



Methods for safety signal detection in healthcare databases: a literature review

Mickael Arnaud, Bernard Bégaud, Nicolas Thurin, Nicholas Moore, Antoine Pariente & Francesco Salvo

To cite this article: Mickael Arnaud, Bernard Bégaud, Nicolas Thurin, Nicholas Moore, Antoine Pariente & Francesco Salvo (2017): Methods for safety signal detection in healthcare databases: a literature review, Expert Opinion on Drug Safety, DOI: [10.1080/14740338.2017.1325463](https://doi.org/10.1080/14740338.2017.1325463)

To link to this article: <http://dx.doi.org/10.1080/14740338.2017.1325463>



Accepted author version posted online: 11 May 2017.
Published online: 15 May 2017.



Submit your article to this journal [↗](#)



Article views: 2



View related articles [↗](#)



View Crossmark data [↗](#)

REVIEW



Methods for safety signal detection in healthcare databases: a literature review

Mickael Arnaud^{a,b}, Bernard Bégaud^{a,b,c}, Nicolas Thurin^{a,b,d}, Nicholas Moore^{a,b,c,d}, Antoine Pariente^{a,b,c,d} and Francesco Salvo^{a,b,c}

^aUniversity of Bordeaux, Bordeaux, France; ^bBordeaux Population Health Research Centre, Pharmacoepidemiology team, INSERM UMR1219, Bordeaux, France; ^cCHU Bordeaux, Service de Pharmacologie Médicale, Bordeaux, France; ^dCIC Bordeaux

ABSTRACT

Introduction: With increasing availability, the use of healthcare databases as complementary data source for drug safety signal detection has been explored to circumvent the limitations inherent in spontaneous reporting.

Areas covered: To review the methods proposed for safety signal detection in healthcare databases and their performance.

Expert opinion: Fifteen different data mining methods were identified. They are based on disproportionality analysis, traditional pharmacoepidemiological designs (e.g. self-controlled designs), sequence symmetry analysis (SSA), sequential statistical testing, temporal association rules, supervised machine learning (SML), and the tree-based scan statistic. When considering the performance of these methods, the self-controlled designs, the SSA, and the SML seemed the most interesting approaches. In the perspective of routine signal detection from healthcare databases, pragmatic aspects such as the need for stakeholders to understand the method in order to be confident in the results must be considered. From this point of view, the SSA could appear as the most suitable method for signal detection in healthcare databases owing to its simple principle and its ability to provide a risk estimate. However, further developments, such as automated prioritization, are needed to help stakeholders handle the multiplicity of signals.

ARTICLE HISTORY

Received 13 December 2016
Accepted 27 April 2017

KEYWORDS

Drug safety;
pharmacoepidemiology;
pharmacovigilance; signal
detection; data mining

1. Introduction

Since the early 1970s, spontaneous reporting (SR) has been the cornerstone of signal detection in drug safety surveillance [1–3]. SR works well for rare and acute adverse drug reactions (ADR), such as bullous eruptions, agranulocytosis, or hepatotoxicity [4]. Most drug withdrawals during the last decades were based on SR, which often provided the only information available [5,6]. Nevertheless, SR is plagued by certain limits such as under- or selective reporting or the absence of information about the actual number of exposed patients [7–10] that can potentially hamper or delay the identification of these safety signals. It also appears poorly efficient in identifying ADRs concerning events which are not a priori evocative of drug causation: pulmonary infections related to proton pump inhibitors [11,12] or myocardial infarction induced by rofecoxib [13–15]. Although signal detection based on SR is still improving [16–19], the availability of large health-care databases, which allows to follow cohorts of several million persons, opens opportunities, complementary to SR, for drug safety surveillance and signal detection [20,21].

Several initiatives have thus emerged worldwide: ‘the Exploring and Understanding Adverse Drug Reactions by integrative mining of clinical records and biomedical knowledge (EU-ADR)’ [22] and ‘the Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium’ [23]

projects in Europe, the Sentinel System [24,25] and the Observational Medical Outcomes Partnership (OMOP) [26,27] in the United States, or again the Asian Pharmacoepidemiology Network (AsPEN) in Asia [28,29].

This perspective of setting up new post-marketing monitoring systems using the information available in health-care databases gave birth to a broad range of approaches. The purpose of this paper was to provide a commented overview of the methods proposed so far for signal detection in health-care databases.

2. Methods proposed for safety signal detection in health-care databases

All the data-mining methods proposed for drug safety signal detection in health-care databases are detailed in the following sections; their strengths and weaknesses are summarized in Table 1. They are classified in the following categories: disproportionality analysis, traditional pharmacoepidemiological designs, sequence symmetry analysis (SSA), sequential statistical testing, temporal association rules (TAR), supervised machine learning (SML), and tree-based scan statistic. Although other and more complex classifications could be as valuable, we use this one as it reflects the fundamental differences in the design of the methods and respects the thinking of their designers.

Article highlights

- For over a decade, there has been a strong push from stakeholders and researchers in pharmacovigilance to use the information available in healthcare databases for safety signal detection.
- Different data mining methods have been developed and tested for that purpose, and they are based on: disproportionality analysis, traditional pharmacoepidemiological designs (e.g. self-controlled designs), sequence symmetry analysis (SSA), sequential statistical testing, temporal association rules, supervised machine learning (SML), and tree-based scan statistic.
- Although a comparison across studies using different reference standards is a difficult exercise, the performance of these methods has been assessed. The results suggest that self-controlled designs, SSA, and SML seem the most interesting approaches for safety signal detection in healthcare databases.
- From a pragmatic point of view, the SSA could be most suitable owing to, for example, its simple principle that allows to understand the method and to trust its results.
- The future challenge should consist in developing a prioritization method to help stakeholders to handle the multitude of signals detected. This appears crucial for making safety signal detection a routine activity.

This box summarizes key points contained in the article.

2.1. The disproportionality analysis approach

Disproportionality analysis methods were originally developed on SR databases. In this approach, all spontaneous reports are displayed in the form of a large contingency table with dimensions corresponding to all the drugs and events encountered at least once. The information is then aggregated into a 2×2 contingency table to compute the ratio of the observed-to-expected count of reports associated with each drug-event association. Four main methods are currently used in SR databases (e.g. the gamma Poisson shrinker [GPS] is used by the United States Food and Drug Administration [48]) and differ with respect to the manner in which disproportionality is measured, and in which low counts are accounted for in the analysis [48–51].

2.1.1. SR-like methods

The main issue when transposing these methods to healthcare databases was to determine how to generate drug safety reports from longitudinal data. Curtis *et al.* considered, for each patient, each month of the observation time as a

Table 1. Strengths and limitations of the methods tested for safety signal detection on health-care databases.

Method	Strengths	Limitations
Disproportionality analysis		
SR-like methods [30]	Easy to implement Can incorporate shrinkage for preventing detecting some spurious signals related to very rare events	Does not provide risk estimates Loss of information due to aggregated data
LGPS-LEOPARD [31]	Provides risk estimates Easy to implement Uses shrinkage for preventing detecting spurious some signals related to very rare events	Unable to handle numerous confounders Sensitive to protopathic and indication biases Loss of information due to aggregated data Unable to handle numerous confounders Sensitive to protopathic and indication biases (theoretically addressed when associated with the LEOPARD method)
Traditional pharmacoepidemiological designs		
New user cohort design [32]	Provides risk estimates Allows controlling for high-dimensional confounding	Needs very large dataset to have enough power to detect signals related to rare events Difficulties to determine the settings as these are not supposed to be standardized for all the drug-event associations screened
Matched case-control design [33]	Provides risk estimates Allows controlling for some confounders thanks to matching and nesting	Needs very large dataset to have enough power to detect signals related to rare events Does not control for numerous confounders
SCCS design [34]	Provides risk estimates Robust to confounders that are stable over time Allows controlling for high-dimensional time-varying confounding	Difficulties to determine the settings as these are not supposed to be standardized for all the drug-event associations Theoretically inappropriate for chronic drug use and for nonrecurrent events Difficulties to determine the settings as these are not supposed to be standardized for all the drug-event associations Sensitive to protopathic and indication biases
CC design [35]	Provides risk estimates Robust to confounders that are stable over time	Theoretically inappropriate for chronic drug use and for nonrecurrent events Does not address time-varying confounding Difficulties to determine the settings as these are not supposed to be standardized for all the drug-event associations
SCC design [36]	Provides risk estimates Robust to confounders that are stable over time	Sensitive to protopathic and indication biases Theoretically inappropriate for chronic drug use and for nonrecurrent events Does not address time-varying confounding Difficulties to determine the settings as these are not supposed to be standardized for all the drug-event associations
Sequence symmetry analysis [37]	Provides risk estimates Robust toward confounders that are stable over time Easy to understand and to implement	Sensitive to protopathic and indication biases Does not address time-varying confounding Inapplicable for death

(Continued)

Table 1. (Continued).

