VALUE IN HEALTH 16 (2013) 1054–1062

Available online at www.sciencedirect.com

ScienceDirect

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

Comparative Effectiveness Research/Health Technology Assessment (HTA)

Treatment Dynamics of Newly Marketed Drugs and Implications for Comparative Effectiveness Research

Joshua J. Gagne, PharmD, ScD[1,], Katsiaryna Bykov, PharmD, MS[1], Richard J. Willke, PhD[2], Kristijan H. Kahler, RPh, PhD[3], Prasun Subedi, PhD[4], 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]*Center for Health Economics, Epidemiology, and HTA, Global Market Access, Primary Care Business Unit, Pfizer, Inc., New York, NY, USA;* [3]*Outcomes Research Methods & Analytics, Novartis Pharmaceuticals Corporation, East Hanover, NJ, USA;* [4]*Worldwide Policy, Pfizer, Inc., New York, NY, USA*

ABSTRACT

Objectives: Clinicians and payers require rapid comparative effectiveness (CE) evidence generation to inform decisions for new drugs. We empirically assessed treatment dynamics of newly marked drugs and their implications for conducting CE research. **Methods:** We used claims data to evaluate five drug-outcome pairs: 1) raloxifene (vs. alendronate) and fracture; 2) risedronate (vs. alendronate) and fracture; 3) simvastatin plus ezetimibe fixed-dose combination (simvastatin + ezetimibe) (vs. simvastatin alone) and cardiovascular events; 4) rofecoxib (vs. nonselective nonsteroidal anti-inflammatory drugs [ns-NSAIDs]) and myocardial infarction; and 5) rofecoxib (vs. ns-NSAIDS) and gastrointestinal bleed. We examined utilization dynamics in the early marketing period, including evolving utilization patterns, outcome risk among those treated with new versus established drugs, and prior treatment patterns that may indicate treatment resistance or intolerance. We addressed these challenges by replicating active CE monitoring with sequential matched cohort analysis. **Results:** Patients initiating new drugs were more likely to have used other drugs for the same indication in the past, but the majority of patients in all new drug cohorts were treatment naive (82.0% overall). Patients initiating rofecoxib had higher predicted baseline risk of gastrointestinal bleed than did patients initiating ns-NSAIDs. Patients initiating risedronate and alendronate had similar predicted baseline risks of fracture, while those initiating raloxifene and simvastatin + ezetimibe had lower risks of outcomes of interest relative to their comparators. Prospective monitoring yielded results consistent with expectation for each example. **Conclusions:** Many challenges to assessing the CE of new drugs are borne out in empirical data. Attention to these challenges can yield valid CE results.

Keywords: effectiveness, new drugs, prospective monitoring, validity.

Introduction

Optimizing patient-centered outcomes requires more than knowing whether a medication works better than placebo in highly protocolized clinical trial settings [1]. At the time of approval, however, few new drugs have been compared with active alternatives [2] and among those that have, a limited set of alternatives is used even when many potential treatment options exist [3]. Patients, clinicians, and payers therefore do not have all the evidence required for fully informed treatment decisions involving new drugs.

Several mechanisms exist to generate comparative effectiveness (CE) evidence in the early marketing period, but they carry important limitations. Large head-to-head trials of multiple drugs tend to be costly and require many years to complete. The generalizability of results of open-label extensions of phase III trials and indirect comparisons or network meta-analyses of efficacy trials [4] are limited to populations enrolled in the preapproval trials [5], which do not tend to reflect populations of patients that use the drugs in the postmarketing setting [6–8].

Once approved, new drugs are often consumed by many thousands of patients despite the lack of evidence of so-called real-world effectiveness. Patients' prescription-filling histories and health care encounters are captured in near real time in payers' and providers' electronic health care databases. Analyzing these data as they accrue offers great potential for providing continuous information support for decision makers and other stakeholders in a "learning health care system" [9]. These data can provide in-the-moment insight into the comparative effectiveness and safety of new drugs as experience with the products grows [10,11], which could form the basis of coverage with evidence development strategies, including risk-sharing arrangements [12], and are being used in the US Food and Drug Administration's Sentinel Initiative to assess the safety of newly approved drugs [13]. Using observational data to determine the comparative effectiveness of new

* *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, USA.

E-mail: jgagne1@partners.org.

1098-3015/$36.00 – see front matter Copyright © 2013, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

http://dx.doi.org/10.1016/j.jval.2013.05.008

drugs in the early marketing period is challenging [5,11]. In particular, selective prescribing of new drugs may lead to confounding by indication and patients initiating a new drug may be more likely to have failed a prior treatment [5].

By using data covering the early marketing periods of four example drugs (rofecoxib, raloxifene, risedronate, and fixed-dose simvastatin plus ezetimibe combination), we empirically evaluated aspects of treatment dynamics in the early marketing period that can give rise to confounding in observational comparative effectiveness studies. Specifically, we sought to examine 1) the utilization trends of newly marketed drugs, 2) whether users of new drugs are "sicker" than users of more established therapies, 3) whether users of new drugs are more likely to have failed prior treatments, and 4) whether it is possible to find patients initiating alternative drugs who are comparable to patients initiating the new drugs. We then applied a recently proposed ensemble of methods to emulate active monitoring beginning at market authorization of each drug of interest [5,14,15] and assessed whether the approach addressed these sources of confounding.

