Using Electronic Health Care Records for Drug Safety Signal Detection

A Comparative Evaluation of Statistical Methods

Martijn J. Schuemie, PhD, Preciosa M. Coloma, MD, MS,* Huub Straatman, MS,† Ron M. C. Herings, PhD,*† Gianluca Trifirò, MD, PhD,*‡ Justin Neil Matthews, MS,§ David Prieto-Merino, PhD,§ Mariam Molokhia, PhD,‖ Lars Pedersen, PhD,¶ Rosa Gini, PhD,# Francesco Innocenti, MS,#** Giampiero Mazzaglia, MD, PhD,** Gino Picelli,†† Lorenza Scotti, MS,‡‡ Johan van der Lei, MD, PhD,* and Miriam C. J. M. Sturkenboom, PhD**

Background: Drug safety monitoring relies primarily on spontaneous reporting, but electronic health care record databases offer a possible alternative for the detection of adverse drug reactions (ADRs).

Objectives: To evaluate the relative performance of different statistical methods for detecting drug-adverse event associations in electronic health care record data representing potential ADRs.

Research Design: Data from 7 databases across 3 countries in Europe comprising over 20 million subjects were used to compute the relative risk estimates for drug-event pairs using 10 different methods, including those developed for spontaneous reporting systems, cohort methods such as the longitudinal gamma poisson shrinker, and case-based methods such as case-control. The newly developed method "longitudinal evaluation of observational profiles of adverse events related to drugs" (LEOPARD) was used to remove associations likely caused by protopathic bias. Data from the different databases were combined by pooling of data, and by meta-analysis for random effects. A reference standard of known ADRs and negative controls was created to evaluate the performance of the method.

Measures: The area under the curve of the receiver operator characteristic curve was calculated for each method, both with and without LEOPARD filtering.

Results: The highest area under the curve (0.83) was achieved by the combination of either longitudinal gamma poisson shrinker or case-control with LEOPARD filtering, but the performance between methods differed little. LEOPARD increased the overall performance, but flagged several known ADRs as caused by protopathic bias.

Conclusions: Combinations of methods demonstrate good performance in distinguishing known ADRs from negative controls, and we assume that these could also be used to detect new drug safety signals.

Key Words: drug safety, active surveillance, electronic health record, method evaluation

(*Med Care* 2012;50: 890–897)

From the *Department of Medical Informatics, Erasmus University Medical Center, Rotterdam; †PHARMO Institute, Utrecht, The Netherlands; ‡Department of Clinical and Experimental Medicine and Pharmacology, University of Messina, Messina, Italy; §Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine; ‖Department of Primary Care and Public Health Sciences, Kings College London, UK; ¶Department of Clinical Epidemiology, Aarhus University Hospital, Århus Sygehus, Denmark; #Agenzi Regionali di Sanità della Toscana; **Health Search, Italian College of General Practitioners; ††Pedianet, Societa' Servizi Telematici SRL; and ‡‡Department of Statistics, Università di Milano-Bicocca, Italy.

On behalf of the EU-ADR consortium.

Supported by the European Commission FP7 Program (FP7/2007-2013) under grant no. 215847 (the EU-ADR Project).

The authors declare no conflict of interest.

Reprints: Martijn J. Schuemie, PhD, Department of Medical Informatics, Erasmus University Medical Center Rotterdam, P.O. Box 2040, 3000 CA, Rotterdam, The Netherlands. E-mail: m.schuemie@erasmusmc.nl.

Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Website, www.lww-medicalcare.com.

ISSN: 0025-7079/12/5010-0890

The modern drug legislation about postmarketing drug safety monitoring was prompted over 40 years ago as a result of the dramatic teratogenic effects of thalidomide in clinical practice.[1] Since then, the mainstay of drug safety surveillance has been the collection of spontaneous reports of suspected adverse drug reactions (ADRs).[2,3] However, a number of recent high-impact drug safety issues (eg, cardiovascular risk with rofecoxib and rosiglitazone) require rethinking of the way safety monitoring is conducted.[4] Spontaneous reporting systems (SRSs) have inherent limitations that hamper safety signal detection[5] such as under-reporting and biases due to selective reporting.[6] The percentage of ADRs being reported by health professionals varies between 1% and 10% of those actually occurring in clinical practice,[7–13] and this problem occurs both in primary care and in the hospital.[14,15] A recent study in Spain showed that less than two thirds of ADRs recorded in electronic medical records were actually reported to the Spanish Pharmacovigilance System.[14] As a consequence,

SRSs may not always guarantee timely and correct signal detection[16] and alternative data sources have to be explored.

One alternative approach is the use of electronic health care record (EHR) databases for drug safety signal detection. These databases, such as electronic medical record and administrative claims databases, have been most commonly used to confirm or refute potential signals flagged by spontaneous reporting or other surveillance systems. However, appropriate use of these databases may have an enormous potential for earlier detection of drug safety signals because of the availability of large numbers of time-stamped medical records from routine clinical practice.[17,18] Several international initiatives have recently embarked on developing such postmarketing surveillance systems: in the United States, the Sentinel Initiative was established with the Mini-Sentinel (http://mini-sentinel.org) and Observational Medical Outcomes Partnership (OMOP, http://omop.fnih.org) as pilot initiatives, and in Europe, the PROTECT (Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium, http://imi-protect.eu) and EU-ADR (Exploring and Understanding Adverse Drug Reactions. http://euadr-project.org) projects were started.

