VALUE IN HEALTH 18 (2015) 1057–1062

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

Comparative Effectiveness Research/Health Technology Assessment (HTA)

Comparison of Benefit-Risk Assessment Methods for Prospective Monitoring of Newly Marketed Drugs: A Simulation Study

Joshua J. Gagne, PharmD, ScD[1]*, *Mehdi Najafzadeh, PhD*[1], *Niteesh K. Choudhry, MD, PhD*[1], *Katsiaryna Bykov, PharmD, MS*[1], *Kristijan H. Kahler, PhD*[2], *Diane Martin, PhD*[2], *James R. Rogers*[1], *Sebastian Schneeweiss, MD, ScD*[1]

[1]*Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA;* [2]*Novartis Pharmaceuticals Corporation, East Hanover, NJ, USA*

ABSTRACT

Objectives: To compare benefit-risk assessment (BRA) methods for determining whether and when sufficient evidence exists to indicate that one drug is favorable over another in prospective monitoring. **Methods:** We simulated prospective monitoring of a new drug (A) versus an alternative drug (B) with respect to two beneficial and three harmful outcomes. We generated data for 1000 iterations of six scenarios and applied four BRA metrics: number needed to treat and number needed to harm (NNT\NNH), incremental net benefit (INB) with maximum acceptable risk, INB with relative-value–adjusted life-years, and INB with quality-adjusted life-years. We determined the proportion of iterations in which the 99% confidence interval for each metric included and excluded the null and we calculated mean time to alerting. **Results:** With no true difference in any outcome between drugs A and B, the proportion of iterations including the null was lowest for INB with relative-value–adjusted life-years (64%) and highest for INB with quality-adjusted life-years (76%). When drug A was more effective and the drugs were equally safe, all metrics indicated net favorability of A in more than 70% of the iterations. When drug A was safer than drug B, NNT\NNH had the highest proportion of iterations indicating net favorability of drug A (65%). Mean time to alerting was similar among methods across the six scenarios. **Conclusions:** BRA metrics can be useful for identifying net favorability when applied to prospective monitoring of a new drug versus an alternative drug. INB-based approaches similarly outperform unweighted NNT\NNH approaches. Time to alerting was similar across approaches.

Keywords: benefit-risk assessment, comparative effectiveness, net benefit, safety.

Introduction

New drugs usually enter the market on the basis of evidence from relatively small, short, placebo-controlled randomized trials [1]. The information required for market approval is not necessarily sufficient for coverage and treatment decisions, where fully informed decisions might consider the longer-term comparative safety and effectiveness of a new medication versus existing alternatives [2]. To fill this evidence gap, stakeholders—including payers themselves—are beginning to use routinely collected electronic health care data to conduct prospective, active monitoring of new drugs [3].

To date, approaches for prospective, active drug monitoring have focused primarily on drug safety and very little on comparative effectiveness [4]. Coverage and treatment decisions require considerations of both harms and benefits; yet, to our knowledge, combined benefit-risk assessment (BRA) has not been evaluated in the context of prospective drug monitoring.

Conflict of interest: J. J. Gagne receives research support from Harvard Catalyst, the Food and Drug Administration (FDA), the Agency for Healthcare Research & Quality, and the Patient-Centered Outcomes Research Institute. K. H. Kahler is an employee of Novartis Pharmaceuticals Corporation. D. Martin was an employee of Novartis Pharmaceuticals Corporation during the conduct of this study. S. Schneeweiss is Principal Investigator of the Harvard-Brigham Drug Safety and Risk Management Research Center funded by the FDA. His work is partially funded by grants/contracts from the Patient-Centered Outcomes Research Institute, the FDA, and the National Heart, Lung and Blood Institute. He is a consultant to WHISCON, LLC, and to Aetion, Inc., a software manufacturer of which he also owns shares. He is principal investigator of investigator-initiated grants to the Brigham and Women's Hospital from Novartis and Boehringer Ingelheim, unrelated to the topic of the study.

* *Address correspondence to:* Joshua J. Gagne, Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, 1620 Tremont Street, Suite 3030, Boston, MA 02120.