Methods

Data

We used 12 years (1994–2005) of medical and pharmacy claims data from Medicare beneficiaries in New Jersey and Pennsylvania who were enrolled in pharmaceutical assistance programs in these states (the Pharmacy Assistance for the Aged and Disabled [PAAD] in New Jersey and the Pharmacy Assistance Contract for the Elderly [PACE] in Pennsylvania). Both PAAD and PACE provide medications with no formulary restrictions and at minimal expense to elderly individuals with low income but who do not meet the Medicaid annual income threshold. PACE and PAAD data are linked to Medicare Parts A and B data.

Patients

We identified initiators of four drugs of interest beginning at their market authorization and initiators of an active comparator drug or class for each: 1) rofecoxib versus nonselective nonsteroidal anti-inflammatory drugs (ns-NSAIDs), 2) raloxifene versus alendronate, 3) risedronate versus alendronate, and 4) simvastatin plus ezetimibe fixed-dose combination (Vytorin; simvastatin + ezetimibe) versus simvastatin alone. Each pair was chosen to create a set of examples with new drugs that represent slightly different types of advances on the comparator. For example, rofecoxib offered a more selective mechanism of action over ns-NSAIDs; raloxifene offers a different mechanism of action than does alendronate; risedronate is a follow-on product in the same class as alendronate; and simvastatin plus ezetimibe is a combination product containing an established therapy and is thought to be more effective for some patients, though conclusive evidence is lacking. Pairs were also selected to ensure that the comparator had been available for some time prior to the authorization of the new drug.

Outcomes

Our outcomes of interest were myocardial infarction (MI) and gastrointestinal (GI) bleeding for the rofecoxib versus ns-NSAIDs example, a composite fracture end point for the raloxifene versus alendronate and risedronate versus alendronate examples, and a composite cardiovascular event end point for the simvastatin plus ezetimibe fixed-dose combination versus simvastatin alone. The composite fracture outcome involved fractures at the hip, humerus, pelvis, and radius, and the cardiovascular event outcome comprised MI, cerebrovascular events (i.e., ischemic and hemorrhagic stroke), and acute coronary syndromes with revascularization. We defined all outcomes by using claims-based algorithms. Validation studies comparing the algorithms to medical chart reviews have found positive predictive values of 94% for MI [16], 88% for GI bleeds [17], 86% for cerebrovascular events [18], and between 93% and 98% for each fracture outcome [19,20].

For each example, we followed patients for the outcome(s) of interest by using an intention-to-treat approach beginning the day after the initiation of their index drug and for a maximum of 180 days. We censored patients at the first of the following events: 1) occurrence of the outcome of interest, 2) death, or 3) end of study period (December 31, 2005).

Analysis

We examined trends in utilization by plotting the number of initiators of the new drugs, the comparators, and similar drugs in each calendar quarter starting 1 year before the new drugs' market authorization. We defined initiators as those patients with no prior use of the index drug in the preceding 180 days [21]. We characterized prior treatment patterns by calculating the proportion of patients initiating the new drugs and their respective comparators who had exposure to other drugs used to treat the same condition in the 180 days before index drug initiation. Specifically, we calculated the proportion of patients in the new drug group who had used the comparator and vice versa, the proportion of patients in each group who had used various alternatives (e.g., celecoxib and valdecoxib for the rofecoxib example, other statins for the simvastatin + ezetimibe example, and other bisphosphonates or calcitonin for the raloxifene and risedronate examples), and the proportion of patients in each group who were treatment naive.

We tabulated demographic and clinical characteristics of patients initiating the new drugs in the very early marketing period (i.e., the first 6 months following market authorization) and compared the characteristics to those initiating the active comparator during the same time. We identified a large number of predefined covariates for each example, including patient demographics (e.g., age, gender, and race), health service utilization variables (e.g., number of physician visits, number of hospitalizations, number of unique drugs dispensed, and a comorbidity score that combines the Charlson and Elixhauser indices [22]), and specific risk factors for the outcome(s) of interest in each example. All covariates were ascertained during the 180 days preceding the index prescription date and are listed in the Appendix Tables in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2013.05.008. To compare the baseline risk of the outcomes of interest between treatment groups, we estimated a disease risk score, defined as a patient's likelihood of experiencing the out come of interest conditional on baseline covariates [23–26]. Follow ing the approach of Glynn et al. [26], we developed the disease risk score model for each example among patients exposed to the comparator drug prior to the market authorization of the new drug. We then applied the resulting model coefficients in the form of a prediction rule to all patients in both treatment groups after the introduction of the new drug. We considered all predefined covariates listed in the Appendix Tables in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2013.05.008.

Finally, we replicated prospective comparative effectiveness monitoring by analyzing each example as if new data became available on a quarterly basis following the introduction of each new drug. We updated estimates of effect over time as experience with the new drug grows [14,15,27]. We divided the databases into sequential data sets defined by claims occurring in each calendar quarter. For each example, patients who initiated the new drug of interest within 6 months of its market authorization formed monitoring period one. Subsequent monitoring periods were defined by 3-month intervals. We 1:1 matched initiators of the new drug in each monitoring period to initiators of the

comparator product with index dates in the same period by using a traditional propensity score (PS) and a high-dimensional PS (hd-PS). PSs are summary scores that reduce the dimensionality of confounding by combining potential confounders into a single scalar variable [28]. The hd-PS is a semi-automated algorithm that identifies and selects for inclusion in the PS model a large number of variables that behave like confounders based on empirical associations with the exposure and outcome of interest [29]. While not necessarily structural confounders, these empirically identified variables may act as confounders because of chance imbalances with the exposure and outcome of interest in the database. The traditional PS model included all baseline covariates listed in the Appendix Tables in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2013.05.008. The hd-PS model included all variables in the Appendix along with the empirically identified covariates [29]. We estimated both hazard ratios and cumulative incidence differences (i.e., risk differences) to compare event frequencies between treatment groups. We reestimated the effect estimates and their 95% confidence interval after adding each sequential data set for each monitoring period.

