Hypothesis-free signal detection in healthcare databases: finding its value for pharmacovigilance

Andrew Bate (iD)**, Ken Hornbuckle, Juhaeri Juhaeri** (iD)**, Stephen P. Motsko and Robert F. Reynolds**

Knowledge about a medicine evolves during its lifecycle, as evidence about the product is generated during the pre-clinical, clinical, and post-marketing phases. Pharmacovigilance systems are designed to identify emerging 'signals' using multiple data sources, particularly as use of the products expands post-approval in healthcare systems and as treatment pathways evolve. Signals may arise from one or multiple sources and may suggest a potentially new causal association or new aspect of a known risk of a medicine.[1] Although companies and regulators use multiple data sources to detect emerging signals, spontaneous reports continue to be the cornerstone of signal detection for most marketed drugs and vaccines, as they have been since the 1960s.[2,3] For example, a total of 19 out of 21 drugs withdrawn in France between 1998 and 2004 were reported to be based on spontaneous individual case safety reports (ICSRs). Among these product withdrawals, 12 were solely based on spontaneous reports, 6 were in combination with results from observational studies, and the remaining two were based on randomized controlled trial (RCT) evidence and animal studies, respectively.[4] Another study showed that 8 out of 11 product withdrawals between 1999 and 2001 in the UK and US were based on early information from spontaneous reports, for another product solely RCTs were identified as the recorded evidence and for the other two products evidence used to support their withdrawal could not be found in any of the identified documents.[5] Spontaneous reports are particularly useful for identifying rare and idiosyncratic safety concerns and acute events.[6,7] Some literature suggests spontaneous reports are most useful for newly marketed products, specifically within 3 years after launch,[8–10] whereas other studies indicated they are also useful for older products.[8,9] A large body of literature has documented

the well-known limitations of spontaneous reporting systems (SRS), including underreporting, stimulated reporting, and other reporting biases, the lack of a population denominator to calculate rates, and frequently poor information in terms of both quality and quantity.[3,6]

As paper-based reporting systems transitioned to electronic systems and regulatory requirements for reporting have increased, the number of reports received over time by pharmaceutical companies and regulators have risen significantly and further compounded these limitations. Quantitative methods, or measures of disproportionality, were introduced around the turn of the millennium[11] with the goal of making the large numbers of reports more manageable and informative[3,12] by taking advantage of the data volume to assist in identifying new safety signals.[13] These methods provide a systematic approach to filter through drug–event pairs and focus the clinical review of drug–event pairs most likely to represent emerging 'signals of suspected causality' as CIOMS VIII terms it.[3,14] However, as the scenarios where reports are required by regulators for collection have also expanded (e.g. reports from patient support marketing programs, reports of events occurring during time of drug exposure without suspected causality), there has not been a corresponding improvement in the time between approval and detection of new signals and the duration remained relatively long.[15] Fukazawa *et al.*[15] showed that, using the FDA AERS database, the median duration between approval and detection of signals was 9 years, between approval and label change 10 years. The same study also showed that FDA AERS was less effective for older drugs, that is, more than 5 years in the market. This is not surprising as there are proportionally fewer reports of older drugs and the

Correspondence to:
Andrew Bate
Division of Translational
Medicine, Department of
Medicine, NYU School of
Medicine, 462 1st Avenue,
NY10016, New York, USA
**andrew.bate@pfizer.com;
ajwb@mail.com**

Ken Hornbuckle
Global Patient Safety,
Eli Lilly and Company,
Indianapolis, IN, USA

Juhaeri Juhaeri
Juhaeri Juhaeri, Medical
Evidence Generation,
Sanofi US, Bridgewater,
NJ, USA

Stephen P. Motsko
Global Patient Safety,
Eli Lilly and Company,
Indianapolis, IN, USA

Robert F. Reynolds
Department of
Epidemiology, Tulane
University School of
Public Health and Tropical
Medicine, New Orleans,
LA, USA

established weaknesses of spontaneous reporting make such reports less likely to impact the well-established benefit–risk profiles of such medicines. This research as well as our experience at pharmaceutical companies where hundreds of thousands to millions of reports are processed annually, suggests that the increased volume of reports has had a limited impact on the approved product label or assessments of the overall benefit–risk of a medicine.