E-mail: jgagne1@partners.org.

1098-3015$36.00 – see front matter Copyright © 2015, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

http://dx.doi.org/10.1016/j.jval.2015.08.015

Simultaneously incorporating benefits and risks into the same active monitoring framework can aid decision makers in determining whether and when sufficient evidence exists to indicate that one drug is favorable over another [5]. The consequences of decision making about a new drug without all relevant data available can be substantial. On the one hand, delayed adoption of a drug with a net favorable benefit-risk profile compared with existing alternatives can result in worse health outcomes for patients and potentially greater total health care costs as a result of those outcomes. On the other hand, prospective BRA monitoring could reveal unfavorable benefit-risk balances that were not evident on the basis of premarketing data alone.

We conducted a simulation study to compare the performance of BRA metrics for calculating the net benefits and risks of medications in a prospective monitoring framework, with the objective of determining which approaches identify net favorability most quickly and accurately.

Methods

Selection of BRA Techniques

We previously reviewed BRA techniques and proposed a unified framework for classifying these methods [6]. In brief, existing BRA methods share substantial commonality and many can be generalized using a single formula. Metrics differ primarily in whether they consider the duration of impact of health outcomes and in the ways in which they weight outcomes, such as by using patient-derived preference weights (e.g., stated preferences and utilities), weights obtained from other perspectives, or no weights. Using this framework, we selected metrics that are amenable to quantitative prospective benefit-risk monitoring. Criteria included the ability to accommodate multiple beneficial and multiple harmful outcomes and that the metric result in a single numeric index that quantitatively summarizes the relative benefits and risks of one product versus another.

Metrics that met our criteria were the number needed to treat and the number needed to harm (NNT|NNH) and three methods that rely on the incremental net benefit (INB) framework: INB with maximum acceptable risk (MAR), INB with relative-value–adjusted life-years (RVALYs), and INB with quality-adjusted life-years (QALYs). NNT|NNH is an unweighted metric that compares the number of patients who need to be treated in order for one of those patients to benefit from treatment to the number of patients who need to be treated for one to be harmed. INB-MAR, INB-RVALYs, and INB-QALYs are all weighted metrics that differ slightly with respect to the types of weights that they use and whether they consider the duration of impact of the outcomes. MAR weights indicate patients' willingness to trade-off harmful and beneficial outcomes of a treatment and are based on stated preference methods such as discrete choice experiments (DCEs). INB-MAR does not explicitly incorporate the duration of impact of the outcomes. Stated preferences can also be used to calculate INB in terms of RVALYs, which also accommodates inclusion of duration of impact of the outcomes. As with INB-RVALYs, INB-QALYs can account for the duration of impact of different outcomes but use utility weights for given health states rather than stated preferences. Utility weights can be obtained from standard gamble or time trade-off methods.

Simulated Prospective Monitoring Framework

We simulated a setting in which a new drug enters the market, and we are interested in using prospectively collected longitudinal electronic health care data to determine as quickly as possible whether the new drug has a favorable benefit-risk profile, with respect to multiple beneficial and harmful outcomes, as compared to an existing alternative. We simulated periodic updating of the electronic health care database over time after approval of the new drug by generating data for 20 sequential cohorts of drug A and B initiators, which, with quarterly data updating, would be equal to prospective monitoring for 5 years [7].

We generated data for simulated patients exposed to the new drug and patients exposed to the comparator in a 1:1 ratio on the basis of hypothetical propensity score (PS) matching. The PS-matched design enables simultaneous monitoring of multiple outcomes within the same PS-matched cohorts without the need for further confounding adjustment [8]. Note that although we assumed a 1:1 matched cohort design, we did not actually simulate covariates, confounding, and PSs. Rather, we simulated cohorts of the same size and devoid of confounding by measured factors, as would be expected in a 1:1 PS-matched design. We have used this approach to monitor the safety and comparative effectiveness of several drugs [7,9–12], including rofecoxib versus nonselective nonsteroidal anti-inflammatory drugs with respect to gastrointestinal bleed and myocardial infarction (MI) outcome.