Results

Among Medicare beneficiaries in New Jersey and Pennsylvania, we identified 65,370 initiators of rofecoxib between the second quarter of 1999 and the third quarter of 2004, 13,173 initiators of raloxifene between 1998 and the end of 2005, 23,341 initiators of risedronate between the second quarter of 2000 and the end of 2005, and 8,288 initiators of the simvastatin + ezetimibe combination between the third quarter of 2002 and the end of 2005 (Table 1).

Utilization

Rofecoxib was rapidly taken up in the early marketing period, reaching a peak in the second quarter of 2000 (Fig. 1). The number of new users per quarter decreased thereafter until 2004 when the drug was withdrawn from the market. Raloxifene exhibited a similar pattern, with an initial rapid uptake followed by a period of stabilization and subsequent decline (Fig. 2). The erosion in raloxifene use appeared to correspond to the introduction and increased use of risedronate. We also observed a period of rapid uptake and stabilization of risedronate, but this was followed by continued growth corresponding to the approval of a once-weekly formulation in 2002 (Fig. 2). A similar uptick was observed with alendronate when the once-weekly formulations were approved at the end of 2000.

Prior Treatment Patterns

As compared to patients initiating the more established therapies, patients initiating the new drugs were more likely to have been previously exposed to drugs in the other comparison groups (Table 1). This was most pronounced among patients initiating the simvastatin + ezetimibe combination where 40% had been exposed to simvastatin alone in the prior 180 days as compared with 3% of patients initiating simvastatin alone with prior use of the simvastatin + ezetimibe combination. Among patients who initiated rofecoxib, 9% had had prior exposure to ns-NSAIDs versus 1% of patients initiating ns-NSAIDs with prior rofecoxib exposure.

Patients initiating the new drugs were also more likely to have been exposed to other drugs for the same indication in the prior 180 days. Among patients initiating rofecoxib, 17% had had a prescription for celecoxib or valdecoxib in the preceding 180 days versus 13% of patients among the ns-NSAID group. As a result, fewer patients in the new drug group were treatment naive as compared to patients in the active comparator groups. In all instances except the simvastatin + ezetimibe example, greater than 70% of new drug initiators were treatment naive in the 180 days prior to index drug initiation.

Differences in Patient Characteristics and Baseline Risk

In general, patients initiating rofecoxib exhibited signs of being sicker or at higher risk of outcomes of interest (Fig. 3) and had higher predicted baseline probabilities of both GI bleed and MI within 6 months as compared to patients initiating ns-NSAIDs (Fig. 4). Patients initiating risedronate and alendronate were nearly identical on patient characteristics and baseline risk factors (Figs. 3 and 4). Patients initiating raloxifene and simvastatin + ezetimibe were generally healthier and had lower predicted baseline probabilities of outcomes of interest than did patients initiating alendronate and simvastatin alone, respectively (Figs. 3 and 4). Differences in predicted baseline probabilities of outcomes between treatment groups did not materially change across periods for each example. Despite differences in baseline risk factors and predicted outcome probabilities, we were able to find hd-PS matches for the majority of treatment-naive patients in each period for each example (see the Appendix Figure in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2013.05.008). hd-PS matching made the treatment groups more similar on baseline risk of each outcome (Fig. 4).

Table 1 – Patterns of use of drugs for the same indication in the 180 days prior to the initiation of each of four new drugs and their comparators.

	Total no. of initiators	Prior use of drug(s) in other comparison group, n (% of total)	Prior use of other drugs for same indication, n (% of total)	Treatment naive, n (% of total)
Rofecoxib	65,370	6,107 (9.3)	11,347 (17.4)	45,932 (70.3)
ns-NSAIDs	96,259	1,146 (1.2)	12,789 (13.3)	81,788 (85.0)
Raloxifene	13,173	499 (3.8)	1,414 (10.7)	11,160 (84.7)
Alendronate	62,699	403 (0.6)	5,402 (8.6)	56,755 (90.5)
Risedronate	23,341	1,189 (5.1)	2,668 (11.4)	19,239 (82.4)
Alendronate	45,075	202 (0.5)	4,193 (9.3)	40,527 (89.9)
Simvastatin + ezetimibe	8,228	3,310 (40.2)	2,107 (25.6)	2,811 (34.2)
Simvastatin alone	14,590	409 (2.8)	2,941 (20.2)	11,240 (77.0)

ns-NSAIDs, nonselective nonsteroidal anti-inflammatory drugs.

