the Mini-Sentinel data environment. Many of these contributions suffer from avoidable methodologic shortcomings. Careful consideration of potential biases is needed when interpreting data from these studies to inform decision-making.⁴

To guide patients, clinicians, and other decision-makers correctly, it is essential to adhere to sound study design and analytic choices when generating early evidence on the comparative safety and effectiveness of new medications in routine care. Fundamental choices that will help to mitigate confounding and other biases in the conduct of nonrandomized studies have been well described in the pharmacoepidemiologic literature.²⁷ These design and analytic choices apply equally to studies comparing new oral anticoagulants with warfarin and studies directly comparing new oral anticoagulants with each other.

The new user design offers a number of crucial advantages in database studies, which outweigh the loss of precision by excluding prevalent users. Restricting the study population to new users of the drug of interest and new users of a comparator agent implicitly ensures that both groups were recently evaluated by a physician who decided that the underlying condition had reached a state where a pharmacologic treatment should be initiated.¹⁵ Therefore, such patients are likely to be more similar in observable and unobservable characteristics than when comparing incident users versus nonusers or versus ongoing users of another drug. The clear temporal sequence in an incident user design ensures baseline covariates at study entry are assessed before treatment initiation and not affected by the treatment itself.¹⁷ Identifying two active treatment groups further reduces the chances of immortal time bias, a mistake that most frequently emerges if future records are used to define earlier exposure status in health care databases, particularly when defining a “nonuser” comparison group.⁵ Because of the well defined starting point of inception cohorts, it is possible to assess whether and in what form hazards vary over time by stratifying on duration of treatment. A new-user design avoids under-ascertainment of events occurring soon after therapy begins,¹³ which is particularly useful when studying newly marketed drugs. It avoids comparing populations

predominantly composed of first-time users of a newly marketed drug with a population predominantly composed of prevalent users of the old drug.²⁸

Confounding is a particular threat to the validity of observational studies, especially when comparing initiators of a new drug with initiators of a more established therapy. Electronic health care databases often include information on many potential confounders or proxies for them.²⁹ Propensity score analysis has emerged as a convenient and effective tool for adjusting large numbers of confounders, especially when outcomes are rare.³⁰ When implemented correctly, the approach ensures that the comparisons are made between patients who are homogeneous in terms of measured sociodemographics, health care utilization characteristics, and medical history, including proxies for treatment indication. In an incident user cohort design, a PS is the estimated probability of starting one medication versus another, conditional on all observed pretreatment patient characteristics. It has been shown that the estimated PS is better than the true PS for confounding control, assuming the correct PS model is known.^{23,31}

Estimating the PS using logistic regression is straightforward, and strategies for variable selection are well described.³² Once estimated, there are several options to utilize the PS in a second step to adjust confounding, including stratification based on quantiles of the score, regression modeling including the PS, or matching on the PS.³³ It is essential to exclude patients in the extreme PS ranges where there is little clinical ambivalence in treatment choice.²² These tails of the PS distribution often harbor extreme patient scenarios that are not representative of the majority of patients in clinical practice, and keeping them in the analyses may lead to less clinically relevant findings.^{34,35} Exclusion of patients in areas of nonoverlap of the PS can be accomplished through matching with an appropriately tight caliper or PS adjustment after symmetric or asymmetric trimming. In situations where exposure is rare, such as in the very early marketing period, disease risk scores estimated in a recent historical population offer a useful alternative or complementary approach.^{36–38} While PSs address only measured confounders, instrumental variable approaches can produce unbiased effect estimates even in the presence of unmeasured confounders if several assumptions are fulfilled.³⁹ Calendar time has recently been proposed as an instrumental variable for comparative safety and effectiveness studies,⁴⁰ and may be particularly useful in the early marketing period. However, further work is needed to evaluate the performance and utility of instrumental variable approaches in this setting, which can be characterized by relatively small numbers of patients exposed

to the new drug(s) and changes in the characteristics of the patients using them over time.

