Drug Saf (2013) 36:13–23
DOI 10.1007/s40264-012-0002-x

SHORT COMMUNICATION

A Reference Standard for Evaluation of Methods for Drug Safety Signal Detection Using Electronic Healthcare Record Databases

Preciosa M. Coloma · Paul Avillach · Francesco Salvo · Martijn J. Schuemie · Carmen Ferrajolo · Antoine Pariente · Annie Fourrier-Réglat · Mariam Molokhia · Vaishali Patadia · Johan van der Lei · Miriam Sturkenboom · Gianluca Trifirò

Published online: 23 November 2012

Abstract

Background The growing interest in using electronic healthcare record (EHR) databases for drug safety surveillance has spurred development of new methodologies for signal detection. Although several drugs have been withdrawn postmarketing by regulatory authorities after scientific evaluation of harms and benefits, there is no definitive list of confirmed signals (i.e. list of all known adverse reactions and which drugs can cause them). As there is no true gold standard, prospective evaluation of signal detection methods remains a challenge.

Objective Within the context of methods development and evaluation in the EU-ADR Project (Exploring and Understanding Adverse Drug Reactions by integrative mining of clinical records and biomedical knowledge), we propose a surrogate reference standard of drug-adverse event associations based on existing scientific literature and expert opinion.

Methods The reference standard was constructed for ten top-ranked events judged as important in pharmacovigilance. A stepwise approach was employed to identify which, among a list of drug-event associations, are well recognized (known positive associations) or highly unlikely ('negative controls') based on MEDLINE-indexed publications, drug product labels, spontaneous reports made to the WHO's pharmacovigilance database, and expert opinion. Only drugs with adequate exposure in the

On behalf of the EU-ADR Consortium.

Electronic supplementary material The online version of this article (doi:10.1007/s40264-012-0002-x) contains supplementary material, which is available to authorized users.

P. M. Coloma (✉) · M. J. Schuemie · C. Ferrajolo · V. Patadia · J. van der Lei · M. Sturkenboom · G. Trifirò
Department of Medical Informatics, Erasmus Medical Center, Postbus 2040, 3000 CA, Rotterdam, The Netherlands
e-mail: p.coloma@erasmusmc.nl

P. Avillach
LESIM, ISPED, University Bordeaux 2, Bordeaux, France

P. Avillach
LERTIM, EA 3283, Faculté de Médecine, Université Aix Marseille 2, Marseille, France

F. Salvo · C. Ferrajolo
Campania Regional Center for Pharmacovigilance and Pharmacoepidemiology, Department of Experimental Medicine, Pharmacology Section, Second University of Naples, Naples, Italy

F. Salvo · A. Pariente · A. Fourrier-Réglat
Inserm U657, Pharmacology Department, Bordeaux, France

F. Salvo · A. Pariente · A. Fourrier-Réglat
CHU Bordeaux, Bordeaux, France

A. Pariente · A. Fourrier-Réglat
Department of Pharmacology, University of Bordeaux, Bordeaux, France

M. Molokhia
Primary Care and Population Sciences, Kings College, London, UK

V. Patadia
Astellas Pharmaceuticals, Deerfield, IL, USA

M. Sturkenboom
Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands

G. Trifirò
Department of Clinical and Experimental Medicine and Pharmacology, Section of Pharmacology, University of Messina, Messina, Italy

EU-ADR database network (comprising ≈60 million person-years of healthcare data) to allow detection of an association were considered. Manual verification of positive associations and negative controls was independently performed by two experts proficient in clinical medicine, pharmacoepidemiology and pharmacovigilance. A third expert adjudicated equivocal cases and arbitrated any disagreement between evaluators.

Results Overall, 94 drug-event associations comprised the reference standard, which included 44 positive associations and 50 negative controls for the ten events of interest: bullous eruptions; acute renal failure; anaphylactic shock; acute myocardial infarction; rhabdomyolysis; aplastic anaemia/pancytopenia; neutropenia/agranulocytosis; cardiac valve fibrosis; acute liver injury; and upper gastrointestinal bleeding. For cardiac valve fibrosis, there was no drug with adequate exposure in the database network that satisfied the criteria for a positive association.

Conclusion A strategy for the construction of a reference standard to evaluate signal detection methods that use EHR has been proposed. The resulting reference standard is by no means definitive, however, and should be seen as dynamic. As knowledge on drug safety evolves over time and new issues in drug safety arise, this reference standard can be re-evaluated.

1 Background

The growing interest in the utility of electronic healthcare records (EHRs) for drug safety surveillance has spurred the development of new methodologies for quantitative and automated signal detection. Timely detection of safety signals remains a challenge because no single technique ensures identification of *all* drug-related adverse events, whether signal detection is done using spontaneous reports [1] or using healthcare records [2]. Generation of false alarms similarly constitutes a public health hazard, not only overwhelming regulatory agencies and diverting already scarce resources, but also triggering unwarranted warnings or even drug market withdrawals [3]. Thus, proper evaluation of signal detection methodologies calls for the creation of a reference standard, the purpose of which is to better define the predictive value of these new techniques, as well as their added value to the current pharmacovigilance armamentarium.

