The role of signal detection in Pharmacovigilance

The European landscape

Alexandra C. Păcurariu

519883-L-sub01-bw-Pacurariu
Processed on: 5-6-2018

The work described in this thesis was conducted at the department of Medical Informatics, within the Interdisciplinary Processing of Clinical Information (IPCI) department at the Erasmus University Medical Center, Rotterdam, the Netherlands and at the Dutch Medicines Evaluation Board (CBG-MEB), Utrecht, the Netherlands.
Financial support for the publication of this thesis was generously provided by the IPCI group and Medicines Evaluation Board.

For consistency reasons, some terms may have been standardized through the text. As a consequence, the text may differ from the published articles.

Cover & illustrations designed by Marija Nikolić
Lay-out by Legatron Electronic Publishing, Rotterdam, the Netherlands
Printing by Ipskamp Printing, Enschede, the Netherlands

ISBN: 978-94-028-0990-9

The Role of Signal Detection in Pharmacovigilance

De rol van signaaldetectie in Geneesmiddelenbewaking

Proefschrift

ter verkrijging van de graad van doctor aan de Erasmus Universiteit Rotterdam
op gezag van de rector magnificus
Prof.dr. H.A.P. Pols
en volgens besluit van het College voor Promoties

De openbare verdediging zal plaatsvinden op
vrijdag, 6 juli 2018 om 9.30 uur

door

Alexandra Cristina Păcurariu

geboren te Boekarest (Roemenie)

Erasmus University Rotterdam

PROMOTIECOMMISSIE

Promotor:	Prof.dr. M.C.J.M. Sturkenboom
Copromotor:	Dr. S.M.J.M. Straus
Overige leden:	Prof.dr. H.G.M. Leufkens Prof.dr. E.P. van Puijenbroek Prof.dr. B.H.Ch. Stricker

To my parents,

To Marius.

TABLE OF CONTENTS

Chapter 1
Introduction

Pharmacovigilance is the science and activities related to the detection, assessment, understanding and prevention of adverse effects or any other drug-related problems [1]. The etymological origin of 'pharmacovigilance' is a mix between the greek word *pharmakon* (drug) and the latin word *vigilare* (to guard). While medicines are tested in clinical trials before a marketing authorisation is granted, trials are necessarily limited in time and in the number and type of patients enrolled [1-3]. As a consequence, the possibilities to detect adverse drug reactions (ADRs) prior to marketing are limited. In particular, ADRs which are rare, with a long latency or those that occur only in specific patient groups or in specific contexts of administration are hard to detect. In addition, once on the market, medicines can be used both according to and outside the approved indications (off-label), by patients with multiple comorbidities and simultaneously treated with other medications. The 'real world' usage of drugs is more complex, unpredictable and dynamic than the experimental situation, and some ADRs cannot be observed in the experimental setting and will only be visible after approval. This leads to an uncertainty at approval stage that needs to be dealt with.

The uncertainty with regards to the safety profile of a drug cannot be completely avoided. However, there are ways to minimize it and this can be done through continuous monitoring along the entire product lifecycle. One of the pharmacovigilance processes in place to achieve this continuous monitoring is signal management. The signal management process is a 'set of activities performed to determine whether, based on an examination of individual case safety reports, aggregated data from active surveillance systems or studies, literature information or other data sources, there are new risks associated with an active substance or a medicinal product or whether risks have changed' [4]. In this thesis we focus on the European signal management process, referencing signal management in other parts of the world as needed, for context.

SIGNAL MANAGEMENT IN EUROPE

The European pharmacovigilance legislation adopted in 2010 [5] and operational since 2012 describes the signal management process. The legislation is supported by the 'Guideline on good pharmacovigilance practices Module IX on Signal management' [6], which offers guidance on each step from signal management process and describes its application in the European Union (EU) regulatory network, including each stakeholder role.

In Europe, the stakeholders involved in the signal management process include patients, healthcare professionals, marketing authorisation holders (MAHs), national competent authorities, the European Medicines Agency (EMA) and scientific committees such as Pharmacovigilance Risk Assessment Committee (PRAC). The PRAC is a scientific committee of EMA that includes members from each Member State, independent EC appointed experts, a patient representative and a healthcare professionals representative. The role of PRAC in signal management includes prioritisation of confirmed signals for further discussion, assessment and recommendations for actions [6]. In addition, PRAC is involved in regular revision of signal detection methodologies.

Signal management is often described as a sequential process (as shown in Figure 1) [7,8] with the following activities: detection, validation, confirmation, analysis and prioritisation, assessment and recommendation for action [9].

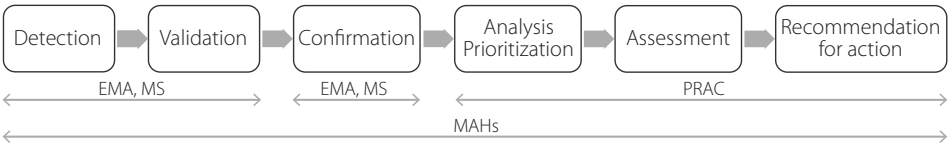


Figure 1: Signal management process and stakeholders' responsibilities as described in the legislation[1]

EMA=European Medicines Agency; PRAC=Pharmacovigilance Risk Assessment Committee; MAH=marketing Authorisation Holders

The first step, **signal detection**, is the process of monitoring safety data for information that 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, that is judged to be of sufficient likelihood to justify verificatory action*' [8]. Spontaneous reporting systems are an important focus for signal detection. In addition to spontaneous reports, a signal can arise from a wide variety of sources, such as aggregated data from active surveillance systems, studies, scientific literature or other data sources. In Europe, EudraVigilance database maintained by the EMA [11] is the largest database of spontaneous reports and an important source of signals. National competent authorities have systems in place to perform signal detection in their national databases as well as in EudraVigilance. This monitoring is required by the law. In parallel, marketing authorisation holders are screening their databases and from November 2017 onwards they will also monitor EudraVigilance through public access. The MAHs have the legal obligation to continuously monitor their products [6] and to report any signal found to the competent authorities. Only signals that require urgent attention will be reported via a standalone notification, namely an 'emerging safety issue' [6]. If is not urgent, the signal can be reported within the regular periodic safety reports, if the time between signal validation and report submission is less than 6 months. Alternatively, the signal can be communicated via update of product information or risk management plans, together with the proposed regulatory action.

The second step of the signal management process is **signal validation**. In this step, the detected signal is evaluated to verify whether there is sufficient evidence to justify further analysis [9]. At the end of the validation phase, it has to be decided if the association is at least a reasonable possibility and if the signal represents a new safety issue or a new aspect of a known safety issue. Various sources of information can be used during the validation phase. Each source has different relevance and weight depending on the signal at hand. The most commonly used data sources are: spontaneous reports, product information, scientific literature, drug utilisation data and regulatory documents such as risk management plans and periodic safety update reports. The third step, **signal confirmation** is a procedural step that identifies those signals that will be discussed at the

[1] Adapted from SCOPE Work Package 5 Signal Management-*Best Practice Guide*

next PRAC meeting [6,7]. The fourth step, **prioritisation**, ranks the signals' based on their impact on patients health, potential public health impact and strength of evidence. High priority means urgent attention and management without delay [6]. Sometimes signals with potential high media attention are prioritised in order to communicate the results of the assessment as early as possible.

The fifth step, **signal assessment** is the assessment of all scientific evidence available related to the signal [9]. The aim is to identify the next step: regulatory action, additional data collection or signal refutation [12]. As for validation, a wide range of data sources can be used and their relevance depends on the signal studied. The most commonly used data sources are: spontaneous reports, scientific literature, the application dossier, expert consultation and information provided by MAHs [12]. Additionally, pharmacoepidemiological studies can be performed in order to evaluate a signal. The studies can be requested to the MAHs or performed by the regulators themselves.

Despite of the clear signal management process, there are still questions that require further scientific investigation. This thesis aims to answer some of these questions and focuses on three areas of potential improvement: data sources, methods of detection, and signal prioritisation process. Since data sources and detection methods are interdependent, they will be discussed together.

DATA SOURCES AND METHODS OF DETECTION

Spontaneous reporting systems

The first important area for improvement is represented by the data sources for signal detection. The most important source for signal detection is spontaneous reporting systems, established since the 1960s [8]. Spontaneous reporting systems are passive public health surveillance systems, based on reporting of potential adverse events by healthcare professionals and patients. The core data elements collected within spontaneous reporting systems are established by international agreements [12] and include: an identifiable patient, an ADR, at least one suspect medicinal product and an identifiable reporter. The lack of any of these four elements means that the report is considered incomplete and does not qualify for reporting to the competent authorities. Other non-mandatory but relevant data elements are: patient demographics (age, gender, medical history), drug-related information (e.g., dose, strength, start and stop date and indication), ADR-related information (seriousness, outcome, duration, reaction start date and end date, dechallenge and rechallenge) [12].

Spontaneous reports are collected at regional, national and international level. Relevant examples of international spontaneous databases are Vigibase® maintained by the WHO Uppsala Monitoring Center in Sweden (WHO-UMC) [13], the FDA Adverse Event Reporting System (FAERS), and the Vaccine Adverse Event Reporting System (VAERS) in US [14], and EudraVigilance maintained by the European Medicines Agency (EMA) in Europe [11]. Established in 1968, Vigibase is the largest database of its kind in the world, with over 15 million reports of suspected ADRs [13] worldwide.

Established one year later, in 1969, FAERS contained at the end of 2016 more than 8.5 milion reports from USA and the rest of the world, on products licensed in the USA [15], while EudraVigilance, has a data collection dating back to 1995 and a size of 6.7 million reports worldwide (also at the end of 2016) [11]. In addition to the international databases, national or regional databases exist, maintained by national competent authorities. Also, each marketing authorisation holder is obliged to have its own internal database for the products owned. Due to reporting rules, the content of the industry owned, regional and international databases are overlapping to some extent in terms of contained reports. As far as we are aware, the degree of overlap is not described in the literature and is very much dependent on the ADR and drug under investigation [16]. A recent study showed that, in some situations, small national databases can reveal signals that are not identified in the larger international databases [17]. The most straightforward explanation for this finding is that the ability to observe signals depends on the background of the database, which differs among the existing databases [18].

The spontaneous reporting systems have both advantages and disadvantages. Their advantages include: large catchment population, low cost, and coverage of virtually all licensed drugs [19]. The most important limitations are: under-, over- and duplicate reporting, missing and incomplete data, lack of denominator data and unknown causality [8]. Underreporting is one of the most notorious limitations and is very hard to overcome since nothing can be done in the absence of data. Although it is difficult to provide an accurate estimate of the level of underreporting, a review [20] has shown that it may be as large as 90%, even for serious events. Public campaigns on ADR reporting can increase the reporting rate, however this may also lead to skewed reporting and false positive signals [21].Ultimately, some researchers argued that spontaneous reports are flawed and we should look for better alternatives [22,23]. One of these alternatives is the electronic healthcare data[2].

Electronic healthcare records[2]

'Electronic healthcare records (EHRs) is an organized set of healthcare data or collection of files available by computer through electronic format. It is derived from a raw electronic healthcare database. EHRs include administrative claims and electronic medical record data' [24]. Electronic medical records constitute a collection of medical records from general practitioners or specialists gathered in the office, clinic, or hospital and are used for diagnosis and treatment. Administrative claims data were the first automated databases used for population-based research and they were first established in North America in the 1980's. They consist of the billing codes that physicians, pharmacies, hospitals, and other health care providers submit for reimbursement of costs to payers [25]. Claims databases usually contain information on medical procedures, and dispensed drugs from primary care, hospitals and pharmacies.

[2] Terminology in this area is often unclear with the term electronic healthcare records and electronic medical records often used interchangeably.

In 1999, Vioxx® (rofecoxib), a non-steroidal anti-inflammatory drug, was approved and quickly gained popularity due to its selective mechanism of action that promised less gastro-intestinal haemorrhagic events. Five years later, in 2004, the drug was voluntarily withdrawn from the market [26]; the data safety monitoring board of the Adenomatous Polyp Prevention on Vioxx, (APPROVe) trial, recommended study termination due to an increased cardiovascular risk observed in the treated arm [27]. At the moment of withdrawal, millions of people were already exposed to the drug. Using actual data on the penetration of rofecoxib in the market, it has been calculated that if the medical records of 100 million patients would have been available for safety monitoring, the adverse cardiovascular effect would have been discovered in just three months after marketing [28]. This safety incident accelerated the investigation of alternative sources and methods for generating better evidence on safety of medicines in the post-marketing setting.

Initiatives started to develop in parallel in the USA and Europe. In Europe, one of the first initiatives, the Exploring and Understanding Adverse Drug Reactions (EU-ADR) project (2008) has created a multi-database network of EHRs across several European countries. The databases are a mix of primary care and administrative databases. Within the network, several detection methods have been tested, some traditional ones and some invented specifically for this purpose (see further details under Signal detection methods) [29]. In the same year, the Observational Medical Outcomes Partnership (OMOP) [30] was initiated in the USA, with a similar aim: to build a network formed of administrative databases and on top of it an alternative surveillance system. Very much focused on the data mining methodology, OMOP has organized a methods competition to facilitate development and evaluation of novel approaches for identifying drug safety issues in EHR [31]. OMOP has empirically evaluated the performance of various analytical methods and established a shared resource so that the broader research community can collaborate. An OMOP-EU-ADR comparison showed similar results in terms of methods performance: self-controlled designs, achieved higher performance than other methods [32]. When the OMOP project ended, it transitioned to the Observational Health Data Sciences and Informatics (OhDSI) community, a network of researchers sharing tools and methods to learn from health data [30]. In 2009, the Sentinel Initiative started in the USA, with a legal mandate to create a new post-marketing surveillance system [33]. As of September 2017, Sentinel has built a distributed database covering data on more than 223 million subjects. It was the source of 137 assessments of products, conditions, product-outcome pairs. Although Sentinel initially focused on signal refinement and validation, they they recently also started to explore new data mining methods for EHRs, as for example the tree temporal scan data mining method [34].

In 2010, the public private research project Pharmacoepidemiological Research on Outcomes of Therapeutics by a European ConsorTium (PROTECT) [35] started in Europe. A collaborative project between regulators, research organisations and pharmaceutical companies, run between September 2009 and February 2015 and had a work package dedicated to signal detection.