The EU-ADR project aims to develop a monitoring system using 7 health care databases in 3 different European countries in a distributed database architecture.[19] In this article, we apply a wide range of statistical methods to the EHR data in EU-ADR, and compare the relative performance on the task of distinguishing known drug-adverse event associations from negative controls.

METHODS

Data Sources

The databases in EU-ADR and their characteristics are shown in Table 1, and more detail can be found in appendix C (Supplemental Digital Content 1, http://links.lww.com/MLR/A313).

All databases coded drugs using the World Health Organization's Anatomical Therapeutic Chemical (ATC) classification system. Events were extracted using a variety of structured and free text queries.[19] Although all statistical analyses were performed at a central site, data aggregation was performed locally using a tool developed in Java within the EU-ADR project called Jerboa before combining data across databases, to ensure patient confidentiality.

Reference Standard

An independent group of researchers constructed a reference standard set of known ADRs and drug-event associations unlikely to represent an ADR (negative controls) by first selecting the following 10 events from a list of 23 events ranked on the basis of importance in pharmacovigilance[20]: bullous eruptions (BE), acute renal failure (ARF), anaphylactic shock (AS), acute myocardial infarction (AMI), rhabdomyolysis (RHABD), pancytopenia (PANCYTOP), neutropenia (NEUTROP), cardiac valve fibrosis (CARDFIB), acute liver injury (ALI), and upper gastrointestinal bleeding (UGIB). For each event, a stepwise approach was used to identify which among a list of drug-event associations are well recognized (known associations) or highly unlikely (negative controls) on the basis of MEDLINE-indexed publications, drug product labels, spontaneous reports made to the World Health Organization pharmacovigilance database systems, and expert opinion. Only drugs with adequate exposure in the EU-ADR database network were considered. Manual verification of positive and negative associations was performed independently by 2 experts proficient in clinical medicine, pharmacoepidemiology, and pharmacovigilance. A third expert adjudicated equivocal cases and arbitrated any disagreement between evaluators. The construction of this reference set is described in detail in previous publications[21,22] and in appendix D (Supplemental Digital Content 1, http://links.lww.com/MLR/A313).

Statistical Methods

An overview of all methods is shown in Table 2.

SRS Methods

Signal detection methods originally developed for SRSs can also be used on EHR data by transforming the data to a format suitable for these methods. For this transformation, we assume that the occurrence of the event of interest during a period of drug exposure constitutes a potential association between the drug and the event. The number of occurrences of a particular drug-event pair in the

TABLE 1. Participating Databases and their Characteristics

	Pedianet (Italy)	Health Search (Italy)	Lombardy Regional (Italy)	ARS (Italy)	IPCI (The Netherlands)	PHARMO (The Netherlands)	Aarhus (Denmark)
Subjects	140,000 children	1,000,000	10,000,000	4,000,000	750,000	1,280,000	2,000,000
Type of database	General practice pediatric database	General practice database (no children)	Record linkage system with: Registry inhabitants Regional drug dispensation records Hospital claims database	Record linkage system with: Registry inhabitants Regional drug dispensation records Hospital claims database Death registry	General practice database	Record linkage system with: Registry inhabitants Regional drug dispensation records Hospital claims database Lab values	National health registry with: Registry inhabitants Regional drug dispensation records Hospital claims database Lab values Death registry

health care records is used as if it is a report count from a SRS.

- Proportional reporting ratio (PRR) is the ratio of the proportion of all reported cases of the event of interest among people exposed to a particular drug compared with the corresponding proportion among people exposed to all drugs.[23]
- Reporting odds ratio (ROR) is the reformulation of the PRR as an odds ratio.[24]
- Gamma poisson shrinker (GPS) also determines the disproportionality of reports for a particular drug compared with all exposure, but uses a Bayesian model to shrink relative risk estimates when less data are available.[25] The prior distribution is established empirically using data of all drug-event pairs.
- Bayesian confidence propagation neural network (BCPNN) works similar to GPS, in that it also uses a Bayesian model to shrink estimates of risk. Typically, the output of a BCPNN is expressed as the Information Component; the logarithm of the ratio between the observed and the expected number of reports for a particular drug-event pair.[26]

TABLE 2. Overview of the Methods Evaluated Here

	Bayesian	Ranking Criteria for AUC
SRS methods		
PPR		PRR
ROR		ROR
GPS	Yes	Point estimate of the RR
BCPNN	Yes	IC
Cohort methods		
IRR		IRR
LGPS	Yes	Point estimate of the IRR
BHM	Yes	Point estimate of the IRR
Case-based methods		
Matched CC		β (odds ratio estimate)
SCCS		β (relative risk estimate)
Elimination of protopathic bias		
LEOPARD		Eliminate if $P < 0.5$

The methods are described in more detail in appendix B (Supplemental Digital Content 1, http://links.lww.com/MLR/A313).