2 Signal Detection in the Context of Pharmacovigilance

The WHO has defined 'signal' as "reported information on a possible causal relationship between an adverse event and a drug, the relationship being unknown or incompletely documented" [4]. An updated and more encompassing definition has been proposed recently based on a systematic review of how the term is being applied in current pharmacovigilance: a signal represents information that arises from one or multiple sources which suggests a new potentially causal association, or a new aspect of a known association, between an intervention and an event or set of related events, either adverse or beneficial, and is judged to be of sufficient likelihood to justify verificatory and remedial actions [5]. Although a 'gold standard' of confirmed signals, i.e. *causal* drug-adverse event associations, does not exist, a reference standard of recognized associations based on existing published scientific literature, regulatory actions (e.g. labelling changes or withdrawal of marketing authorization), as well as expert opinion, may serve as a suitable surrogate. In this study we describe a reference standard that was put together in the context of methods development within the EU-ADR Project (Exploring and Understanding Adverse Drug Reactions by integrative mining of clinical records and biomedical knowledge'; http://www.euadr-project.org), which aims to exploit information from various EHR databases in Europe to produce a computerized integrated system for the early detection of drug safety signals [6]. This reference standard was developed for the primary purpose of evaluating performance of methods for signal detection using EHR.

3 Methodology

The EU-ADR network currently comprises anonymous healthcare data from eight established European databases located in four countries (Denmark, Italy, The Netherlands and the UK) [7]. Clinical and drug dispensing/prescription data used for this paper represent data from 19,647,445 individuals with 59,929,690 person-years (PYs) of follow-up.

4 Adverse Events

In the EU-ADR Project we have chosen an event-based approach to active drug safety surveillance, focusing on events considered to be important from a pharmacovigilance and public health perspective. For the construction of this reference standard, we considered the following top ten events which have been selected from a list of 23 events ranked on the basis of importance in pharmacovigilance using predefined criteria: (i) bullous eruptions; (ii) acute renal failure; (iii) anaphylactic shock; (iv) acute myocardial infarction; (v) rhabdomyolysis; (vi) aplastic anaemia/pancytopenia; (vii) neutropenia/agranulocytosis; (viii) cardiac valve fibrosis; (ix) acute liver injury; and (x) upper gastrointestinal bleeding [8].

5 Drug Selection

The procedure employed in the construction of the reference standard is outlined in Fig. 1. It was first necessary to ensure that the drug-event associations to be included in the reference standard are identifiable in clinical practice and could be investigated in the EU-ADR network. That is, there should be adequate exposure to the drugs to permit detection of an association with the adverse event of interest, if present. In another publication we described the sample size calculations used to derive the total amount of PYs of drug exposure required to detect an association between a drug and a particular event over varying magnitudes of relative risk (RR), using one-sided significance level $\alpha = 0.05$ and power of 80 %, given pooled population-based incidence rates (IR) estimated directly within the EU-ADR network [2]. For this reference standard we employed in the calculations an RR of at least two for all events except for rhabdomyolysis, bullous eruptions and anaphylactic shock, where we used an RR of at least 4. The latter was done to account for the very low background IR of these events in the population (2.5/100,000 PYs for rhabdomyolysis, 5.7/100,000 PYs for anaphylactic shock and 5.9/100,000 PYs for bullous eruptions). A series of steps was subsequently employed to select the positive drug-event associations and 'negative controls' among those potentially eligible (i.e. drugs with an adequate amount of exposure to detect the association of interest) [see Fig. 1].

6 Information Retrieval from Published Literature

To streamline the scientific literature search, we utilized a tool developed within the EU-ADR Project that automatically searches MEDLINE-indexed publications concerning adverse drug reactions (ADRs) [9]. A subset of MEDLINE was downloaded (via PubMed) and imported into a database including all the citations from December 1952 to February 2010 with the 'adverse effects' Medical Subject Heading (MeSH) subheading. For each citation, the PubMed identification (PMID), MeSH descriptors, major/minor subheadings, substances, date of creation of the citation, as well as publication type, were obtained. Co-occurrence of the drug (from 'substances' OR 'MeSH heading' fields) and the event (under the subheading 'adverse effects') in a citation were noted. Drug codes in the WHO Anatomical Therapeutic Chemical (ATC) classification were first mapped to MeSH headings or supplementary concept records using standardized concept unique identifiers from the Unified Medical Language System (UMLS) [10]. Drugs from the 'substances' field were taken into account only if their pharmacological action was qualified by the subheading 'adverse effects'. Taking the pharmacological action as an additional element for consideration was an attempt to establish a link between the adverse event of interest and the drug in the context of drug safety and not just a co-occurrence in a MEDLINE citation. This becomes particularly important when more than one drug is mentioned in the citation [10].