Base-Case Example

We modeled our simulation study on a hypothetical comparison of two antiplatelet agents (drugs A and B) for which we were interested in comparing two beneficial outcomes—MI reduction and ischemic stroke reduction—and three harmful outcomes—hemorrhagic stroke, other major bleeding, and minor bleeding. We used estimates from the literature to develop a base-case scenario, in which we modeled a cumulative incidence of 7% for MI, 1% for ischemic stroke, 0.4% for hemorrhagic stroke, 2.4% for other major bleeding, and 2.4% for minor bleeding [13–17]. We derived MAR and RVALY weights for these outcomes from a previously conducted DCE [18], and we calculated QALYs from previously published EuroQol five-dimensional questionnaire utilities [19–22]. Weights for each INB metric and assumed durations of impact of each outcome are listed in Table 1.

Scenarios

We simulated data under six scenarios (Table 2). In scenario 1 ("null scenario"), we set the average true underlying relative risk (RR_{true}) of each outcome to 1 (i.e., $\ln[RR_{true}] = 0.00$), representing a case in which there is no true difference in any outcome between the two drugs of interest. In each additional scenario, we simulated different combinations of drug A being less, equally, or more effective than drug B plus drug A being safer, equally safe, or less safe than drug B. For scenarios in which drug A was more effective, we selected RR estimates for the two beneficial outcomes from distributions in which the mean RR indicated an approximately 18% benefit of drug A versus drug B (i.e., $\ln[RR_{true}] = -0.20$). Note that although we refer to these as beneficial outcomes, they are beneficial in the sense that the drugs are intended to reduce their incidence (i.e., prevent MI or stroke) such that a reduction is favorable. When drug A was safer than drug B, we used the same 18% reduction in all three harmful outcomes. When drug A was less effective than drug B, we simulated an approximately 18% increase in the occurrence of each beneficial outcome (i.e., $\ln[RR_{true}] = 0.17$). We used the same 18% increase in all three harmful outcomes when drug A was less safe than drug B. In scenarios 2, 3, and 6, drug A dominates drug B regardless of the weights used. In scenarios 4 and 5, there is no single truth because the relative favorability of drugs A and B depends on the weights used.

For simplicity, we omitted scenarios that are mirror images of scenarios 2 to 6 because the results would be redundant. For example, the mirror image of scenario 6 would be one in which drug A is both less effective and less safe than drug B.

Table 1 – Weights and durations of impact for each simulated outcome.

Outcome	Outcome type	Utility weights from EQ-5D (INB-QALY)	Preference weights from DCE (INB-MAR and INB-RVALY)*	Duration of outcome impact
Myocardial infarction	Beneficial	0.84	0.09	20 y
Ischemic stroke	Beneficial	0.60	0.07	20 y
Hemorrhagic stroke	Harmful	0.60	0.07	20 y
Major bleeding	Harmful	0.46	–0.03	1 mo
Minor bleeding	Harmful	0.80	–0.01	1 mo

DCE, discrete choice experiment; EQ-5D, EuroQol five-dimensional questionnaire; INB, incremental net benefit; QALY, quality-adjusted life-year; MAR, maximum acceptable risk; RVALY, relative-value–adjusted life-year.
*Numbers represent preference weights for a 1% increase in probability of each outcome derived from a previous DCE study.

Data Generation

For each of the six scenarios, we generated data for 1000 sets of 20 sequential cohorts. In each of the 6000 iterations, we randomly sampled a baseline outcome incidence for each of the five outcomes among drug B initiators (R0) from a log-normal distribution to yield the base-case cumulative incidences described above. We used a log-normal distribution to ensure that the baseline cumulative incidence was always positive and nonzero. We then selected a true underlying log risk ratio ($\ln[RR_{true}]$) from a normal distribution with mean parameters described above, which permits both negative and positive log risk ratios and therefore risk ratios bound by zero and infinity. We exponentiated the $\ln(RR_{true})$ to obtain the true underlying risk ratio (RR_{true}) and multiplied this by R0 to obtain the underlying event incidence among drug A patients (R1).