The EHRs are attractive due to their comprehensive clinical information, large coverage and their longitudinal dimension. In addition, they do not contain duplicates, as spontaneous reports do, they are not affected by under- or over-reporting and the information about exposure periods and clinical events is more valid and complete [36]. Despite these advantages, the EHRs are methodologically challenging for signal detection, as they were not built with this objective in mind. Hypothesis-free exploration is a secondary use for these data sources. Since they do not contain the presumed causality inherent in the spontaneous reports, the found associations are more prone to confounding. This could be solved by proper adjustment, unfortunately it is limited by the lack of information on lifestyle factors (main confounders) in many databases. Another limitation of EHRs is the lack of validation of diagnostic coding. In case of multi-database networks, heterogeneity in database structure, language and coding of drugs and diseases add to the challenges [37].

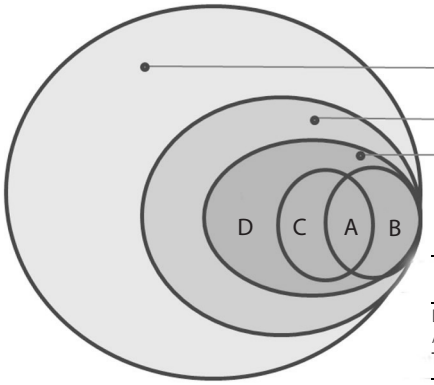
Signal detection methods

Signal detection started with the manual review of individual spontaneous case reports by trained medical professionals [38]. This method is still applied and may be very effective when numbers of cases are not too large. In 1960, in Canada, Ed Napke developed a system to collect and classify ADRs, the so -called the 'pigeonhole system' [39]. This system contained a storage unit with many small compartments, classified according to the drug and the reported reaction in which the reports were stored. In this way, it was easier to observe an increase in a certain type of reports. Afterwards, in 1974, Finney suggested statistical approaches for observing signals of ADRs [40]. He introduced the idea of using a 2×2 table for comparing the proportion of reports for a particular drug suspected for a certain ADR with the proportion of that reaction observed with the other drugs from the database (see Figure 2). This concept was named by Finney 'reaction proportion signalling' and was later updated and refined by Evans and renamed as 'proportional reporting ratio (PRR)' [41], a term which is established and still used nowadays.

Disproportionality analysis, as conceived by Finney and Evans, is the classical statistical approach to detect signals in large spontaneous databases and consists of calculating an observed-to-expected events ratio. The observed-to-expected ratio establishes if a specific ADR-drug combination is reported more frequently than expected in the untreated population. This ratio is based on an 'artificial denominator' since in spontaneous reporting systems there are no untreated patients, only patients treated with other drugs which experienced a certain ADR. These patients form the denominator, see Figure 2 [42].

The most common disproportionality methods are: PRR [41] and reporting odds ratio (ROR) [43], see Table 2. Their estimates are easy to calculate, however the results tend to become unstable when the number of events is small, resulting in potentially high estimates with wide confidence intervals. This instability led to the development of more advanced detection techniques based on Bayesian statistics. The Bayesian techniques try to adjust for uncertainty in the data by shrinking the estimates depending on the amount of data available [44]. The commonly used Bayesian methods are the Multi item Gamma Poisson Shrinker (MGPS) [44] and the Bayesian Confidence Propagation Neural

Network (BCPNN) [45]. An overview of the most common used methods and their calculation is presented in Table 1.



	Adverse event of interest	Other adverse events	Total
Drug of interest	A	B	a+b
All other drugs*	C	D	c+d
Total	a+c	b+d	a+b+c+d

Figure 2: A 2x2 contingency table for a drug-adverse event combination, in spontaneous reporting databases[3]

*Drugs may be reported as suspected or as concomitant medication.

Table 1: An overview of the common methods in signal detection

Name	Point estimate	Confidence interval	Institutions which use it	Advantages and Disadvantages
Frequentist methods				
ROR	ad/bc	$95\%CI=e^{\ln(ROR)\pm1.96\sqrt{(\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d})}}$	EMA, MEB	(+) Easy to apply and interpret (–) Higher rate of false positives (–) Unreliable at low numbers
PRR	a/(a+b)/c/(c+d)	$95\%CI=e^{\ln(ROR)\pm1.96\sqrt{(\frac{1}{a}-\frac{1}{a+b}+\frac{1}{c}-\frac{1}{c+d})}}$		
Bayesian methods				
IC*	$\log_2 \frac{a(a+b+c+d)}{(a+c)(a+d)}$		WHO	(+) Higher specificity (–) Lower sensitivity
EBGM	$\frac{a(a+bc+d)}{(a+c)(a+d)^{\dagger}}$		FDA, MHRA	(–) Harder to apply and interpret

BCPNN=Bayesian Confidence Propagation Neural Network; EBGM=Empirical Bayes Geometric Mean; IC=Information component; PRR=Proportional reporting ratio; ROR=Reporting Odds Ratio;
* based on BCPNN aproach [45]
† Complex formula, not entirely reproduced above. It is based on the Multi item Gamma Poisson Shrinker [46]

[3] Adapted from Poluzzi Elisabetta ER. Data Mining Techniques in Pharmacovigilance: Analysis of the Publicly Accessible FDA Adverse Event Reporting System (AERS), PhD thesis. 2012.

With regards to the signal detection methods applied to EHR, one option is to fit the EHR database into a spontaneous reports data structure and use the disproportionality methods presented above. However, the disproportionality methods are 'cross-sectional' and they do not take into account the longitudinal dimension of the data. Therefore, the second option is to use detection methods that can incorporate the longitudinal dimension of EHRs. One such example is the Longitudinal Gamma Poisson Shrinker (LGPS). LGPS is an adaptation of the Bayesian Gamma Poisson Shrinker but uses person time rather than case counts for the estimation of the expected number of events [47]. In this method, the incidence rate ratio for a ADR during treatment with a specific drug is calculated by multiplying the duration of exposure with the incidence ratio for the ADR when not exposed [47]. An underlying assumption here is that the risk does not vary with the time.

Another way to incorporate longitudinal information in the analysis is to use an extension/ adaptation of traditional epidemiological designs as: new user cohort [48], case control [49] or the self-controlled case series method [50]. These designs compare the incidence rate of the ADR during exposed time with the incidence rate during unexposed time. Since both designs use a comparator group to estimate the incidence rate, between-person confounding is an issue to be addressed when using these methods.

Also borrowed from the pharmacoepidemiology field, the 'self-controlled case series' [51] estimate the drug-ADR association using only information on cases. Each case acts as its own control, automatically adjusting for between -person confounding. This method is widely used for studying safety of vaccines that are administered to large cohorts and for which it is hard to find an unvaccinated group as comparison [50]. Temporal pattern discovery is another 'self-controlled method' specifically created for EHRs, invented by Noren et al. [52]. The objective is to identify interesting or unusual temporal patterns between the occurrence of an event and the administration of a prescribed drug. These unusual occurrence patterns are an indication for a potential signal. Again, only information on cases is used.

A full range of methods designed for data mining in EHR were tested and described by Schuemie [29] and Ryan [53].

The PROTECT initiative in Europe explored signal detection methods tailored to EHRs [19]. Their analysis was performed in The Health Improvement Network (THIN) database of longitudinal electronic health records from general practices in the UK [20], where they implemented a self-controlled cohort analysis with temporal Pattern Discovery [21]. The method identifies outcomes which have increased rates soon after initiation of treatment (temporally associated). The tested method was more conservative than the epidemiological studies, highlighting a lower number of drug adverse event pairs [22].

In the other project, the PROTECT group has explored the use of different medical term groupings when mining the database and use of subgrouping and stratification techniques in signal detection.

They also compared the existing signal methods across a wide range of databases to find out which one performs better and under which circumstances [35]. Based on the work done by PROTECT, changes were introduced in 2016 in signal management activities in Europe: the change of of statistical measure from PRR to ROR, change of applied thresholds, stratification by age groups and by region where reports originated [54]. Preliminary analyses have shown that these changes increase performance of the signal detection process [55].

Statistical signal detection methods function like a diagnostic system that needs to discriminate between true and false signals. The performance of methods is measured by this discrimination capacity. Ideally, the signal detection methods will be highly sensitive (e.g., identifying all existing signals) and at the same time highly specific (e.g., correctly dismissing all the false associations). Unfortunately, improving both sensitivity and specificity at the same time is not possible without incorporating extra information in the methods and since available information is usually limited, a trade-off between specificity and sensitivity needs to be made. Due to the cautionary approach in pharmacovigilance, a premium is put on sensitivity over specificity, e.g., we prefer to pick as many signals as possible with the cost of having more false positives. Many studies have compared the performance of statistical signal detection methods. In general, performance is similar, except when the number of reports is very low [56,57]. The implementation of methods in a specific database seems to be important for performance. Some of the implementation decisions that influence performance are: the type of database; including suspect or suspect and concomitant drugs in the denominator; choice of reference groups; control for confounding; level of hierarchy within the medical dictionary and choice of threshold and precision estimate [58].

Signal prioritisation

A third area in signal management that can benefit from further research and evidence based recommendations is signal prioritisation. Prioritisation is a decision-making process aimed to focus attention on signals with a high likelihood to be real and with a high public health impact, while at the same time discards the false positive signals. Prioritisation is necessary in signal management due to the high number of potential signals generated by data mining [59] and helps allocation of resources in the most optimal way.

During prioritisation, clinical, pharmacological and regulatory data are evaluated and weighted and a variety of variables can be considered as prioritisation factors. However, there is no general consensus as to how prioritisation should be done, mainly because different variables might have different importance depending on the evaluated signal. The Implementing Regulation mentions three prioritisation factors: the novelty of the drug, the strength of the association and the seriousness of the reaction [9]. The CIOMS report also mentions the novelty of the drug, seriousness, and in addition, a high and rapidly increasing disproportionality score [8]. Standardization of the prioritisation process, by using already constructed prioritisation frameworks, might help reduce the subjectivity. A summary of the main existing prioritisation tools is presented below.

The Regulatory Pharmacovigilance Prioritisation System has been developed, validated and implemented within the UK national competent authority. It prioritises signals according to four categories of factors: health consequences, strength of evidence, regulatory obligations and public perceptions. A second prioritisation tool, multi criteria decision analysis (MCDA) was developed by Levitan et al. [58] and takes into account medical impact (50% weight), strength of evidence (40% weight) and novelty of event (10% weight). The model was tested against expert group judgment and the agreement between the model and expert opinion was found to be moderate. A third instrument is the vigiRank model developed by Caster et al. [59], a prioritisation algorithm that accounts mainly for reports quality and content. The variables considered for inclusion capture different aspects of strength of evidence, focusing on quality and clinical content of individual reports, as well as trends in time and geographic spread. Public health impact was not considered by this algorithm. Finally, Coloma et al. [27] published a prioritisation exercise, albeit based on signals for EHRs, considering public health importance, novelty and biologic plausibility.

Once the prioritisation is complete, further signal strengthening and assessment is conducted and a recommendation for action is taken accordingly. The recommendation for action is more often an update of the product information with the newly discovered ADR. But actions can range from conducting additional studies to direct communication to healthcare professionals or even product withdrawal.

AIMS AND OUTLINE OF THIS THESIS

This thesis aims to present an overview of the current signal management process and explore how this may be improved from a scientific and a regulatory perspective, addressing especially three previously identified key areas: data sources, methods for detection and prioritisation. The work described in this thesis is based on various data sources both European and US based, mostly spontaneous reporting systems but also EHRs, as presented in Table 2.

Table 2: Summary of data sources used in this thesis

Chapter	Topic	Data source	Type	Setting	Size	Inception year
Chapter 3.1	Signal detection	EU-ADR	Network of electronic healthcare record databases	Denmark, Italy, Netherlands, UK	~30 million patients	2008
Chapter 3.1/6	Signal detection and prioritisation	Eudra-Vigilance	Spontaneous reporting database	Worldwide*	~6.7 million cases	1995
Chapter 3.2	Signal assessment	THIN	Electronic medical records; primary care	UK	~15.6 million patients	2002
Chapter 4	Signal detection	FAERS	Spontaneous reporting database	US	~9 million cases	1969

* Under the condition that the drug associated with the ADR has marketing authorisation in Europe; EU-ADR=Exploring and Understanding adverse drug reactions; FAERS=FDA Adverse Event Reporting System; THIN=The Health Improvement Network; UK=United Kingdom.

The thesis is divided into seven chapters. The first chapter provides a general introduction and context to the pharmacovigilance field. In Chapter 2 we provide an overview of the signal detection process in Europe; we describe the safety signals brought to the PRAC during the first 18 months of its operation and the actions taken in relation to those.

In **Chapter 3** we explore the use of EHRs in signal management. Firstly, as a tool for generation of new signals. The aim was to investigate in which particular situations EHR-based signal detection systems may provide an added value to spontaneous reporting systems, focusing on a limited set of adverse events of considerable importance in pharmacovigilance. Secondly, we explore EHR use for validation/assessment of existing signals, through hypothesis testing exercises. We chose to study the association between triptans and ischemic colitis, a signal that led to a regulatory action (update of the product information) but remained to be evaluated by a pharmacoepidemiological study. In **Chapter 4** we investigate if the performance of signal detection could be improved through age stratification and adjustment, with a special focus on paediatric signal detection.

In **Chapter 5** we provide an overview of the current prioritisation criteria as well as a brief description of their validity and performance. In **Chapter 6** we explore a more risk-based monitoring, based on the usage of the drug before and after authorization. We test the hypothesis that the number of patients exposed to the drug is a predictor of how quickly safety issues will be identified for that product in the initial period after authorisation.

Finally, in the last chapter, **Chapter 7**, a summary of findings, discussion and future perspectives are presented, as well as derived recommendations.

REFERENCES

1 Vandenbroucke JP. Observational research, randomised trials, and two views of medical science. PLoS Med. 2008;5(3):e67.

2 Amery WK, ISPE. Why there is a need for pharmacovigilance. Pharmacoepidemiol Drug Saf. 1999 Jan;8(1):61–4.

3 Stricker BHC, Psaty BM. Detection, verification, and quantification of adverse drug reactions. BMJ. 2004 Jul 3;329(7456):44–7.