7 Selection of Known Positive Drug-Event Associations

The drug-event associations were ranked according to the number of MEDLINE citations with co-occurrence of the drug and the adverse event of interest. For the pool of positive drug-event associations, we considered those with the highest number of citations. This meant that more published evidence was available on these associations. Citations may refer to case reports, observational studies, clinical trials, reviews or meta-analyses. The type of publication was taken into account in the evaluation of the evidence regarding each drug-adverse event association, as subsequently described. Supplementary information was obtained from the Summary of Product Characteristics or product labels [11–16]. The aim was to select five drugs that are positively associated with each event of interest. Whenever possible, drugs belonging to different classes were included in the pool. However, the need for minimizing ambiguity (i.e. by selecting strong and well substantiated drug-adverse event associations) took precedence over the need for diversity in terms of drug class. Except for fixed-dose combinations, drug preparations with more than one active substance were excluded from the pool.

8 Selection of 'Negative Controls'

A drug-event association was considered for the pool of 'negative controls' if there were no MEDLINE citations with co-occurrence of the drug and the event of interest and if there was no explicit mention of such adverse event in the drug product label. The pool of 'negative controls' was further evaluated using the WHO spontaneous reporting database (VigiBase™) to exclude associations flagged as a potential signal using standard data mining methodology. The list of potential signals from VigiBase™ (including data up to the fourth quarter of 2010) was generated using the Oracle Health Sciences Empirica™ Signal tool (courtesy of Astellas Pharmaceuticals, Deerfield, IL, USA). Bayesian disproportionality analysis was performed using preferred terms mapped to the events of interest [17]. A value greater than 2 for the lower bound of the 90 % confidence interval (CI) of the Empirical Bayes Geometric Mean (EB_{05}) and the presence of at least one report were used as the criteria for flagging a signal [18]. The aim was to likewise obtain five drug-event associations as 'negative controls' per event of interest.

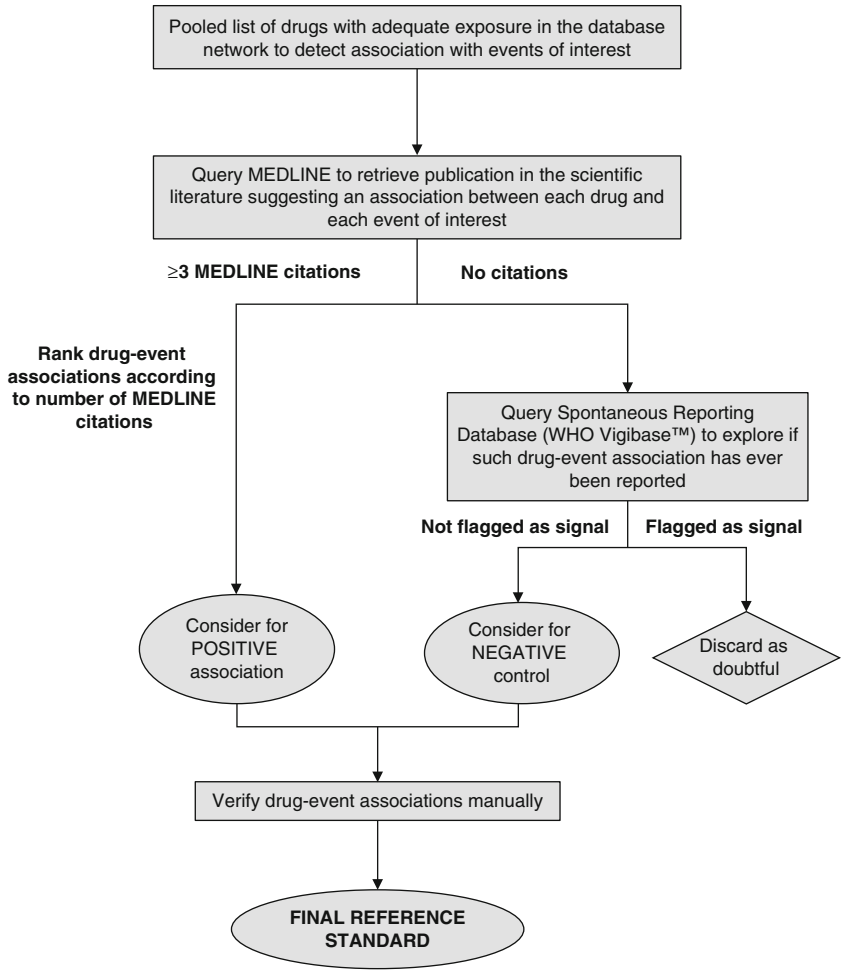

Fig. 1 Flowchart showing the process of the construction of the reference standard

9 Evaluation of the Evidence from Literature

Table 1 shows the scheme that was used as a guide to evaluate evidence from the literature. Manual verification of the positive associations and 'negative controls' was conducted by two physicians with proficiency in clinical medicine, epidemiology and pharmacovigilance. A third expert arbitrated any disagreement between evaluators. The following indices of agreement between evaluators were assessed: (i) proportion of overall agreement; (ii) proportion of specific agreement; and (iii) kappa statistic, κ, for chance-corrected agreement. The earliest date of MEDLINE citation was also noted for each drug-event association.