In routine care, patients change or discontinue treatment at high rates because they lack a noticeable treatment effect or experience early signs of a side effect.⁴¹ The stronger the association between such nonadherence and the outcome, the more an as-treated analysis, which censors at the point of treatment discontinuation, will be biased due to informative censoring.¹⁴ An alternative approach, sometimes called the intention-to-treat approach, follows all patients for a fixed time period, carrying forward the initial exposure status and disregarding any changes in treatment status over time. Such an analysis will suffer bias as a consequence of exposure misclassification, which increases with a longer follow-up period and a shorter average time to discontinuation. In most but not all cases, such misclassification will bias effects towards the null, similar to intention-to-treat analyses in randomized trials.¹⁴ Viewed separately, as-treated and intention-to-treat analyses trade different biases, but together give a range of plausible effect estimates.¹⁵

A challenge with the use of administrative data is the potential for misclassification of the outcome. Nondifferential and independent misclassification of the outcome tends to bias the effect estimates towards the null. However, relative measures of effect are expected to be unbiased if the specificity of the disease classification is near perfect and the sensitivity is the same among exposure groups. This can be achieved by using an outcome definition with a high positive predictive value.⁴²

Database studies are also vulnerable to time-related biases since studies are generally implemented in retrospectively collected data. By designing and analyzing the observational study as if it were a randomized controlled trial, but without the randomization, investigators can avoid conditioning on factors that occur during follow-up. For example, including all patients in the analysis who are eligible for the cohort based on information available only at the time of cohort entry avoids this selection bias.

Shortly after a new drug enters the market, studies will tend to observe few exposed events. It is important to calculate and present confidence limits around the point estimate to provide insight into the uncertainty of the findings due to random error. Studies with extremely wide confidence intervals require very cautious and tentative interpretation, particularly in light of other potential biases, which are not incorporated in the confidence intervals. Ideally, an effort would be made to continuously grow balanced cohorts of new users as data refresh and become more plentiful.^{1,43,44}

Conclusion

Overall, the review of the first set of studies analyzing new oral anticoagulants with health care utilization databases showed varying quality. Some design choices likely lead to substantial biases, but most of the mistakes are avoidable by following the standard recommendations for the conduct of pharmacoepidemiologic studies using health care databases. Despite the availability of large health care databases, there was still a lack of precision for most major endpoints at this stage. Monitoring systems that will update multivariable adjusted point estimates using sequential new-user cohort designs may lead the way forward. Taking the results of many of the published results at face value may lead to wrong conclusions, whether overly optimistic or pessimistic, about new oral anticoagulants.

Acknowledgment

This work was funded by the National Library of Medicine (RO1-LM010213) and the National Heart Lung and Blood Institute (RC4-HL106373).

Disclosure

SS is principal investigator of the Brigham and Women's Hospital DEcIDE Center on Comparative Effectiveness Research and the DEcIDE Methods Center (both funded by AHRQ) and of the Harvard-Brigham Drug Safety and Risk Management Research Center (funded by the FDA). He is co-chair of the Mini-Sentinel methods committee, a paid consultant to Whiscon LLC and Booz and Company, and is the principal investigator of investigator-initiated contracts from Novartis, Pfizer, and Boehringer Ingelheim to the Brigham and Women's Hospital that include medications discussed in this paper. JJG is an investigator in the FDA-funded Mini-Sentinel pilot project. KFH has no conflicts of interest in this work. The opinions expressed here are those of the authors and not necessarily in agreement with the FDA.

References

1. Schneeweiss S, Gagne JJ, Glynn RJ, Ruhl M, Rassen JA. Assessing the comparative effectiveness of newly marketed medications: methodological challenges and implications for drug development. *Clin Pharmacol Ther*. 2011;90(6):777–790.
2. Connolly SJ, Ezekowitz MD, Yusuf S, et al. Dabigatran versus warfarin in patients with atrial fibrillation. *N Engl J Med*. 2009;361(12):1139–1151.
3. Wittkowsky AK. Effective anticoagulation therapy: defining the gap between clinical studies and clinical practice. *Am J Manag Care*. 2004;10(Suppl 10):S297–S306.
4. Gagne JJ, Bykov K, Willke RJ, Kahler KH, Subedi P, Schneeweiss S. Treatment dynamics of newly marketed drugs and implications for comparative effectiveness research. *Value Health*. 2013; In press.