4 EMA guideline on good pharmacovigilance practices (GVP) Annex I -Definitions. 2012.

5 The European Parliament and the Council of the European Union. Directive 2010/84/EU of the European Parliament and of the Council of 15 December 2010 amending, as regards pharmacovigilance, Directive 2001/83/EC on the community code relating to medicinal products for human use [Internet]. 2014 [cited 2014 Apr 5]. Available from: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:348:0074:0099:EN:PDF

6 EMA. Guideline on good pharmacovigilance practices, Module IX – Signal management. 2012.

7 SCOPE Joint Action [Internet]. Available from: http://www.scopejointaction.eu/

8 CIOMS Working Group. Practical aspects of signal detection in pharmacovigilance . Geneva; 2010.

9 Commission Implementing Regulation (EU) No 520/2012 [4] on the Performance of Pharmacovigilance Activities Provided for in Regulation (EC) No 726/2004 and Directive 2001/83/EC. 2012 Jun;

10 Hauben M, Aronson JK. Defining "signal" and its subtypes in pharmacovigilance based on a systematic review of previous definitions. Drug Saf Int J Med Toxicol Drug Exp. 2009;32(2):99–110.

11 2015 Annual Report on EudraVigilance for the European Parliament, the Council and the Commission. 2016 [cited 2016 Oct 9]. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Report/2016/03/WC500203705.pdf

12 EMA. Module VI – Management and reporting of adverse reactions to medicinal products. 2012.

13 Lindquist M. VigiBase, the WHO Global ICSR Database System: Basic Facts. Drug Inf J. 2008 Sep;42(5):409–19.

14 FDA. Safety reporting requirements for human drug and biological products: proposed rule. Federal Register 2003 Mar 14; 67 (50) 483-6.

15 Wong CK, Ho SS, Saini B, Hibbs DE, Fois RA. Standardisation of the FAERS database: a systematic approach to manually recoding drug name variants. Pharmacoepidemiol Drug Saf. 2015 May 1.

16 Poluzzi Elisabetta ER. Data Mining Techniques in Pharmacovigilance: Analysis of the Publicly Accessible FDA Adverse Event Reporting System (AERS). PhD thesis. 2012.

17 Poluzzi E, Raschi E, Godman B, Koci A, Moretti U, Kalaba M, et al. Pro-Arrhythmic Potential of Oral Antihistamines (H1): Combining Adverse Event Reports with Drug Utilization Data across Europe. PLoS ONE [Internet]. 2015 [cited 2015 Nov 10];10(3). Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4364720/

18 Almenoff JS, Pattishall EN, Gibbs TG, DuMouchel W, Evans SJW, Yuen N. Novel Statistical Tools for Monitoring the Safety of Marketed Drugs. Clin Pharmacol 38 Ther. 2007 May 30;82(2):157–66.

19 Almenoff J, Tonning JM, Gould AL, Szarfman A, Hauben M, Ouellet-Hellstrom R, et al. Perspectives on the use of data mining in pharmaco-vigilance. Drug Saf Int J Med Toxicol Drug Exp. 2005;28(11):981–1007.

20 Hazell L, Shakir SAW. Under-reporting of adverse drug reactions : a systematic review. Drug Saf Int J Med Toxicol Drug Exp. 2006;29(5):385–96.

21 Moore N, Hall G, Sturkenboom M, Mann R, Lagnaoui R, Begaud B. Biases affecting the proportional reporting ratio (PPR) in spontaneous reports pharmacovigilance databases: the example of sertindole. Pharmacoepidemiol Drug Saf. 2003 Jun;12(4):271–81.

22 Coloma PM, Trifirò G, Patadia V, Sturkenboom M. Postmarketing Safety Surveillance : Where does Signal Detection Using Electronic Healthcare Records Fit into the Big Picture? Drug Saf Int J Med Toxicol Drug Exp. 2013 Feb 2;

23 McNaughton R, Huet G, Shakir S. An investigation into drug products withdrawn from the EU market between 2002 and 2011 for safety reasons and the evidence used to support the decision-making. BMJ Open. 2014 Jan 15;4(1):e004221.

24 Goettsch. Glossary of Definitions of Common Terms [Internet]. Available from: http://www.imi-getreal.eu/News/ID/54/Update-GetReal-Glossary.

25 Strom BL, Kimmel SE, Hennessy S, editors. Pharmacoepidemiology. 5th ed. Chichester, West Sussex, UK: Wiley-Blackwell; 2012.

26 Krumholz HM, Ross JS, Presler AH, Egilman DS. What have we learnt from Vioxx? BMJ. 2007 Jan 20;334(7585):120–3.

27 Bresalier RS, Sandler RS, Quan H, Bolognese JA, Oxenius B, Horgan K, et al. Cardiovascular events associated with rofecoxib in a colorectal adenoma chemoprevention trial. N Engl J Med. 2005 Mar 17;352(11):1092–102.

28 McClellan M. Drug safety reform at the FDA--pendulum swing or systematic improvement? N Engl J Med. 2007 Apr 26;356(17):1700–2.

29 Schuemie MJ, Coloma PM, Straatman H, Herings RMC, Trifiro G, Matthews JN, et al. Using electronic health care records for drug safety signal detection: A comparative evaluation of statistical methods. Med Care. 2012;50(10):890–7.

30 Observational health data science and informatics [Internet]. Available from: https://ohdsi.org/

31 Observational Medical Outcomes Partnership [Internet]. 2013 [cited 2013 Mar 22]. Available from: http://omop.fnih.org/

32 Schuemie MJ, Gini R, Coloma PM, Straatman H, Herings RMC, Pedersen L, et al. Replication of the OMOP experiment in Europe: evaluating methods for risk identification in electronic health record databases. Drug Saf. 2013 Oct;36 Suppl 1:S159-169.

33 Platt R, Wilson M, Chan KA, Benner JS, Marchibroda J, McClellan M. The new Sentinel Network--improving the evidence of medical-product safety. N Engl J Med. 2009 Aug 13;361(7):645–7.

34 Yih. Pilot of Self-controlled Tree-temporal Scan Analysis for Gardasil Vaccine,. Sentinel; 2016.

35 PROTECT website [Internet]. 2013 [cited 2013 Mar 22]. Available from: http://www.imi-protect.eu/

36 Suling M, Pigeot I. Signal Detection and Monitoring Based on Longitudinal Healthcare Data. Pharmaceutics. 2012 Dec 13;4(4):607–40.

37 Coloma PM, Schuemie MJ, Trifirò G, Gini R, Herings R, Hippisley-Cox J, et al. Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: the EU-ADR Project. Pharmacoepidemiol Drug Saf. 2011 Jan;20(1):1–11.

38 Vandenbroucke JP. In defense of case reports and case series. Ann Intern Med. 2001 Feb 20;134(4):330–4.

39 David U. Canada's own "Dr. Safety." Can Med Assoc J. 2011 Nov 22;183(17):E1273–E1273.

40 Finney DJ. Systemic signalling of adverse reactions to drugs. Methods Inf Med. 1974 Jan;13(1):1–10.

41 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 Oct;10(6):483–6.

42 Hauben M, Madigan D, Gerrits CM, Walsh L, Van Puijenbroek EP. The role of data mining in pharmacovigilance. Expert Opin Drug Saf. 2005 Sep;4(5):929–48.

43 Rothman KJ, Lanes S, Sacks ST. The reporting odds ratio and its advantages over the proportional reporting ratio. Pharmacoepidemiol Drug Saf. 2004 Aug;13(8):519–23.

44 DuMouchel W. Bayesian Data Mining in Large Frequency Tables, with an Application to the FDA Spontaneous Reporting System. Am Stat. 1999;53(177–190).

45 Bate A, Lindquist M, Edwards IR, Olsson S, Orre R, Lansner A, et al. A Bayesian neural network method for adverse drug reaction signal generation. Eur J Clin Pharmacol. 1998 Jun;54(4):315–21.

46 DuMouchel W. Bayesian Data Mining in Large Frequency Tables, with an Application to the FDA Spontaneous Reporting System. Am Stat. 1999;53(177–190).

47 Schuemie MJ. Methods for drug safety signal detection in longitudinal observational databases: LGPS and LEOPARD. Pharmacoepidemiol Drug Saf. 2011 Mar;20(3):292–9.

48 Ryan PB, Schuemie MJ, Gruber S, Zorych I, Madigan D. Empirical performance of a new user cohort method: lessons for developing a risk identification and analysis system. Drug Saf. 2013 Oct;36 Suppl 1:S59-72.

49 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 Oct;36 Suppl 1:S73-82.

50 Petersen I, Douglas I, Whitaker H. Self controlled case series methods: an alternative to standard epidemiological study designs. BMJ. 2016 Sep 12;354:i4515.

51 Whitaker HJ, Hocine MN, Farrington CP. The methodology of self-controlled case series studies. Stat Methods Med Res. 2009 Feb;18(1):7–26.

52 Norén GN, Hopstadius J, Bate A, Star K, Edwards IR. Temporal pattern discovery in longitudinal electronic patient records. Data Min Knowl Discov. 2009 Nov 17;20(3):361–87.

53 Ryan PB, Madigan D, Stang PE, Overhage JM, Racoosin JA, Hartzema AG. Empirical assessment of methods for risk identification in healthcare data: results from the experiments of the Observational Medical Outcomes Partnership. Stat Med. 2012 Dec 30;31(30):4401–15.

54 Wisniewski AFZ, Bate A, Bousquet C, Brueckner A, Candore G, Juhlin K, et al. Good Signal Detection Practices: Evidence from IMI PROTECT. Drug Saf. 2016 Mar 7;

55 Kurz X. Advancing regulatory science, advancing regulatory practice. Pharmacoepidemiol Drug Saf. 2017 Jun 1;26(6):722–6.

56 Candore G, Juhlin K, Manlik K, Thakrar B, Quarcoo N, Seabroke S, et al. Comparison of Statistical Signal Detection Methods Within and Across Spontaneous Reporting Databases. Drug Saf. 2015 Jun;38(6):577-87.

57 van Puijenbroek EP, Bate A, Leufkens HGM, Lindquist M, Orre R, Egberts ACG. A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions. Pharmacoepidemiol Drug Saf. 2002;11(1):3–10.

58 Bate A, Evans SJW. Quantitative signal detection using spontaneous ADR reporting. Pharmacoepidemiol Drug Saf. 2009;18(6):427–436.

59 Seabroke S, Wise L, Waller P. Development of a novel regulatory pharmacovigilance prioritisation system: an evaluation of its performance at the UK Medicines and Healthcare products Regulatory Agency. Drug Saf. 2013 Oct;36(10):1025–32.

60 Levitan B, Yee CL, Russo L, Bayney R, Thomas AP, Klincewicz SL. A model for decision support in signal triage. Drug Saf Int J Med Toxicol Drug Exp. 2008;31(9):727–35.

61 Caster O, Juhlin K, Watson S, Norén GN. Improved Statistical Signal Detection in Pharmacovigilance by Combining Multiple Strength-of-Evidence Aspects in vigiRank : Retrospective Evaluation against Emerging Safety Signals. Drug Saf Int J Med Toxicol Drug Exp. 2014 Aug;37(8):617–28.

Chapter 2

Signal detection in Europe: current status

2.1

A description of signals during the first 18 months of the European Pharmacovigilance Risk Assessment Committee

Alexandra C Păcurariu
Sabine M Straus
Gianluca Trifirò
Martijn J Schuemie
Rosa Gini
Ron Herings
Giampiero Mazzaglia
Gino Picelli
Lorenza Scotti
Lars Pedersen
Peter Arlett
Johan van der Lei
Miriam C Sturkenboom
Preciosa M Coloma

Drug Saf. 2014 Dec;37(12):1059-66. doi: 10.1007/s40264-014-0240-1.

ABSTRACT

Background and Objectives: New pharmacovigilance legislation in the EU has underlined the importance of signal management, giving the European Medicines Agency's newly-established Pharmacovigilance Risk Assessment Committee (PRAC) the mandate to oversee all aspects of use of medicinal products including detection, assessment, minimization and communication relating to the risk of adverse reactions. In this study we describe the signals as brought to the PRAC during the first 18 months of its operation and the ensuing regulatory actions.

Methods: Data was collected from publicly available sources, for the period September 2012–December 2013, classified according to predefined rules and described using the appropriate descriptive statistics. Suspected adverse drug reactions (ADRs) were categorized using the Medical Dictionary for Regulatory Affairs (MedDRA) terms and drug names were mapped to the Anatomical Therapeutic Chemical (ATC) codes.

Results: During the study period, 125 signals concerning 96 medicinal products were discussed by the PRAC. The majority of signals were triggered by spontaneous reports (72%) and the median drug age (since marketing authorization) for drugs that prompted a signal was 12.3 years, significantly less compared to drugs that had no signal within the same period (19.7 years). The mean time until a decision was reached by PRAC was 2.5 months, with 42.8% of all decisions taken during the first meeting. The decisions to start a referral and to send a direct healthcare professional communication took the least amount of time (1.8 months and 1.7 months, respectively).

Conclusions: The importance of spontaneous reporting in signal detection and monitoring of safety issues throughout the entire lifecycle of a medical product is confirmed by our study. The new role of the PRAC contributed to a better coordination of real-time signal management via more prompt assessment and decision-making. If sustained, this may well optimize the safe and effective use of medical products.

INTRODUCTION

Pre-approval clinical research is primarily focused on establishing efficacy and its limitations with regards to identifying risks are well known and described previously [1-3]. Only after market exposure and use in every day practice more information on the full benefit risk profile will be identified.

An important cornerstone in further clarifying the risk profile of a medical product post-marketing is the detection of 'signals', that is, *'information which arise from one or multiple sources (including observation and experiments), which suggest a new potentially causal association or a new aspect of a known association, between an intervention and a set of related events, either adverse or beneficial, which would command regulatory, societal or clinical attention, and is judged to be of sufficient likelihood to justify verificatory action*' [4]. In pharmacovigilance we are primarily concerned with safety signals. Safety signals may arise anytime during the drug lifecycle but they are expected to occur more frequently in the first years of marketing [4]. However, also after several years new adverse events can arise.

Within the context of the new pharmacovigilance legislation in the European Union (EU), this key initial stage in the pharmacovigilance process is now duly recognized and specific responsibilities and interactions between stakeholders have been laid down in several guidance documents [5,6]. According to the current legislation, the marketing authorization holders, the European Medicines Agency (EMA) and national competent authorities *'should continuously monitor the data available in the EudraVigilance database'* [6,7].