10 Results

The amount of drug exposure required to detect a potential signal in the EU-ADR database network for each of the events of interest is shown in Table 2. Overall, there were 893 drugs (i.e. unique ATC codes, 5th level chemical substance) with enough exposure to permit detection of an association with at least one of the ten events of interest. Out of the 893 drugs, the following are the number (i.e. count) of drugs for which there were at least three MEDLINE citations with co-occurrence of the drug and the corresponding event: acute liver injury, 21; acute myocardial infarction, 52; acute renal failure, 51; anaphylactic shock, 26; bullous eruptions, 47; cardiac valve fibrosis, 2; neutropenia/agranulocytosis, 30; aplastic anaemia/pancytopenia, 21; rhabdomyolysis, 8; upper gastrointestinal bleeding, 54. Close to 1,200 abstracts and, when necessary, the full-text journal articles pertaining to all ten events were reviewed to arrive at a shortlist of potential positive associations and 'negative controls'. Specific citations in drug product labels concerning 'undesirable effects', 'warnings', and 'adverse reactions' were used to further restrict the shortlist of associations. Table 3 shows how the manual evaluation of a positive association for acute liver injury with valproic acid and for upper gastrointestinal bleeding with indometacin were done. The complete evaluation for all the positive drug-adverse event associations of interest can be found in Appendix 1 (Online Resource 1).

Table 1 Levels of evidence used in the evaluation of drug safety information from the literature

Level of evidence	Description
I	Evidence from at least one (properly designed) randomized controlled trial or meta-analysis
II	Evidence from at least one observational study (e.g. cohort, case-control, case-crossover, self-controlled case series) OR from at least three published case reports from different sources and concerning different patients
III	Evidence from not more than two published case reports OR from unpublished reports in pharmacovigilance databases and no further substantiation in the literature
IV	Included in drug label (SPC) but no case reports or published studies
V	No evidence from published literature or from WHO spontaneous reporting database and not mentioned in the SPC

Recommendations: Levels I and II → positive association; Levels III and IV → cannot be determined → disregard as doubtful; Level V → 'negative control'

SPC summary of product characteristics

Table 2 Amount of drug exposure required to detect a potential signal in the EU-ADR database network for the events of interest

Event	Required exposure (person-years)	No. of drugs with sufficient exposure to detect association and with ≥3 MEDLINE citations
Acute liver injury	32,769	21
Acute myocardial infarction	4,706	52
Acute renal failure	30,397	51
Anaphylactic shock	21,733	26
Bullous eruptions	20,823	47
Cardiac valve fibrosis	13,604	2
Neutropenia/agranulocytosis	82,697	30
Aplastic anaemia/ pancytopenia	77,192	21
Rhabdomyolysis	49,593	8
Upper gastrointestinal bleeding	12,028	54

The final reference standard consisted of 94 drug-event associations, which included 44 positive associations and 50 'negative controls' related to the ten events of interest. Table 4 lists the positive associations, including the corresponding level of evidence. The majority of positive associations were based on Level II evidence. The associations for which there was Level I evidence included that of NSAIDs and of heparin with upper gastrointestinal bleeding, the association of the statins with rhabdomyolysis, and the association of coxibs and rosiglitazone with acute myocardial infarction. All 'negative controls', by definition, have Level V evidence and are listed in Table 5. Both positive and 'negative control' associations comprised 68 unique drugs (i.e. ATC 5th level) belonging to 42 different pharmacological subgroups (i.e. ATC 3rd level). Only four drugs having sufficient exposure in the database network satisfied the criteria for a positive association with rhabdomyolysis, all of them being HMG-CoA reductase inhibitors (statins). Fibrates, as a class (ATC 4th level, chemical subgroup), comprised enough exposure to detect an association with rhabdomyolysis, but the individual drugs did not. For cardiac valve fibrosis, no drug with adequate exposure met the criteria for a positive association after review of the literature.

11 Inter-Evaluator Agreement

The indices for agreement were computed across all drug-event pairs evaluated (179 drug-event pairs), including those that eventually did not get included in the final reference standard. The proportion of overall agreement (the proportion of cases for which both evaluators agreed across all evaluation categories) was 0.93 (95 % CI 0.89, 0.97). The proportions of specific agreement were as follows: (i) 'positive' agreement 0.96 (95 % CI 0.93, 0.98); and (ii) 'negative' agreement 0.90 (95 % CI 0.89, 0.90). There were three instances where one evaluator considered a drug-event association 'undetermined' while the other considered it a positive association (paracetamol [acetaminophen]-anaphylactic shock, bromocriptine-acute myocardial infarction and aspirin [acetylsalicylic acid]-bullous eruptions). Of these three instances only one was eventually included in the reference standard after arbitration (paracetamol-anaphylactic shock). There was a single case where one evaluator marked the association 'undetermined' while the other marked it as 'negative control' (prednisone-neutropenia/ agranulocytosis). Arbitration was done by a third expert. There was no disagreement between evaluators in the final list of 'negative control' associations. The chance-corrected agreement kappa coefficient, κ, was 0.83 (unweighted, 95 % CI 0.74, 0.92).