Method	Strengths	Limitations
Sequential statistical testing		
MaxSPRT [38,39]	Maintains type I error at 0.05 across multiple testing	Does not provide risk estimates Loss of information due to aggregated data Inappropriate for chronic exposure and for very rare events Loses information when using matching, otherwise requires large historical data Sensitive to protopathic and indication biases Computational load
CSSP [40]	Maintains type I error at 0.05 across multiple testing Works well for rare events	Does not provide risk estimates Unable to handle numerous confounders Sensitive to protopathic and indication biases Difficulties to maintain type I error at 0.05 across multiple testing when there are many strata computational heaviness
Temporal association rule		
MUTARA/HUNT [41,42]	Incorporates a filter that prevents detecting expected signals	Does not provide risk estimates Sensitive to protopathic and indication biases Does not have a natural threshold for discriminating positive to negative signals Inapplicable for death
TPD [43]	Robust to confounders that are stable over time Calibrated for systematic differences between time-at-risk and control periods Uses shrinkage for preventing detecting some spurious signals related to very rare events	Does not provide risk estimates Difficulties to address time-varying confounding Inapplicable for death
Fuzzy logic [44]	Based on imputability criteria for signal detection	Does not provide risk estimates Difficulties to define the imputability criteria for an automated use Sensitive to protopathic and indication biases Does not have a natural threshold for discriminating positive to negative signals
Supervised machine learning		
[45,46]	Based on Bradford-Hill's causality criteria that make the method more robust to the detection of false positive signals Performance for signal detection should improve with increased data	Does not provide risk estimates Needs both large data and a large reference set for training efficiently the random forest model Needs to set up one random forest model per drug screened Sensitive to protopathic and indication biases Computational load
Tree-based scan statistic [47]		
	Maintains type I error at 0.05 across multiple testing Tests simultaneously different event definitions	Does not provide risk estimates Unable to handle confounders Sensitive to protopathic and indication biases Inapplicable for the death event

SR: spontaneous report; LGPS: longitudinal gamma Poisson shrinker; LEOPARD: longitudinal evaluation of observational profiles of adverse events related to drugs; SCCS: self-controlled case series; CC: case crossover; SCC: self-controlled cohort; maxSPRT: maximized sequential probability ratio test; CSSP: conditional sequential sampling procedure; TAR: temporal association rules; MUTARA: mining the unexpected TARs given the antecedent; HUNT: highlighting TARs negating TARs; TPD: temporal pattern discovery.

pseudo-report including drug exposures and events which occurred during this month, or drug exposures without any event associated, or events which occurred in the absence of drug exposure [52]. For Choi *et al.* [53,54] and Kim *et al.* [55], the observation time starts with drug prescription and ends after 12 weeks; only events which occurred in this period are considered as 'reports.' Zorych *et al.* proposed to consider the duration of drug exposure as the observation time, whatever its duration, and experimented three ways to build the 2×2 contingency table used for the disproportionality analyses [30]. They first considered each patient only once in the contingency table. Classification depended whether or not the patient had experienced the event during drug exposure. The two other approaches corresponded to that experimented in Curtis *et al.* [52], and in Choi *et al.* [53,54] and Kim *et al.* [55].

2.1.2. Longitudinal gamma Poisson shrinker

Schuemie proposed an alternative method in which the exposure and non-exposure periods are expressed in patient-days to exploit better the information available in longitudinal data

[56]. This approach was tested in combination with the GPS and called longitudinal GPS (LGPS). The LGPS includes a minimalist adjustment on age and sex to control for basic confounders. The author proposed to combine the LGPS with the longitudinal evaluation of observational profiles of adverse events related to drugs (LEOPARD) method aiming to discard spurious signals related to a protopathic or indication bias. The principle is to compare rates of prescriptions during a fixed-time window before and after the occurrence of the event; if this rate is greater after the event than before, LEOPARD considers the association as related to protopathic or indication bias [56].

2.2. The traditional pharmacoepidemiological approach

OMOP, and to a lesser extent EU-ADR, has tested several methods based on traditional pharmacoepidemiological study designs [31,35,57–59]. These designs, which have been extensively used for ad hoc studies, consist in a two-step process: (1) to identify prospectively or retrospectively two groups of patients based on

exposures (cohort approach) or events (case based), and (2) to compare the rate of the drug-event association in these groups. Statistical tools are usually available with these designs to control for putative confounders (e.g. co-prescriptions).

2.2.1. New user cohort design

The basic principle of the new user cohort designs is (1) to follow prospectively cohorts of patients from the start of a first drug exposure: a first cohort would include patients newly exposed to the drug of interest, while another cohort includes patients newly exposed to another drug (generally, a drug sharing the same indication) and (2) to compare the rate of occurrence of the event(s) of interest in these two cohorts.

This approach has been extensively explored, as the cohort design provides many solutions for addressing confounders [31,32,35] such as adjusting incidence rate ratio for age and sex by using Mantel-Haenszel adjustment [31], or using propensity scores or high-dimensional propensity scores to weigh a Cox proportional hazards model or to adjust by means of a classical logistic regression model [32,35], or to match patients in both cohorts. Penalized logistic regression models were also tested (e.g. the lasso regression [60]); these models handle a large number of covariates by selecting those with the highest confounding ability and by including them in a classical logistic regression model [32,35].

2.2.2. Matched case-control designs

The basic principle of matched case-control designs is (1) starting from a given date to analyze retrospectively prior drug exposure(s) among two groups of subjects matched on confounders (e.g. age, sex). The first group includes patients who have experienced the event of interest (i.e. the 'cases'), and the second patients free of this event (i.e. the 'controls'), (2) to compare the odds of exposure to the drug(s) of interest in these two groups.

The availability of large health-care databases popularized the use of case-control designs nested in a cohort of patients, which improves the comparability across groups. Though the use of propensity scores or disease risk scores to control for numerous confounders is theoretically possible, these two statistical tools have apparently not been tested in signal detection using a matched case-control design [31,33,35]. In addition to the classical matched case-control design, a more original and complex approach was also tested, called multi-set case-control estimation, which enables to estimate odds ratio simultaneously for multiple events and multiple drug exposures [35].

2.2.3. Self-controlled designs

The self-controlled designs differ from the previous ones in that only one cohort of patients is considered and each patient is his/her own control. The effect of a drug on the occurrence of an event is measured for each patient by comparing the event rate in exposed periods to that in unexposed periods. Self-controlled designs implicitly control for all time-invariant (e.g. chronic comorbidities) and patient-invariant confounders (e.g. genetic risk factors).

Three different self-controlled designs have been investigated in the framework of safety signal detection [34–36]. The

self-controlled case series (SCCS) design considers only patients who have been both exposed to the drug and have experienced the event of interest at least once. As patients are followed prospectively, this design can be considered as a cohort analysis. Statistical tools through penalized regression models were developed to apply high-dimensional multivariate adjustment with the SCCS to control for time-varying confounders (e.g. acute diseases) [34,35].