Until recently, there were no widespread alternative data sources for large scale routine monitoring of the safety profile of products post-launch in the real world setting. There is growing recognition that electronic healthcare databases (for example, insurance claims, electronic medical records, and registry databases), together referred to as real-world data (RWD),[16] may offer an alternative to SRS. These longitudinal healthcare databases are a rich source of information about the benefits and risks of medical products, as they are longitudinal with anonymized and routinely collected information on millions of patients over years. Given the characteristics of these data sources compared with spontaneous reports databases, for example, duration and completeness of follow up, ease of estimating the population at risk, better ascertainment of exposure and outcomes, and more relevant comparison groups, over time they have the potential to enhance or even replace current signal detection activities within pharmacovigilance systems.

We focus herein on challenges and opportunities to what we term hypothesis-free signal detection in RWD: searching for potential signals across multiple drug/vaccine–adverse event pairs simultaneously and irrespective of whether there is an index of suspicion.[17] Many of the issues we raise will also apply to the more focused activity of surveillance on predefined outcomes in RWD, an area where there is more scientific activity.[18,19] We do not discuss spontaneous reporting signal detection methods at any more length given how accepted it is, other than to point out how these inform our perspectives on healthcare databases signal detection. Similarly, we consider other existing data streams such as query log (search engine) data[20] and social media[21,22] out of scope, although we recognize these are sometimes categorized as RWD.

RWD in pharmacovigilance: not only signal evaluation

Healthcare databases have been used for epidemiological studies for decades in the evaluation of *a priori* defined hypotheses (signal evaluation) and to further investigate potential safety signals identified from other sources (signal refinement).[23,24]

However, despite the ubiquity of real-world databases for years, their use for hypothesis-free signal detection has been limited. Why is this, given the well-known limitations of SRS and the potential benefits of using real-world databases for signal detection? One of the reasons is that until recently access to such data was costly and limited given the proprietary nature of many databases; thus, access was primarily driven by the need to answer specific research questions or assess specific safety concerns, that is, an *a priori* hypothesis. Timeliness of the data has also been an issue with substantial delays from the date of an encounter to when it is available in an analytic database. These delays have limited the usefulness of these data for signal detection that relies on timely data and rapid analyses. In addition, data quality and the non-random variability on which data are captured or missing across databases makes rapid assessment of multiple outcome pairs across databases extremely difficult, at least currently. All of the above, until recently, made signal detection in RWD far from straightforward operationally and fraught with methodologic pitfalls.

However, with increased access to data through more affordable subscription models, faster computing and data refresh times, common data models (CDMs),[25] and similar enabled software solutions for rapid analytics, the conditions for testing signal detection in RWD has never been better. Compared with spontaneous reports, real-world databases (both electronic medical record databases and insurance claims databases) have several characteristics, some of which are potentially favorable, to enhance signal detection efforts,[1] including the following.

1. The ability to provide accurate and informative risk estimates, for example, incidence rates, relative risks or incidence rate ratios, as a result of being able to quantify the population at risk and outcomes at baseline and follow-up intervals.

2. Data capture on patients before and after exposure, often for many years. This is in contrast to the cross-sectional nature of most spontaneous reports.
3. Information on comparators, which permits matching (design) or statistical adjustment using epidemiologic study approaches.
4. Reliable data capture for those variables systematically collected for billing or clinical management purposes, even though systems were not developed for research purposes. This is in contrast to the voluntary nature of spontaneous reports.
5. There is no suspicion of causality in coded (structured) data. This provides the ability to detect signals statistically that may never have resulted in a voluntary report, but also presents a potential weakness when compared to spontaneous reports, particularly those with compelling causal rationales (e.g. dechallenge/rechallenge).
6. RWD sources are heterogeneous in their data collection standards and methods, for example, the type of healthcare data captured, stored, and accessible.
7. RWD sources increasingly exist internationally but they are not as pervasive in coverage as spontaneous report systems and until recently mostly covered populations in North America and Europe.