BCPNN indicates Bayesian confidence propagation neural network; BHM, Bayesian hierarchical model; CC, case-control; GPS, gamma poisson shrinker; IC, information component; IRR, incidence rate ratio; LEOPARD, longitudinal evaluation of observational profiles of adverse events related to drugs; LGPS, longitudinal gamma poisson shrinker; PRR, proportional reporting ratio; ROR, reporting odds ratio; SCCS, self-controlled case series; SRS, spontaneous reporting system.

Cohort Methods

One of the limitations of the SRSs and their methods is that only numerator data are available, that is the number of people who are exposed to drugs and have the event of interest. What is missing is the denominator data: the number of people who are exposed to the drugs. In longitudinal databases, this information is readily available, along with the duration of drug exposure. This information is used in cohort methods.

- Incidence rate ratio (IRR) is the ratio between the incidence rate during exposure to the drug and a background incidence rate. A Mantel-Haenszel test is used to test the differences between the incidence rates, correcting for age and sex.
- Longitudinal gamma poisson shrinker (LGPS) is an adaptation of the GPS to longitudinal data, and applies Bayesian shrinkage to the IRR. It was developed in the EU-ADR project.[27]
- Bayesian hierarchical model (BHM) uses a full Bayesian approach to perform shrinkage of risk estimates, but instead of using a single prior distribution for all drugs, priors are also created for classes and superclasses of drugs. The BHM combines statistical models for the observations, given the parameters (likelihood) and the parameters themselves (priors). In the application reported here, the incidence rate is modeled using a Poisson process, and the priors as a hierarchy (using guidance from Gelman[28]). The groupings forming the hierarchy are decided a priori on the basis of the criteria of similarity between drugs; in this case, we have used ATC coding levels based on organ/systems and therapeutic or chemical characteristics. Berry and Berry[29] used a similar hierarchical approach, with a hierarchy based on related outcomes rather than drugs. The BHM shrinks the original "frequentist" estimates to yield an updated posterior distribution of each individual drug to the group mean and reduces its variance. This is because the posterior considers both the data provided by the drug and by the other drugs in the same group. Shrinkage is stronger for drugs with an initial large variance (less information) and larger effects. These novel methods can offer key advantages by reducing the likelihood of false-positive or false-negative results obtained from the data. Although the BHM is grouped here with the cohort methods, it can also be applied to other types of relative risk estimates.

Case-based Methods

Several analytical epidemiological methods start with persons with the disease or event of interest (cases) and compare these with a sample of the population that gives rise to the cases (ie, the controls) to evaluate differences in the exposure status. As case-based methods are more efficient in terms of data needs (exposure assessed only at one point in time), they allow for easier adjustments of confounding factors.

- Matched case control (CC) starts with all cases, and finds for every case a predefined number of controls (in our experiments 2 controls per case), where controls should have the same age and sex as the case. For both cases and controls, the exposure to drugs is determined at the calendar time of the event (also known as the index date). A logistic regression condition on the case sets is performed to determine the effect size (odds ratio) of exposure to a drug. To adjust for comorbidity and overall patient health status, we included the drug count[30] in the regression, which is the number of different drugs (distinct ATC codes) the subject was exposed to in 1 year before the event date, until 1 month before the event date.
- Self-controlled case series (SCCS) investigates the association between acute outcomes and transient exposures, whereby cases are used as their own controls. In essence, the SCCS is a Poisson regression conditioned on

the patient.[31] Only information of cases is used in this analysis; all other persons are ignored.

Other Types of Methods

One other method not categorized elsewhere remains:

- Longitudinal evaluation of observational profiles of adverse events related to drugs (LEOPARD)[27] attempts to detect protopathic bias. Protopathic bias occurs when a drug is prescribed for an early manifestation of a disease that has not yet been detected diagnostically. For every drug-event combination, the number of prescriptions initiated in the 25 days before the event is compared with the number of prescriptions starting in the 25 days after the event. If the number of prescriptions increases after the event date, this is an indication that the drug is used to treat the event or a precursor of the event, rather than cause it. This is assessed using a binomial test. For example, omeprazole is a proton pump inhibitor indicated for the treatment of esophageal disorders and gastric ulcers, and its association with UGIB (IRR = 3.9) is likely due to protopathic bias. Nine hundred seventy-eight prescriptions of omeprazole were initiated in the 25 days before a UGIB, and 3459 were initiated in the 25 days after. This increase ($P < 0.001$) in prescriptions indicates that the signal is probably caused by protopathic bias. A signal is considered to be caused by protopathic bias if the P value is < 0.5.

Settings Common to All Methods

For all methods, these specifications were used to define exposures and outcomes:

- *Incident events*: Only the first occurrence of an event was considered; patient time after an event was censored. The main reason for this is that in EHR data, it is often difficult to distinguish between a recurrence of an event and a reference made to events that occurred earlier.
- *Run-in period of 365 days*: To determine that an even is incident event, some patient time has to be available before the event occurred. Hence, during the first year of observation, subjects were not considered for events or exposure counts, but events during this so-called run-in period were used to determine whether later events were truly incident events. This run-in period was omitted for children younger than 1 year at the start of observation.
- *Exposure window definition*: Exposure to a drug was defined as the estimated length of the drug prescription in days, excluding the first day of the prescription. The duration was calculated on the basis of the prescribed daily dosage in some databases (IPCI, PHARMO, Pedianet) and on the basis of the defined daily dose and the quantity prescribed in other databases (Lombardy Regional, Aarhus, Health Search, ARS). If 2 prescriptions of the same drug overlapped in time, the exposure was assumed to start the day after the first day of the first prescription and end on the last day of the last prescription. Please note that this could lead to underestimation of the true exposure window.
- *Age stratification*: Whenever appropriate, age was stratified in 5-year age ranges.
- *Independence of drug risks*: Currently, every drug-event pair is evaluated separately; co-medication is not taken into account.