The case-crossover (CC) design is similar to the SCCS but uses a case-control approach in the sense that drug exposure is explored retrospectively. However, contrary to the SCCS, no specific statistical tool has been yet developed to control for time-varying confounders [35].

The self-control cohort (SCC) design differs from the SCCS and the CC: it considers all the exposed patients whether or not they have experienced the event of interest. In this design, incidence rate ratios correspond to the ratio of event incidence rates during/after versus before the start of drug exposure. As for the CC, the SCC cannot handle time-varying confounders, except by stratification [35,36].

2.3. The SSA method

The basis of this method was introduced by Petri *et al.* in 1988 [61] and was conceptualized by Hallas in 1996 [37]. Its aim is to compare the sequence of the initiation of two drug exposures A and B within a given time-window, where drug exposure A is the drug exposure of interest and drug exposure B used as a surrogate for the potential adverse event. If drug exposure A induces the prescription of drug exposure B as a consequence of an ADR, the number of patients that initiated drug exposure A first and drug exposure B in second is expected to exceed the number of patients that initiated drug exposure B before drug exposure A.

This *crude sequence ratio* is, by essence, not affected by confounders that are stable over time, but sensitive to changes in prescribing trends. For instance, if reimbursements for drug exposure A increased during the study period while those for drug exposure B remained stable, this trend would, by itself, result in an excess of sequences where drug exposure A precedes drug exposure B; this could hamper the detection of a potential signal. Hallas proposed an adjustment for correcting such temporal trends by dividing the crude sequence ratio by the *null-effect sequence ratio*, which is the sequence ratio that would have been expected from the trends in drug use if drug exposures A and B were independent [37].

Tsiropoulos *et al.* slightly modified the computation of the null-effect sequence ratio to account for shorter observation time-windows between the initiations of the two studied drugs [62]. The same authors also validated the use of hospitalization diagnoses instead of drug reimbursements for adverse event selection [62].

2.4. The sequential statistical testing approach

The group of sequential statistical testing methods aims to test sequentially (e.g. on a monthly basis) the null hypothesis – the event rate is higher among exposed patients

compared to unexposed – on prospective cohort data, as it could be done in a routine signal-detection activity. Each new analysis takes into account the number of new patients exposed and unexposed to the drug of interest since the last analysis, and the increment in exposure time for patients already included in the previous analysis. A signal is raised if the test statistic exceeds a predefined critical value, which is chosen so that the overall type I error is maintained at $\alpha = 0.05$ across the multiple tests to reduce the generation of false positives.

2.4.1. The maximized sequential probability ratio test

Brown *et al.* tested the maximized sequential probability ratio test (maxSPRT), which was already implemented in vaccine safety monitoring [38]. This method consists in applying sequentially the log-likelihood ratio (LLR) test statistic, which is computed using a large cohort of historical controls. Kulldorf *et al.* suggested to collect events simultaneously from exposed and unexposed patients to fit better situations in which the use of a cohort of historical controls is not suitable, as for example for signal detection of newly marketed drugs [39]. In the proposed approach, an exposure matching with a fixed matching ratio (1:M) is used to control for confounding. Cook *et al.* proposed a generalized LLR test statistic, which is computed considering all the patients instead of only those who experienced the event, as this provides a more reliable estimate when the event of interest becomes frequent [63].

2.4.2. The conditional sequential sampling procedure

The conditional sequential sampling procedure (CSSP) was developed by Li *et al.* [40] because maxSPRT failed to handle chronic drug exposures [63]. In this approach, the population is first stratified in two groups depending on whether the persons are exposed or not. The two groups are stratified a second time according to the categories of each confounder considered in the analysis. The cumulative drug exposure and the number of events are then computed, within each stratum, since the previous analysis. The total number of adverse events only observed among the persons exposed up to the time analysis corresponds to the test statistic and is compared to a critical value computed using a CSSP and corresponding to the number of events that would be expected to occur considering all the patients in each stratum.

2.5. The TAR approach

In the context of signal detection, TAR algorithms consider the following two rules: (1) the event must follow drug exposure and (2) the event must occur during a prespecified time-window (i.e. the period considered at risk). For a given drug, all potential events are mined sequentially, and a correlation score is computed using a measure of interestingness.

2.5.1. MUTARA/HUNT

Jin *et al.* proposed the MUTARA (mining the unexpected TARs given the antecedent) algorithm that incorporates a third rule specifying that the event must occur ‘unexpectedly’ [41]. For each user of the drug of interest, a reference period is set before the start of drug exposure. If the event studied is

observed in that particular period, the authors consider that it is expected to also find this event after the start of exposure. In that case, it is considered that the event has few chances to be an ADR; thus, all its occurrences are excluded from the patient data. Computation of the correlation score is then performed in the filtered data.

As this measure of interestingness appeared to be prone to detect spurious signals due to protopathic or indication bias [41], the authors incorporated a new metric corresponding to the ratio of the rank of the signal based on the correlation score calculated in the whole data to that calculated in the filtered data. The method was renamed ‘highlighting unexpected TARs negating TARs (HUNT)’ [42].

2.5.2. Temporal pattern discovery

Norén *et al.* implemented the temporal pattern discovery (TPD) method, which considers several control periods prior to the start of drug exposure in order to adjust for systematic variability in event rates over time [43]. These control periods consider both the same patient and another patient using a drug sharing similar indication with the drug of interest. As measure of interestingness, TPD uses the ratio of the expected-to-observed ratio for the time-at-risk period to that for the control period; this ratio is computed for each control period and the minimal value is then selected. To make them more robust against random variability when event counts become small, the ratios are transformed: (1) a constant is added both to the nominator and the denominator in order to pull the ratios toward a value of 1 (i.e. absence of association) and (2) a base 2 logarithm is applied to the transformed ratios to make the distribution more regular [43].

2.5.3. Fuzzy logic rule based

An approach combining TAR algorithm with fuzzy logic was proposed by Ji *et al.* to add the degree of causality between the drug and the event in the TAR definition [44,64,65]. A first individual score of causality is computed for each case of the drug-event association using fuzzy rules. This score takes into account, temporality, existence of other explanations, dechallenge, and rechallenge. For instance, the fuzzy rule *temporality* considers the temporal association as *likely*, *possible*, or *unlikely* depending on the duration between the start of the exposure and the event occurrence. Similarly, the fuzzy rule *dechallenge* concludes in *likely* if the patient is still alive after the discontinuation of the drug of interest. A score is attributed to each value of the fuzzy rules; for example, each *likely* value corresponds to a score of 1. All the individual scores are combined to provide a global score of causality for each drug-event association and then included in the computation of the measure of interestingness.

2.6. The SML approach

The basis of the SML algorithms can basically be divided in two parts (Figure 1). The first consists in training a classifier (e.g. a random forest model) by using a reference set that includes drug-event associations a priori known as being related or not-related. For each association, a vector of

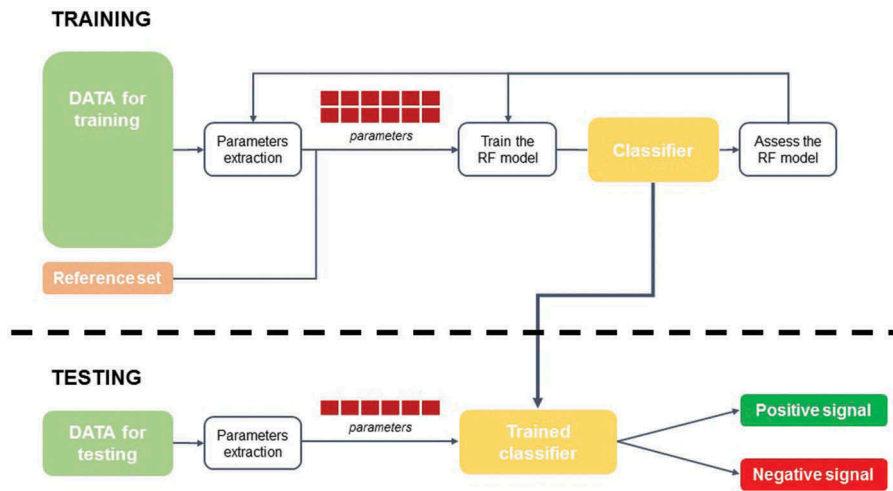


Figure 1. Basis of the supervised machine learning algorithms.