Despite these potential advantages, to date, and to the best of the authors' knowledge, there is limited routine use of RWD to systematically detect signals potentially associated with medical products. A few public–private initiatives have evaluated the value, feasibility, and utility of observational databases to identify safety issues using various methods, including projects such as the Observational Medical Outcomes Partnership (OMOP),[26] the Innovative Medicines Initiative's (IMI) PROTECT,[27] Exploring and Understanding Adverse Drug Reactions (EU-ADR),[28] and the Asian Pharmacoepidemiology Network (AsPEN).[29] Each of these projects has contributed to our understanding of the potential of using RWD for signal detection. With the shift in pharmacovigilance toward a more RWD centric approach, we note that companies and regulators are actively testing signal detection methods in RWD. Further work is needed, however, before implementing these activities routinely. The relative performance of methods applied to RWD must be

clarified and, importantly, so must the relative value of signals from RWD *versus* spontaneous reports in terms of signal type, timeliness, and likelihood to yield information that informs patients and physicians through the approved product label or overall benefit–risk assessment.

Comparing methods for signal detection in RWD

With the emergence of 'Big Data' and other data science concepts there is widespread attention on further enabling appropriate harvesting of data for secondary use in general.[30,31]

The optimal study design and analyses should always be selected based on the appropriateness of the research question (e.g. fit for purpose). Nevertheless, with signal detection, one cannot design bespoke analyses for individual drug–event pairs, given the general objective at hand. That said, is it possible to employ a similar analytic approach for groups of drug–event pairs with specific characteristics? The need for generic approaches, coupled with the complexity and heterogeneity of RWD sources, makes the effective application of signal detection to such data challenging. The solution to the appropriate analytical method for signal detection in RWD may not be a 'one size fits all'; rather, multiple methods may need to be used in parallel for different drug–adverse event combination groupings.

There have been several methods proposed and tested for signal detection, including the cohort study with propensity score matching, high-dimensional propensity score matching (HDPS), self-controlled case series study (SCCS), tree-based scan statistic, self-controlled cohort analysis with the use of temporal pattern discovery, and Bayesian approaches.[32–42] Existing methods have been applied 'as is' or modified slightly for use in RWD signal detection; for example, traditional epidemiologic focused surveillance adapted for hypothesis-free use[41] and prescription symmetry analysis[43]; methods historically more associated with disease epidemiology have been implemented,[35] informatics approaches,[44,45] SRS-based methods,[46] and approaches novel for this application[42]; as have a combination of the above approaches, as well as 'advanced analytics' such as Q-methodology[47] and other machine learning and artificial intelligence approaches.[44,48,49] Use of these methods will improve over time as
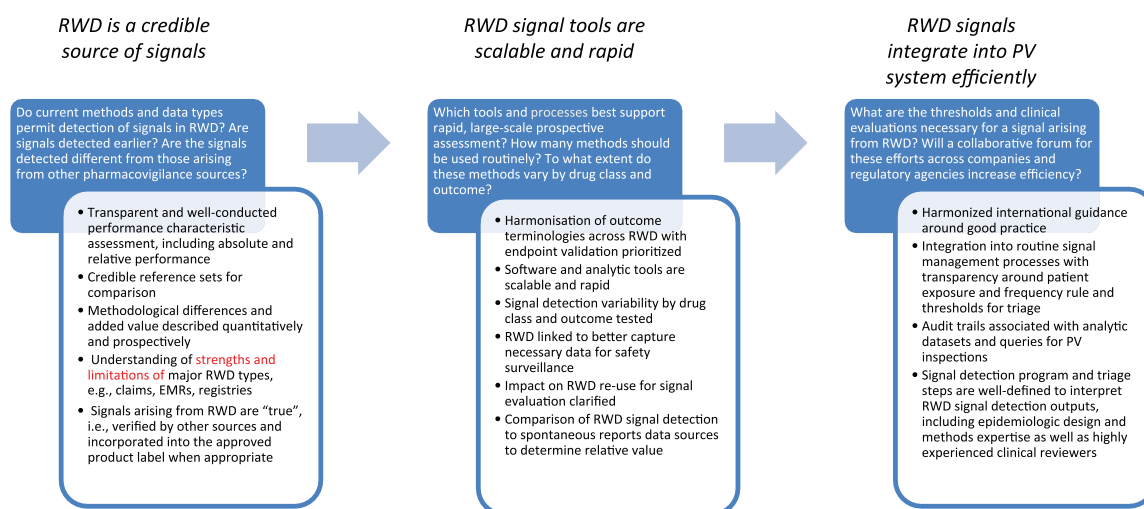
Figure 1. Implementing signal detection in RWD: the necessary steps.

experience with each approach accrues. For example, propensity score matching for signal detection might theoretically be improved by the use of calendar-specific propensity scores, use of multiple comparison groups, and data visualization to better understand characteristics of the study populations, although method testing would be essential to determine whether such changes would translate into meaningful routine performance improvement.