Table 3 Example summary of manual evaluation of positive drug-event associations for valproic acid and indometacin

ATC code	Drug name	Event type	No. of MEDLINE notices	Labelled as AE in SPC [Yes/No]? (Source and label section)
N03AG01	Valproic acid	Acute liver injury	*Total no. of citations = 31* Review[a] = 1 Clinical trial = 1 (RCT) Epidemiological study = 1 (cohort study) Case reports[b] = 28 (1 citation involving 3 cases, 1 citation involving 5 cases, 1 citation reviewing 31 cases, 2 other citations with literature review)	Yes DailyMed[c] (boxed warning, adverse reactions) eMC[d] (special warnings and precautions for use, undesirable effects) Micromedex[e] (adverse reactions)
M01AB01	Indometacin	Upper gastrointestinal bleeding	*Total no. of citations = 45* Review = 13 Clinical trial = 16 (9 RCTs) Epidemiological study = 5 (1 case control and 4 cohort studies) Case reports = 11	Yes eMC[d] (undesirable effects) Micromedex (adverse reactions)

AE adverse event, *ATC* Anatomical Therapeutic Chemical, *eMC* electronic medicines compendium, *RCT* randomized controlled trial, *SPC* summary of product characteristics

[a] Review refers to both systematic and narrative reviews

[b] Case reports involve only one case pertinent to the drug of interest, unless specified

[c] Website for drugs currently marketed and approved by the US FDA (http://dailymed.nlm.nih.gov/)

[d] For drugs licensed in the UK (http://www.medicines.org.uk)

[e] The Micromedex family of international databases provides full-text drug and substance information (http://www.thomsonhc.com/micromedex2/)

Table 4 Positive drug-event associations

Event	Positive associations		
	ATC code	Name	Level of evidence
Acute liver injury	N03AF01	Carbamazepine	II
	N03AG01	Valproic acid	II
	M01AX17	Nimesulide	II
	J01CR02	Amoxicillin and clavulanic acid	II
	A07EC01	Sulfasalazine	II
Acute myocardial infarction	M01AH02	Rofecoxib	I
	A10BG02	Rosiglitazone	I
	G03AA07	Levonorgestrel and estrogen	II
	N02CC01	Sumatriptan	II
	M01AH03	Valdecoxib	I
Acute renal failure	C09AA01	Captopril	II
	M01AE01	Ibuprofen	II
	N02BE01	Paracetamol (acetaminophen)	II
	J01MA02	Ciprofloxacin	II
	N05AN01	Lithium	II
Anaphylactic shock	B01AC06	Aspirin (acetylsalicylic acid)	II
	N02BE01	Paracetamol (acetaminophen)	II
	J01CA04	Amoxicillin	II
	J01MA02	Ciprofloxacin	II
	M01AB05	Diclofenac	II
Bullous eruptions	N03AF01	Carbamazepine	II
	J01EE01	Sulfamethoxazole and trimethoprim	II
	N03AX09	Lamotrigine	II
	M04AA01	Allopurinol	II
	C03CA01	Furosemide	II
Cardiac valve fibrosis	No drug with sufficient exposure that satisfies criteria for True Positive		
Neutropenia/agranulocytosis	H03BB02	Thiamazole	II
	B01AC05	Ticlopidine	II
	C09AA01	Captopril	II
	N03AF01	Carbamazepine	II
	N03AG01	Valproic acid	II
Aplastic anaemia/pancytopenia	B01AC05	Ticlopidine	II
	N03AF01	Carbamazepine	II
	H03BB02	Thiamazole	II
	M04AA01	Allopurinol	II
	C09AA01	Captopril	II
Rhabdomyolysis	C10AA07	Rosuvastatin	I
	C10AA05	Atorvastatin	I
	C10AA03	Pravastatin	I
	C10AA01	Simvastatin	I
Upper gastrointestinal bleeding	N02BA01/B01AC06	Aspirin	I
	M01AB01	Indometacin	I
	B01AB01	Heparin	I
	H02AB06	Prednisolone	II
	M01AE01	Ibuprofen	I