The first part consists of training a classifier by using a reference set that includes drug-event associations known to be related or unrelated. For each association, a vector of pre-determined parameters, which correspond to proxies of the association, is extracted from the sample of data used for training. All the vectors constitute the input data of the classifier, which is trained using a re-sampling method and an impurity criterion to select the parameters that provide the best ability for the correct identification of the associations included in the reference dataset. The second part consists to extract from the sample of data used for testing the selected parameters for each possible drug-event association, and to then apply the trained classifier to identify those associations that could be new ADRs.

predetermined parameters corresponding to proxies for this association is extracted from the sample of data used for training. All the vectors constitute the input data of the classifier, which is programmed to select the best parameters that identify the true associations of the reference set by using a resampling method and an impurity criterion. The second part consists in extracting the selected parameters for each drug-event association screened from the sample of data used for testing and to apply the trained classifier to predict those associations that could be new ADRs.

Reps *et al.* proposed a first SML version in which the parameters required for the random forest model classifier were six risk-ratio values generated from six simple cohort studies [45]. Although these cohort studies used identical at-risk population and a fixed time-at-risk period, different control populations and control periods were considered. No statistical adjustment for controlling for confounders was used. Three additional parameters were included to indicate the deviation of the strength of the association when varying the setting control.

The same authors proposed another SML version, which included parameters inspired from Bradford-Hill's causality criteria to make the method less prone to detect spurious ADRs [46]. For each drug-event association screened, 27 parameters were computed based on the age, sex of the patients, the drug dosage, and the number of co-medications: 17 parameters referred to 5 of the Bradford-Hill's criteria – strength, temporality, experimentation, biological gradient, and specificity – while the other 10 indicated the deviation of some of the previous parameters when varying the event definition according to the International Classification of Diseases, 10th revision (ICD-10).

2.7. The tree-based scan statistic method

The fundamental principle of the tree-based scan statistic method proposed by Kulldorff *et al.* is to map a tree according

to the basis of the hierarchical structure of classifications used for coding events [47]: *the root* corresponds to the broadest definition of a given event, *the nodes* correspond to the different sublevel definitions, *the leaves* correspond to the codes with the finest definitions, and *the branches* link the three elements together (Figure 2a). For each leaf, the observed and the age and sex-adjusted expected number of exposed persons who experienced the selected event is computed. Then, all the possible samples of a given root–node–leaf event pathway are tested simultaneously using the LLR test statistic (Figure 2b). The method uses Monte-Carlo-based p values to formally adjust the p values for multiple testing due to the many overlapping definitions of the events to maintain the overall type I error at $\alpha = 0.05$ [66].

3. Performance assessment

The performance for signal detection of the methods above presented is detailed in Table 2 and briefly discussed below.

The OMOP and EU-ADR collaborative projects mainly assessed the methods based on disproportionality analyses and traditional pharmacoepidemiological designs, and concluded that the former ones were less efficient, notably when compared to self-controlled designs, which achieved the best performance [31,35,57,59]. They also compared the methods based on sequential statistical approach, which appeared to perform worse than a random signal detection [35]. The TPD method seemed to perform quite well [57,59] but the work of Reps *et al.* showed that the TAR algorithms performed as poorly as disproportionality based methods [70]. However, all these results should be considered with caution as the reference sets used for the computations suffered from several limitations such as their small size [31,35], the fact that some drug classes were too heterogeneous to be pooled (e.g. antibiotics) [35], or that adverse events were inadequately defined [57,59]. More details

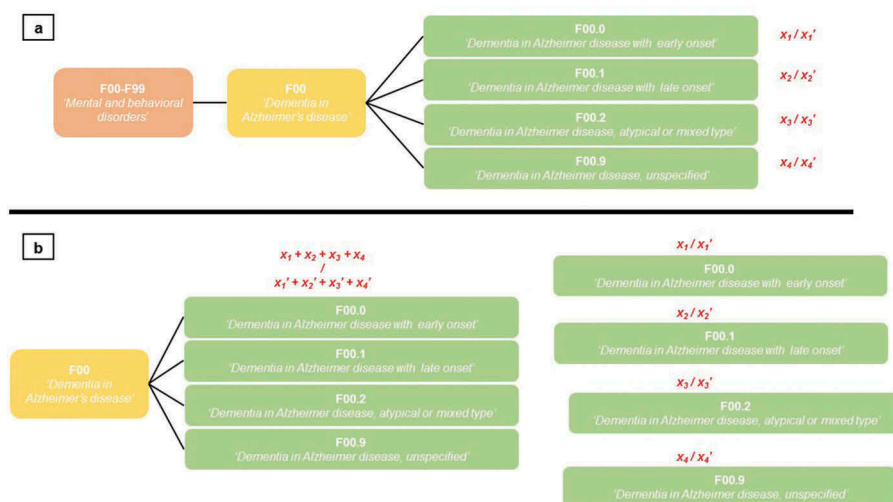


Figure 2. Example of application of the tree-based scan method to the Alzheimer's disease.

The root is in brown; the node in yellow; the leaves in green; the branches in solid black lines;

X_1, X_2, X_3, X_4 correspond to the observed numbers of patients that experienced the event;

X_1', X_2', X_3', X_4' correspond to the expected numbers of patients that should have experienced the event.

The basis of tree-based scan method is to map the ICD-10 classification on the model of a tree. The part of that tree related to the Alzheimer's disease is drawn here: the root corresponds to the broadest definition of this disease; the nodes correspond to the different sublevel definitions; the leaves correspond to the most specific diagnoses related to this disease; and the branches link the three elements together (Figure 2a).

All the associations between a given drug exposure and each potential element of the tree are tested simultaneously using the log likelihood ratio test statistic to assess different levels of the Alzheimer's disease definition (Figure 2b).

about these methodological limitations can be found elsewhere [72–75].

Concerning SSA, it showed values of sensitivity and specificity suggesting that it performs as well as, or even better than, the self-controlled designs [71]. Other studies provided information that strengthens the confidence in the generalization of the results of signal detection based on SSA in health-care databases [76–79]. A simulation study demonstrated that the SSA provided reliable effect estimates when one varied the prevalence of use and the trend of use of the drug implied in the signal [77], while two other studies indicated that this method could produce similar effect estimates in health-care databases covering larger and various populations [76,78]. A more recent study even led to discover new safety signals, which appear biologically plausible, and is supported by the presence of few case reports (e.g. histamine antagonists and heart failure) [79].

Reps *et al.* assessed separately the performance of the two SML versions in the THIN database, both providing values of performances that exceeded those of the self-controlled designs [45,46], even if they might have been inflated by the reference sets used. Indeed, for the version using the Bradford-Hill's criteria [46], the authors built a reference set that included only 10% of true associations; such imbalance between true and false associations automatically increases the method's ability for prediction [80]. For the other SML version [45], the reference set included as true associations commonly observed adverse events whatever the drug used (e.g. nausea), and events that are unlikely to be related to drugs as false associations (e.g. dog bite).

Unlike the other methods, the tree-based scan statistic method was not tested with a reference set, but it is noteworthy that one study applied this method to detect new

safety signals in health-care databases and identified a signal that justified further investigation [47].