There have been only a limited number of studies[26,48,50–52] examining the performance (such as sensitivity and specificity) of methods, and only some studies conducted comparative evaluation of more than one method.[35,38] No method has emerged as superior across the range of different exposure and outcomes (with varying covariates of focus) from which emerging signals need to be detected. Research from OMOP, although not explicitly focused on signal detection, identified the self-controlled approach as demonstrating the highest performance characteristics as compared with other study designs, when looking indiscriminately at a large number of different exposure and outcomes with no attempt *a priori* to optimize the method for a specific drug–outcome pair.[26] The self-controlled study design has several strengths, including the elimination of confounding factors and selection bias, because a person serves as their own control, as well as no need to select a comparator product. This design is best used to study acute or transient outcomes that occur shortly after exposure and not appropriate for fatal outcomes. It seems clear that for some

types of exposure and outcomes other approaches for example cohort based, will be needed. There is only one, recent review[49] evaluating methods for signal detection in RWD. However the conclusions on relative method performance are necessarily limited as the authors themselves emphasize, owing to the dearth of articles specifically comparing methodologies.

Challenges to implementing signal detection in RWD

We have identified three broad challenges that must be addressed prior to widespread implementing of RWD signal detection into the signal management system: establish RWD is a credible source of signals; demonstrate that tools for RWD signal detection are scalable and rapid; and characterize the best conditions and means to integrate these signals into the overall pharmacovigilance system efficiently and in a way demonstrated to enhance value. In Figure 1, the key questions and actions proposed are described, ordered in a manner that is most likely to lead to the adoption of RWD for signal detection.

There are additional considerations, not identified in Figure 1, which are also important for operationalizing signal detection in RWD as part of a signal management system. These include the following.

Signal evaluation in the same database(s) as signal detection? Some contend[53] that it is critically important that signal detection, or hypothesis

generation, should always be conducted independently using different data sources from those used for signal evaluation, or hypothesis testing epidemiology studies. With a finite number of databases at the field's disposal (albeit increasing) this may be impractical – particularly given overlapping studies from different groups and the necessity of also conducting feasibility work before initiating studies. Walker proposed[54] that if different methods are applied then the same data can in practice be used for two separate analyses (i.e. near-independence between the two) and that this type of data reuse may contribute to better understanding a hypothesis and is distinct from re-using data with the same methods.[55] An alternative, given the significant power one often has access to for a given drug-AE pair, might be that, for a given analysis, a database would be partitioned, one part to be used for signal detection and the other for hypothesis testing studies.[56] More work is needed to assess how either approach could be used in practice.

Concordance across multiple data streams

Criteria for weighting of outputs from analyses across different RWD data types, as well as other data sources in the pharmacovigilance system, is critical. Complex decision making heuristics for taking into account contradictory evidence across data sets will need to be addressed. In practice, results may vary across different data sources; in other words, Signals of Disproportionate Recording (SDRs) in some data sources may not be SDRs in other data sources or the magnitude of the associations may be different and approaches to integrate results from different data sources, for example with weighting, need be evaluated. The complexity will come in part from the necessity to weigh outputs from the different databases differently, as exposure and outcome varies.

Repeated analyses

One needs to consider the repeated iterative nature of interrogations of different data sources for different drugs and at different points in time. As the amount of data in a database increases over time, signal detections for a single drug in a single database will need to be repeated over time as new signals may be detected and the magnitude of the old signals may change overtime

(stronger associations). For this reason, various issues with repeated testing will need to be addressed. For example, if a signal is defined based on *p*-value, should it be adjusted? One school of thought suggests that no adjustment is needed as this is not a clinical trial. Others advocate some kind of statistical adjustments. Another issue with repeated analysis is how to combine data (for the same signal) across databases and at different points in time. One solution is to apply meta-analyses approaches to creating a single composite measure based on different point estimates and variances.