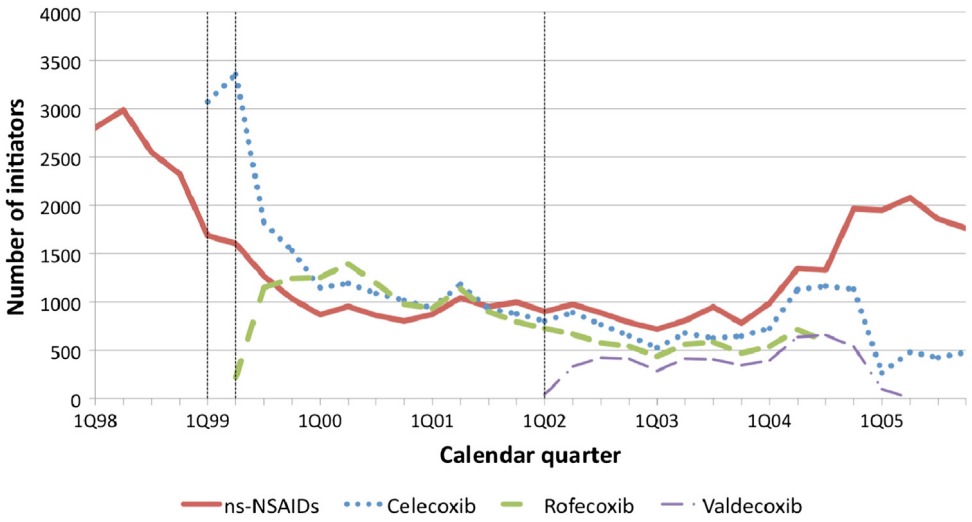

Fig. 1 – Utilization of nonsteroidal anti-inflammatory medications among Medicare beneficiaries in two US states (1998–2005). Dotted vertical lines indicate the first quarter in which prescriptions for each drug appeared in the database. ns-NSAIDs, nonselective nonsteroidal anti-inflammatory drugs.

Emulated Prospective Monitoring

Figure 5 demonstrates the results of the rofecoxib versus ns-NSAID comparison on both MI and GI bleed outcomes, replicating sequential analyses as if the data accrued prospectively, using an hd-PS–matched sequential incident user cohort approach. Overall, as compared to ns-NSAIDS, rofecoxib was associated with a lower rate of GI bleed (6-month risk difference −0.00082; 95% confidence interval −0.00232 to 0.00069; corresponding to 8 fewer GI bleed events for every 10,000 patients treated with rofecoxib instead of ns-NSAIDs) and a higher rate of MI (6-month risk difference 0.00229; 95% confidence interval 0.00019–0.00438; corresponding to 23 excess MIs for every 10,000 patients treated with rofecoxib instead of ns-NSAIDs). The direction and magnitude of these results became evident during the monitoring process several years before the last period. Although the crude estimate for GI bleed was elevated among rofecoxib initiators (Fig. 6), reflecting the underlying confounding by higher risk status, the fully adjusted results suggest a protective effect of the drug, consistent with expectation.

In unadjusted analyses, we also observed lower rates of outcomes among raloxifene and simvastatin + ezetimibe initiators, reflecting confounding made apparent by the large differences in baseline outcome risks (Fig. 6). After hd-PS matching we observed no difference in fracture rates among patients treated with raloxifene versus alendronate and no differences in cardiovascular event rates among patients treated with simvastatin + ezetimibe versus simvastatin alone. Neither PS matching nor hd-PS matching materially changed the effect estimate for risedronate because this association was not very confounded.

Discussion

Observational studies of treatment effects are vulnerable to various biases, particularly when conducted for new drugs shortly

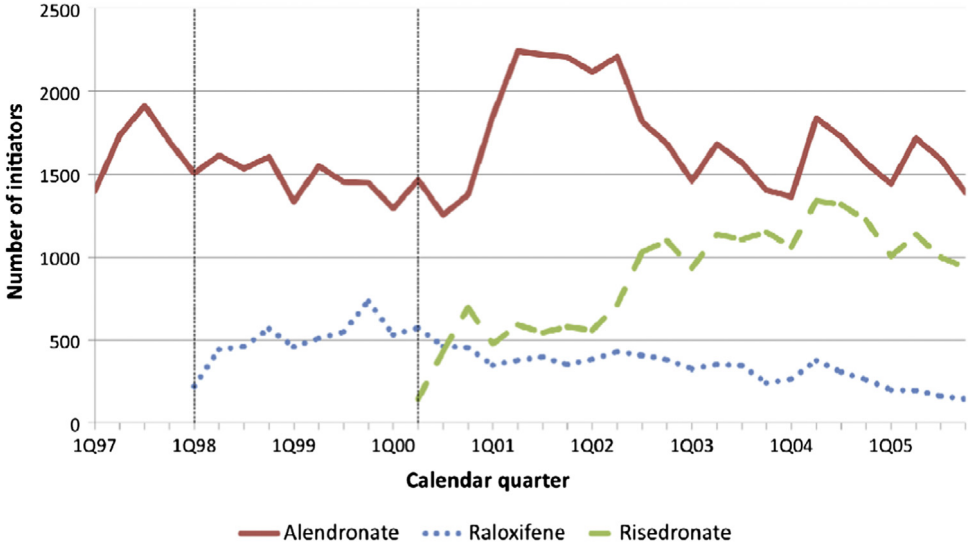

Fig. 2 – Utilization of alendronate, raloxifene, and risedronate among Medicare beneficiaries in two US states (1997–2005). Dotted vertical lines indicate the first quarter in which prescriptions for each drug appeared in the database.