4. Conclusion

A wide range of data-mining methods has been explored for safety signal detection in health-care databases. Methods currently used in pharmacovigilance – e.g. disproportionality analyses – or in pharmacoepidemiology – e.g. traditional study designs combined with statistical tools for controlling for confounding – have been adapted and tested. More specifically, the availability of large amount of longitudinal data provided the opportunity to develop dedicated methods such as SSA or sequential statistical testing methods. This new opportunity for signal detection in pharmacovigilance research has also attracted the attention on data mining methods used in other scientific fields such as artificial intelligence and gave birth to methods based on TAR algorithms, the SML approach, and the tree-based scan statistic. Although numerous studies attempted to assess the performance of these methods, consensus about the best approach to use is far from being reached.

5. Expert opinion

Comparing the retrieved data-mining methods described is difficult, as their performance was evaluated using different reference standards. For instance, to identify events of interest, OMOP developed an algorithm based on ICD-9 diagnostic hospitalization codes, and/or diagnostic or therapeutic procedure codes, and/or laboratory results [27]; conversely, AsPEN considered drug use as a surrogate for adverse events identification [78]. Nonetheless, from the performance assessments,

Table 2. Performance for safety signal detection.

Study	Main settings used for the performance assessment	Results
Murphy <i>et al.</i> [35]	<p><i>Data source:</i> 6 private claims databases and 4 private electronic health-care records databases covering a population of 190 million persons living in the United States</p> <p><i>Reference set:</i> 9 positive and 44 negative controls, covering</p> <ul style="list-style-type: none"> – <i>Exposure:</i> angiotensin-converting enzyme inhibitors, amphotericin B, antibiotics, antiepileptics, benzodiazepines, beta-blockers, bisphosphonates, tricyclic antidepressants, typical antipsychotics, warfarin – <i>Events:</i> angioedema, aplastic anemia, acute liver injury, bleeding, myocardial infarction, hip fracture, mortality after myocardial infarction, renal failure, gastrointestinal ulcer hospitalization <p><i>Main performance criteria:</i> AUC</p>	<p><i>SR-like:</i> 0.63</p> <p><i>New user design</i></p> <ul style="list-style-type: none"> – <i>Weighted Cox model:</i> 0.47 – <i>HDPS:</i> 0.68 – <i>Bayesian regression model:</i> 0.64 <p><i>Matched case-control design</i></p> <ul style="list-style-type: none"> – <i>Classical:</i> 0.61 – <i>Multi-set estimations:</i> 0.57 <p><i>Self-controlled design</i></p> <ul style="list-style-type: none"> – <i>SCCS:</i> 0.57 – <i>CC:</i> 0.61 – <i>SCC:</i> 0.53 <p><i>MaxSPRT:</i> 0.23</p> <p><i>CSSP:</i> 0.38</p> <p><i>TPD:</i> 0.65</p>
Ryan <i>et al.</i> [57]	<p><i>Data source:</i> 4 private claims databases and 1 private electronic health-care records database covering a population of 74 million persons living in the United States</p> <p><i>Reference set:</i> 165 positive ADRs and 234 negative controls</p> <ul style="list-style-type: none"> – <i>Exposure:</i> 183 drugs, more details in Ref. [67] – <i>Events:</i> acute liver injury, acute myocardial infarction, acute kidney injury, upper gastrointestinal bleeding <p><i>Main performance criteria:</i> AUC</p>	<p><i>SR-like:</i> 0.53</p> <p><i>LGPS–LEOPARD:</i> 0.58</p> <p><i>New user design (HDPS):</i> 0.54</p> <p><i>Classical matched case-control design:</i> 0.69</p> <p><i>SCCS:</i> 0.71</p> <p><i>SCC:</i> 0.81</p> <p><i>TPD:</i> 0.75</p>
Schuemie <i>et al.</i> [31]	<p><i>Data source:</i> 4 public claims databases and 3 public electronic health-care records databases covering a total population of 19 million persons living in Italy, Netherlands, or Denmark</p> <p><i>Reference set:</i> 165 positive ADRs and 234 negative controls</p> <ul style="list-style-type: none"> – <i>Exposure:</i> 68 drugs, more details in Ref. [68] – <i>Events:</i> bullous eruption, acute renal failure, anaphylactic shock, acute myocardial infarction, rhabdomyolysis, pancytopenia, neutropenia, cardiac valve fibrosis, acute liver injury, upper gastrointestinal bleeding <p><i>Main performance criteria:</i> AUC</p>	<p><i>SR-like:</i> 0.72</p> <p><i>LGPS–LEOPARD:</i> 0.83</p> <p><i>New user design</i></p> <ul style="list-style-type: none"> – <i>Without statistical adjustment:</i> 0.76 – <i>HDPS:</i> 0.77 <p><i>Classical matched case-control design:</i> 0.75</p> <p><i>SCCS:</i> 0.76</p>
Schuemie <i>et al.</i> [59]	<p><i>Data source:</i> 3 public claims databases and 3 electronic health-care records databases covering a population of 11 million persons living in Italy, Netherlands, or Denmark</p> <p><i>Reference set:</i> 165 positive ADRs and 234 negative controls</p> <ul style="list-style-type: none"> – <i>Exposure:</i> 183 drugs, more details in Ref. [69] – <i>Events:</i> acute liver injury, acute myocardial infarction, acute kidney injury, upper gastrointestinal bleeding <p><i>Main performance criteria:</i> AUC</p>	<p><i>SR-like:</i> 0.60</p> <p><i>LGPS–LEOPARD:</i> 0.59</p> <p><i>New user design (HDPS):</i> 0.61</p> <p><i>Classical matched case-control design:</i> 0.59</p> <p><i>SCCS:</i> 0.67</p> <p><i>SCC:</i> 0.75</p> <p><i>TPD:</i> 0.67</p>
Reps <i>et al.</i> [70]	<p><i>Data source:</i> the THIN electronic health-care records covering a population of 11 million persons living in the United Kingdom</p> <p><i>Reference set:</i> NA</p> <ul style="list-style-type: none"> – <i>Exposure:</i> drugs belonging to the following drug classes: nonsteroidal anti-inflammatory drugs, tricyclic antidepressants, penicillins, quinolones, calcium channel blocker drugs, and sulfonylureas – <i>Events:</i> all ICD-9 codes – <i>Validation of drug-event pairs:</i> ADRs listed or not in the British National Formulary <p><i>Main performance criteria:</i> AUC</p>	<p><i>SR-like:</i> 0.55</p> <p><i>TPD:</i> 0.57</p> <p><i>MUTARA:</i> 0.60</p> <p><i>HUNT:</i> 0.57</p>
Reps <i>et al.</i> [45]	<p><i>Data source:</i> the THIN electronic health-care database covering a population of 11 million persons living in the United Kingdom</p> <p><i>Reference set:</i> 64 positive ADRs and 141 negative ADRs</p> <ul style="list-style-type: none"> – <i>Exposure:</i> three drugs belonging to the penicillin drug class – <i>Events:</i> NA <p><i>Main performance criteria:</i> AUC</p>	<i>SML using an ensemble of simple studies:</i> 0.81
Reps <i>et al.</i> [46]	<p><i>Data source:</i> the THIN database covering a population of 11 million persons living in the United Kingdom</p> <p><i>Reference set:</i> 405 positive ADRs and 3844 negative controls</p> <ul style="list-style-type: none"> – <i>Exposure:</i> angiotensin-converting enzyme inhibitors, amphotericin B, antibiotics, antiepileptics, benzodiazepines, beta-blockers, bisphosphonates, tricyclic antidepressants, typical antipsychotics, warfarin – <i>Events:</i> events possibly related to angioedema, aplastic anemia, acute liver injury, bleeding, myocardial infarction, hip fracture, mortality after myocardial infarction, renal failure, gastrointestinal ulcer hospitalization <p><i>Main performance criteria:</i> AUC</p>	<i>SML using the Bradford-Hill's criteria:</i> 0.86

(Continued)

Table 2. (Continued).