5. Suissa S. Immortal time bias in pharmaco-epidemiology. *Am J Epidemiol.* 2008;167(4):492–499.
6. Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology.* 2009;20(4):488–495.
7. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol.* 2005;58(4):323–337.
8. Washington C, Toe D, Burke S. Real-world effectiveness and utilization patterns of rivaroxaban, dabigatran, and warfarin in patients with atrial fibrillation in a managed care population. Poster presented at: Academy of Managed Care Pharmacy's 25th Annual Meeting & Expo; April 3–5, 2013; San Diego, CA. *Coventry Health Care.*
9. Trask N, Kotadiya P, Kohn M, et al. Adherence measurement and incidence of bleeding and systemic embolism with dabigatran in a Medicaid population. Poster presented at: University of Massachusetts Center for Clinical & Translational Science 4th Annual Research Retreat; May 8, 2013; Worcester, MA. University of Massachusetts Medical School – Clinical Pharmacy Services.
10. Nanchanatt G, Brunetti L, Farber K, Noorigan W, Jan S. Comparison of the incidence of stroke and bleeding complications in patients using dabigatran vs rivaroxaban for stroke prevention in nonvalvular atrial fibrillation. Poster presented at: Academy of Managed Care Pharmacy's 25th Annual Meeting & Expo; April 3–5, 2013; San Diego, CA. Horizon Blue Cross Blue Shield of New Jersey; Ernest Mario School of Pharmacy, Rutgers University.
11. Charland SL, Malone DC, Agatep BC, et al. Real-world risk of thromboembolism or bleeding with warfarin or dabigatran therapy in patients with atrial fibrillation. *Circulation.* 2012;126:A16370.
12. Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology.* 2004;15(5):615–625.
13. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol.* 2003;158(9):915–920.
14. Hernan MA, Hernandez-Diaz S. Beyond the intention-to-treat in comparative effectiveness research. *Clin Trials.* 2012;9(1):48–55.
15. Schneeweiss S. A basic study design for expedited safety signal evaluation based on electronic health care data. *Pharmacoepidemiol Drug Saf.* 2010;19(8):858–868.
16. Thelus R, Villines TC, Coster TS. Dabigatran versus warfarin among patients with atrial fibrillation: real-world post-market results. *Circulation.* 2012;126:A14877.
17. Rothman KJ. *Modern Epidemiology.* 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2008.
18. Sorensen R, Gislason G, Torp-Pedersen C, et al. Dabigatran use in Danish atrial fibrillation patients in 2011: a nationwide study. *BMJ Open.* 2013;3:5.
19. Larsen TB, Rasmussen LH, Skjøth F, et al. Efficacy and safety of dabigatran etexilate and warfarin in 'real world' patients with atrial fibrillation: a prospective nationwide cohort study. *J Am Coll Cardiol.* 2013;61(22):2264–2273.
20. Solomon DH, Lunt M, Schneeweiss S. The risk of infection associated with tumor necrosis factor alpha antagonists: making sense of epidemiologic evidence. *Arthritis Rheum.* 2008;58(4):919–928.
21. Rosenbaum PR, Rubin DB. The central role of propensity scores in observational studies for causal effects. *Biometrika.* 1983;70(1):41–55.
22. Sturmer T, Rothman KJ, Avorn J, Glynn RJ. Treatment effects in the presence of unmeasured confounding: dealing with observations in the tails of the propensity score distribution – a simulation study. *Am J Epidemiol.* 2010;172(7):843–854.
23. Platt R, Carnahan RM, Brown JS, et al. The US Food and Drug Administration's Mini-Sentinel program: status and direction. *Pharmacoepidemiol Drug Saf.* 2012;21 Suppl 1:1–8.
24. About Mini-Sentinel [webpage on the Internet]. Silver Spring, MD: US Food and Drug Administration; 2013. Available from: http://www.mini-sentinel.org/about_us/. Accessed August 1, 2013.