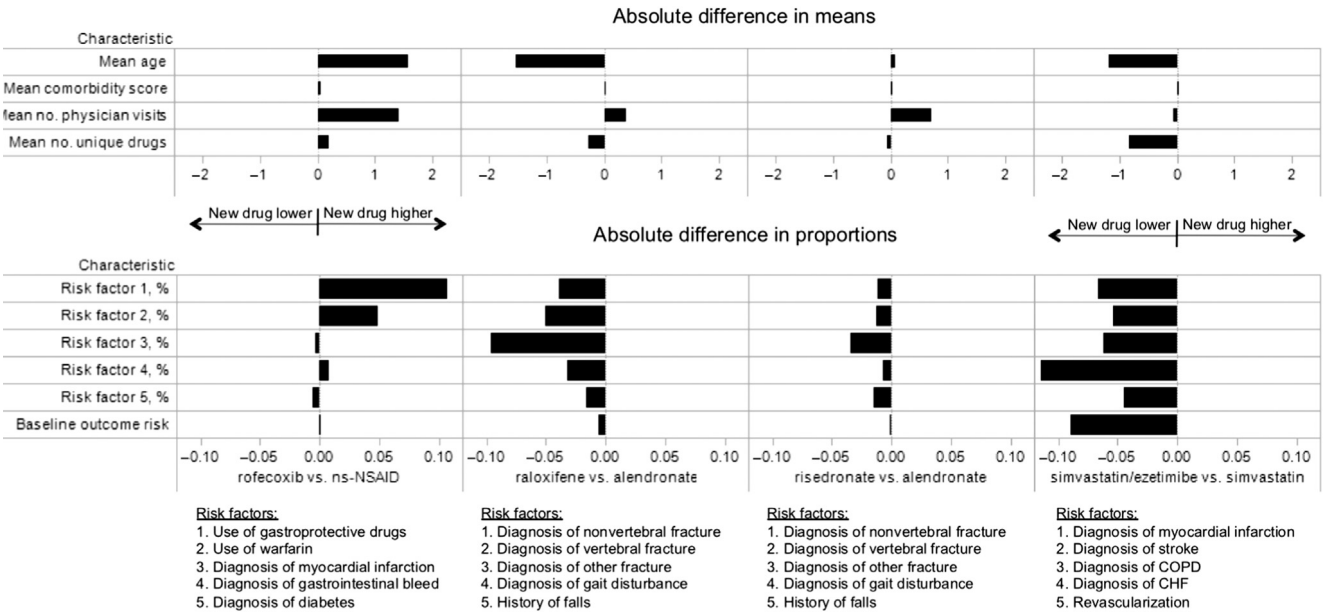

Fig. 3 – Balance in key baseline characteristics prior to matching among initiators in the first 6 months following each new drug's market authorization. CHF, congestive heart failure; COPD, chronic obstructive pulmonary disease; ns-NSAIDs, nonselective nonsteroidal anti-inflammatory drugs.

after their marketing authorization, when examining anticipated or intended outcomes, and when using claims data that are not collected for research purposes. Across four examples we found that the utilization of new drugs does not follow a predictable pattern, that new drug initiators are more likely to have been exposed to other drugs that can be used to treat the same indication, and that new drug initiators can differ with respect to important confounders and baseline outcome risk from patients initiating more established therapies. These dynamics can pose threats to observational studies aimed at generating comparative effectiveness and safety information in the early marketing phase. However, an hd-PS–matched incident user cohort design can overcome many of these challenges to produce valid and transparent comparative effectiveness evidence as experience with the new drugs grows in near real time, providing continuous decision-support information. Despite the many differences in initiators of new drugs versus the more established therapies, we were able to identify adequate matches for the majority of new drug initiators, even in the time immediately following market authorization. The approach may be particularly useful for monitoring prespecified drug-outcome pairs in routine surveillance activities, including the US Food and Drug Administration's developing Sentinel System [13–15,27,30].

Confounding is a key threat to generating valid comparative evidence from observational data, particularly for anticipated and intended outcomes for which prognoses are incorporated into the treatment selection decision. Patients initiating rofecoxib had a higher average baseline risk of GI bleed than did patients initiating ns-NSAIDs, reflecting the preferential prescribing of cyclooxygenase inhibitors to those at risk for gastrotoxicity. PS matching and hd-PS matching moved the unadjusted estimate of GI bleeding risk in the direction consistent with what has been observed in randomized trials [31]. Differences in the predicted risk of MI, which likely would not have been anticipated at the time of prescribing, were less pronounced, though rofecoxib patients were slightly older than ns-NSAID patients and age is a strong risk factor for MI. PS matching and hd-PS matching enabled us to examine multiple outcomes among the same matched cohorts while simultaneously adjusting for confounding of both associations.

Predicted baseline risks of fracture among patients initiating risedronate versus alendronate were nearly identical, reflecting very similar use of drugs that are perceived to be similar and demonstrating the value of a well-selected active comparator. Comparing drugs that share similar treatment selection processes can greatly reduce the potential for confounding [10]. Patients initiating raloxifene and simvastatin + ezetimibe had lower predicted probabilities of outcomes at baseline relative to patients initiating their comparators. These differences were removed by PS matching and, even more so, by hd-PS matching.

Patients initiating a new drug instead of a more established therapy may be more likely to have tried other drugs for the same indication but have failed on them or not tolerated them well. Indeed, we observed that patients initiating a new drug were more likely to have experience with other drugs for the same indication. However, the majority of patients initiating each of the new drugs and their active comparators were treatment naive for that indication. The exception was the simvastatin + ezetimibe fixed-dose combination where many patients had experience with simvastatin, ezetimibe, or both. Nevertheless, we were able to identify enough patients in each sequential period to enable monitoring among those who were treatment naive for each example.

By combining various design and analytic features [32], the hd-PS–matched incident user approach to active monitoring addresses many of the challenges in assessing the comparative effectiveness and safety of drugs in the period shortly after their market authorization. Focusing on incident users, and more specifically patients who are treatment naive, overcomes concerns about patients initiating the new drug being more likely to be resistant or intolerant to treatment and helps ensure that the analysis answers the most relevant question for decision makers [33,34]. Using active comparators with as similar indications and use as possible reduces the degree to which confounding by indication can bias results. As expected, we observed the best marginal balance in baseline covariates and predicted baseline outcome risk between initiators of drugs from the same class. PS matching and hd-PS matching further enhanced balance on baseline risk factors and facilitated simultaneous balancing of predicted baseline risk of multiple outcomes.