ATC Anatomical Therapeutic Chemical

Table 5 'Negative control' associations

Event	ATC code	Name
Acute liver injury	R03AC13	Formoterol
	S01ED05	Carteolol
	G04CA03	Terazosin
	N04BA02	Levodopa and decarboxylase inhibitor
	C01DA02	Glyceryl trinitrate
Acute myocardial infarction	A10AD01	Insulin (human)
	B03AA07	Ferrous sulfate
	J01CR02	Amoxicillin and clavulanic acid
	J05AB11	Valaciclovir
	C10AB04	Gemfibrozil
Acute renal failure	R01AD09	Mometasone
	H03AA01	Levothyroxine sodium
	R06AX26	Fexofenadine
	N04BA02	Levodopa and decarboxylase inhibitor
	B03AA07	Ferrous sulfate
Anaphylactic shock	N06AX11	Mirtazapine
	H03AA01	Levothyroxine sodium
	C02AC01	Clonidine
	C02CA04	Doxazosin
	N05BA04	Oxazepam
Bullous eruptions	C01BC03	Propafenone
	C07AB03	Atenolol
	R03BB01	Ipratropium bromide
	R03BB04	Tiotropium bromide
	C08CA02	Felodipine
Cardiac valve fibrosis	N06AB08	Fluvoxamine
	L04AX03	Methotrexate
	C09CA04	Irbesartan
	C03CA01	Furosemide
	G03CA03	Estradiol
Neutropenia/agranulocytosis	C07AA07	Sotalol
	H03AA01	Levothyroxine sodium
	C10AA05	Atorvastatin
	C01DA14	Isosorbide mononitrate
	G04CA02	Tamsulosin
Aplastic anaemia/pancytopenia	C09CA04	Irbesartan
	C10AA04	Fluvastatin
	S01EE01	Latanoprost
	S01ED01	Timolol
	R06AX27	Desloratadine
Rhabdomyolysis	G03CA03	Estradiol
	C02CA04	Doxazosin
	A10BB12	Glimepiride
	S01ED01	Timolol
	C01DA02	Glyceryl trinitrate
Upper gastrointestinal bleeding	R06AX26	Fexofenadine
	C10AA01	Simvastatin
	S01EC03	Dorzolamide
	L02AE03	Goserelin
	N05CF01	Zopiclone

ATC Anatomical Therapeutic Chemical

12 Discussion

In this study we present a novel approach to identify a surrogate 'gold standard' for drug safety signal detection using a systematic and rigorous methodology, applied across various data sources and which could be extended to examine other drug-event associations. We put together a list of drug-adverse event associations known to be true and drug-event associations considered to be unlikely based on current published scientific literature, drug product labels, spontaneous ADR reports and expert opinion. Although the rationale for creating this reference standard is to have one single index against which signal detection methods (as applied to EHR data) can be tested, this reference standard can be re-evaluated and adapted to different settings as needed.

In evaluating the evidence from the literature we only considered associations that were reported with use of the drug in therapeutic doses, which is consistent with the definition of an ADR [19]. For aspirin, citations referring to both cardiovascular prophylactic (low dose) and analgesic doses were considered. We considered, aside from case reports that described the clinical characteristics leading to suspicion of an ADR, publications that proposed (or elucidated) biological mechanisms for the associations. Such publications came in the form of both narrative reviews and systematic reviews. We likewise considered associations that were described in the context of drug-drug interactions (e.g. aplastic anaemia resulting from the synergistic interaction between azathioprine and allopurinol) [20]. For the event acute renal failure, we disregarded associations that arose from rhabdomyolysis leading to renal failure, but considered the reverse situation (i.e. associations for rhabdomyolysis that resulted in renal failure). While randomized controlled trials (RCTS) and meta-analyses are considered supreme with respect to level of evidence, this is more true for evidence regarding efficacy, not so much safety, of interventions [21–24]. This is apparent in Table 4, where most of the evidence pertaining to the positive associations came from observational studies and case reports (or reviews). The associations with Level I evidence are those that are well known (e.g. association of the NSAIDs and heparin with upper gastrointestinal bleeding) or well investigated, either because of controversy or public health impact (e.g. the association of the statins with rhabdomyolysis, and the association of coxibs and rosiglitazone with acute myocardial infarction). Interestingly, but perhaps not surprisingly, the most widely-investigated association was that between aspirin and upper gastrointestinal bleeding (259 MEDLINE citations overall, see Appendix [Online Resource 1]). Most of the publications related to this association, including clinical trials, described the drug as a comparator to other drugs that are presumed (and proven) to confer a lower risk of the event.

There have been previous attempts to develop a reference standard with which data mining methods for safety signal detection can be evaluated, 'rules of evidence' being devised ad hoc [25–27]. In the creation of this reference standard we employed a systematic approach incorporating various sources of drug safety information, the process designed to be transparent and reproducible, thus also making it easier to update. Different sources have varying comprehensiveness and accuracy with regards to documenting drug-adverse event associations. Because RCTs may be restricted to specific populations and lack statistical power to detect rare events, they must be supplemented by non-experimental studies and other types of evidence, including case reports [21–24]. Rare or idiosyncratic events (e.g. bullous eruption such as Stevens-Johnson syndrome) and events occurring after chronic exposure (e.g. cardiac valvulopathy) are unlikely to be identified in clinical trials, but rather in case reports or observational studies.