Study	Main settings used for the performance assessment	Results
Wahab <i>et al.</i> [71]	<p><i>Data source:</i> the DVA public claims database covering a population of 250,000 persons living in Australia</p> <p><i>Reference set:</i> 44 positive ADRs and 121 negative controls</p> <ul style="list-style-type: none"> – <i>Exposure:</i> perindopril, ramipril, irbesartan, amlodipine, carvedilol, metformin, strontium ranelate, rabeprazole, risperidone, donepezil, raloxifene, tramadol, meloxicam, celecoxib, venlafaxine, citalopram, mirtazapine, escitalopram, sertraline – <i>Events:</i> 28 events, more details in [71] <p><i>Main performance criteria:</i> Se, Sp, PPV, NPV</p>	SSA: Se = 61%; Sp = 93%; PPV = 77%; NPV = 87%

SR: spontaneous report; HDPS: high dimensional propensity score; SCCS: self-controlled case series; CC: case crossover; SCC: self-controlled cohort; maxSPRT: maximized sequential probability ratio test; CSSP: conditional sequential sampling procedure; TPD: temporal pattern discovery; LGPS: longitudinal gamma Poisson shrinker; LEOPARD: longitudinal evaluation of observational profiles of adverse events related to drugs; TAR: temporal association rules; MUTARA: mining the unexpected TARs given the antecedent; HUNT: highlighting TARs negating TARs; SML: supervised machine learning; SSA: sequence symmetry analysis; PPV: positive predictive value; NPV: negative predictive value; AUC: area under the receiver operating characteristics curve; ADR: adverse drug reactions; Se: sensitivity; Sp: specificity; ICD: international classification of diseases.

the self-controlled designs, the SSA, and the SML approach seem the most interesting candidates for safety signal detection in health-care databases. The two former approaches provided a slightly lower performance for safety signal detection than the latter approach, but it is noteworthy that they were assessed much more intensively. The self-controlled designs [31–36,56–59] and the SSA [37,61,62,71,76–79] have been assessed using different health-care databases, both empirically and through simulation studies. By contrast, the two versions of the SML approach have been studied once and using one health-care database [45,46]. The evidence supporting the performance of the self-controlled designs and SSA appears therefore much more reliable.

Nevertheless, stating about the best method to use between these three should look beyond the only statistical characteristics by considering pragmatic aspects. In the perspective of routine signal detection, the choice of what methods to use should be based on the three following questions:

- Does the method achieve the goals expected for signal detection?
- Is the method understandable (e.g. principle, meaning of the results) by stakeholders?
- Does the method provide a guidance to help stakeholders to handle the massive amount of detected signals expected in certain situations?

5.1. Does the method achieve the goals expected for signal detection?

By definition, signal detection consists in screening all drug-event associations recorded in a database, without a priori assumptions regarding their potential relationship, in order to highlight those that could be considered as true associations, i.e. signals with a special interest to those that are unexpected ADRs (in nature or in frequency) [81]. In contrast with the SSA and the SML approaches, the self-controlled designs, and more generally the traditional pharmacoepidemiological designs, were originally developed to assess specific drug-event associations, with settings (e.g. time-at-risk period) tuned to the knowledge about this association. An application to a large-scale screening considering all possible types of drug-event associations will

probably not appear soon, especially as this would imply to override some assumptions required when applying the self-controlled designs such as the event needs to be recurrent, and its occurrences must be independent [82].

5.2. Is the method understandable by stakeholders?

All signal detection methods were designed to be highly sensitive, in the sense that the thresholds for detection are set low in order to avoid missing a real drug safety concern. Once a signal is detected, its assessment is crucial and is generally made by persons qualified in pharmacovigilance, pharmacoepidemiology, by clinicians, or decision maker with a medical background, but seldom by statisticians. Therefore, to be able to adequately decide which signals should be further investigated, a prerequisite is that these stakeholders understand at least the basic principles of the method which generated them. In that perspective, the SML approach, which is based upon complex statistical concepts, has good chance to be viewed as a 'black box.' The results provided could thus have a lower acceptability compared to self-controlled designs, which are well-known methods, and to the SSA, the principle of which is straightforward to understand.

5.3. Does the method provide a guidance to help stakeholders handle the massive amount of signals expected in certain situations?

Although signal detection is not yet performed routinely in health-care databases, in that context, thousands of signals would be reasonably expected, which would inevitably congest and slow down the signal strengthening and decision-making processes. For instance, Reps *et al.* detected up to 67,000 signals with MUTARA [70]. To handle this tremendous amount of signals, stakeholders should be guided in determining what signals are worthy of further investigation. From this point of view, one may assume that methods providing a risk estimate for each drug-event association are more appealing, than those providing probabilities of association. Indeed, only the former quantify the potential strength of the association, which is probably the most relevant information to consider for decision-making. Considering this criterion, the SML

approach could be considered less appropriate for routine safety signal detection in health-care databases.

5.4. Conclusion

When combining the three aspects discussed above, the SSA appears to be the most suitable method for signal detection in health-care databases. It is dedicated to longitudinal data and developed for large-scale and standardized applications. Its simple and understandable concept allows non-statisticians to grasp it rapidly, and to be confident in the results provided. It includes self-controlled analyses, which are likely to reduce the detection of false positives. It is of course plagued by some limitations. The SSA is per se unable to prevent the detection of spurious signals related to protopathic or indication biases, a limitation inherent to signal detection in longitudinal data, but solutions to discard these signals have been developed. For instance, Avillach *et al.* implemented a method able to retrieve well-established drug-event associations from the Medline database [67]. The comparison of these associations with the signals detected could rule out spurious signals related to protopathic or indication biases. As many drugs tend to be used in specific sequences when diseases progress, the SSA could lead to flag up some fallacies associations. The adjustment of the sequence symmetry estimations for trend prescribing over time is likely to minimize this issue, even if not fully canceling its effects. Likewise, medical events that affect the likelihood of future prescriptions of the drug can bias the sequence symmetry estimation toward the identification of a risk. This issue should not be perceived as a matter of concern, as it will not impact sensitivity. Anyhow, a careful clinical reviewing of the detected signal is absolutely needed to take full advantage of automated safety signal detection. The inability of the SSA to capture ‘death’ event should not be seen as a major limitation, as it can assess other risks (hospitalization or other hard outcomes [79]), which provide valuable information from the public health perspective.

5.5. Perspectives

Signal detection from health-care database is already possible but is not yet used for routine surveillance of drug safety. In this perspective, a further challenge will be represented by methods helping for signal management, as developing complementary methods for signal prioritization. These could be based on criteria usually considered as relevant during the decision-making process. Health-care databases could provide automatically the potential public health impact of signals through information such as incidence of drug use or the incidence of events, which cannot be provided by SR, and that could help in selecting signals justifying a more thorough evaluation. Such a prioritization method is thus crucial to imagine routine safety signal detection activities.

Funding

This paper was not funded.

Declaration of interest

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

References

Papers of special note have been highlighted as either of interest (*) or of considerable interest (**) to readers.