In each of the 20 sequential cohorts in each iteration, we generated the observed number of events among drug B initiators using a binomial distribution, with probability R0 for the outcome and number of trials *N*, which was the number of matched drug A and B initiators in the given cohort. We used a separate binomial distribution, with probability R1 and number of trials *N*, to generate the number of events among drug A initiators in each cohort. We set the number of matched pairs (*N*) to 500 in the first sequential cohort and increased *N* linearly across the 20 cohorts (to 10,000 in period 20) to simulate the uptake of drug A over time. Additional details of the simulation framework can be found in other published work [7].

BRA Analysis

We cumulated the number of events for each outcome among drug A and B initiators prospectively across the sequential cohorts. At each simulated data update, we used the cumulative number of events and the relevant weights to calculate each of the four BRA metrics of interest. We bootstrapped 99% confidence intervals (CIs) for each of the 20 sequential estimates in each of the 6000 iterations by regenerating the observed events across the 20 sequential cohorts 1000 times on the basis of the selected R0 and R1 parameter values. We used 99% CIs as a more conservative approach than nominal 95% CIs given the sequential design.

In each scenario, we determined the proportion of iterations in which the 99% CI for each metric included the null (indicating no net favorability of either method), was entirely above the null (indicating net favorability of drug A over drug B), and was entirely below the null (indicating net favorability of drug B over drug A). In scenarios 2 to 6, we calculated the mean time to alerting, which we defined as the number of sequential cohorts accrued when the CI first excluded the null.

Results

Null Scenario (Scenario 1)

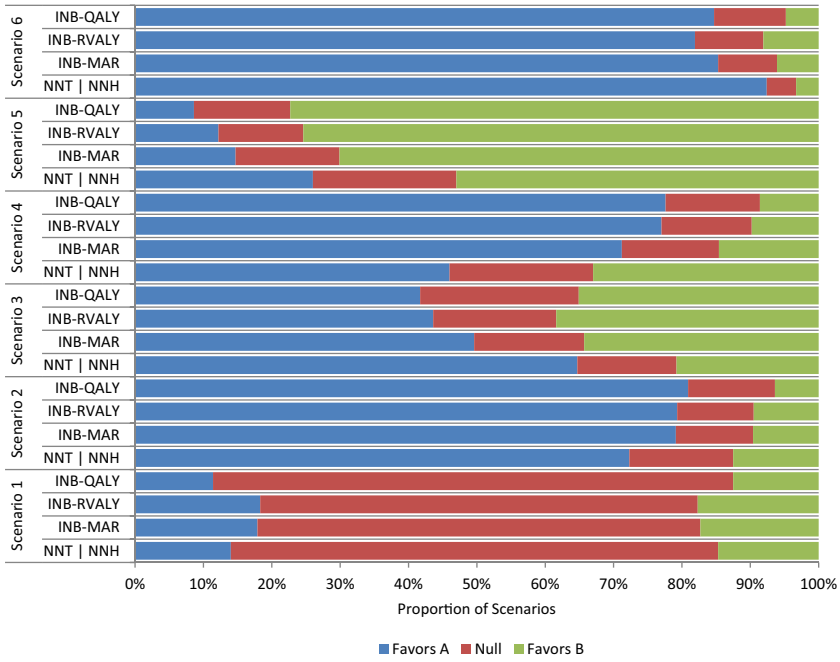
In the null scenario, in which the underlying R0 and R1 distributions were the same for each of the five outcomes, the 99% CIs for all four metrics included the null in at least 60% of the iterations (Fig. 1). The proportion of iterations including the null was lowest for INB-RVALYs (64%) and INB-MAR (65%) and highest for INB-QALYs (76%). The proportion of iterations favoring A versus B was approximately balanced for each metric.

Drug A More Effective, Equally Safe (Scenario 2)

When drug A was more effective than drug B and the drugs were equally safe, all four metrics indicated net favorability of A in more than 70% of the iterations (Fig. 1). NNT\NNH had the lowest proportion of iterations indicating net favorability of drug A (72%). Proportions of iterations indicating net favorability of drug A were near 80% for the three INB-based approaches: 81% for INB-QALYs and 79% for INB-MAR and INB-RVALYs. Among iterations in which the metrics indicated net favorability of drug A, INB-MAR and INB-RVALYs had the fastest time to alerting, which, on average, occurred in the fifth period (Fig. 2). NNT\NNH and INB-QALYs had average alert times in the sixth period.