Complementary nature of RWD

It is important to note that no system is perfect and, even if adequate approaches to the above issues are found, RWD may not be ready to replace SRS, at least not at this point in time, and both sources are complementary to each other. Despite limitations mentioned above, SRS may for example well continue to be very useful, or even better than RWD despite it becoming ever more ubiquitous, for detection of rare events in the early years after launch. RWD may be superior to SRS for events that require a longer induction period or those that are acute, especially when the self-control case series is used. Some evaluations suggest such differential performance of the data streams[28,57] but more evaluations are needed to see how the two sources complement each other, when and in what situations RWD is superior to SRS and should be used instead of SRS, and *vice versa*. There may even be situations in the future where neither RWD nor SRS is most suitable. For example, given the large amount of data and the high speed of the accumulation of data in social media, this data source may at some point in the future be found to be even faster to detect certain specific events for newly launched drugs than both RWD and SRS.

The development and implementation of signal detection capabilities within a signal management system needs to take a practical approach that evolves through knowledge and experience before adoption as a best practice to the field of drug safety. The initial approach should be pragmatic, have realistic expectations that should not cause an over-reaction to the process or to the results generated from signal detection efforts. The evolution of a robust process will take time to develop,

learn, and demonstrate its value to the overall signal management system. It is important to have individuals with experience, in pharmacovigilance, drug safety surveillance, but also epidemiology and healthcare databases, involved in the process to apply clinical judgement in the interpretation of potential signals identified *via* signal detection capabilities. There should also be areas of opportunities to collaborate and share learnings from these efforts with key stakeholders. The ultimate goal should be the creation of a rapid, scalable, and value-added approach to effectively supplement and enhance current signal detection capabilities.

Conclusion

The use of real-world healthcare databases for signal detection offers an opportunity to enhance current signal detection activities. It could first be augmented to the current system and eventually replace outmoded approaches that fail to add value in a modern pharmacovigilance system. Experience to date suggests a distinct role for SRS remains when compared with RWD for signal detection.[51] We expect this will evolve in the ensuing years. Signal detection in RWD is likely to make some spontaneous reporting processes redundant and to elucidate those data sources, such as patient support programs,[21] where active monitoring overall adds little to the approved product label, benefit–risk knowledge or patient safety. The circumstances when signal detection in RWD will complement or replace current activities require further investigation.

A large-scale approach, that is, most drugs for multiple events, where the best method for each outcome pair is applied is unlikely. For example, self-controlled designs may be expected to be beneficial for signal detection in acute outcomes where exposure is short term and accurately measured, so accurate time between exposure and outcomes can be identified and the within patient control brings benefit in terms of confounding control.[33] When considering cancer outcomes, one would need to consider cohort or case control studies as explored in recent research.[58] Rather selecting the best approach across a broad range of outcomes for a drug class or outcomes across drug classes is likely to be a necessity. As a result, outputs with such a large-scale approach will contain more false-positives than formal hypothesis testing studies. This

means signals emerging from RWD must be triaged into routine signal management processes and undergo the essential clinical review phase and comparison to other surveillance outputs.[51]

To reach the goal of a pharmacovigilance system enhanced by signal detection in RWD, explicit guidance on how signal detection is best performed in RWD and indeed other emerging data streams,[27] must be a priority for collaborative work across stakeholders. We envision a forum with participation from across industry, regulatory, and other government agencies and patient/healthcare provider advocacy groups to collaborate on answering questions about methodologies and the best ways to operationalize signal detection in RWD in a manner that is efficient and avoids redundancy. The FDA have recently held a public meeting focused on signal detection methods in the Sentinel Network.[59] This is an important start but will require transparent, extensive foundational research and collaboration with stakeholders about practical approaches to be successful. Developing a harmonized approach and pooling resources is also critical and likely the best means to resolve the current uncertainty about the contribution signal detection in RWD will make to safety surveillance strategies. Eventually guidance documents, similar to those completed for SRS in an FDA-industry collaborative white paper[2] and in the CIOMS VIII,[1] reflecting an harmonized regulatory view globally will be needed. Signal detection in RWD holds promise for enhancing overall signal detection capability of pharmacovigilance systems. To harness this promise, in the coming years, it will be critical that we all work together to identify the best way to advance signal detection in RWD to ensure patient safety and protect public health.