25. Mini-Sentinel. *Modular Program Report.* Silver Spring, MD: US Food and Drug Administration. Available from: http://www.mini-sentinel.org/work_products/Assessments/Mini-Sentinel_Modular-Program-Report_MS3_MPR31-Part-2_Dabigatran-Warfarin-GIH-ICH.pdf. Accessed August 20, 2013.
26. Southworth MR, Reichman ME, Unger EF. Dabigatran and postmarketing reports of bleeding. *N Engl J Med.* 2013;368(14):1272–1274.
27. Strom BL. *Pharmacoepidemiology.* 5th ed. Hoboken, NJ: J Wiley; 2013.
28. Moride Y, Abenheim L. Evidence of the depletion of susceptibles effect in non-experimental pharmacoepidemiologic research. *J Clin Epidemiol.* 1994;47(7):731–737.
29. Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology.* 2009;20(4):512–522.
30. Rubin DB. Estimating causal effects from large data sets using propensity scores. *Ann Intern Med.* 1997;127(8 Pt 2):757–763.
31. Robins JM, Mark SD, Whitney KN. Estimating exposure effects by modelling the expectation of exposure conditional on confounders. *Biometrics.* 1992;48(2):479–495.
32. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable selection for propensity score models. *Am J Epidemiol.* 2006;163(12):1149–1156.
33. Glynn RJ, Schneeweiss S, Sturmer T. Indications for propensity scores and review of their use in pharmacoepidemiology. *Basic Clin Pharmacol Toxicol.* 2006;98(3):253–259.
34. Kurth T, Walker AM, Glynn RJ, et al. Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect. *Am J Epidemiol.* 2006;163(3):262–270.
35. Lunt M, Solomon D, Rothman K, et al. Different methods of balancing covariates leading to different effect estimates in the presence of effect modification. *Am J Epidemiol.* 2009;169(7):909–917.
36. Arbogast PG, Ray WA. Use of disease risk scores in pharmacoepidemiologic studies. *Stat Methods Med Res.* 2009;18(1):67–80.
37. Arbogast PG, Kaltenbach L, Ding H, Ray WA. Adjustment for multiple cardiovascular risk factors using a summary risk score. *Epidemiology.* 2008;19(1):30–37.
38. Glynn RJ, Gagne JJ, Schneeweiss S. Role of disease risk scores for comparative effectiveness with emerging therapies. *Pharmacoepidemiol Drug Saf.* 2012;21 Suppl 2:138–147.
39. Brookhart MA, Rassen JA, Schneeweiss S. Instrumental variable methods in comparative safety and effectiveness research. *Pharmacoepidemiol Drug Saf.* 2010;19(6):537–554.
40. Mack CD, Glynn RJ, Brookhart MA, et al. Calendar time-specific propensity scores and comparative effectiveness research for stage III colon cancer chemotherapy. *Pharmacoepidemiol Drug Saf.* 2013;22(8):810–818.
41. Benner JS, Glynn RJ, Mogun H, Neumann PJ, Weinstein MC, Avorn J. Long-term persistence in use of statin therapy in elderly patients. *JAMA.* 2002;288(4):455–461.
42. Kelsey JL, Whittemore AS, Evans AS, Thompson WD. *Methods in Observational Epidemiology.* 2nd ed. New York, NY: Oxford University Press; 1996.
43. Gagne JJ, Glynn RJ, Rassen JA, et al. Active safety monitoring of newly marketed medications in a distributed data network: application of a semi-automated monitoring system. *Clin Pharmacol Ther.* 2012;92(1):80–86.
44. Gagne JJ, Rassen JA, Walker AM, Glynn RJ, Schneeweiss S. Active safety monitoring of new medical products using electronic health care data: selecting alerting rules. *Epidemiology.* 2012;23(2):238–246.

Comparative Effectiveness Research**Dovepress****Publish your work in this journal**

Comparative Effectiveness Research is an international, peer reviewed open access journal focusing on comparative effectiveness of health care including preventative health care strategies, diagnostic strategies, diagnostic technology, medical devices, drugs, medical technology, health systems and organization. The manuscript management system

is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <http://www.dovepress.com/comparative-effectiveness-research-journal>