Table 2 – Description of six simulated scenarios.

Scenario no.	Description	Risk ratio for each benefit	Risk ratio for each harm
1	Null scenario	1.00	1.00
2	Dug A more effective, equally safe	0.82	1.00
3	Dug A safer, equally effective	1.00	0.82
4	Drug A less safe, more effective	0.82	1.18
5	Drug A less effective, safer	1.18	0.82
6	Drug A more effective, safer	0.82	0.82



Scenario 1, Null scenario; Scenario 2, Drug A more effective, equally safe; Scenario 3, Drug A safer, equally effective; Scenario 4, Drug A more effective, less safe; Scenario 5, Drug A safer, less effective; Scenario 6, Drug A more effective, safer

NNT|NNH, number needed to treat and number needed to harm; INB-MAR, incremental net benefit with maximum acceptable risk; INB-RVALY, INB with relative-value adjusted life years; INB-QALY, INB with quality-adjusted life years

Fig 1 – Proportion of iterations indicating favorability of drug A, drug B, or no net favorability (null) for each of four benefit-risk assessment methods in six scenarios. INB, incremental net benefit; QALY, quality-adjusted life-year; MAR, maximum acceptable risk; RVALY, relative-value–adjusted life-year. (Color version of figure available online).

Drug A Safer, Equally Effective (Scenario 3)

When drug A was safer and drugs A and B were equally effective, NNT|NNH had the highest proportion of iterations indicating net favorability of drug A (65%), followed by INB-MAR (50%; Fig. 1). When indicating net favorability of drug A, NNT|NNH, INB-MAR, and INB-RVALYs all had mean alerting times in the seventh period and INB-QALYs had a mean alerting time in the eight period (Fig. 2).

Drug A More Effective, Less Safe (Scenario 4)

When drug A was more effective but less safe than drug B, the three INB-based methods indicated net favorability of drug A in more than 70% of the iterations: 71% for INB-MAR, 77% for INB-RVALYs, and 78% for INB-QALYs (Fig. 1). NNT|NNH indicated net favorability of drug A in 46% of the iterations, net favorability of drug B in 33% of the iterations, and no net favorability of either drug in 21% of the iterations. When metrics indicated net favorability of drug B, mean alert times were either eight or nine periods (Fig. 2). Mean time to alerting of net favorability of drug A varied from five periods for INB-RVALYs to eight periods for NNT|NNH.

Drug A Safer, Less Effective (Scenario 5)

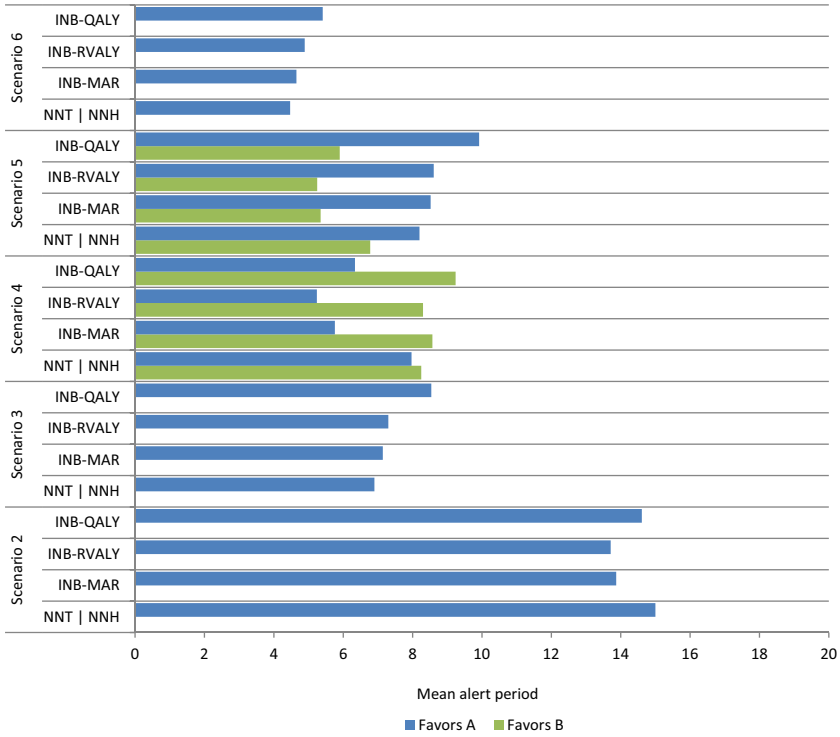
When drug A was safer but less effective than drug B, NNT|NNH indicated net favorability of drug A in 26% of the iterations and of drug B in 53% of the iterations (Fig. 1). INB-MAR indicated net favorability of drug A in 15% of the iterations and of drug B in 70% of the iterations. INB-RVALYs indicated net favorability of drug A in 12% of the iterations and of drug B in 74% of the iterations. INB-QALYs indicated net favorability of drug A in 9% of the iterations and of drug B in 77% of the iterations. Mean time to alerting of net favorability of drug A ranged from eight periods for NNT|NNH to 10 periods for INB-QALYs (Fig. 2). Mean time to alerting of net favorability of drug B ranged from five periods for INB-MAR and INB-RVALYs to seven periods for NNT|NNH.