The PRAC [8] at the EMA has a central role in scientific advice and decision making in relation to signal management. The mandate of the Pharmacovigilance Risk Assessment Committee covers all aspects of the risk management of the use of medicinal products for human use including the detection, assessment, minimisation and communication relating to the risk of adverse reactions, having due regard to the therapeutic effect of the medicinal product for human use, the design and evaluation of post-authorization safety studies and pharmacovigilance audit [6,7,9]. For signal management, the PRAC has an important role in the prioritization of potentially new safety issues, evaluating the underlying data and making recommendations regarding the regulatory actions that should be taken [5,9].

OBJECTIVES AND HYPOTHESIS

The objective of this study was to characterize the signals as brought to the PRAC during the first 18 months of its operation (September 2012–December 2013) and the ensuing regulatory actions. Within this characterization we focused on factors related to the lifecycle of a drug (e.g., time since marketing authorization).

METHODS

Data collection

Data regarding the safety signals was extracted from the publicly available information on the website of the EMA containing PRAC meeting minutes and recommendations on safety signals [10,11]. Suspected adverse drug reactions (ADRs) were categorized using the Medical Dictionary for Regulatory Affairs (MedDRA terminology, version 16.1), an international medical terminology developed under the auspices of the International Conference on Harmonization (ICH) [12]. International nonproprietary names of drugs were mapped to the Anatomical Therapeutic Chemical (ATC) codes according to the World Health Organization (WHO) classification and first authorization date was collected from the European Union Reference Date (EURD) list, when not available from the EMA website or other regulatory resources. For class effects, the oldest substance was used as a reference for calculating the time since marketing authorization. Data on medical product exposure was likewise obtained from PRAC meeting minutes.

Classification of variables

Signals are classified, by default, to be derived from either EU spontaneous reporting systems (Eudravigilance or national) or other sources. However, in order to provide more detailed information on the source of the signals, we employed the following classification for this study: (1) spontaneous case reports; (2) clinical trials; (3) observational studies; (4) published case reports or case-series; and (5) others (see Figure 1).

Signals with a mixed source: spontaneous and published case reports were classified as spontaneous. Communications from other regulatory authorities outside the EU were tracked, when possible, in order to determine the original signal source. The medical products were classified according to the type of authorization: centralized authorization (i.e., single marketing authorization across all EU countries) or national authorization (i.e., the product is authorized at a national level in one or more member states).

PRAC recommendations were collected and categorized according to the action taken. After a signal is brought to the PRAC for discussion, these are the possible outcomes: (1) no action; (2) request for further data; or (3) immediate action. Further evaluation in an ongoing periodic safety update report (PSURs) assessment was considered a separate and temporary outcome. The recommendation for cumulative reviews to be provided in future PSURs was considered routine pharmacovigilance. The regulatory actions that can be taken after a signal is discussed include: (1) update of summary of product characteristics (SmPC and PL); (2) direct communication to healthcare professionals; (3) update of the risk management plans; (4) suspension/withdrawal from the market; or (5) re-evaluation of benefit risk profile through a referral procedure. It is possible to have more than one regulatory action per signal. In this study, we considered an action as immediate if the decision was taken in the first PRAC meeting. Signals for which the outcome was not available in the month after the end of our study period (i.e., January 30, 2014) were labeled as ongoing.

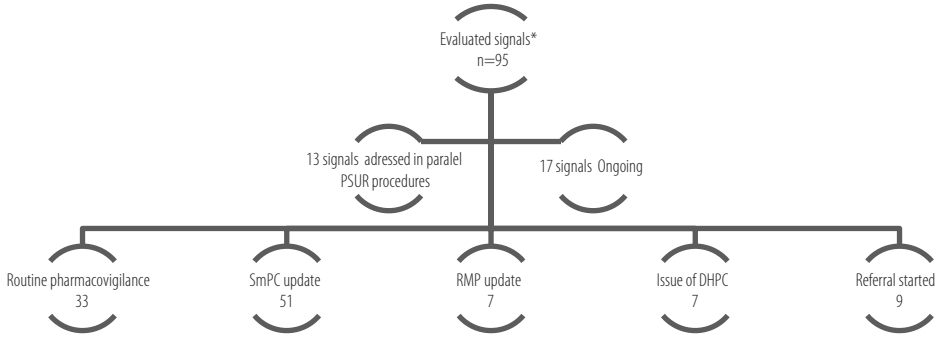


Figure 1: Workflow of signals at PRAC

SmPC=Summary of Product Characteristics; RMP=Risk Management Plan, DHPC=Dear Healthcare Professional communication; PSUR=Periodic Safety Update Report/ *More than one recommendation per signal possible

Data analysis

Descriptive statistics appropriate to the type of variables were used to describe the characteristics of signals discussed at PRAC during the study period.

We also tested the hypothesis that drugs that had signals in the study period were 'younger' than those that did not have signals. For this purpose we compared the drugs which had at least one signal on the PRAC agenda during the study period with a set of controls that were drugs monitored during the same period but that did not yield any signal considered at PRAC. These controls were chosen from the signal work-sharing list [13] and from the list of centralized products monitored by the EMA. In the case of signal work-sharing list, to correct for potential variations in applying monitoring methodologies between countries, drugs were matched on Lead Member State (i.e., country responsible for monitoring of a particular drug) to ensure that they underwent the same screening process.

RESULTS

During the study period September 2012–December 2013, PRAC 125 signals were discussed by PRAC, for 96 different drugs. Among the 125 signals, 15 were follow-ups from the previous Pharmacovigilance Working Party (i.e., former scientific group that handled signals at EMA before establishment of PRAC) discussion.

A descriptive analysis of all signals discussed at PRAC is presented in Table 1. The majority of signals were triggered by spontaneous reports (72%), followed by clinical trials (8%) and observational studies (8%). Ten signals (8%) originated from regulatory authorities outside Europe. The most

frequently discussed signals were related to skin and subcutaneous tissue disorders (12.8%), nervous system disorders (10.4%), cardiac disorders (6.4%) and immune system disorders (6.4%).

Table 1: Characteristics of signals discussed at PRAC September 2012–December 2013

Variables	
Signals count	125*
Medicinal products count	96
Identifier count (%)	
European Medicines Agency	65 (52.0%)
Individual Member States	
The Netherlands	15 (12.0%)
United Kingdom	15 (12.0%)
Other	30 (24%)
Italy	7 (5.6%)
Sweden	5 (4.0%)
France	4 (3.2%)
Time since marketing authorization years (%)	
Median (range)	12.3 (0.54–67.9)
≤5 years	20 (21.1%)
5-10 years	20 (21.1%)
≥10–15 years	16 (16.8%)
≥15 years	39 (41.0%)
Type of authorization† count (%)	
Centralized	49 (51.0%)
National	41 (42.7%)
Mixed	6 (6.3%)
Signals of special interest count (%)	
Drug Interaction	13 (10.6%)
Medication error	2 (1.6%)
Off-label use	2 (1.6%)
In utero exposure	2 (1.6%)
Accidental exposure	1 (0.8%)
Source, count (%)	
Spontaneous cases	90 (72.0%)
Randomized controlled trials	10 (8.0%)
Observational (post marketing) studies	10 (8.0%)
Literature case reports	8 (6.4%)
Other	7 (5.6%)

Table 1: *Continued*

Variables	
EU spontaneous reporting systems	94 (87%)
Other sources	14 (13%)
System Organ Class‡, count (%)	
Skin and subcutaneous tissue disorders	16 (12.8%)
Nervous system disorders	13 (10.4%)
Cardiac disorders	8 (6.4%)
Immune system disorders	8 (6.4%)
Blood and lymphatic system disorders	7 (5.6%)
Investigations	7 (5.6%)
Vascular disorders	7 (5.6%)
Other	59 (47.2%)
Drug class, count (%)	
Antineoplastic and Immunomodulators	25 (26%)
Nervous system	20 (20.8%)
Anti-infective for Systemic Use	13 (13.5%)
Alimentary Tract And Metabolism	5 (5.2%)
Other	33 (34.4%)

ATC=anatomic therapeutic chemical classification. *Three signals were not counted for the following reasons: two were considered a duplication of the same signal for a different vaccine strain (primary ovarian failure and complex regional pain syndrome with HPV vaccines) and another one (boceprevir and drug interaction with quetiapine) was extended (considered class effect) from an already discussed signal. †Centralized authorization=a single marketing authorization that is valid in all European Union countries, National authorization=the product is authorized and marketed in one or more member state(s), Mixed=a combination of centralized and national authorization; ‡System Organ Class= classification of an adverse reaction according to its etiology and manifestation site in MedDRA terminology.

The median time since the first marketing authorization in a European country for the drugs discussed at the PRAC was 12.3 years (range=0.5–67.9), with 42.2% being less than 10 years on the market (see Figure 2).

Exposure data was available for 75% of drugs, however it was variously reported as either number of patients (42%) or person-years (33%) and across different time periods and therefore not directly comparable between drugs. From the comparable data, the median cumulative exposure since marketing authorization until signal date was 2.1 million patients (range=0.003 to 320 million patients), the majority of drugs (67.6%) having an exposure of less than 10 million patients.

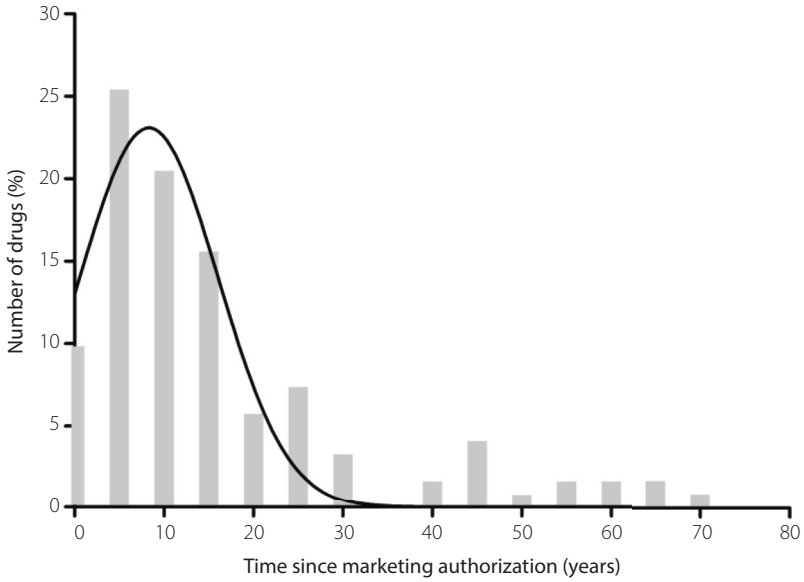


Figure 2: Time since marketing authorization across drugs that had a signal on the PRAC agenda

Table 2 summarizes PRAC final recommendations regarding signals and the mean time from first discussion until decision. Signals under ongoing evaluation (n=17, 13.6%) and those addressed in parallel procedures were excluded from the time analysis since no final outcome was reached for those at the time of writing of this article. The mean time-to-PRAC decision for a signal was 2.5 months (95%CI=2.0–3.2) with 42.8% of all decisions taken during the first meeting (i.e., immediate action). We performed a sensitivity analysis where we included the decision to address the signals in ongoing PSURs in the calculation; for this we obtained a mean time-to-PRAC decision of 2.2 months (95%CI=1.7–2.7) with 53.8% immediate actions taken.

The decisions to start a referral and to communicate a safety issue via direct healthcare professional communication took the least amount of time (1.8 months and 1.7 months, respectively). These results should be considered in the context of the fact that PRAC conducts meetings on a monthly basis.

For 57.2% of the signals, additional information was requested after the first discussion in the PRAC either from marketing authorization holders via a cumulative review (n=65) or from member states, in the form of non-urgent information request (n=8). The cumulative reviews were submitted either within 30 or 60 days, or addressed during an ongoing periodic safety updates report procedure (see Figure 3).

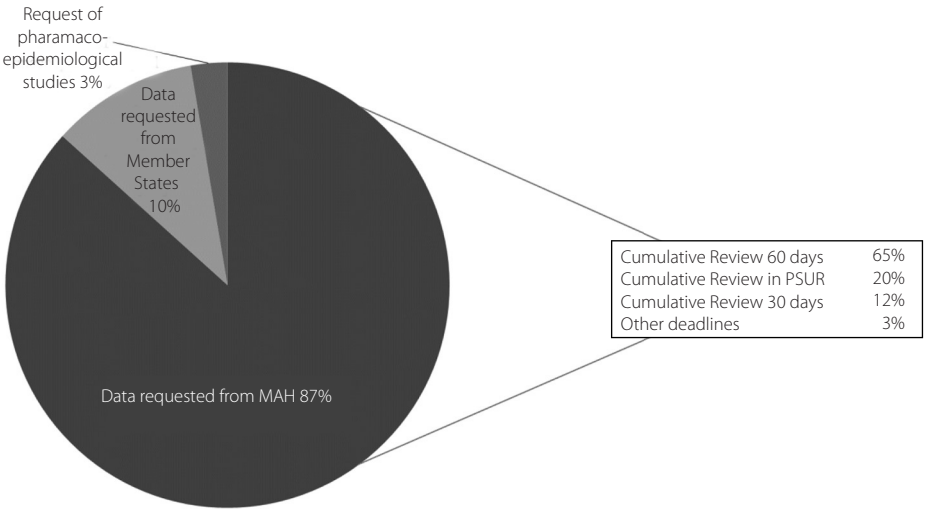


Figure 3: Type of additional information requested during signal assessment at PRAC.
MAH=Marketing authorisation holder, PRAC=Pharmacovigilance Risk Assessment Committee; PSUR=Periodic Safety Update Report

In the second part of the study, the hypothesis that drugs with signals are 'younger' (i.e., have been on the market more recently) on average than drugs without signals was tested. The comparison between drugs with signals (n=96) and without (n=894) at the end of the monitoring period showed that the drug age was significantly lower for drugs that had identified safety issues in the period (median=12.3 vs. 19.7 years, p=0.01, Mann-Whitney U test).

DISCUSSION

At the time of approval, knowledge of the full benefit-risk profile of any new drug is incomplete due to well-known limitations of pre-approval research. Throughout a drug's lifecycle, (serious) safety issues may emerge and while market approval may mark the end of drug development, it also marks the start of continuous evaluation of benefits and risks. The results of our study reaffirm the important role of spontaneous reporting in detection of signals and continuous need for monitoring since safety issues are also identified later in the life cycle of a drug.
For the interpretation of the results, it is important to keep in mind that the signals discussed at PRAC and hence considered in our study represent only a subset of all signals discussed in the regulatory framework.