LEOPARD was considered to be potentially complimentary to all methods, and was therefore applied as a filter to the output of each method. LEOPARD can be applied at the level of the individual drug, but it can also be applied at the drug class level. LEOPARD appears to be better at detecting protopathic bias when drugs are grouped within the same pharmacological subgroup (ie, the same first 4 ATC digits, which usually means the same indication) (Schuemie MJ. unpublished results). Signals that are flagged by LEOPARD, either at the individual or at the group level, were ranked lower in the list of signals than signals that were not flagged when calculating the AUC.

Combination of Databases

The information from the different databases was combined to generate a single score per drug-event pair per method. In principle, there are 2 approaches: pooling of the data as if the databases together form 1 large database or computing the score per database and using meta-analysis techniques to combine the scores. We have tested both data pooling and meta-analysis for risk estimates, assuming random effects. The latter used weighting by inverse variance (both within and between database variance).

Performance Metrics

Typically, the output of a signal detection method is turned into a binary decision (positive or negative) using a threshold, for instance that the relative risk be larger than 2. By comparing the binary outcomes of the method to the reference standard, sensitivity and specificity can be computed. However, sensitivity and specificity can be traded off by varying the threshold, and comparing individual values of sensitivity and specificity is therefore not informative. Typically, for method comparison, the receiver-operator characteristics (ROC) curve is plotted, showing all values of sensitivity and specificity. (Note that this implies varying the threshold from the smallest relative risk to the largest relative risk found for the reference set.) Such a curve can subsequently be summarized into 1 statistic: the area under the ROC Curve (AUC). The AUC indicates the overall performance of a method, independent of any threshold. An AUC of 0.5 indicates random performance; an AUC of 1.0 indicates a perfect performance. The measure used to calculate the AUC for each method is shown in Table 2.

RESULTS

In total, 146,830,906 patient-years of follow-up data concerning 20,042,652 subjects from 3 European countries from 1997 to 2010 were included in the study.

Overall Performance of Methods

Figure 1 shows the performance of the different methods on the reference standard, using meta-analysis for random effects and data pooling.

All methods perform better than random baseline, and LEOPARD filtering for protopathic bias always improved