Drug A More Effective, Safer (Scenario 6)

When drug A was both more effective and safer than drug B, all metrics indicated net favorability of drug A in more than 80% of the iterations: 82% for INB-RVALYs, 85% for INB-MAR and INB-QALYs, and 92% for NNT|NNH (Fig. 1). On average, all metrics indicated net favorability of drug A versus B by the fifth period (Fig. 2).

Discussion

Our simulation study demonstrates the feasibility of using BRA metrics for simultaneous prospective monitoring of the comparative effectiveness and safety of new drugs. Consistent with previous theoretical work [6], our results suggest that existing BRA methods share substantial similarity. Overall, the four metrics that we examined produced similar results, both with respect to whether and when they generate alerts indicating net favorability of one drug over another.



Scenario 2, Drug A more effective, equally safe; Scenario 3, Drug A safer, equally effective; Scenario 4, Drug A more effective, less safe; Scenario 5, Drug A safer, less effective; Scenario 6, Drug A more effective, safer

NNT|NNH, number needed to treat and number needed to harm; INB-MAR, incremental net benefit with maximum acceptable risk; INB-RVALY, INB with relative-value adjusted life years; INB-QALY, INB with quality-adjusted life years

Note: mean time to alerting is not presented for either drug in Scenario 1 (Null scenario) and for drug B in scenarios 2, 3, and 6 because alerting in these instances would indicate false positive alerting

Fig. 2 – Mean time to alerting for each of four benefit-risk assessment methods in five scenarios. (Color version of figure available online).

As expected, in certain situations, NNT|NNH tended to perform less well than did metrics that incorporate weights for different outcomes. For example, when drug A dominated drug B by being more effective and equally safe, the INB-based metrics indicated net favorability of drug A in approximately 80% of the iterations as compared with 72% for NNT|NNH. In scenarios in which drug A dominated drug B by being safer, NNT|NNH had the highest proportions of iterations indicating net favorability of drug A versus drug B when compared with the other metrics. This is because the NNT|NNH implicitly assumes equal weights among all outcomes and equal durations of impact of all outcomes. The beneficial outcomes, however, were generally more common (e.g., MI incidence of 7%) and had longer-lasting impacts (e.g., 20 years for MI and stroke vs. 1 month for major and minor bleeding). Thus, although NNT|NNH correctly identified net favorability of drug A in these situations, it likely exaggerated the net benefit.