Conflict of interest statement

While Andrew Bate and Robert Reynolds are employees of Pfizer the work was conducted outside their employment and the contents herein represent their personal opinions. Juhaeri Juhaeri is an employee and stockholder of Sanofi, Ken Hornbuckle and Stephen P Motsko are employees and stockholders of Eli Lilly and company.

ORCID iDs
Andrew Bate [iD] https://orcid.org/0000-0003-3151-3653

Juhaeri Juhaeri [iD] https://orcid.org/0000-0001-8235-8388

References

1. Council for International Organizations of Medical Sciences. Practical aspects of signal detection in pharmacovigilance: report of CIOMS Working Group VIII. Geneva: Council for International Organizations of Medical Sciences; 2010.

2. Almenoff J, Tonning JM, Gould AL, *et al.* Perspectives on the use of data mining in pharmaco-vigilance. *Drug Saf* 2005; 28: 981–1007.

3. Bate A and Evans SJ. Quantitative signal detection using spontaneous ADR reporting. *Pharmacoepidemiol Drug Saf* 2009; 18: 427–436.

4. Olivier P and Montastruc JL. The nature of the scientific evidence leading to drug withdrawals for pharmacovigilance reasons in France. *Pharmacoepidemiol Drug Saf* 2006; 15: 808–812.

5. Clarke A, Deeks JJ and Shakir SA. An assessment of the publicly disseminated evidence of safety used in decisions to withdraw medicinal products from the UK and US markets. *Drug Saf* 2006; 29: 175–181.

6. Rawlins MD. Spontaneous reporting of adverse drug reactions. I: the data. *Br J Clin Pharmacol* 1988; 26: 1–5.

7. Ellenberg SS and Braun MM. Monitoring the safety of vaccines: assessing the risks. *Drug Saf* 2002; 25: 145–152.

8. Alvarez Y, Hidalgo A, Maignen F, *et al.* Validation of statistical signal detection procedures in eudravigilance post-authorization data: a retrospective evaluation of the potential for earlier signalling. *Drug Saf* 2010; 33: 475–487.

9. Bailey S, Singh A, Azadian R, *et al.* Prospective data mining of six products in the US FDA Adverse Event Reporting System: disposition of events identified and impact on product safety profiles. *Drug Saf* 2010; 33: 139–146.

10. Hochberg AM and Hauben M. Time-to-signal comparison for drug safety data-mining algorithms vs. traditional signaling criteria. *Clin Pharmacol Ther* 2009; 85: 600–606.

11. Hauben M and Noren GN. A decade of data mining and still counting. *Drug Saf* 2010; 33: 527–534.

12. Lindquist M, Edwards IR, Bate A, *et al.* From Association to Alert - a revised approach to International Signal Analysis. *Pharmacoepidemiol Drug Saf* 1999; 8: S15–S25.

13. Edwards IR. Adverse drug reactions: finding the needle in the haystack. *BMJ* 1997; 315: 500.

14. Bate A, Lindquist M, Edwards IR, *et al.* A Bayesian neural network method for adverse drug reaction signal generation. *Eur J Clin Pharmacol* 1998; 54: 315–321.

15. Fukazawa C, Hinomura Y, Kaneko M, *et al.* Significance of data mining in routine signal detection: analysis based on the safety signals identified by the FDA. *Pharmacoepidemiol Drug Saf* 2018; 27: 1402–1408.

16. Bate A, Juniper J, Lawton AM, *et al.* Designing and incorporating a real world data approach to international drug development and use: what the UK offers. *Drug Discov Today* 2016; 21: 400–405.

17. Harpaz R, Vilar S, Dumouchel W, *et al.* Combing signals from spontaneous reports and electronic health records for detection of adverse drug reactions. *J Am Med Inform Assoc* 2013; 20: 413–419.

18. Dore DD, Seeger JD and Arnold Chan K. Use of a claims-based active drug safety surveillance system to assess the risk of acute pancreatitis with exenatide or sitagliptin compared to metformin or glyburide. *Curr Med Res Opin* 2009; 25: 1019–1027.

19. Nelson JC, Cook AJ, Yu O, *et al.* Challenges in the design and analysis of sequentially monitored postmarket safety surveillance evaluations using electronic observational health care data. *Pharmacoepidemiol Drug Saf* 2012; 21(Suppl. 1): 62–71.