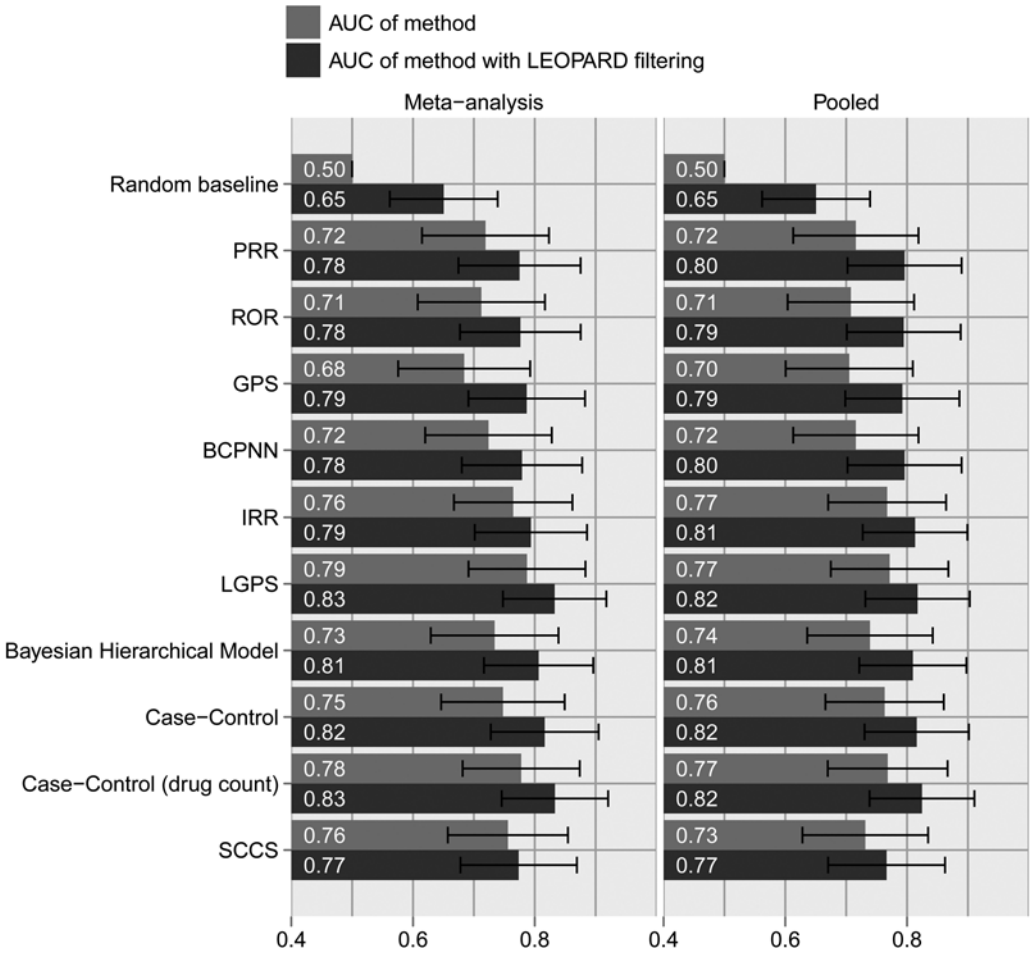

FIGURE 1. AUC for all methods, with and without LEOPARD filtering. Combination across databases was performed using meta-analysis for random effects (left panel) and pooling (right panel). Error bars indicate the 95% confidence interval. The ROC curves are included in the online appendix (Supplemental Digital Content 1, http://links.lww.com/MLR/A313). AUC indicates area under the ROC curve; BCPNN, Bayesian confidence propagation neural network; GPS, gamma poisson shrinker; IRR, incidence rate ratio; LEOPARD, longitudinal evaluation of observational profiles of adverse events related to drugs; LGPS, longitudinal GPS; PRR, proportional reporting ratio; ROC, receiver-operator characteristics; ROR, reporting odds ratio; SCCS, self-controlled case series.

the overall performance, but less so for methods that are already performing well. The performance of methods does not differ much. In general LGPS and case-control adjusting for drug count seem to slightly outperform the other methods, and the SRS-based methods have lower performance, although this is certainly not statistically significant.

Signal Detection Using LGPS and LEOPARD

Figure 2 shows that most of the known ADRs have an estimated IRR higher than 1. Unfortunately, this is also the case for a large number of negative controls, but many of these are flagged by LEOPARD as protopathic bias. This reduction in false positives comes at a price: several of the known ADRs are also flagged as protopathic bias. Notably, ciprofloxacin and ARF, which is a known ADR, has an estimated IRR of 13.98, but because there are more prescriptions starting in the 25 days after ARF diagnoses than in the 25 days before (631 and 574 prescriptions, respectively), this signal is rejected by LEOPARD. The strongest false positive not discarded by LEOPARD is fexofenadine and ARF. Figure 3 shows that most other methods indicate a low relative risk instead.

If we were to use a threshold value of IRR >1.5, and were to remove signals flagged by LEOPARD, LGPS using meta-analysis for random effects would achieve a sensitivity of 0.73 and a specificity of 0.88.

DISCUSSION

In general, the performance of the methods is high, with the best performing method achieving an area under the ROC curve of 0.83. When using a default threshold of IRR >1.5, a sensitivity and a specificity of 0.73 and 0.88, respectively, are achieved. On the one hand, this is not surprising, as the reference standard was limited to drugs with a large amount of exposure. This probably explains why the