INB-MAR and INB-RVALYs, which use similarly derived weights but differ with respect to whether they consider the duration of impact of the outcomes, performed similarly to each other. These INB approaches also performed similarly to INB-QALY, which considers the duration of impact of the outcomes but uses utility weights to calculate QALYs rather than preference weights derived from DCEs. In general, our results suggest that given the choice, stakeholders should opt for metrics that incorporate weights and account for the duration of outcome impact. We observed little difference between INB-RVALYs and INB-QALYs, which both account for the duration of outcome impact but use different types of weights. Although additional work is needed to examine correlations between stated preferences and utilities, we expect that these approaches result in similar rankings of outcomes. Thus, the decision to use INB-RVALYs or INB-QALYs may be best based on practical considerations, such as whether DCE results or utilities exist for a set of outcomes of interest or the practicability of obtaining these.

In general, time to alerting did not differ substantially across metrics. This is likely because the calculation of each metric is based on the same number of events for each outcome and it is the number of events that drives the CI estimation. Thus, the selection of BRA metrics for prospective benefit-risk monitoring should be based on the most appropriate weighting scheme for a particular monitoring activity.

Our simulation study has several limitations. First, we simulated data under the assumptions of no residual confounding and no misclassification of study variables. Although these assumptions are generally not tenable in real-world data studies, we do not expect confounding or misclassification biases to

differentially affect certain metrics and not others. As with any simulation study, the results of our study depend on the scenarios that we modeled. We created six scenarios. In one scenario (no. 1), we simulated no true difference between drugs A and B. In three scenarios (nos. 2, 3, and 6), we simulated situations in which drug A dominates drug B regardless of the outcome incidences and weights used. These scenarios allowed us to compare the performance of the four BRA metrics in situations in which the ground truth is known. Scenarios 4 and 5, however, represent situations in which either drug A or drug B could be considered net favorable, depending on one's preference for the given outcomes. Although we cannot make statements about which metrics perform best in these scenarios, they do highlight situations in which different methods could yield different results. In addition, in an attempt to compare the BRA metrics in scenarios with parameters reflective of a real-world example, the magnitudes of simulated benefits and risks were rather modest and may not resemble effects observed in other examples. Although we expect the overall discriminative ability of the metrics to improve in situations with more extreme effects, whether this changes the relative performance of the metrics should be investigated. Furthermore, the outcome-generating model used log-normal distributions to ensure that baseline cumulative incidences were greater than zero. Although this approach has been used in other studies using a similar simulation approach [7], it is possible that the results could be sensitive to the use of other distributions (e.g., gamma), which should be investigated in subsequent studies.

To make clear comparisons across the four methods without introducing substantial variation, we limited the scope of the simulation study to a base-case example with two beneficial outcomes, three harmful outcomes, fixed weights for these outcomes for each metric, and a fixed number of patients across 20 sequential cohorts in each iteration. We also used a single threshold based on a 99% CI for determining favorability. Although we based our model on realistic estimates and literature-derived weights and we varied the underlying incidences for each outcome and RRs for each outcome between drug A and drug B initiators across 1000 iterations for each scenario, future work should explore the effects of additional variation on the performance of BRA metrics for prospective drug monitoring.

Finally, although the prospective BRA monitoring framework is particularly suited for use by stakeholders who capture data on their own population, the results of these activities can be used to inform decision making in other settings, such as pharmacy and therapeutics or formulary committees. Because the analyses and data underlying the conduct of prospective monitoring can be complex, future efforts are needed to demonstrate the use of these analyses in clinical practice and the utility of the results of prospective BRA for informing health technology assessments.

In conclusion, we found that BRA metrics can be useful for identifying net favorability when applied to prospective monitoring of a newly marketed drug versus an existing alternative. When selecting a BRA metric for prospective monitoring, stakeholders should consider those that incorporate weights for outcomes of interest and the duration of impact of those outcomes. Prospective BRA monitoring may help stakeholders identify whether and when sufficient evidence has accrued to determine that a new drug has a benefit-risk profile that is better or worse than that of an alternative, which may support timely decision making in the postapproval setting.

Source of financial support: This study was funded by an unrestricted research grant from Novartis Pharmaceuticals Corporation to the Brigham and Women's Hospital.