The most frequent source of signals discussed by the PRAC was spontaneously reported ADRs (72%). This is in line with studies from the United States [13], where spontaneous reports were also found to be the most frequent source. Within signals from spontaneous sources, 10.6% of them had multiple origins: spontaneous reports and literature, while another 6.5% were identified exclusively from

published case reports, which emphasizes the importance of continuously monitoring scientific publications [3,14]. For a few signals arising from spontaneous reports, the evidence was based on one single case report (e.g., nomegestrol acetate-deep vein thrombosis, human papillomavirus vaccine-bronchospasm and erlotinib-pancreatitis).

The drug age in our study was significantly lower for medicines with a signal as compared to those without (median=12.3 vs. 19.7 years, p=0.01). Two papers regarding Food and Drug Administration (FDA) safety-related drug label changes reported that the safety actions occurred at a median of 11 [15], and 10 years [16] after initial approval, similar to our results. Another paper investigating regulatory actions for biologicals reported the mean time to a safety related regulatory action to be 3.7 years [17]. This shorter time is probably explained by the fact that this study only included biological drugs, which have an essentially different safety profile and are thus more intensively monitored post-marketing via post-authorization safety studies and/or registries.

While signals occur more frequently for younger drugs (see Figure 2), safety issues still appear for drugs that have been on the market for more than 50 years (e.g., cloroquine, thiopental, codeine, and triamcinolone).

This is probably due to change in patterns of utilization for these drugs, better implementation of safety monitoring, increased awareness in relation to certain safety issues as well as finalization of long-term observational studies. Such an example is the signal for codeine and life-threatening toxicity in CYP2D6 ultra-rapid metabolizers, a safety issue which only occurs in a small sub-population. The findings in our study are in line with the results of Mol et al. who showed that 27% of serious safety issues were communicated to healthcare professionals ten or more years after approval [18]. Another contributing factor to the identification of signals for old drugs is that some signals are not new from a scientific point of view but they can appear to be so from a regulatory perspective (e.g., when a certain adverse reaction is listed in the summary of product characteristics in some countries but not in others).

The most frequent recommendation was a change in the product information and this is similar to what has been reported in relation to the post-marketing safety surveillance decisions taken in the US [13,19].
The mean time from signal the identification until a PRAC decision was taken 2.4 months. A timeframe of 21 months from signal detection to action has been reported by Hochberg et al. [20] for the FDA's system, although the data are not directly comparable, since there is also a time-lapse between decision and actual implementation, which we did not take into account.

According to our analysis, the PRAC decision making process is efficient; especially in case of serious concerns leading either to referral or DHPC dissemination, which were handled even more expeditiously (see Table 2). This is in line with a recent study that described the PRAC activities since

its initiation and reported some process indicators which showed that the system is more structured, faster and with a more risk-proportionate approach [9].

A limitation of our work might be that only signals discussed at PRAC were considered, although there are other regulatory pathways through which signals can be handled (e.g., PSURs) so we analyzed only a fraction of the available information. Limited availability and heterogeneity of exposure data precluded further analysis of this variable, therefore we recommend increased standardization in its reporting, although we acknowledge the difficulties of acquiring accurately consistent exposure data at the European level.

In conclusion, the importance of spontaneous reporting in signal detection and monitoring of safety issues throughout the entire lifecycle of a medical product is confirmed by our study. The new role of the PRAC contributed to a better coordination of real-time signal management via more prompt assessment and decision-making. If sustained, this may well optimize the safe and effective use of medical products.

REFERENCES

1 Amery WK, ISPE. Why there is a need for pharmacovigilance. Pharmacoepidemiol. Drug Saf. 1999;8:61–4.

2 Vandenbroucke JP. Observational research, randomised trials, and two views of medical science. PLoS Med. 2008;5:e67.

3 Stricker BHC, Psaty BM. Detection, verification, and quantification of adverse drug reactions. BMJ. 2004;329:44–7.

4 CIOMS Working Group. Practical aspects of signal detection in pharmacovigilance . Geneva; 2010.

5 EMA. Guideline on good pharmacovigilance practices, Module IX – Signal management. 2012.

6 Commission Implementing Regulation (EU) No 520/2012. 2012;

7 Directive 2010/84/EU of the European Parliament and of the Council [Internet]. [cited 2014 Apr 5]. Available from: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:348:0074:0099:EN:PDF

8 Countdown to July 2012: the Establishment and Functioning of the PRAC, EMA/315258/2012 [Internet]. [cited 2014 Apr 15]. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Other/2012/07/WC500129301.pdf

9 Arlett P, Portier G, de Lisa R, Blake K, Wathion N, Dogne J-M, et al. Proactively managing the risk of marketed drugs: experience with the EMA Pharmacovigilance Risk Assessment Committee. Nat. Rev. Drug Discov. 2014;

10 Questions & answers on signal management [Internet]. 2014. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Other/2013/09/WC500150743.pdf

11 PRAC meeting minutes [Internet]. [cited 2014 Apr 15]. Available from: http://www.ema.europa.eu/ema/index.jsp?curl=pages/about_us/document_listing/document_listing_000353.jsp&mid=WC0b01ac05805a21cf

12 PRAC recommendations on safety signals [Internet]. [cited 2014 Apr 15]. Available from: http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/document_listing/document_listing_000375.jsp

13 MedDRA, Introductory Guide MedDRA Version 15.1, MSSO-DI-6003-15.1.0. 2012.

14 Ishiguro C, Hall M, Neyarapally GA, Dal Pan G. Post-market drug safety evidence sources: an analysis of FDA drug safety communications. Pharmacoepidemiol. Drug Saf. 2012;21:1134–6.

15 Pontes H, Clément M, Rollason V. Safety Signal Detection: The Relevance of Literature Review. Drug Saf. Int. J. Med. Toxicol. Drug Exp. 2014;

16 Moore TJ, Singh S, Furberg CD. THe fda and new safety warnings. Arch. Intern. Med. 2012;172:78–80.

17 Lester J, Neyarapally GA, Lipowski E, Graham CF, Hall M, Dal Pan G. Evaluation of FDA safety-related drug label changes in 2010. Pharmacoepidemiol. Drug Saf. 2013;22:302–5.

18 Giezen TJ, Mantel-Teeuwisse AK, Straus SMJM, Schellekens H, Leufkens HGM, Egberts ACG. Safety-related regulatory actions for biologicals approved in the United States and the European Union. JAMA J. Am. Med. Assoc. 2008;300:1887–96.

19 Powers A, Cook G. POtential safety signals and their significance. Arch. Intern. Med. 2012;172:72–3.

20 Hochberg AM, Reisinger SJ, Pearson RK, O'Hara DJ, Hall K. Using data mining to predict safety actions from FDA adverse event reporting system data. Drug Inf. J. 2007;41:633–43.

21 Edwards IR. Good pharmacovigilance practice and the curate's egg. Drug Saf. Int. J. Med. Toxicol. Drug Exp. 2012;35:429–35.

Chapter 3

Electronic healthcare records for signal generation and validation – two possible uses

3.1

Useful interplay between spontaneous ADR reports and electronic healthcare records in signal detection

Alexandra C Păcurariu
Sabine M Straus
Gianluca Trifirò
Martijn J Schuemie
Rosa Gini
Ron Herings
Giampiero Mazzaglia
Gino Picelli
Lorenza Scotti
Lars Pedersen
Peter Arlett
Johan van der Lei
Miriam C Sturkenboom
Preciosa M Coloma

Drug Saf. 2015 Dec;38(12):1201-10. doi: 10.1007/s40264-015-0341-5.

ABSTRACT

Background and Objectives: Spontaneous reporting systems (SRSs) remain the cornerstone of post-marketing drug safety surveillance despite their well-known limitations. Judicious use of other available data sources is essential to enable better detection, strengthening and validation of signals. In this study we investigate the potential of electronic healthcare records (EHRs) to be used alongside SRS as an independent system, with the aim to improve signal detection.

Methods: A signal detection strategy, focused on a limited set of adverse events deemed important in pharmacovigilance, was performed retrospectively in two data sources: (1) EU-ADR database network and (2) EudraVigilance database using data between 2000 and 2010. Five events were considered for analysis: (1) acute myocardial infarction (AMI); (2) bullous eruption; (3) hip fracture; (4) acute pancreatitis; and (5) upper gastrointestinal bleeding (UGIB). Potential signals identified in each system were verified using current published literature. The complementarity of the two systems to detect signals was expressed as percentage of unilaterally identified signals out of the total confirmed signals. As a proxy for the associated costs, the number of signals that need to be reviewed to detect one true signal (number needed to detect: NND) was calculated. The relationship between background frequency of events and capability of each system to detect signals was also investigated.

Results: The contribution of each system to signal detection appeared to be correlated with the background incidence of the events, being directly proportional to the incidence in EU-ADR and inversely proportional in EudraVigilance. EudraVigilance was particularly valuable in identifying bullous eruption and acute pancreatitis (71% and 42% of signals correctly identified from the total pool of known associations, respectively) while EU-ADR was most useful in identifying hip fractures (60%). Both systems contributed reasonably well in identification of signals related to UGIB (45% in EudraVigilance, 40% in EU-ADR), but only fairly for signals related to AMI (25% in EU-ADR, 20% in EudraVigilance). The costs associated with detection of signals were variable across events; however, it was often more costly to detect safety signals in EU-ADR than in EudraVigilance (median NND=7 vs. 5).

Conclusions: An EHR-based system may have additional value for signal detection, alongside already established systems, especially in the presence of adverse events with high background incidence. While SRS appeared to be more cost-effective overall, for some events the costs associated with signal detection in EHR might be justifiable.

INTRODUCTION

Spontaneous reporting systems (SRSs) have long been the foundation of post-marketing drug safety surveillance [1]. Despite their broad utilization, information found in such systems is sometimes limited [2] and as a consequence, decisions based solely on data from these systems often need to take into account significant uncertainties [3]. In many instances it is not sufficient to use a single source of information to understand a particular drug safety issue [4] and there is a recognized need to put together, in an efficient way, most, if not all, available relevant sources.

Following the drug safety concerns surfacing between 2004 and 2007, many discussions started whether we can have a more pro-active approach to signal detection instead of relying on passive surveillance systems. Both in Europe and the USA it was explored whether electronic healthcare record (EHR) databases that comprise detailed data collected longitudinally and routinely in actual care for large scale populations [5] may be used for post-marketing safety surveillance. EHRs have been primarily used for signal evaluation studies; however, in recent years, various projects have explored ways of using them as an additional source for signal detection systems, e.g. OMOP [6,7] PROTECT [8] and EU-ADR [9,10].

To date, only two studies [11,12] have tried to combine both sources in order to support the signal detection process, while the majority of the available research focused rather on comparing the two systems in terms of overall performance and usefulness [13-16].

The aim of this study was to investigate in which particular situations EHR-based signal detection systems may provide an added value to already existing SRS, focusing on a limited set of adverse events of considerable importance in pharmacovigilance. To express this added value, we used performance indicators, including percentage of unilaterally identified signals and sensitivity to describe the "gains" as well as number NND for the "costs" associated with signal detection.

METHODS

Design

A signal detection strategy focused on a limited set of adverse events was performed retrospectively in two database systems: (1) EU-ADR and (2) EudraVigilance from January 1, 2000 to January 1, 2010. These are described separately below. The two systems were considered individually and the most sound event definitions possible and implementation of signal detection methods were taken into account in each database independently in order to optimize the performance of each. In this study we used the term "signal" to refer to a signal of disproportionate reporting (SDRs) as defined in CIOMS VIII [1] in the context of signal detection in EudraVigilance and equivalent to a statistically significant drug-adverse event association which meets a specific threshold of increased risk in the context of EU-ADR. All drugs captured in either of the two systems were considered. Drugs not identifiable at

the fifth level of the WHO Anatomical Therapeutic Chemical (ATC) classification system, as well as herbal supplements, were excluded.

Events of interest

We considered the following five events, selected from a list of events previously identified as important based on expert judgment and predefined criteria [17]: (1) acute myocardial infarction (AMI); (2) bullous eruption (BE); (3) hip fracture; (4) acute pancreatitis; and (5) upper gastrointestinal bleeding. These events were chosen because of their diversity in etiology, background incidence, and drug-attributable risk -attributes which we consider might impact the performance of the two systems. To investigate the possible correlation between the signal detection performance of each system and the frequency of the events, we ranked our events of interest according to empirically determined background frequency (i.e., incidence rate in the general population). These incidence rates were derived from the EU-ADR network, in order to maintain the same base population across events, which allowed for a more meaningful comparison [9].

SRS: EudraVigilance

As exemplar for SRS, we used EudraVigilance, a web-based information system launched in December 2001 and designed to manage information on suspected adverse drug reactions (ADRs) which are reported for drugs licensed in Europe. The total number of individual reports as of December 2013 was 4.5 million, with 38% cases originating from the EU and 62% from the rest of the world [18].

Capturing events of interest

In EudraVigilance, suspected ADRs are coded using the Medical Dictionary for Regulatory Activities (MedDRA®), an international medical terminology developed under the auspices of the International Conference on Harmonisation (ICH) [19]. For capturing the events of interest we used adapted searches derived ad hoc from standardized MedDRA® queries, [20] similar to the approach used by Patadia et al. [16].

Method of signal detection

Signal detection in EudraVigilance was performed using the proportional reporting ratio (PRR) method [21], previously validated in this database by Alvarez et al. [22]. Only cases received within the study period (January 1, 2000 to January 1, 2010) were considered for identification of signals.

The threshold chosen to define a signal was a lower limit of the confidence interval of the PRR greater than 1 together with at least 3 cases reported with the investigated association [23]. No further adjustment was done for possible confounding variables.

EHR-based system: EU-ADR

As exemplar for EHR, we used EU-ADR, a computerized system designed to detect potential ADRs and built on a network of established databases from various European countries [9]. Data from seven databases in three countries (Denmark, Italy, the Netherlands) were used in this study. EU-ADR includes both population-based primary care databases (Integrated Primary Care Information database (IPCI, Netherlands), Health Search/CSD Patient and Pedianet (Italy)), and record-linkage systems (Aarhus University Hospital Database (Denmark), the PHARMO Network (Netherlands), and the regional Italian claims databases of Lombardy and Tuscany). The source population covered by the database network is approximately 20 million patients. Drug exposure in EU-ADR was identified from prescription or dispensing data (depending on the database) using ATC codes. Prescriptions with the same ATC code where the start date of one prescription precedes the end date of the other prescription were merged into a single episode of drug use, starting at the beginning of the first prescription, and ending at the end of the last prescription. Periods of concomitant drug use were labelled as separate episodes. Only current exposure (within 30 days of an event of interest) was considered [9]. The characteristics of the EU-ADR network have been extensively described elsewhere [9,24].