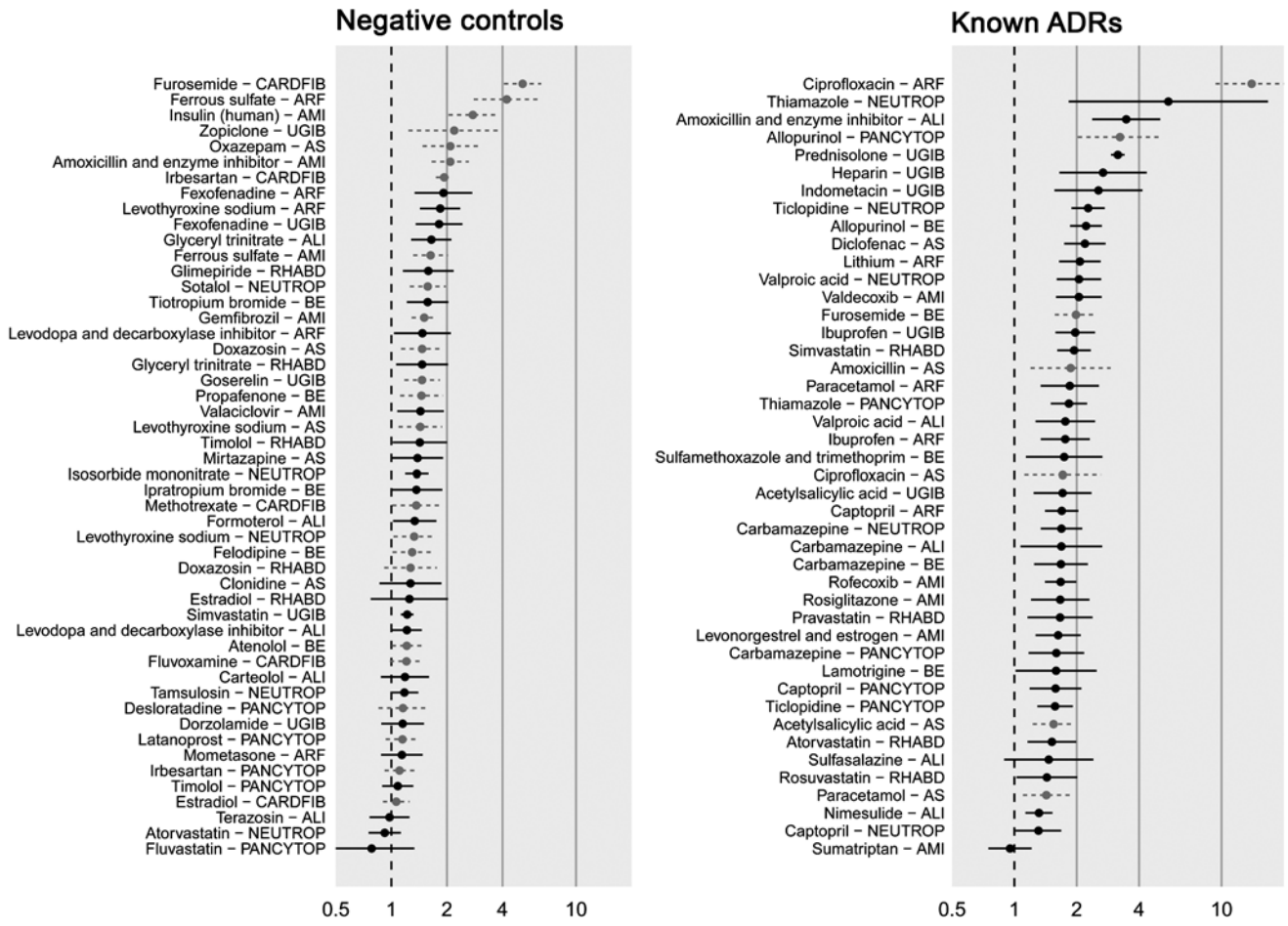

FIGURE 2. Risk estimates for all drug-outcome combinations in the reference set using LGPS and using meta-analysis for random effects to combine estimates across databases. Error bars indicate the 95% confidence interval. Gray markers and dashed lines indicate that a drug-event pair has been flagged as a protopathic bias by LEOPARD. AMI indicates acute myocardial infarction; ALI, acute liver injury; ARF, acute renal failure; AS, anaphylactic shock; BE, bullous; CARDFIB, cardiac valve fibrosis; LEOPARD, longitudinal evaluation of observational profiles of adverse events related to drugs; LGPS, longitudinal gamma poisson shrinker; NEUTROP, neutropenia; PANCYTOP, pancytopenia; RHAB, rhabdomyolysis; UGIB, upper gastrointestinal bleeding.

Bayesian methods are not performing much better than frequentist methods, as these methods are designed to deal with sparse data. On the other hand, the high performance is surprising given the fact that most of these methods are very simple: the best performing methods do not take any potential confounding factors into account other than age and sex. It is expected that by including potential confounders in the analysis, the performance of the method could be improved further, but the data needed for this are not uniformly available across databases. For example, diabetes is a risk factor for myocardial infarction, and diabetic patients will tend to be exposed to antidiabetic drugs. Diabetes is therefore a confounder between myocardial infarction and antidiabetics, and should ideally be included in the analysis. However, diabetes will be coded differently in different databases in EU-ADR, and time-consuming harmonization would be needed to extract these data in a uniform way. As there can be different potential confounders for every drug-event combination, many such variables would need to be extracted, which is currently not feasible. One possible solution could be the development of techniques for adjusting for confounding without the use of harmonized covariates, for instance using summary statistics such as propensity scores instead.[32] Another simplification is that the methods currently consider each drug independently, which could lead to harmless drugs being implicated because of frequent coprescribing with drugs that do cause an ADR. As data on most drug prescriptions are available, this problem could potentially be solved by adapting the methods to include all drugs in 1 analysis.

Filtering signals for protopathic bias using LEOPARD has a positive effect on the overall performance, but some of the known ADRs are incorrectly flagged as protopathic bias. For example, ciprofloxacin is known to be associated with ARF,[33–35] and we indeed find that ciprofloxacin users have an increased risk of ARF. However, the data also show that subjects are more likely to receive a prescription after an ARF when compared with the period preceding the event,