REFERENCES

[1] Goldberg NH, Schneeweiss S, Kowal MK, Gagne JJ. Availability of comparative efficacy data at the time of drug approval in the United States. JAMA 2011;305:1786–9.
[2] Schneeweiss S, Gagne JJ, Glynn RJ, et al. Assessing the comparative effectiveness of newly marketed medications: methodological challenges and implications for drug development. Clin Pharmacol Ther 2011;90:777–90.
[3] Payers step in with 'real-world' comparative effectiveness research. MediMedia USA, 2011. Available from: http://www.managedcaremag.com/archives/1106/1106.cer.html. [Accessed October 6, 2014].
[4] Gagne JJ. You can observe a lot (about medical products) by watching (those who use them). Epidemiology 2013;24:700–2.
[5] Gagne JJ, Fireman B, Ryan PB, et al. Design considerations in an active medical product safety monitoring system. Pharmacoepidemiol Drug Saf 2012;21(Suppl. 1):32–40.
[6] Najafzadeh M, Schneeweiss S, Choudhry NK, et al. A unified framework for classification of methods for benefit-risk assessment. Value Health 2015;18:250–9.
[7] Gagne JJ, Rassen JA, Walker AM, et al. Active safety monitoring of new medical products using electronic healthcare data: selecting alerting rules. Epidemiology 2012;23:238–46.
[8] Rassen JA, Solomon DH, Glynn RJ, Schneeweiss S. Simultaneously assessing intended and unintended treatment effects of multiple treatment options: a pragmatic "matrix design". Pharmacoepidemiol Drug Saf 2011;20:675–83.
[9] Gagne JJ, Bykov K, Willke RJ, et al. Treatment dynamics of newly marketed drugs and implications for comparative effectiveness research. Value Health 2013;16:1054–62.
[10] 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;92:80–6.
[11] Wahl PM, Gagne JJ, Wasser TE, et al. Early steps in the development of a claims-based targeted healthcare safety monitoring system and application to three empirical examples. Drug Saf 2012;35:407–16.
[12] Gagne JJ, Rassen JA, Choudhry NK, et al. Near-real-time monitoring of new drugs: an application comparing prasugrel versus clopidogrel. Drug Saf 2014;37:151–61.
[13] The effect of low-dose warfarin on the risk of stroke in patients with nonrheumatic atrial fibrillation. The Boston Area Anticoagulation Trial for Atrial Fibrillation Investigators. N Engl J Med 1990;323:1505–11.
[14] Risk factors for stroke and efficacy of antithrombotic therapy in atrial fibrillation: analysis of pooled data from five randomized controlled trials. Arch Intern Med 1994;154:1449–57.
[15] Connolly SJ, Ezekowitz MD, Yusuf S, et al. Dabigatran versus warfarin in patients with atrial fibrillation. N Engl J Med 2009;361:1139–51.
[16] Patel MR, Mahaffey KW, Garg J, et al. Rivaroxaban versus warfarin in nonvalvular atrial fibrillation. N Engl J Med 2011;365:883–91.
[17] Granger CB, Alexander JH, McMurray JJ, et al. Apixaban versus warfarin in patients with atrial fibrillation. N Engl J Med 2011;365:981–92.
[18] Najafzadeh M, Gagne JJ, Choudhry NK, et al. Patients' Preferences in Anticoagulant Therapy: A Discrete Choice Experiment. Circ Cardiovasc Qual Outcomes 2014;7:912–9.
[19] Levin LA, Wallentin L, Bernfort L, et al. Health-related quality of life of ticagrelor versus clopidogrel in patients with acute coronary syndromes—results from the PLATO trial. Value Health 2013;16:574–80.
[20] Pickard AS, Johnson JA, Feeny DH, et al. Agreement between patient and proxy assessments of health-related quality of life after stroke using the EQ-5D and Health Utilities Index. Stroke 2004;35:607–12.
[21] Spiegel BM, Chiou CF, Ofman JJ. Minimizing complications from nonsteroidal antiinflammatory drugs: cost-effectiveness of competing strategies in varying risk groups. Arthritis Rheum 2005;53:185–97.
[22] Shah SV, Gage BF. Cost-effectiveness of dabigatran for stroke prophylaxis in atrial fibrillation. Circulation 2011;123:2562–70.