Capturing events of interest

Definitions for each event of interest were previously constructed by a team of experts and based on those, queries were performed in each database of the network, using the corresponding diagnosis coding schemes: International Classification of Diseases 9th and 10th revision and International Classification of Primary Care, supplemented with additional criteria as laboratory values and unstructured free text searches, where applicable. Results were subsequently pooled across all databases [25]. The events acute myocardial infarction and upper gastrointestinal bleeding were previously validated in the databases concerned [26,27].

Method of signal detection

For EU-ADR, a signal detection method specifically developed for EHR data was used: Longitudinal Gamma Poisson Shrinker (LGPS) [28]. LGPS is a cohort-based method, adapted from a Bayesian method (DuMouchel's Gamma Poisson Shrinker, that uses person-time rather than case counts for the estimation of the expected number of events. Previous evaluation against other signal detection methodologies showed that LGPS is the best performing method in this database system [29]. We applied a threshold to the LGPS risk ratio (RR_{LGPS}) of a lower limit of the 95% credible interval (95%CI)>1 [28]. After LGPS, we applied a second method, Longitudinal Evaluation of Observational Profiles of Adverse events Related to Drugs (LEOPARD), which adjusts for possible protopathic bias and improves performance. LEOPARD is based on the comparison of rates of drug prescriptions initiated within a fixed window (±25 days) prior to and after the occurrence of an event, based on the assumption that an increase in the number of prescriptions started after an event relative to the number of prescriptions started prior to the event is an indication of protopathic bias. From a statistical perspective, this is a binomial test which compares the distributions of prescription before and after the event [28].

Verification of signals

We reviewed currently available literature in order to determine which drug-adverse event associations identified in the dataset represent already known associations. In contrast to the approach used in previous studies, verification was performed for all drug-event associations, irrespective of whether a potential signal was flagged or not by any of the signal detection methods used in either EudraVigilance or EU-ADR.

An automatic tool developed within the EU-ADR, which searches MEDLINE-indexed publications concerning adverse drug reactions [30,31], was used to qualify the drug-event associations as 'ADR' (i.e., already known to be true) or 'non-ADRs'. For each drug-event association MEDLINE citations with co-occurrence of the drug and the adverse event of interest were extracted and manually reviewed by two independent evaluators with experience in pharmacovigilance and pharmacoepidemiology. For the list of ADRs, we considered only those with at least three independent PubMed citations that showed a potentially causal association. The lowest level of evidence accepted was 3 case reports mentioning the occurrence of a specific adverse event in individual patients exposed to the specific drug. The discrepancies in assessment were resolved through discussion. All the associations for which not enough evidence was found in the literature (i.e., <3 confirmatory articles) were considered 'non-ADRs'. For these, a random sample of 5% of drug-event associations for each of the five events of interest was manually reviewed by the two evaluators. In addition, if at least two drugs belonging to the same therapeutic class (defined by common first 5 digits from ATC code) had a positive association with a certain event, a class effect was suspected and additional manual review of the summary of product characteristics (SmPC) was done to see if other drugs in the class were also associated with that specific event. Figure 1 shows a schematic workflow of the verification process.

For purposes of this study, drug-event associations that have been suggested to be ADRs by the criteria described above are assumed to be ADRs, otherwise, these associations s are assumed to be non-ADRs.

Performance indicators

In order to assess the complementarity of the systems and to calculate the costs associated with identifying potential signals from different sources we used the following indicators:

Percentage of unilaterally identified signals- this is a variation of sensitivity (recall) metrics which uses as numerator the number of true associations identified in one system that were not identified in the other. We considered this variable useful in quantifying the incremental value of each system.

Sensitivity was calculated as the proportion of associations correctly identified by the method out of the total pool of ADRs known to be true from the literature [51].

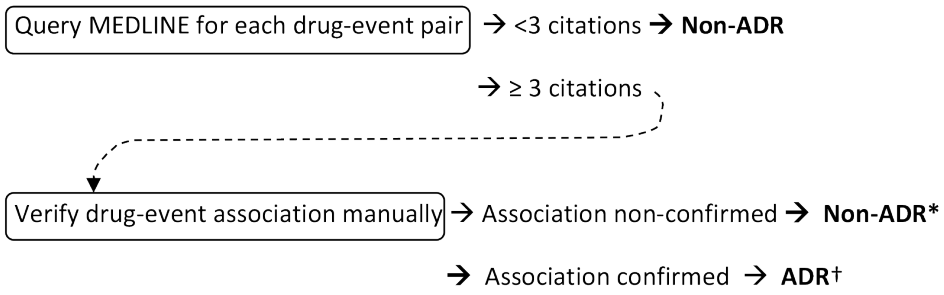


Figure 1: Workflow of verification of signals

* A 5% random sample was double-checked by a second reviewer
† If a class effect is suspected additional manual check is performed and inclusion as an ADR for other class members is performed

NND, originally described by Hauben et al. [52] in the context of signal detection within SRS, was used as a proxy to express the costs associated with each source of signals. This represents the number of signals that would have to be reviewed to detect a single signal that was proven to be true.

Spearman's correlation coefficient was calculated to determine whether there was a correlation between background incidence of the events and each system's capability to detect signals.

RESULTS

From an initial dataset of 5,536 drug-event associations overall, 1,490 (27%) potential signals were detected in either EudraVigilance or EU-ADR (Figure 2). Upon signal verification, the ratio of ADRs to non-ADRs varied from 1:6 for acute pancreatitis to 1:19 for hip fracture.

The therapeutic classes comprising the majority of potential signals identified in EudraVigilance were: agents acting on the renin-angiotensin system, antivirals for systemic use and antithrombotic agents, while for EU-ADR these were anti-asthmatics, psychoanaleptics and antiepileptics (see Figure 2). The percentage of potential signals identified in both systems ranged from 2% to 24%.

The median sensitivity for detecting signals across all events in EudraVigilance was 42% (range 20%–71%) and for EU-ADR 27% (range 23%–60%), with the values depending on the event of interest (see Figure 3). Acute myocardial infarction was the hardest to detect among all five events, with 65% of known AMI associations from literature not flagged in either database system. Hip fracture and bullous eruption seemed to be the easiest to identify overall, with 21% and 28% of known associations remaining undetected. From a system perspective, the most easily identified events

in SRS were: bullous eruption, acute pancreatitis and upper gastro-intestinal bleeding, while in EHR these were: hip fracture, UGIB and AMI.

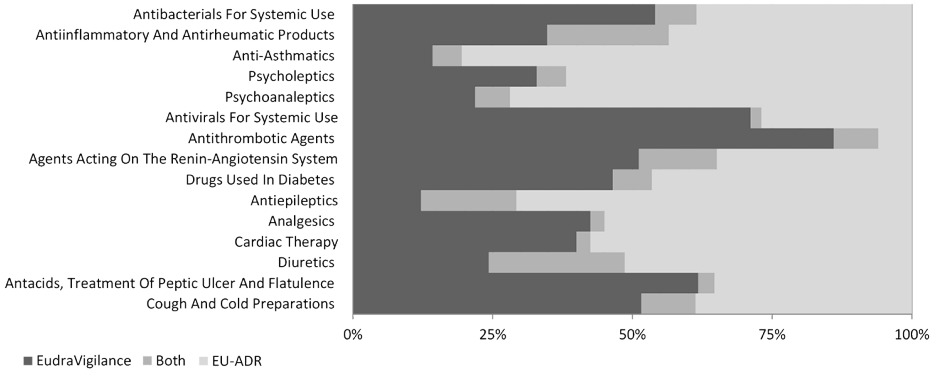


Figure 2: Distribution of potential signals in EudraVigilance or EU-ADR grouped by ATC therapeutic subgroup.

Note: only classes with >30 potential signals are shown

The median sensitivity for detecting signals across all events in EudraVigilance was 42% (range 20%–71%) and for EU-ADR 27% (range 23%–60%), with the values depending on the event of interest (see Figure 3). Acute myocardial infarction was the hardest to detect among all five events, with 65% of known AMI associations from literature not flagged in either database system. Hip fracture and bullous eruption seemed to be the easiest to identify overall, with 21% and 28% of known associations remaining undetected. From a system perspective, the most easily identified events in SRS were: bullous eruption, acute pancreatitis and upper gastro-intestinal bleeding, while in EHR these were: hip fracture, UGIB and AMI.

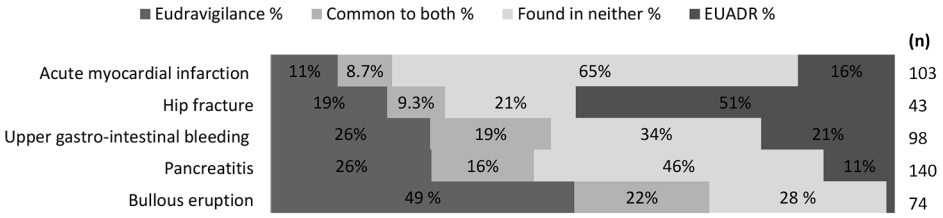


Figure 3: Contribution of each system to signal detection (i.e., % of 'positive' associations detected out of the total 'positive associations' pool in the literature)

n=total number of true associations in the dataset; found in neither= the association was not highlighted as a signal in any of the databases during the signal detection process; due to round-up the total sometimes surpasses 100%

The background incidence of the events, obtained from EU-ADR, was plotted against the percentage of unilaterally identified signals. The contribution of each database to signal detection appears to be correlated with the background incidence of the events, being positively although non-significantly correlated in EU-ADR (R=0.7, p=0.18) and inversely and significantly correlated in EudraVigilance (R=-1, p<0.01) (see Figure 4).

The associated costs were expressed as the number of signals that would need to undergo review and further investigation for one true safety issue to be identified. The costs associated with detecting signals, expressed by NND, were highly variable across events. With the exception of bullous eruption, it seems to be more 'costly' to detect safety signals in EU-ADR than in EudraVigilance, with a median NND across all events of 7 vs. 5. The most 'costly' event in EudraVigilance is bullous eruption (NND=8) and the least 'costly' are UGIB and acute pancreatitis (NND=2). In EU-ADR, the most costly signals to detect are those related to hip fracture (NND=9) and acute myocardial infarction (NND= 7) while the least costly are those related to pancreatitis and bullous eruption (NND=3), see Figure 5.

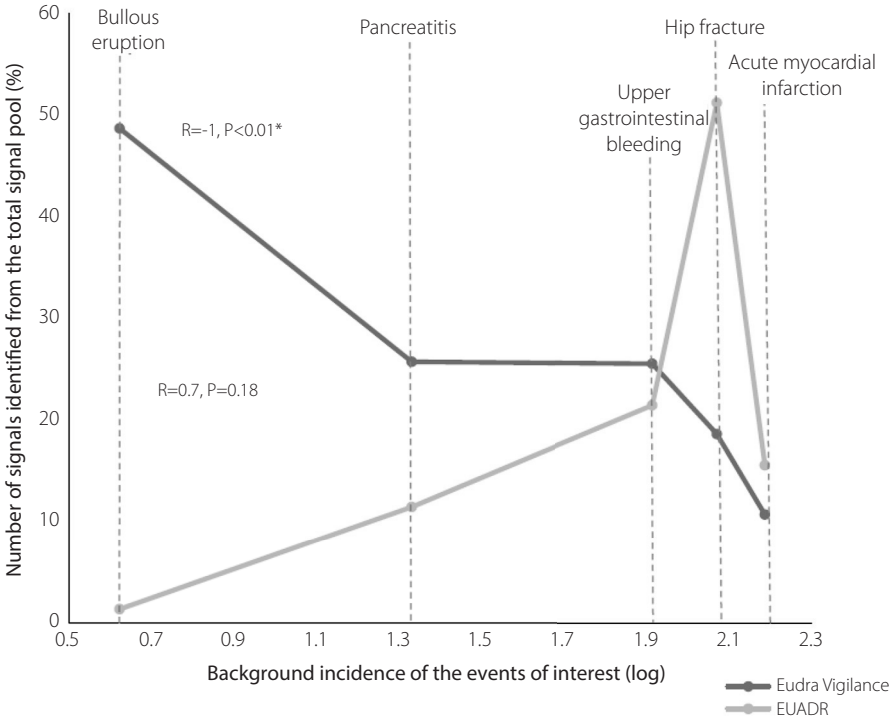


Figure 4: Correlation between background incidence of events and contribution of each system to signal detection

The background incidences of the events, estimated from EU-ADR data, pooled across all databases are (per 100,000 person-years): bullous eruption=4.2, pancreatitis=21.4, upper GI bleeding=82.2, hip fractures=117.7, acute myocardial infarction=153.7. Identified signals refer to signals proven to be known ADRs; R=Spearman's correlation coefficient

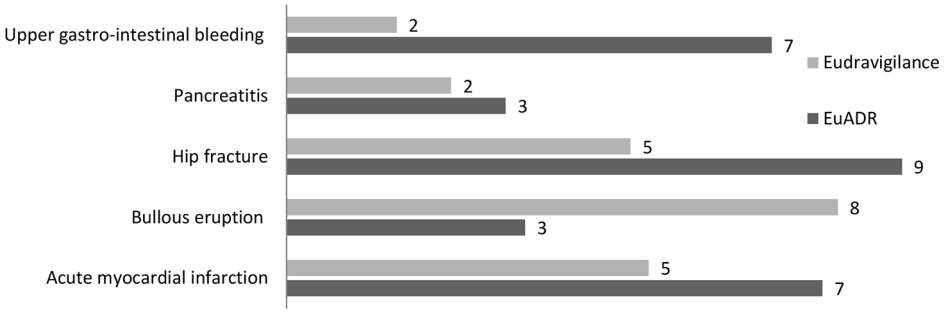


Figure 5: Number needed to detect a true association

Note: This represents the 'cost' per signal, for example in case of AMI detection in EudraVigilance, 5 signals need to be investigated in order to detect one true association.

DISCUSSION

In this study we aimed to investigate the additional value of EHR for signal detection on top of THE traditional spontaneous reporting system. We focused on five different adverse events deemed to be important in pharmacovigilance: bullous eruption, acute myocardial infarction, acute pancreatitis, hip fracture and upper gastrointestinal bleeding.