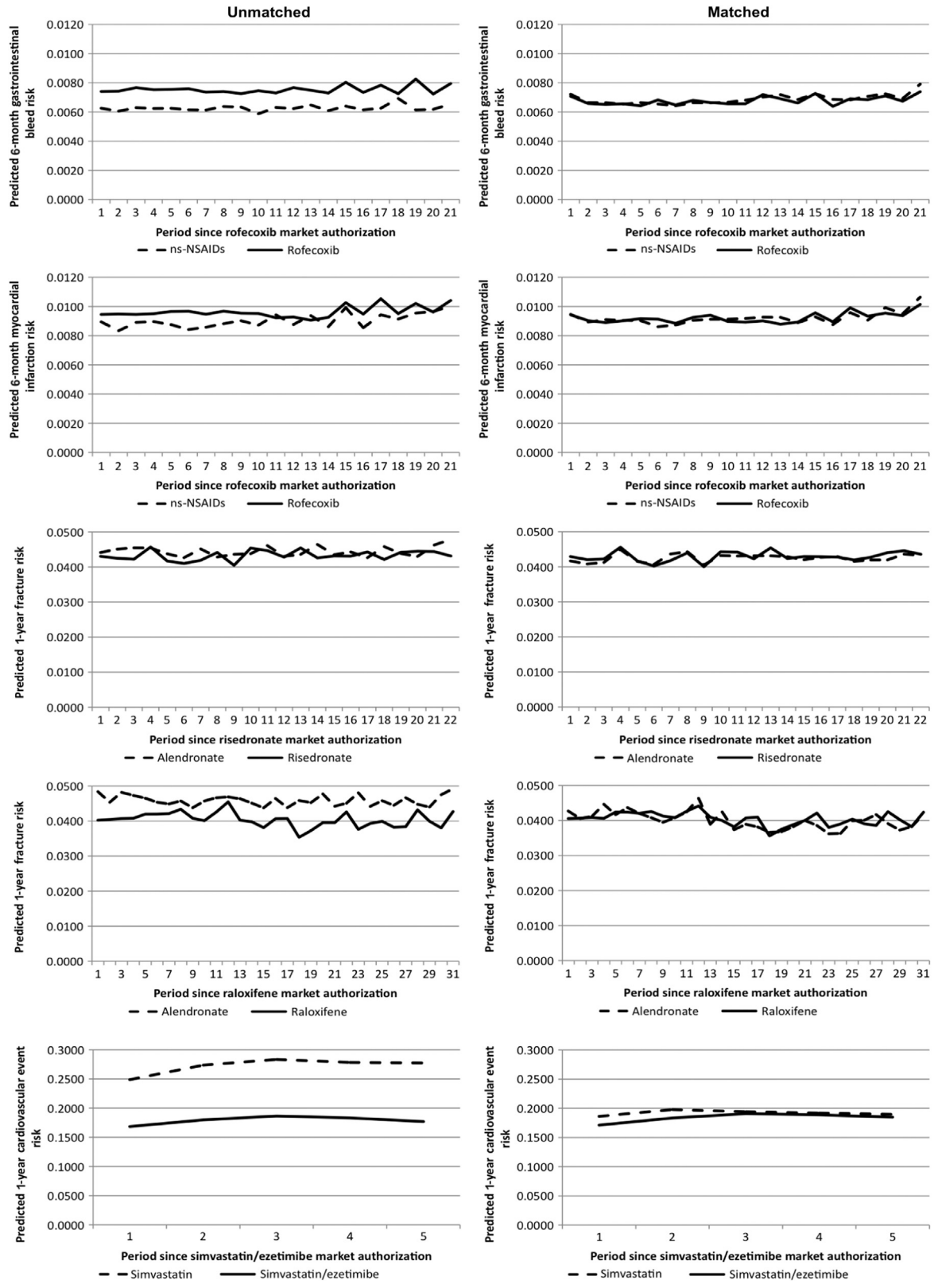

Fig. 4 – Baseline predicted probability of each outcome among unmatched and matched cohorts. Each panel displays the predicted baseline probability of the outcome of interest (disease risk score) among initiators of the new drug (solid line) and the active comparator (dashed line) in each calendar quarter following introduction of the new drug. Panels on the left are based on the full cohort, prior to matching, and the panels on the right are after matching on a high-dimensional propensity score. ns-NSAIDs, nonselective nonsteroidal anti-inflammatory drugs.