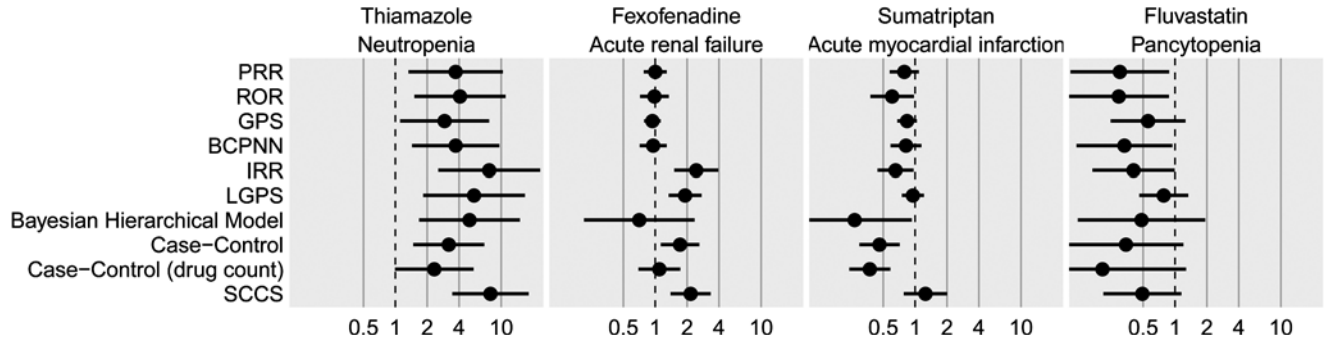

FIGURE 3. Estimates and 95% confidence intervals of the risk estimates as calculated using the different methods for 4 drug-event combinations, using meta-analysis for random effects. Thiamazole and neutropenia is the known ADR with the highest overall LGPS estimate (true positive). Fexofenadine and ARF is the negative control with the highest overall LGPS estimate (false positive). Sumatriptan and AMI is the known ADR with the lowest LGPS estimate (false negative), and fluvastatin and pancytopenia is the negative control with the lowest LGPS estimate (true negative). ADR indicates adverse drug reaction; AMI, acute myocardial infarction; ARF, acute renal failure; BCPNN, Bayesian confidence propagation neural network; GPS, gamma poisson shrinker; IRR, incidence rate ratio; LGPS, longitudinal GPS; PRR, proportional reporting ratio; ROR, reporting odds ratio; SCCS, self-controlled case series.

indicating potential protopathic bias. One possible explanation is that patients with renal failure are at risk of developing various infections, and may therefore be treated with ciprofloxacin, which is a broad-spectrum antibiotic. In fact, for patients with ARF that require dialysis, ciprofloxacin is used in the treatment of dialysis-associated peritonitis.[36] It seems that the output of LEOPARD cannot be used to rule out drug safety signals, but only as an indication that protopathic bias might be present. The strongest false positive for LGPS not eliminated by LEOPARD is fexofenadine and ARF, but interestingly, this signal is not detected by the majority of other methods. Perhaps different methods can be selected for different types of drugs or events to achieve better results; this is something that needs to be explored in future investigations.

We also identified some ADRs that are not picked up by the methods, even before LEOPARD filtering. For example sumatriptan is known to be associated with AMI,[37] but none of the methods find an increased risk for this drug. Sumatriptan is used for the treatment of migraine headaches and may be taken by patients only intermittently and as needed, which can be a long time after the drug is prescribed. As our current definition of the exposure window assumes that exposure starts on the day of prescribing, this could explain why this ADR is not detected. In general, we expect that signal detection using EHRs will perform poorly on all drugs that are taken as needed because of the uncertainty about the drug exposure window.

The results are based on data from 7 different databases, but we did not investigate the performance per database because none of the databases by itself has enough data to detect all drug-event pairs in the reference standard. The need for sufficient data on drug exposure is one of the limitations of using EHRs for drug safety monitoring,[38] and combining databases can overcome this problem to some extent. We investigated 2 methods of combining data: pooling and meta-analysis, and the differences in performance were negligible.

In conclusion, the results from this study indicate that there are several combinations of statistical methods that show good performance in distinguishing the known side effects from negative controls, and these methods can be applied to detect new, previously unknown ADRs or to start a reappraisal of ADRs found by SRSs. These methods and the approach described here should become an essential component of postmarketing drug safety surveillance.

ACKNOWLEDGMENTS

The authors thank Stephen Evans, Annie Fourrier-Reglat, and Silvana Romio for their participation in the many discussions about this research.

REFERENCES

1. Mann R, Andrews E. *Pharmacovigilance*. Chichester, UK: John Wiley & Sons; 2002.
2. Ahmad SR. Adverse drug event monitoring at the Food and Drug Administration. *J Gen Intern Med*. 2003;18:57–60.
3. Olsson S. The role of the WHO programme on International Drug Monitoring in coordinating worldwide drug safety efforts. *Drug Saf*. 1998;19:1–10.
4. Avorn J. Evaluating drug effects in the post-Vioxx world: there must be a better way. *Circulation*. 2006;113:2173–2176.
5. Meyboom RH, Egberts AC, Edwards IR, et al. Principles of signal detection in pharmacovigilance. *Drug Saf*. 1997;16:355–365.
6. Belton KJ. Attitude survey of adverse drug-reaction reporting by health care professionals across the European Union. The European Pharmacovigilance Research Group. *Eur J Clin Pharmacol*. 1997;52:423–427.
7. Alvarez-Requejo A, Carvajal A, Begaud B, et al. Under-reporting of adverse drug reactions. Estimate based on a spontaneous reporting scheme and a sentinel system. *Eur J Clin Pharmacol*. 1998;54:483–488.
8. Eland IA, Belton KJ, van Grootheest AC, et al. Attitudinal survey of voluntary reporting of adverse drug reactions. *Br J Clin Pharmacol*. 1999;48:623–627.
9. De Bruin ML, van Puijenbroek EP, Egberts AC, et al. Non-sedating antihistamine drugs and cardiac arrhythmias-biased risk estimates from spontaneous reporting systems? *Br J Clin Pharmacol*. 2002;53:370–374.
10. Begaud B, Martin K, Haramburu F, et al. Rates of spontaneous reporting of adverse drug reactions in France. *JAMA*. 2002;288:1588.
11. Hazell L, Shakir SA. Under-reporting of adverse drug reactions: a systematic review. *Drug Saf*. 2006;29:385–396.