- Moore N. The past, present and perhaps future of pharmacovigilance: homage to Folke Sjoqvist. *Eur J Clin Pharmacol*. 2013;69(S1):33–41.
- A well-documented review presenting the history and future of safety signal detection.**
- Hauben M, Norén GN. A decade of data mining and still counting. *Drug Saf*. 2010;33(7):527–534.
- Almenoff JS, Pattishall EN, Gibbs TG, et al. Novel statistical tools for monitoring the safety of marketed drugs. *Clin Pharmacol Ther*. 2007;82(2):157–166.
- Bégaud B, Moride Y, Tubert-Bitter P, et al. False-positives in spontaneous reporting: should we worry about them? *Br J Clin Pharmacol*. 1994;38(5):401–404.
- Olivier P, Montastruc J-L. The nature of the scientific evidence leading to drug withdrawals for pharmacovigilance reasons in France. *Pharmacoepidemiol Drug Saf*. 2006;15(11):808–812.
- Paludetto M-N, Olivier-Abbal P, Montastruc J-L. Is spontaneous reporting always the most important information supporting drug withdrawals for pharmacovigilance reasons in France? *Pharmacoepidemiol Drug Saf*. 2012;21(12):1289–1294.
- Bégaud B, Martin K, Haramburu F, et al. Rates of spontaneous reporting of adverse drug reactions in France. *Jama*. 2002;288(13):1588.
- Moride Y, Haramburu F, Requejo AA, et al. Under-reporting of adverse drug reactions in general practice. *Br J Clin Pharmacol*. 1997;43(2):177–181.
- Pariante A, Gregoire F, Fourrier-Reglat A, et al. Impact of safety alerts on measures of disproportionality in spontaneous reporting databases the notoriety bias. *Drug Saf*. 2007;30(10):891–898.
- Hartnell NR, Wilson JP. Replication of the Weber effect using post-marketing adverse event reports voluntarily submitted to the United States Food and Drug Administration. *Pharmacother J Hum Pharmacol Drug Ther*. 2004;24(6):743–749.
- Johnstone J, Nerenberg K, Loeb M. Meta-analysis: proton pump inhibitor use and the risk of community-acquired pneumonia. *Aliment Pharmacol Ther*. 2010 Mar 4;31(11):1165–1177.
- Lambert AA, Lam JO, Paik JJ, et al. Risk of community-acquired pneumonia with outpatient proton-pump inhibitor therapy: a systematic review and meta-analysis. *Plos One*. 2015;10(6):e0128004.
- Avorn J. Evaluating drug effects in the post-vioxx world: there must be a better way. *Circulation*. 2006;113(18):2173–2176.
- Topol EJ. Failing the public health—rofecoxib, Merck, and the FDA. *N Engl J Med*. 2004;351(17):1707–1709.
- Fontanarosa PB. Postmarketing surveillance—lack of vigilance, lack of trust. *Jama*. 2004;292(21):2647.
- Arnaud M, Salvo F, Ahmed I, et al. A method for the minimization of competition bias in signal detection from spontaneous reporting databases. *Drug Saf*. 2016;39(3):251–260.
- Ahmed I, Pariante A, Tubert-Bitter P. Class-imbalanced subsampling lasso algorithm for discovering adverse drug reactions. *Stat Methods Med Res*. 2016. 25 april 2016. published online. DOI:10.1177/0962280216643116.
- Juhlin K, Ye X, Star K, et al. Outlier removal to uncover patterns in adverse drug reaction surveillance - a simple unmasking strategy. *Pharmacoepidemiol Drug Saf*. 2013;22(10):1119–1129.
- Maignen F, Hauben M, Hung E, et al. A conceptual approach to the masking effect of measures of disproportionality. *Pharmacoepidemiol Drug Saf*. 2014;23(2):208–217.

20. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol*. 2005;58(4):323–337.
21. Norén GN, Hopstadius J, Bate A, et al. Safety surveillance of longitudinal databases: methodological considerations. *Pharmacoepidemiol Drug Saf*. 2011;20(7):714–717.
22. Trifirò G, Fourrier-Reglat A, Sturkenboom MC, et al. The EU-ADR project: preliminary results and perspective. *Stud Health Technol Inf*. 2009;148:43–49.
23. Wisniewski AFZ, Bate A, Bousquet C, et al. Good signal detection practices: evidence from IMI PROTECT. *Drug Saf*. 2016;39(6):469–490.
24. 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;21:1–8.
- **A clear presentation of the objectives and the challenges existing when attempting to set up a post-market drug safety framework using the information available on observational health-care databases.**
25. U.S. Food and Drug Administration. Mini-Sentinel. 2014. Available at: <http://www.mini-sentinel.org> [Last accessed 13 December 2016]
26. Stang PE, Ryan PB, Racoosin JA, et al. Advancing the science for active surveillance: rationale and design for the observational medical outcomes partnership. *Ann Intern Med*. 2010;153(9):600–606.
27. Observational medical outcomes partnership. Available at: <http://omop.org> [Last accessed 13 December 2016]
28. AsPEN collaborators, Andersen M, Bergman U, Choi N-K, et al. The asian pharmacoepidemiology network (AsPEN): promoting multi-national collaboration for pharmacoepidemiologic research in Asia. *Pharmacoepidemiol Drug Saf*. 2013;22(7):700–704.
29. Asian pharmacoepidemiological network. Available at: <http://aspenet.asia/index.html> [Last accessed 13 December 2016]
30. Zorych I, Madigan D, Ryan P, et al. Disproportionality methods for pharmacovigilance in longitudinal observational databases. *Stat Methods Med Res*. 2013;22(1):39–56.
- **Study presenting the different ways to apply the disproportionality methods on observational health-care databases.**
31. Schuemie MJ, Coloma PM, Straatman H, et al. Using electronic health care records for drug safety signal detection: a comparative evaluation of statistical methods. *Med Care*. 2012;50(10):890–897.
32. Ryan PB, Schuemie MJ, Gruber S, et al. Empirical performance of a new user cohort method: lessons for developing a risk identification and analysis system. *Drug Saf*. 2013;36(S1):59–72.
33. Madigan D, Schuemie MJ, Ryan PB. Empirical performance of the case-control method: lessons for developing a risk identification and analysis system. *Drug Saf*. 2013;36(S1):73–82.
34. Suchard MA, Zorych I, Simpson SE, et al. Empirical performance of the self-controlled case series design: lessons for developing a risk identification and analysis system. *Drug Saf*. 2013;36(S1):83–93.
35. Murphy SN, Castro V, Colecchi J, et al. Partners healthcare OMOP study report. 2011. Available at: <http://omop.org>. [Last accessed 13 December 2016].
- **Study presenting the first OMOP experiment that implemented, tested, and compared numerous methods: disproportionality analysis, traditional, pharmacoepidemiological designs, TPD, and sequential statistical testing.**
36. Ryan PB, Schuemie MJ, Madigan D. Empirical performance of a self-controlled cohort method: lessons for developing a risk identification and analysis system. *Drug Saf*. 2013;36(S1):95–106.
37. Hallas J. Evidence of depression provoked by cardiovascular medication: a prescription sequence symmetry analysis. *Epidemiology*. 1996;7(5):478–484.
- **Study conceptualizing the SSA method for signal detection.**
38. Brown JS, Kulldorff M, Chan KA, et al. Early detection of adverse drug events within population-based health networks: application of sequential testing methods. *Pharmacoepidemiol Drug Saf*. 2007;16(12):1275–1284.
39. Kulldorff M, Davis RL, Kolczak M, et al. A maximized sequential probability ratio test for drug and vaccine safety surveillance. *Seq Anal*. 2011;30(1):58–78.
40. Li L. A conditional sequential sampling procedure for drug safety surveillance. *Stat Med*. 2009;28(25):3124–3138.
41. Jin H, Chen J, Kelman C, et al. Mining unexpected associations for signalling potential adverse drug reactions from administrative health databases. *Adv Knowl Discov Data Min*. 2006;3918:867–876. Ng W-K, Kitsuregawa M, Li J, Chang K, editors.
42. Jin H, Chen J, He H, et al. Signaling potential adverse drug reactions from administrative health databases. *IEEE Trans Knowl Data Eng*. 2010;22(6):839–853.
43. Norén GN, Hopstadius J, Bate A, et al. Temporal pattern discovery in longitudinal electronic patient records. *Data Min Knowl Discov*. 2010;20(3):361–387.
44. Ji Y, Massanari RM, Ager J, et al. A fuzzy logic-based computational recognition-primed decision model. *Inf Sci*. 2007;177(20):4338–4353.
- **Study presenting an unusual approach based on a TAR algorithm combined with fuzzy logic.**
45. Reips JM, Garibaldi JM, Aickelin U, et al. Signalling paediatric side effects using an ensemble of simple study designs. *Drug Saf*. 2014;37(3):163–170.
46. Reips JM, Garibaldi JM, Aickelin U, et al. A supervised adverse drug reaction signalling framework imitating Bradford Hill's causality considerations. *J Biomed Inform*. 2015;56:356–368.
- **Study presenting an interesting supervised machine-learning approach incorporating Bradford-Hill's criteria.**
47. Kulldorff M, Dashevsky I, Avery TR, et al. Drug safety data mining with a tree-based scan statistic. *Pharmacoepidemiol Drug Saf*. 2013;22(5):517–523.
- **Study presenting an original approach for discovering ADRs based on hierarchical trees: the tree-based scan statistic method.**
48. Dumouchel W. Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. *Am Stat*. 1999 Aug;53(3):177–190.
49. 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(4):315–321.
50. Moore N, Kreft-Jais C, Haramburu F, et al. Reports of hypoglycaemia associated with the use of ACE inhibitors and other drugs: a case/non-case study in the French pharmacovigilance system database. *Br J Clin Pharmacol*. 1997 Nov;44(5):513–518.
51. Evans SJW, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiol Drug Saf*. 2001;10(6):483–486.
52. Curtis JR, Cheng H, Delzell E, et al. Adaptation of bayesian data mining algorithms to longitudinal claims data: cobox safety as an example. *Med Care*. 2008;46(9):969–975.
53. Choi N-K, Chang Y, Choi YK, et al. Signal detection of rosvastatin compared to other statins: data-mining study using national health insurance claims database. *Pharmacoepidemiol Drug Saf*. 2010;19(3):238–246.
54. Choi N-K, Chang Y, Kim J-Y, et al. Comparison and validation of data-mining indices for signal detection: using the Korean national health insurance claims database. *Pharmacoepidemiol Drug Saf*. 2011;20(12):1278–1286.
55. Kim J, Kim M, Ha J-H, et al. Signal detection of methylphenidate by comparing a spontaneous reporting database with a claims database. *Regul Toxicol Pharmacol*. 2011;61(2):154–160.
56. Schuemie MJ. Methods for drug safety signal detection in longitudinal observational databases: LGPS and LEOPARD. *Pharmacoepidemiol Drug Saf*. 2011;20(3):292–299.
57. Ryan PB, Stang PE, Overhage JM, et al. A comparison of the empirical performance of methods for a risk identification system. *Drug Saf*. 2013;36(S1):143–58.
- **Study presenting the second OMOP experiment, which improved the methodology for implementing and assessing the methods.**