Although EudraVigilance identified more signals overall than EU-ADR, 41% (187/458) signals compared to 32% (147/458), as previously shown in other studies [16], this was not unexpected considering that EudraVigilance has a worldwide coverage, whereas EU-ADR covered only three countries. Furthermore, the EudraVigilance system is primarily designed for signal detection.

Patadia et al. [16] performed a time-restricted analysis (before and after a safety communication/ media attention) and showed that this impacts the numbers of signals detected in both data sources in opposite ways (i.e., increase in number of signals in SRS after media attention and decrease in EHR). While this effect might partially explain our findings of higher sensitivity for EudraVigilance, we consider it unlikely to entirely explain the difference since the majority of signals tested did not attract media attention.

The capacity of EU-ADR and EudraVigilance to detect signals was shown to differ depending on the nature of the adverse event investigated. The relatively poor performance of EudraVigilance in detecting hip fractures and AMI might be due to the fact that both events are not usually perceived as being drug-induced and thus often fail to be recognized and reported as ADRs, as previously hypothesized [34]. The suspected ADRs documented in a SRS like EudraVigilance are highly dependent on the reporter's ability to recognize them as such and some characteristics are helpful in this respect: biologic concordance with the drug mechanism of action, short time to onset, positive dechallenge, lack of alternative causes. The adverse events which are not so obviously attributed

to drugs (as they are multifactorial), or which already have a high background incidence, are likely to be poorly captured by spontaneous reports [1,3]. EHR systems do not rely on reporter judgment therefore these events may be better represented in such systems.

On the other hand, there is a very low sensitivity for detecting rare drug-induced events in EU-ADR and this is in line with previous research which showed that, despite the broad coverage of the EU-ADR network (around 20 million of patients) there is simply not enough (statistical) power to identify very rare events in the database [24]. Thus, for rare events that have drug treatment as primary etiology, SRS still seems to be the better solution so far. Our findings are consistent with those of Patadia et al., [34] although different parameters were used to determine the complementarity of the two systems.

We found a correlation between the background incidence of the events and the contribution of each database to signal detection; the correlation was statistically significant for EudraVigilance, but not for EU-ADR, which may be due to the low number of events tested.

The burden associated with screening any data source for signals depends on the number of signals that require further assessment or investigation and the workload involved in each of these investigations. The amount of work needed to confirm or refute a signal is highly variable, ranging from simple product information checks to performing more complex analyses and formal pharmacoepidemiologic studies. Pizzoglio et al. reported in their study a median time of 6 hours for initial assessment of a signal (range 2–26 hours) [35]. We did not collect similar information on time spent on assessment of the signals in our study since a semi-automatic method was used; however, we considered the number of signals which need further investigation to be a reasonable proxy to express the associated costs. We found that for all the events (with the exception of bullous eruption), it is more costly to identify signals within an EHR-based system. However there is a notable difference across events as follows: for hip fracture, where EU-ADR provided the most added value, 80% increase in cost per signal was observed compared to detection in EudraVigilance. On the other hand, for acute pancreatitis, the cost associated with signal detection in EU-ADR was not much higher compared to that in EudraVigilance; therefore, in this case it may be justifiable to use both systems. Because the two systems provided roughly the same contribution to detection of UGIB-related signals, the value of supplementing SRS with EHR-based systems is probably dependent on the types of drugs that can be captured in the particular EHR database. Detecting acute myocardial infarction proved to be costly in both systems; however, the seriousness of the event, its public health impact and the difficulty in detection might justify the extra cost of using both SRS and EHR. For bullous eruption, due to lack of additional gain (only one extra signal identified), it might not be efficient to use EHR as a secondary signal source.

The range of events tested in this study, albeit carefully selected, represents only a small sample of all possible adverse events and therefore limits the external validity of this research. Applicability of our findings to a broader range of events will require further investigation. In addition, the overall

background incidence of the events selected is slightly skewed towards more common events and, if our hypothesis holds true, this might have resulted in a bias favoring EHR performance.

Another limitation was the retrospective nature of the study: we actually tested the capacity of systems not to identify signals but rather already identified safety issues. As Noren et al. pointed out, ideally, the evaluation should be made using emerging safety issues and not well-established ADRs [36]. This is a common limitation in signal detection research, however, due to the difficulties that accompany building a 'reference standard' and long time needed to gather data prospectively.

Moreover, combining the two systems involved many decisions regarding the choice of signal detection method and their implementation (e.g., choice of thresholds and precision estimates) which have a LARGE impact on the results [37]. In addition, it was repeatedly demonstrated that the aggregation level at which data mining is performed influences the results [38,39] and we consider THAT the appropriate level is still an open question. The decision to use SMQs instead of preferred terms (PTs) might have an impact on the results. We acknowledge that use of different signal detection methodologies and reference standards as well as the definition of events used in data mining may lead to different results. Therefore, our results may reflect to some extent the relative strengths of each data source but they will also be influenced by the algorithms chosen. Although we acknowledge that time to signal detection is an important element to take into account, for this study we focused primarily on other quantitative measures such as number of signals and number of false positives which are equally important.

This evaluation did not take into account the fact that an EHR-based surveillance system would require additional work for implementation and subsequent maintenance for the purposes of signal detection (vs. SRS which are already established for such activities) and would thus incur extra costs which are difficult to estimate.

Our study is one of the few performed studies so far which tried to explore how a SRS and an EHR-based system might be used together with the aim of augmenting drug safety surveillance. A previous study by Harpaz et al. [12] had a similar aim but a different strategy, trying to combine information from both data sources at a very early stage in order to improve the ranking of signals by replication of findings. In contrast, we explored the scenarios where use of EHRs can fill the gaps and provide added value to the already existing systems.

CONCLUSION

The more prudent goal in signal detection is the identification of not all but the majority of signals in the most efficient way – with the least time and resource-consuming approach. With this aim in mind, we showed that EHR may complement SRS in certain situations, especially in the presence of adverse events with high background incidence. While SRS appeared to be more cost-effective overall, for some events which are very hard to pick up, the costs associated with additional signal detection in an EHR-based system may be justifiable.

REFERENCES

1 CIOMS Working Group. Practical aspects of signal detection in pharmacovigilance . Geneva; 2010.

2 Almenoff JS, Pattishall EN, Gibbs TG, DuMouchel W, Evans SJW, Yuen N. Novel Statistical Tools for Monitoring the Safety of Marketed Drugs. Clin. Pharmacol. 38 Ther. 2007;82:157–66.

3 Hauben M, Bate A. Decision support methods for the detection of adverse events in post-marketing data. Drug Discov. Today. 2009;14:343–57.

4 Hammad TA, Neyarapally GA, Iyasu S, Staffa JA, Dal Pan G. The future of population-based postmarket drug risk assessment: a regulator's perspective. Clin. Pharmacol. Ther. 2013;94:349–58.

5 Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. J. Clin. Epidemiol. 2005;58:323–37.

6 Stang PE, Ryan PB, Racoosin JA, Overhage JM, Hartzema AG, Reich C, et al. Advancing the science for active surveillance: rationale and design for the Observational Medical Outcomes Partnership. Ann. Intern. Med. 2010;153:600–6.

7 Observational Medical Outcomes Partnership . Available from: http://omop.fnih.org/ [accessed 22 Mar 2014]

8 PROTECT website. Available from: http://www.imi-protect.eu/ [accessed 22 Mar 2014].

9 Coloma PM, Schuemie MJ, Trifirò G, Gini R, Herings R, Hippisley-Cox J, et al. Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: the EU-ADR Project. Pharmacoepidemiol. Drug Saf. 2011;20:1–11.

10 EU-ADR webpage. Available from: http://www.euadr-project.org/ [accessed 22 Mar 2014].

11 Star K, Watson S, Sandberg L, Johansson J, Edwards IR. Longitudinal medical records as a complement to routine drug safety signal analysis. Pharmacoepidemiol. Drug Saf. 2015;

12 Harpaz R, Vilar S, Dumouchel W, Salmasian H, Haerian K, Shah NH, et al. Combing signals from spontaneous reports and electronic health records for detection of adverse drug reactions. J. Am. Med. Inform. Assoc. JAMIA. 2013;20:413–9.

13 Trifirò G, Patadia V, Schuemie MJ, Coloma PM, Gini R, Herings R, 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.

14 Kim J, Kim M, Ha J-H, Jang J, Hwang M, Lee BK, et al. Signal detection of methylphenidate by comparing a spontaneous reporting database with a claims database. Regul. Toxicol. Pharmacol. RTP. 2011;61:154–60.

15 Harpaz R, Dumouchel W, Lependu P, Bauer-Mehren A, Ryan P, Shah NH. Performance of Pharmacovigilance Signal-Detection Algorithms for the FDA Adverse Event Reporting System. Clin. Pharmacol. Ther. 2013;

16 Patadia VK, Schuemie MJ, Coloma P, Herings R, van der Lei J, Straus S, et al. Evaluating performance of electronic healthcare records and spontaneous reporting data in drug safety signal detection. Int. J. Clin. Pharm. 2014;

17 Trifirò G, Pariente A, Coloma PM, Kors JA, Polimeni G, Miremont-Salamé G, et al. Data mining on electronic health record databases for signal detection in pharmacovigilance: which events to monitor? Pharmacoepidemiol. Drug Saf. 2009;18:1176–84.

18 Annual Report on EudraVigilance for the European Parliament, the Council and the Commission, European Medicines Agency, 2014; available from http://www.ema.europa.eu/docs/en_GB/document_library/Report/2014/04/WC500165780.pdf; [accessed 22 Mar 2014].

19 MedDRA, Introductory Guide MedDRA Version 15.1, MSSO-DI-6003-15.1.0. 2012.

20 MedDRA-Introductory Guide for Standardised MedDRA Queries (SMQs) Version 15.0. 2012.

21 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;10:483–6.

22 Alvarez Y, Hidalgo A, Maignen F, Slattery J. Validation of statistical signal detection procedures in eudravigilance post-authorization data: a retrospective evaluation of the potential for earlier signalling. Drug Saf. Int. J. Med. Toxicol. Drug Exp. 2010;33:475–87.

23 Slattery J, Alvarez Y, Hidalgo A. Choosing Thresholds for Statistical Signal Detection with the Proportional Reporting Ratio. Drug Saf. Int. J. Med. Toxicol. Drug Exp. 2013;

24 Coloma PM, Trifirò G, Schuemie MJ, Gini R, Herings R, Hippisley-Cox J, et al. Electronic healthcare databases for active drug safety surveillance: is there enough leverage? Pharmacoepidemiol. Drug Saf. 2012;21:611–21.

25 Avillach P, Coloma PM, Gini R, Schuemie M, Mougin F, Dufour J-C, et al. Harmonization process for the identification of medical events in eight European healthcare databases: the experience from the EU-ADR project. J. Am. Med. Inform. Assoc. [Internet]. 2012 [cited 2013 Feb 12];

26 Valkhoff VE, Coloma PM, Masclee GMC, Gini R, Innocenti F, Lapi F, et al., EU-ADR Consortium. Validation study in four health-care databases: upper gastrointestinal bleeding misclassification affects precision but not magnitude of drug-related upper gastrointestinal bleeding risk. J. Clin. Epidemiol. 2014;67:921–31.

27 Coloma PM, Valkhoff VE, Mazzaglia G, Nielsson MS, Pedersen L, Molokhia M, et al., EU-ADR Consortium. Identification of acute myocardial infarction from electronic healthcare records using different disease coding systems: a validation study in three European countries. BMJ Open. 2013;3.

28 Schuemie MJ. Methods for drug safety signal detection in longitudinal observational databases: LGPS and LEOPARD. Pharmacoepidemiol. Drug Saf. 2011;20:292–9.

29 Schuemie MJ, Coloma PM, Straatman H, Herings RMC, Trifirò G, Matthews JN, et al. Using electronic health care records for drug safety signal detection: a comparative evaluation of statistical methods. Med. Care. 2012;50:890–7.

30 Bauer-Mehren A, van Mullingen EM, Avillach P, Carrascosa M del C, Garcia-Serna R, Piñero J, et al. Automatic Filtering and Substantiation of Drug Safety Signals. PLoS Comput Biol. 2012;8:e1002457.

31 Avillach P, Dufour J-C, Diallo G, Salvo F, Joubert M, Thiessard F, 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. JAMIA. 2012;

32 Zhou X-H, Obuchowski NA, McClish DK. Statistical Methods in Diagnostic Medicine. New York, NY: John Wiley; 2002

33 Hauben M, Vogel U, Maignen F. Number Needed to Detect. Pharm. Med. 2012;22:13–22.

34 Patadia VK, Coloma P, Schuemie MJ, Herings R, Gini R, Mazzaglia G, et al., EU-ADR consortium. Using real-world healthcare data for pharmacovigilance signal detection - the experience of the EU-ADR project. Expert Rev. Clin. Pharmacol. 2015;8:95–102.

35 Pizzoglio V, Ahmed I, Auriche P, Tuber-Bitter P, Haramburu F, Kreft-Jaïs C, et al. with French Network of Pharmacovigilance Centres. Implementation of an automated signal detection method in the French pharmacovigilance database: a feasibility study. Eur. J. Clin. Pharmacol. 2012;68:793–9.

36 Norén GN, Caster O, Juhlin K, Lindquist M. Zoo or savannah? Choice of training ground for evidence-based pharmacovigilance. Drug Saf. 2014;37:655–9.

37 Bate A, Evans SJW. Quantitative signal detection using spontaneous ADR reporting. Pharmacoepidemiol. Drug Saf. 2009;18:427–36.

38 Pearson RK, Hauben M, Goldsmith DI, Gould AL, Madigan D, O'Hara DJ, et al. Influence of the MedDRA hierarchy on pharmacovigilance data mining results. Int. J. Med. Inf. 2009;78:e97–103.

39 Hill R., Hopstadius J., Lerch M., Noren N., An attempt to expedite signal detection by grouping related adverse reaction terms, 24th European Medical Informatics Conference (MIE 2012), August 2012, Pisa; Drug Saf 2012; 35 (12)

3.2

Association of ischemic colitis with triptans treatment – a nested case-control study in United Kingdom, using The Health Improvement Network primary care data

Alexandra C Păcurariu
Remy DC Francisca
Swabra Nakato
Miriam Sturkenboom
Maria de Ridder
Sabine M Straus

Submitted

ABSTRACT

Background and Objectives: Ischemic colitis is the most common form of intestinal ischemia. In addition to other aetiologies, Ischemic colitis can also be caused by drugs such as alosetron and cocaine. An association between the use of triptans and Ischemic colitis has been hypothesised but not yet proven. The objective of this study was to estimate the risk of ischemic colitis during triptans' use in a migraine cohort.