20. Colilla S, Tov EY, Zhang L, *et al.* Validation of new signal detection methods for web query log data compared to signal detection algorithms used with FAERS. *Drug Saf* 2017; 40: 399–408.

21. Caster O, Dietrich J, Kürzinger ML, *et al.* Assessment of the utility of social media for broad-ranging statistical signal detection in pharmacovigilance: results from the WEB-RADR project. *Drug Saf* 2018; 41: 1355–1369.

22. Kürzinger ML, Schück S, Texier N, *et al.* Web-based signal detection using medical forums data in France. *J Med Internet Res* 2018; 20: e10466.

23. Robb MA, Racoosin JA, Sherman RE, *et al*. The US Food and Drug Administration's Sentinel Initiative: expanding the horizons of medical product safety. *Pharmacoepidemiol Drug Saf* 2012; 21(Suppl. 1): 9–11.

24. Woodcock J, Behrman RE and Dal Pan GJ. Role of postmarketing surveillance in contemporary medicine. *Annu Rev Med* 2011; 62: 1–10.

25. Xu Y, Zhou X, Suehs BT, *et al*. A comparative assessment of observational medical outcomes partnership and mini-sentinel common data models and analytics: implications for active drug safety surveillance. *Drug Saf* 2015; 38: 749–765.

26. Ryan PB, Madigan D, Stang PE, *et al*. Empirical assessment of methods for risk identification in healthcare data: results from the experiments of the Observational Medical Outcomes Partnership. *Stat Med* 2012; 31: 4401–4415.

27. Wisniewski AF, Bate A, Bousquet C, *et al*. Good signal detection practices: evidence from IMI PROTECT. *Drug Saf* 2016; 39: 469–490.

28. Pacurariu AC, Straus SM, Trifirò G, *et al*. Useful interplay between spontaneous ADR reports and electronic healthcare records in signal detection. *Drug Saf* 2015; 38: 1201–1210.

29. AsPEN collaborators, Andersen M, Bergman U, *et al*. The Asian Pharmacoepidemiology Network (AsPEN): promoting multi-national collaboration for pharmacoepidemiologic research in Asia. *Pharmacoepidemiol Drug Saf* 2013; 22: 700–704.

30. Bate A, Reynolds RF and Caubel P. The hope, hype and reality of Big Data for pharmacovigilance. *Ther Adv Drug Saf* 2018; 9: 5–11.

31. Hauben M, Reynolds R and Caubel P. Deconstructing the pharmacovigilance hype cycle. *Clin Ther* 2018; 40: 1981–1990. e3.

32. Norén GN, Bergvall T, Ryan PB, *et al*. Empirical performance of the calibrated self-controlled cohort analysis within temporal pattern discovery: lessons for developing a risk identification and analysis system. *Drug Saf* 2013; 36(Suppl. 1): S107–S121.

33. Zhou X, Douglas IJ, Shen R, *et al*. Signal detection for recently approved products: adapting and evaluating self-controlled case series method using a US claims and UK electronic medical records database. *Drug Saf* 2018; 41: 523–536.

34. Cai B, Murugesan S, Geier J, *et al*. Applying High Dimensional Propensity Score (HDPS) in a exploratory data analysis with a US claims database for recent medicinal products. *Pharmacoepidemiol Drug Saf* 2014; 23: 13–14.

35. Brown JS, Petronis KR, Bate A, *et al*. Drug adverse event detection in health plan data using the Gamma Poisson Shrinker and comparison to the tree-based scan statistic. *Pharmaceutics* 2013; 5: 179–200.

36. Bate A, Edwards IR and Edwards J. Knowledge finding in IMS Disease Analyser Mediplus UK database - Effective data mining in longitudinal patient safety data. *Drug Saf* 2004; 27: 917–918.

37. Norén GN, Hopstadius J, Bate A, *et al*. Temporal pattern discovery in longitudinal electronic patient records. *Data Min Knowl Discov* 2010; 20: 361–387.

38. Schuemie MJ. Methods for drug safety signal detection in longitudinal observational databases: LGPS and LEOPARD. *Pharmacoepidemiol Drug Saf* 2011; 20: 292–299.