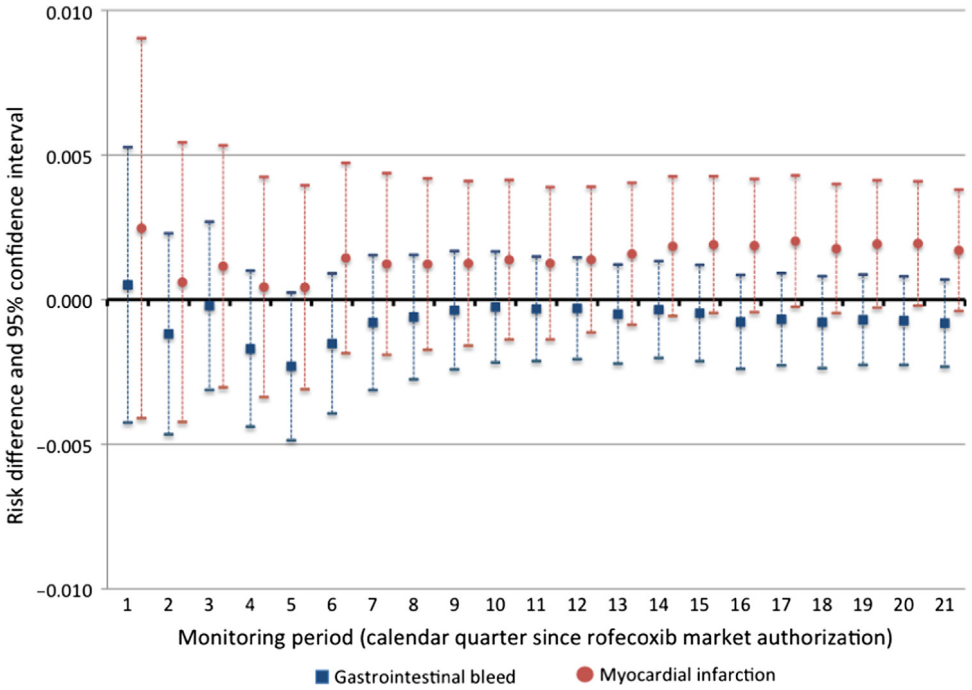

Fig. 5 – Replicated prospective monitoring comparing rofecoxib and non-selective non-steroidal anti-inflammatory drugs. Each point represents a risk difference for gastrointestinal bleed (blue) or myocardial infarction (red) among initiators of rofecoxib matched by high-dimensional propensity score to initiators of nonselective nonsteroidal anti-inflammatory drugs. Risk differences are cumulative in that they are based on all data that had accrued in the database up to and including that monitoring period. Dotted lines represent the 95% confidence intervals.

While our empirical assessments corroborated many concerns that have been raised about comparing the safety and effectiveness of newly marketed drugs in observational data, our proposed approach to active monitoring in this setting appears to mitigate potential bias related to these concerns. Nevertheless, we evaluated only four outcomes, among four new drugs intended to treat chronic conditions, with three active comparators in two states' low-income Medicare prescription assistance populations. Therefore, our findings may not always generalize to other examples. In particular, appropriate active comparators may not be available for some new drugs, especially those that are the first approved for a given indication, which will limit the ability to address confounding. In addition, the Medicare beneficiaries that we studied received prescription drug coverage through pharmaceutical benefits programs that impose few financial barriers to drug access. As such, utilization and channeling of new drugs may

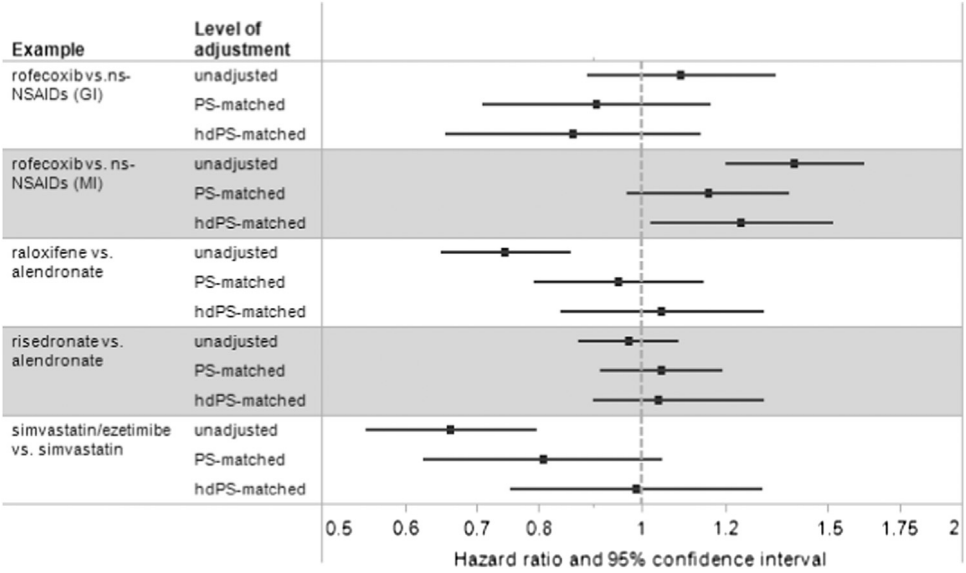

Fig. 6 – Overall hazard ratios and 95% confidence intervals for each of five examples using three levels of confounding adjustment. hdPS, high-dimensional propensity score; GI, gastrointestinal bleed; MI, myocardial infarction; ns-NSAIDs, nonselective nonsteroidal anti-inflammatory drugs; PS, propensity score.

differ in other systems, such as when payers require higher co-payments for new drugs or exclude them entirely from a formulary.

The limitations of claims data are well described [10,11]. The data do not include information on whether patients actually consumed their dispensed medications, on over-the-counter medication use, or on other clinically important potential confounders such as smoking status, physical activity, body mass index, or indication for drug use, such as bone density and cholesterol levels. In addition, we compared the results of each active monitoring example to expectations based on prior evidence and clinical judgment. Large randomized trials, however, do not exist for each comparison. This further highlights the need for observational comparative effectiveness research but also makes it difficult to establish benchmarks. Moreover, clinical trials and observational studies often include different patient populations. To the extent that treatment effect heterogeneity exists among patient subgroups, results of the different study types may not be comparable even if each is perfectly internally valid for its specific population. Nevertheless, after addressing sources of potential bias, our five assessments yielded results consistent with our *a priori* expectations.

In conclusion, understanding the comparative effectiveness and safety of new drugs requires an ensemble of data from multiple sources and various analytic approaches [5]. Observational studies in electronic health care databases can provide one important piece of this evidence base, particularly in the early marketing period of a new drug when data from other sources are lacking. Analyses that use rigorous design and analytic features can overcome or reduce threats to validity and provide rapid and useful information to stakeholders.

Acknowledgments

The authors thank Professor Adrian Towse, Director, Office of Health Economics, for valuable discussions that led to the commencement of this project.