There was only one disagreement between evaluators in the final list of positive associations ('undetermined' vs 'positive' for the association paracetamol-anaphylactic shock; arbitration resulted in positive association). There was no disagreement between evaluators in the final list of 'negative control' associations. Although this high overall agreement between evaluators indicates that the resulting reference standard fulfills the pre-determined criteria, as the definitions of positive associations and 'negative controls' are based on existing knowledge at the time of this review, these associations (especially the 'negative controls') may be refuted as new data come along [28]. Hence, this reference standard should be considered dynamic and will need periodic re-evaluation. Adoption of this reference standard for use by other investigators can validate its applicability in other settings and will facilitate its further improvement.

While a reference standard, however rigorously constructed, may be able to permit evaluation and comparison of methods for signal detection, a method shown to successfully detect known drug-adverse events associations is not a guarantee that such method will also be able to detect signals, i.e. new, currently unknown drug-event associations (problem of contemporary comparison) [29].

13 Limitations

Since the selection of drugs for the reference standard was dependent on the presence of adequate exposure to detect an association within the EU-ADR network (i.e. drugs that are more frequently used in the population were more

likely to be chosen), this reference standard may not be as useful for evaluation in situations where the drug use patterns are expected to be different. In particular, the EU-ADR database network is unable to capture information on drugs that are primarily used in hospitals or specialist centres (e.g. anti-cancer drugs), and for this reason such drugs have not been included in the reference standard. This criterion also precluded the inclusion of known associations with drugs that have been withdrawn from the market for a long time before the accrual of healthcare data in the databases. Because of this there was no drug that could be used as a positive reference for the event cardiac valve fibrosis; the use of the appetite suppressants fenfluramine and phentermine, as well as the dopamine agonists pergolide and cabergoline, were inadequately documented or no longer captured in the databases because of the decline in use (or eradication in practice) of these drugs [30]. The choice as to which drug-event pairs can be considered for the positive associations was primarily established on the basis of the number of publications (i.e. number of MEDLINE citations with co-occurrence of the drug and the event of interest). This meant that drugs that have been on the market longer—or were involved in high-profile or controversial issues—had a higher chance of being included in the reference standard.

Finally, the availability of a surrogate 'gold standard' is only one component of the evaluation process for signal detection methodologies [3, 31]. Other issues that need to be considered in performance evaluation of these methods include standardization of event definitions, establishment of reliable and consistent criteria for adjudicating causality and expectedness of adverse events, as well as understanding variations in database content and quality.

14 Conclusions

A unique strategy for the construction of a reference standard to evaluate drug safety signal detection methodologies using EHR has been proposed. This reference standard should be considered dynamic, and as knowledge on drug safety evolves over time and new issues in drug safety arise, this reference standard can be periodically re-evaluated. Our proposed strategy represents a novel contribution to pharmacovigilance, with opportunities for adaptation to evaluate harms and benefits for other suspected ADRs.

Acknowledgments This research has been funded by the European Commission Seventh Framework Programme (FP7/2007-2013) under grant no. 215847—The EU-ADR Project. The funding agency had no role in the design and conduct of the study, the collection and management of data, the analysis or interpretation of the data, and preparation, review or approval of the manuscript. The authors thank the anonymous reviewers and Anders Ottosson of Astra Zeneca for their valuable comments and insights. Mariam Molokhia has previously received grants from AstraZeneca, Pfizer and the Serious Adverse Events Consortium (not for profit collaboration of industry and academia) for studies of ADRs. Miriam Sturkenboom is running a research group that occasionally performs studies for pharmaceutical companies according to unconditional grants. These companies include AstraZeneca, Pfizer, Lilly and Boehringer. She has also been a consultant to Pfizer, Novartis, Consumer Health, Servier, Celgene and Lundbeck on issues not related to the study.