39. Shapiro S. Case control surveillance. In: Strom BL (ed.) *Pharmacoepidemiology*. Churchill Livingstone: New York, 2000, pp.15–28.

40. Wahab IA, Pratt NL, Wiese MD, *et al*. The validity of sequence symmetry analysis (SSA) for adverse drug reaction signal detection. *Pharmacoepidemiol Drug Saf* 2013; 22: 496–502.

41. Choi NK, Chang Y, Choi YK, *et al*. Signal detection of rosuvastatin compared to other statins: data-mining study using national health insurance claims database. *Pharmacoepidemiol Drug Saf* 2010; 19: 238–246.

42. Whalen E, Hauben M and Bate A. Time series disturbance detection for hypothesis-free signal detection in longitudinal observational databases. *Drug Saf* 2018; 41: 565–577.

43. Hallas J. Evidence of depression provoked by cardiovascular medication: a prescription sequence symmetry analysis. *Epidemiology* 1996; 7: 478–484.

44. Page D, Costa VS, Natarajan S, *et al*. Identifying adverse drug events by relational learning. *Proc Conf AAAI Artif Intell* 2012; 790–793.

45. Jin HW, Chen J, He H, *et al*. Mining unexpected temporal associations: applications in detecting adverse drug reactions. *IEEE Trans Inf Technol Biomed* 2008; 12: 488–500.

46. Zorych I, Madigan D, Ryan P, *et al*. Disproportionality methods for pharmacovigilance in longitudinal observational databases. *Stat Methods Med Res* 2013; 22: 39–56.

47. Watts S and Stenner P. Doing Q methodology: theory, method and interpretation. *Qual Res Psychol* 2005; 2: 67–91.

48. Thiébaut R and Thiessard F. Artificial intelligence in public health and epidemiology. *Yearb Med Inform* 2018; 27: 207–210.

49. Arnaud M, Bégaud B, Thurin N, *et al.* Methods for safety signal detection in healthcare databases: a literature review. *Expert Opin Drug Saf* 2017; 16: 721–732.

50. Norén GN, Hopstadius J, Bate A, *et al.* Safety surveillance of longitudinal databases: results on real-world data. *Pharmacoepidemiol Drug Safety* 2012; 21: 673–675.

51. Arnaud M, Bégaud B1, Thiessard F, *et al.* An Automated System Combining Safety Signal Detection and Prioritization from Healthcare Databases: A Pilot Study. *Drug Saf* 2018; 41: 377–387.

52. Cederholm S, Hill G, Asiimwe A, *et al.* Structured assessment for prospective identification of safety signals in electronic medical records: evaluation in the health improvement network. *Drug Saf* 2015; 38: 87–100.

53. Gould AL. Generating and confirming hypotheses. *Pharmacoepidemiol Drug Saf* 2010; 19: 533–536.

54. Walker AM. Orthogonal predictions: follow-up questions for suggestive data. *Pharmacoepidemiol Drug Saf* 2010; 19: 529–532.

55. Wang SV, Kulldorff M, Glynn RJ, *et al.* Reuse of data sources to evaluate drug safety signals: when is it appropriate? *Pharmacoepidemiol Drug Saf* 2018; 27: 567–569.

56. Toh S, Avorn J, D'Agostino RB Sr, *et al.* Re-using Mini-Sentinel data following rapid assessments of potential safety signals *via* modular analytic programs. *Pharmacoepidemiol Drug Saf* 2013; 22: 1036–1045.

57. Patadia VK, Schuemie MJ, Coloma P, *et al.* Evaluating performance of electronic healthcare records and spontaneous reporting data in drug safety signal detection. *Int J Clin Pharm* 2015; 37: 94–104.

58. Pottegard A, Friis S, Christensen Rd, *et al.* Identification of Associations Between Prescribed Medications and Cancer: A Nationwide Screening Study. *EBioMedicine* 2016; 7: 73–79.

59. Nguyen M and Munoz M. Duke Margolis Center for Health Policy. Implementation of Signal Detection Capabilities in the Sentinel System. Discussion Guide; 2018. Available at: https://healthpolicy.duke.edu/sites/default/files/atoms/files/sentinel_signal_detection_discussion_guide_final_print.pdf (accessed 15 April 2019).

Visit SAGE journals online journals.sagepub.com/home/taw

SAGE journals