Source of financial support: This work was funded by an unrestricted grant from Pfizer, Inc. Although the sponsor was involved in the design, data collection, analysis, and interpretation of the study results, the decision to submit the manuscript for publication was made by the authors independently. Drs. Willke and Subedi are employees of Pfizer. Dr. Kahler is an employee of Novartis. No Pfizer or Novartis products were studied in this research. Dr. Gagne had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Supplemental Materials

Supplemental material accompanying this article can be found in the online version as a hyperlink at http://dx.doi.org/10.1016/j.jval.2013.05.008 or, if a hard copy of article, at www.valueinhealthjournal.com/issues (select volume, issue, and article).

REFERENCES

[1] Avorn J. In defense of pharmacoepidemiology—embracing the yin and yang of drug research. N Engl J Med 2007;357:2219–21.
[2] 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.
[3] Gagne JJ, Choudhry NK. How many "me-too" drugs is too many? JAMA 2011;305:711–2.
[4] Schneeweiss S, Gagne JJ, Patrick AR, et al. Comparative efficacy and safety of new oral anticoagulants in patients with atrial fibrillation. Circ Cardiovasc Qual Outcomes 2012;5:480–6.
[5] 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.
[6] Blosser CD, Huverserian A, Bloom RD, et al. Age, exclusion criteria, and generalizability of randomized trials enrolling kidney transplant recipients. Transplantation 2011;91:858–63.
[7] Gurwitz JH, Col NF, Avorn J. The exclusion of the elderly and women from clinical trials in acute myocardial infarction. JAMA 1992;268:1417–22.
[8] Travers J, Marsh S, Williams M, et al. External validity of randomised controlled trials in asthma: to whom do the results of the trials apply? Thorax 2007;62:219–23.
[9] Greene SM, Reid RJ, Larson EB. Implementing the learning health system: from concept to action. Ann Intern Med 2012;157:207–10.
[10] Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. J Clin Epidemiol 2005;58:323–37.
[11] Schneeweiss S. Developments in post-marketing comparative effectiveness research. Clin Pharmacol Ther 2007;82:143–56.
[12] Neumann PJ, Chambers JD, Simon F, Meckley LM. Risk-sharing arrangements that link payment for drugs to health outcomes are proving hard to implement. Health Aff (Millwood) 2011;30:2329–37.
[13] Platt R, Carnahan RM, Brown JS, et al. The U.S. Food and Drug Administration's Mini-Sentinel program: status and direction. Pharmacoepidemiol Drug Saf 2012;21(Suppl. 1):1–8.
[14] 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.
[15] 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.
[16] Kiyota Y, Schneeweiss S, Glynn RJ, et al. Accuracy of Medicare claims-based diagnosis of acute myocardial infarction: estimating positive predictive value on the basis of review of hospital records. Am Heart J 2004;148:99–104.
[17] Wahl PM, Rodgers K, Schneeweiss S, et al. Validation of claims-based diagnostic and procedure codes for cardiovascular and gastrointestinal serious adverse events in a commercially-insured population. Pharmacoepidemiol Drug Saf 2010;19:596–603.
[18] Tirschwell DL, Longstreth WT Jr. Validating administrative data in stroke research. Stroke 2002;33:2465–70.
[19] Gagne JJ, Patrick AR, Mogun H, Solomon DH. Antidepressants and fracture risk in older adults: a comparative safety analysis. Clin Pharmacol Ther 2011;89:880–7.
[20] Ray WA, Griffin MR, Fought RL, Adams ML. Identification of fractures from computerized Medicare files. J Clin Epidemiol 1992;45:703–14.
[21] Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. Am J Epidemiol 2003;158:915–20.
[22] Gagne JJ, Glynn RJ, Avorn J, et al. A combined comorbidity score predicted mortality in elderly patients better than existing scores. J Clin Epidemiol 2011;64:749–59.
[23] Arbogast PG, Ray WA. Use of disease risk scores in pharmacoepidemiologic studies. Stat Methods Med Res 2009;18:67–80.
[24] Cadarette SM, Gagne JJ, Solomon DH, et al. Confounder summary scores when comparing the effects of multiple drug exposures. Pharmacoepidemiol Drug Saf 2010;19:2–9.
[25] Hansen BB. The prognostic analogue of the propensity score. Biometrika 2008;95:8.
[26] Glynn RJ, Gagne JJ, Schneeweiss S. Role of disease risk scores in comparative effectiveness research with emerging therapies. Pharmacoepidemiol Drug Saf 2012;21(Suppl. 2):138–47.
[27] Gagne JJ, Walker AM, Glynn RJ, et al. An event-based approach for comparing the performance of methods for prospective medical product monitoring. Pharmacoepidemiol Drug Saf 2012;21:631–9.
[28] Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70:41–55.
[29] Schneeweiss S, Rassen JA, Glynn RJ, et al. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. Epidemiology 2009;20:512–22.
[30] Rassen JA, Schneeweiss S. Using high-dimensional propensity scores to automate confounding control in a distributed medical product safety surveillance system. Pharmacoepidemiol Drug Saf 2012;21(Suppl. 1):41–9.

[31] Bombardier C, Laine L, Reicin A, et al. Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. VIGOR Study Group. N Engl J Med 2000;343:1520–8, 2 p following 8.
[32] 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.
[33] Hernan MA, Alonso A, Logan R, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. Epidemiology 2008;19:766–79.
[34] Johnson ES, Bartman BA, Briesacher BA, et al. The incident user design in comparative effectiveness research. Pharmacoepidemiol Drug Saf 2013;22:1–6.