References

1. Amery WK. Signal generation from spontaneous adverse event reports. Pharmacoepidemiol Drug Saf. 1999;8(2):147–50.
2. Coloma PM, Trifirò G, Schuemie MJ et al. On behalf of the EU-ADR Consortium. Electronic healthcare databases for active drug safety surveillance: is there enough leverage? Pharmacoepidemiol Drug Saf. Epub 2012 Feb 8.
3. Hauben M, Reich L. Drug-induced pancreatitis: lessons in data mining. Br J Clin Pharmacol. 2004;58(5):560–2.
4. World Health Organization. Safety of medicines: a guide to detecting and reporting adverse drug reactions [online]. Available from URL: http://whqlibdoc.who.int/hq/2002/WHO_EDM_QSM_2002.2.pdf. Accessed 10 Sep 2011.
5. Hauben M, Aronson JK. Defining 'signal' and its subtypes in pharmacovigilance based on a systematic review of previous definitions. Drug Saf. 2009;32(2):99–110.
6. Trifirò G, Fourrier-Reglat A, Sturkenboom MC, et al. The EU-ADR project: preliminary results and perspective. Stud Health Technol Inform. 2009;148:43–9.
7. Coloma PM, Schuemie MJ, Trifiró G, on behalf of the EU-ADR Consortium, et al. Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: the EU-ADR Project. Pharmacoepidemiol Drug Saf. 2011;20(1):1–11.
8. Trifirò G, Pariente A, Coloma PM. On behalf of the EU-ADR consortium, et al. Data mining on electronic health record databases for signal detection in pharmacovigilance: which events to monitor? Pharmacoepidemiol Drug Saf. 2009;18(12):1176–84.
9. Avillach P, Dufour JC, Diallo G, et al. Design and validation of an automated method to detect known adverse drug reactions in MEDLINE: a contribution to the European EU-ADR project. AMIA 2010 Annual Symposium (2010).
10. Bodenreider O. The unified medical language system (UMLS): integrating biomedical terminology. Nucleic Acids Res. 2004;32 (database issue):D267–70.
11. European Medicines Agency. European public assessment reports [online]. Available from URL: http://www.ema.europa.eu/ema/index.jsp?curl=/pages/medicines/landing/epar_search.jsp&murl=menus/medicines/medicines.jsp&mid=WC0b01ac058001d125. Accessed 20 Sep 2011.
12. DailyMed [online]. Available from URL: http://dailymed.nlm.nih.gov/dailymed/about.cfm. Accessed 20 Sep 2011.
13. Electronic Medicines Compendium (for drugs licensed in the United Kingdom). http://www.medicines.org.uk/emc/. Accessed 20 Sep 2011.
14. Micromedex. https://www.thomsonhc.com/hcs/librarian/. Accessed 20 Sep 2011.
15. RxList [online]. Available from URL: http://www.rxlist.com/. Accessed 20 Sep 2011.
16. Drugbank: open data drug and drug target database [online]. Available from URL: http://www.drugbank.ca/. Accessed 20 Sep 2011.

17. Trifirò G, Patadia V, Schuemie MJ, et al. EU-ADR healthcare database network vs. spontaneous reporting system database: preliminary comparison of signal detection. Stud Health Technol Inform. 2011;166:25–30.
18. DuMouchel W, Smith ET, Beasley R, et al. Association of asthma therapy and Churg-Strauss syndrome: an analysis of postmarketing surveillance data. Clin Ther. 2004;26(7): 1092–104.
19. Edwards IR, Aronson JK. Adverse drug reactions: definitions, diagnosis, and management. Lancet. 2000;356(9237):1255–9.
20. Kennedy DT, Hayney MS, Lake KD. Azathioprine and allopurinol: the price of an avoidable drug interaction. Ann Pharmacother. 1996;30(9):951–4.
21. Black N. Why we need observational studies to evaluate the effectiveness of health care. BMJ. 1996;312(7040):1215–8.
22. Papanikolaou PN, Christidi GD, Ioannidis JP. Comparison of evidence on harms of medical interventions in randomized and nonrandomized studies. CMAJ. 2006;174(5):635–41.
23. Stricker BH, Psaty BM. Detection, verification, and quantification of adverse drug reactions. BMJ. 2004;329(7456):44–7.
24. Ray W. Population-based studies of adverse drug effects. N Engl J Med. 2003;349(17):1592–4.
25. Lindquist M, Ståhl M, Bate A, et al. A retrospective evaluation of a data mining approach to aid finding new adverse drug reaction signals in the WHO international database. Drug Saf. 2000; 23(6):533–42.
26. Hauben M, Reich L. Safety related drug-labelling changes: findings from two data mining algorithms [published erratum appears in Drug Saf 2006; 29 (12): 1191]. Drug Saf. 2004;27(10): 735–44.
27. Hochberg AM, Hauben M, Pearson RK, et al. An evaluation of three signal-detection algorithms using a highly inclusive reference event database. Drug Saf. 2009;32(6):509–25.
28. Waller P. Dealing with uncertainty in drug safety: lessons for the future from sertindole. Pharmacoepidemiol Drug Saf. 2003;12(4): 283–7. (discussion 289–90).
29. Bate A, Edwards IR. Data mining techniques in pharmacovigilance. In: Hartzema AG, Tilson HH, Chan KA, editors. Pharmacoepidemiology and therapeutic risk management. Cincinnati: Harvey Whitney; 2008. p. 239–72.
30. Bhattacharyya S, Schapira AH, Mikhailidis DP, et al. Drug-induced fibrotic valvular heart disease. Lancet. 2009;374(9689): 577–85.
31. Levine JG, Tonning JM, Szarfman A. Reply: the evaluation of data mining methods for the simultaneous and systematic detection of safety signals in large databases: lessons to be learned. Br J Clin Pharmacol. 2006;61(1):105–13 (author reply 115–7).