12. Lopez-Gonzalez E, Herdeiro MT, Figueiras A. Determinants of under-reporting of adverse drug reactions: a systematic review. *Drug Saf*. 2009;32:19–31.
13. Perez Garcia M, Figueras A. The lack of knowledge about the voluntary reporting system of adverse drug reactions as a major cause of underreporting: direct survey among health professionals. *Pharmacoepidemiol Drug Saf*. 2011;20:1295–1302.
14. Gonzalez-Rubio F, Calderon-Larranaga A, Poblador-Plou B, et al. Underreporting of recognized adverse drug reactions by primary care physicians: an exploratory study. *Pharmacoepidemiol Drug Saf*. 2011;20:1287–1294.
15. Pushkin R, Frassetto L, Tsourounis C, et al. Improving the reporting of adverse drug reactions in the hospital setting. *Postgrad Med*. 2010;122: 154–164.
16. Meyboom RH, Hekster YA, Egberts AC, et al. Causal or casual? The role of causality assessment in pharmacovigilance. *Drug Saf*. 1997;17:374–389.
17. McClellan M. Drug safety reform at the FDA—pendulum swing or systematic improvement? *N Engl J Med*. 2007;356:1700–1702.
18. Platt R. *Challenges for the FDA: the future of drug safety, workshop summary*. Washington, DC: National Academy of Sciences; 2007.
19. Coloma PM, Schuemie MJ, Trifiro G, et al. Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: the EU-ADR Project. *Pharmacoepidemiol Drug Saf*. 2011;20:1–11.
20. Trifiro G, Pariente A, Coloma PM, et al. Data mining on electronic health record databases for signal detection in pharmacovigilance: which events to monitor? *Pharmacoepidemiol Drug Saf*. 2009;18:1176–1184.
21. Coloma PM, Avillach P, Schuemie MJ, et al. A Reference Standard for Evaluating Methods for Drug Safety Signal Detection using Electronic Healthcare Record (EHR) Databases. *11th ISOP Annual Meeting*, Istanbul, 2011.
22. Coloma PM, Avillach P, Salvo F, et al. A reference standard for evaluation of methods for drug safety signal detection using electronic healthcare records (EHR) databases. *Drug Saf*. 2012. In print.
23. Evans SJ, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiol Drug Saf*. 2001;10:483–486.
24. Rothman KJ, Lanes S, Sacks ST. The reporting odds ratio and its advantages over the proportional reporting ratio. *Pharmacoepidemiol Drug Saf*. 2004;13:519–523.
25. DuMouchel W. Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. *Am Stat*. 1999; 53:177–190.
26. Norén GN, Bate A, Orre R, et al. Extending the methods used to screen the WHO drug safety database towards analysis of complex associations and improved accuracy for rare events. *Stat Med*. 2006;25: 3740–3757.
27. Schuemie MJ. Methods for drug safety signal detection in longitudinal observational databases: LGPS and LEOPARD. *Pharmacoepidemiol Drug Saf*. 2011;20:292–299.
28. Gelman A. Prior distributions for variance parameters in hierarchical models. *Bay An*. 2006;1:515–534.
29. Berry SM, Berry DA. Accounting for multiplicities in assessing drug safety: a three-level hierarchical mixture model. *Biometrics*. 2004;60: 418–426.
30. Schneeweiss S, Seeger JD, Maclure M, et al. Performance of comorbidity scores to control for confounding in epidemiologic studies using claims data. *Am J Epidemiol*. 2001;154:854–864.
31. Whitaker HJ, Farrington CP, Spiessens B, et al. Tutorial in biostatistics: the self-controlled case series method. *Stat Med*. 2006;25:1768–1797.
32. Rassen JA, Solomon DH, Curtis JR, et al. Privacy-maintaining propensity score-based pooling of multiple databases applied to a study of biologics. *Med Care*. 2010;48:S83–S89.
33. Gerritsen WR, Peters A, Henny FC, et al. Ciprofloxacin-induced nephrotoxicity. *Nephrol Dial Transplant*. 1987;2:382–383.
34. Allon M, Lopez EJ, Min KW. Acute renal failure due to ciprofloxacin. *Arch Intern Med*. 1990;150:2187–2189.
35. Hootkins R, Fenves AZ, Stephens MK. Acute renal failure secondary to oral ciprofloxacin therapy: a presentation of three cases and a review of the literature. *Clin Nephrol*. 1989;32:75–78.
36. Goffin E, Herbiet L, Pouthier D, et al. Vancomycin and ciprofloxacin: systemic antibiotic administration for peritoneal dialysis-associated peritonitis. *Perit Dial Int*. 2004;24:433–439.
37. Ottervanger JP, Wilson JH, Stricker BH. Drug-induced chest pain and myocardial infarction. Reports to a national centre and review of the literature. *Eur J Clin Pharmacol*. 1997;53:105–110.
38. Coloma PM, Trifirò G, Schuemie MJ, et al. Electronic healthcare databases for active drug safety surveillance: Is there enough leverage? *Pharmacoepidemiol Drug Saf*. 2012;21:611–621.