58. Ryan PB, Schuemie MJ. Evaluating performance of risk identification methods through a large-scale simulation of observational data. *Drug Saf.* 2013;36(S1):171–180.
59. Schuemie MJ, Gini R, Coloma PM, et al. Replication of the OMOP experiment in Europe: evaluating methods for risk identification in electronic health record databases. *Drug Saf.* 2013;36(S1):159–169.
- **Study presenting the EU-ADR experiment that implemented and assessed in European data the methods based on disproportionality analysis, traditional pharmacoepidemiological designs.**
60. Tibshirani R. Regression shrinkage and selection via the Lasso. *J R Stat Soc Ser B Methodol.* 1996;58(1):267–288.
61. Petri H, De Vet HCW, Naus J, et al. Prescription sequence analysis: a new and fast method for assessing certain adverse reactions of prescription drugs in large populations. *Stat Med.* 1988;7(11):1171–1175.
- **The first study to have described the SSA, a smart method for safety signal detection in observational health-care databases.**
62. Tsiropoulos I, Andersen M, Hallas J. Adverse events with use of antiepileptic drugs: a prescription and event symmetry analysis. *Pharmacoepidemiol Drug Saf.* 2009;18(6):483–491.
63. Cook AJ, Tiwari RC, Wellman RD, et al. Statistical approaches to group sequential monitoring of postmarket safety surveillance data: current state of the art for use in the mini-sentinel pilot. *Pharmacoepidemiol Drug Saf.* 2012;21:72–81.
64. Ji Y, Ying H, Dews P, et al. A potential causal association mining algorithm for screening adverse drug reactions in postmarketing surveillance. *IEEE Trans Inf Technol Biomed.* 2011;15(3):428–437.
65. Ji Y, Ying H, Tran J, et al. A method for mining infrequent causal associations and its application in finding adverse drug reaction signal pairs. *IEEE Trans Knowl Data Eng.* 2013;25(4):721–733.
66. Kulldorff M, Fang Z, Walsh SJ. A tree-based scan statistic for database disease surveillance. *Biometrics.* 2003;59(2):323–331.
67. Avillach P, Dufour J-C, Diallo G, et al. Design and validation of an automated method to detect known adverse drug reactions in MEDLINE: a contribution from the EU-ADR project. *J Am Med Inform Assoc.* 2013;20(3):446–452.
68. Coloma PM, Avillach P, Salvo F, et al. A reference standard for evaluation of methods for drug safety signal detection using electronic healthcare record databases. *Drug Saf.* 2013;36(1):13–23.
69. Ryan PB, Schuemie MJ, Welebob E, et al. Defining a reference set to support methodological research in drug safety. *Drug Saf.* 2013;36(S1):33–47.
70. Reps JM, Garibaldi JM, Aickelin U, et al. Comparison of algorithms that detect drug side effects using electronic healthcare databases. *Soft Comput.* 2013;17(12):2381–2397.
- **Study presenting and comparing the performance for signal detection of the methods based on TAR algorithms.**
71. 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(5):496–502.
- **Study demonstrating that the SSA performed pretty well for safety signal detection.**
72. Gruber S, Chakravarty A, Heckbert SR, et al. Design and analysis choices for safety surveillance evaluations need to be tuned to the specifics of the hypothesized drug-outcome association. *Pharmacoepidemiol Drug Saf.* 2016;25:973–981.
73. Gagne JJ, Schneeweiss S. Comment on “Empirical assessment of methods for risk identification in healthcare data: results from the experiments of the observational medical outcomes partnership.”. *Stat Med.* 2013;32(6):1073–1074.
74. Hennessy S, Leonard CE. Comment on: “Desideratum for evidence-based epidemiology.”. *Drug Saf.* 2015;38(1):101–103.
75. Hansen RA, Gray MD, Fox BI, et al. How well do various health outcome definitions identify appropriate cases in observational studies? *Drug Saf.* 2013;36(S1):27–32.
76. Pratt NL, Ilomäki J, Raymond C, et al. The performance of sequence symmetry analysis as a tool for post-market surveillance of newly marketed medicines: a simulation study. *BMC Med Res Methodol.* 2014;14(1):1.
77. Pratt N, Andersen M, Bergman U, et al. Multi-country rapid adverse drug event assessment: the Asian pharmacoepidemiology network (AsPEN) antipsychotic and acute hyperglycaemia study. *Pharmacoepidemiol Drug Saf.* 2013;22:915–924.
78. Pratt N, Chan EW, Choi N-K, et al. Prescription sequence symmetry analysis: assessing risk, temporality, and consistency for adverse drug reactions across datasets in five countries. *Pharmacoepidemiol Drug Saf.* 2015;24(8):858–864.
79. Wahab IA, Pratt NL, Ellett LK, et al. Sequence symmetry analysis as a signal detection tool for potential heart failure adverse events in an administrative claims database. *Drug Saf.* 2016;39(4):347–354.
- **Study demonstrating that the SSA method is able to identify unknown ADRs, which are supported by biological plausibility and case reports.**
80. Blagus R, Lusa L. Class prediction for high-dimensional class-imbalanced data. *BMC Bioinformatics.* 2010;11(1):523.
81. The Uppsala Monitoring Centre. Practical Pharmacovigilance. Signal definition. 2013. Available at: <http://www.who-umc.org> [Last accessed 13 December 2016]
82. Whitaker HJ, Paddy Farrington C, Spiessens B, et al. Tutorial in biostatistics: the self-controlled case series method. *Stat Med.* 2006;25(10):1768–1797.