Methods: This is a population based case control study using an UK primary care database. Incident ischemic colitis cases ≥ 18 years of age were matched to controls within the same migraine cohort. Drug use was assessed in the last 24 months before event and odds ratios were calculated by conditional logistic regression, adjusting for known confounders from literature.

Results: Within the migraine cohort of 293,037 patients, we identified 41 incident cases of ischemic colitis. Use of triptans in the previous 12 months increased the risk of ischemic colitis compared to no use (OR=2.29, 95%CI 1.02–5.15). When we split the exposure window in current and past use, the results became non-significant; the risk was highest for the past use category (use 3–12 months before event) (OR=2.57, 95%CI 0.88–7.54). After 12 months, the risk decreased to (OR=1.90, 95% 0.44–8.13).

Conclusions: We observed that triptan use 12 month before the event in a cohort of migraine patients, increases the risk of ischemic colitis. The highest risk appears to occur in the period 12 to 3 months prior to event. Considering the wide confidence interval and the limited number of cases included we were not able to further explore the risk windows. We consider that this finding should not be considered definitive and should be replicated in a more powered study.

INTRODUCTION

Triptans are selective serotoninergic receptor (5-$HT_{1B/1D}$) agonists that have been approved since 1991 for acute treatment of migraine [1]. Although triptans are rather selective for the vascular receptors in the nervous system, they may also cause vascular constriction outside the cerebral vascular bed [2]. Triptans have been reported to induce myocardial ischemia and stroke, albeit in very rare instances [3,4]. In addition, it has been suggested that triptans might also cause vasoconstriction in the splanchnic circulation, since the gastrointestinal cavity is known to have a strong serotoninergic signalling [5]. A few anecdotal reports of ischemic colitis (IC) following use of triptans have been reported [6-8]. All triptans' package inserts mention that colonic ischemia with abdominal pain and bloody diarrhoea has been reported in the post-marketing setting.

IC is the most common form of intestinal ischemia. It usually results from an interruption of colonic blood flow due to low flow states or thromboembolic events [9]. Classic symptoms include lower abdominal pain, diarrhoea and rectal bleeding in patients without a history of inflammatory bowel disorders. The estimated incidence of IC ranges from 4.5 to 44 cases per 100,000 person-years, depending on underlying comorbidities [10,11] and abruptly increases with age. IC is a serious condition, that may lead to bowel infarction, necrosis and may even be fatal [12]. The most common risk factors are severe hypotension, hypercoagulable states, mechanical bowel obstruction, abdominal surgery, inflammatory bowel disease, ischaemic heart diseases, cancer and severe constipation [11-13]. Some medications may also cause IC through vasospasm or vasoconstriction, thrombogenesis or through shunting of blood from mesenteric vessels [14]. For several drugs a possible association with IC was established, for example alosetron, cocaine, ergotamine, opioids, estrogens, taxanes and vinca alkaloids [14,15]. Alosetron, a 5-HT3 antagonist used for the management of irritable bowel syndrome was even withdrawn from the market for this reason [16].

To our knowledge, the association between triptans and IC was investigated only in two studies so far: a case series [6] of seven patients and a case-control study in a USA claims database [15]. The case series has suggested a possible association between triptans and IC and recommended further investigation. The case control study did not find an association, however it included only a very small group of triptans users (16 triptan users out of which 2 IC cases) and only hospital cases.

OBJECTIVE

Given the lack of good quantitative data from general practice, the aim of this study was to assess the risk of IC associated with use of triptans in migraine patients.

METHODS

Data source

We used The Health Improvement Network (THIN) as data source. THIN is a population-based electronic health care records database with data from 562 general practices all over the United Kingdom [17,18]. In the UK, all patients are registered to a general practitioner, who acts as a gatekeeper for secondary care. THIN reflects UK general population and comprise diagnostics from general practitioners, as well as discharge letters, prescriptions outside hospital and some lifestyle related information. It has been demonstrated that the clinical information in THIN is sufficiently accurate for use in epidemiological studies [18]. Investigators had access to the full version of THIN database when creating the study population. The study was approved by IMS Health Committee (SRC Number: 16THIN083).

Design, study population and follow-up

We conducted a case-control study nested in a cohort of patients that were diagnosed with migraine. The migraine cohort consisted of all patients aged ≥18 years who had at least one year of valid data in the database (run-in period) and at least one record of migraine or cluster headache diagnosis. The study period started from 1st January 2003 and ended 31st December 2015. Patients entered the migraine cohort upon the latest of the following dates: start of study period, fulfilling one year of database history, reaching 18 years of age or a diagnosis of migraine. The run-in period was used to check the patient's medical history and to distinguish between incident and prevalent cases of IC. Follow-up ended upon end of study period, transfer out of the practice or diagnosis of IC, whichever date was the earliest.

Cases and controls

The outcome in this study was incident ischemic colitis, which was identified by READ codes. The date of the first recorded IC diagnosis date was used as the index date. Controls were cohort members without a diagnosis of IC prior or at the index date of the case. For each case, a maximum number of 100 controls were drawn from the cohort using the incidence density sampling method [19]. Controls were matched to cases on age (±1 year) and sex. The following exclusion criteria were applied to both cases and controls:

- subjects with a diagnosis of colorectal cancer or inflammatory bowel disease prior to the index date,
- prevalent cases of IC (first diagnosis before cohort entry),
- prevalent users of triptans (first prescription issued before cohort entry),
- one of the following acute conditions recorded within 3 months prior to the index date: acute pancreatitis, sepsis, cardiovascular or hemorrhagic shock, abdominal or aortic surgery, infectious colitis and acute deep vein thrombosis.

Drug exposure

Exposure to triptans was assessed based on prescription drug codes. Cases and controls were classified as triptan users or non-users in the mutually exclusive exposure categories to investigate the hazard shape:

- Current use: any prescription received 3 months before the index date
- Past use: any prescription received > 3 months and <=12 months before the index date
- Distant past use: any prescription received >12 and <=24 months before the index date
- No use (reference): no prescription received in the last 24 months (see Figure 1).

In order to increase the power we aggregated current and past use in one category.

Covariates

The following covariates were considered as risk factors for IC: smoking status (categorised as current smoker, non-smoker and smoking unknown), constipation, hypertension, diabetes type II, ischemic cardiovascular disease, atrial fibrillation, pulmonary embolism, deep vein thrombosis, treatment with opioids, oestrogens or ergotamine [11,12,20]. If one of the diagnoses was recorded within one year before the index date, the patient was classified as having the disease. Study subjects were classified as exposed to drugs if prescriptions were recorded in the one year prior to the index date. All covariates were identified through structured diagnostic and drug codes.

Statistical analysis

The incidence rate of IC in the migraine cohort was calculated by dividing the number of incident cases by the total number of person-years at risk.

The differences of various characteristics between cases and controls were determined by ANOVA for continuous variables and by chi-square or Fisher's exact test for categorical variables.

To estimate the relation between triptan prescriptions and the risk of IC, we estimated the odds ratios using conditional logistic regression. Identification of confounders was performed by a backward selection procedure; confounders were kept in the model if the risk estimate for drug exposure changed more than 10% or if they improved the fit of the model significantly [21]. All analyses were performed using SAS version 9.4.

We conducted a sensitivity analyses extended definition of IC in an attempt to capture more cases and to check the robustness of our estimates.

RESULTS

The initial migraine cohort consisted of 293,037 patients, with a mean age at entry in the cohort of 43.6 years (IQR range 31–54), and 74.3% females (see Table 1). The cohort had a median follow-up of 6.7 years≤ (range: 2.8–10.9 years). Within this cohort, we identified 42 incident cases of IC, which led to an overall incidence rate for IC of 2.3/100,000 person-years. More than 70% of cases occured in the >60 years age category. The percentage of triptan users in the migraine cohort was 32% (94,256 users). Among triptan users, sumatriptan was the most commonly used (62%), followed by rizatriptan (14%) and zolmitriptan (10%).

After exclusion of patients with comorbidities, and exclusion of prevalent triptan users, the remaining 41 cases were matched to 4,005 controls (see Figure 1).

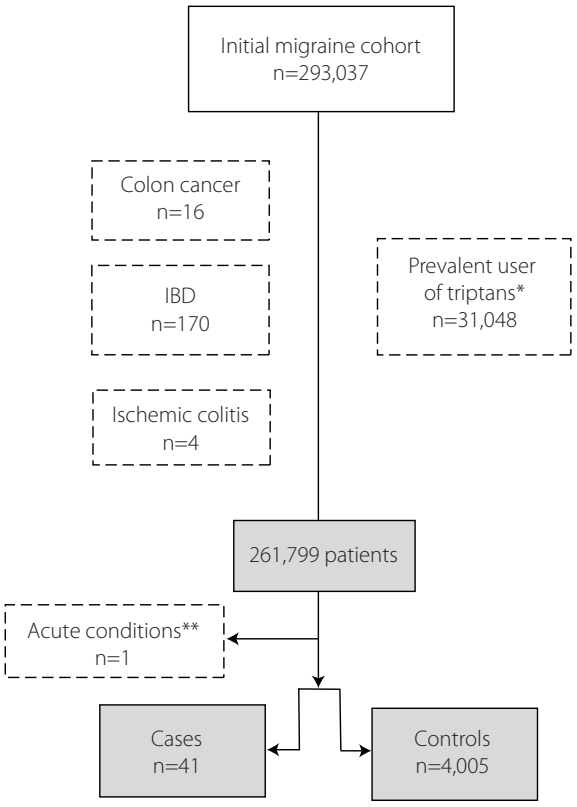


Figure 1: Flowchart for patient's recruitment

* Patients with at least one triptan prescription issued during run-in period; ** Patients with at least one of the following acute conditions recorded in the past 3 months prior to the date of IC for cases or matched index date for the controls: acute pancreatitis, sepsis, cardiovascular or hemorrhagic shock, abdominal or aortic surgery, infectious colitis and acute deep vein thrombosis; IBD=irritable bowel syndrome

Table 1: Characteristics of the initial migraine cohort

Variable	n=293,037
Gender (Female, (%))	74.3%
Age (years) (mean, IQR range)	43.6 (31–54)
≤40 years	45.3%
>40 and <60 years	37.7%
≥60 years	16.9%

The main characteristics of cases and matched controls at index date are presented in Table 2. Being a current smoker increased the risk of ischemic colitis (OR=2.67 (95%CI 1.07–6.67)), as well as being treated with opioids (OR=3.17 (95%CI 1.60–6.28)). The number of patients treated with ergotamine was very low (less than 5 patients) therefore this variable was not further considered in the analysis.

Table 2: Characteristics of cases and matched controls at index date

	Cases	Controls	p-value*
	(n=41)	(n=4,005)	
Gender (Female, (%))	33 (80.49)	3,205 (80.02)	0.941
Age (years) (median, range)	59 (32–95)	59 (31–96)	0.693
BMI (mean ± sd)	26.34±7.33	26.52±5.62	0.836
Smoking status			0.459
Current smoker	9 (21.95)	432 (10.79)	
Non-smoker	12 (29.27)	1,500 (37.45)	
Unknown	20 (48.78)	2,071 (51.71)	
Co-morbidities (n, (%))			
Constipation	13 (31.71)	828 (20.67)	0.083
Hypertension	3 (7.32)	188 (4.69)	0.442
Diabetes type II	2 (4.88)	376 (9.39)	0.427
Ischemic cardiovascular disease	4 (9.76)	201 (5.02)	0.152
Atrial fibrillation	6 (14.63)	286 (7.14)	0.117
Deep vein thrombosis	0 (0)	82 (2.05)	1.000
Triptans use			0.078
Current use	4 (9.75)	241 (6.01)	
Past use	4 (9.75)	166 (4.14)	
Distant past use	2 (4.87)	128 (3.2)	
Co-medication use (n, (%))			
Oestrogens	12 (29.27)	923 (23.05)	0.347
Opioids	27 (65.85)	1623 (40.52)	**0.001**
Duration of follow up in days (median, range)	2,352 (33–4,724)	3,073 (12–4,724)	0.284

*p-values for differences between cases and controls determined by ANOVA and chi-square test or Fisher exact test for categorical covariates where appropriate; ** Bold font indicates significant differences; BMI=Body Mass Index

In the unadjusted analysis, the risk of IC was not increased with current use of triptans (OR=1.97 (95% CI 0.68-5.76)), past use (OR=2.86 (95%CI 0.98–8.31)) or distant past use categories (OR=1.85 (95%CI 0.43–7.87)). Aggregation of current and past use resulted in an OR of 2.42 (95%CI 1.08–5.38).

Adjustment for the selected confounders (smoking status, opioid use and body mass index (BMI)) led to a small decrease in risks for all exposure categories (see Table 3). Aggregation of current and past use resulted in an OR of 2.29 (95%CI 1.02.–5.15). If we used a broader definition of IC, the unadjusted estimate dectreased to 1.74 (95%CI 0.77–3.89).

Table 3: Risk of IC with use of triptans

Triptans exposure	Number of cases/ controls	OR (95%CI)	OR adj* (95%CI)
No use	31/3470	*Reference*	*Reference*
Current+past use	8/407	2.42 (1.08.–5.38)	2.29 (1.02.–5.15)
Current use <3 months	4/241	1.97 (0.68–5.76)	1.94 (0.66–5.72)
Past use (>3 and <=12 months)	4/166	2.86 (0.98–8.31)	2.57 (0.88–7.54)
Distant past use (>12 and <=24 months)	2/128	1.85 (0.43–7.87)	1.90 (0.44–8.13)

*adjusted for smoking status, opioid use and BMI

DISCUSSION

In this case control study nested in a cohort of migraine patients we have observed that triptans' use in the previous year increases the risk of IC. The data suggest that the risk is higher within 3 to 12 months prior to diagnosis, decreasing for more distant exposures. The results were borderline significant.

Strengths of the current study include the population based setting, ensures the allows for identification of all potential IC cases in the population. The controls were derived from the same source population. By nesting our study in a migraine cohort and matching we reduced the potential (un)measurable confounding by indication.

To our knowledge this is the second study that has evaluated this association and the first one conducted in an European database. The previous case control study was performed in an US claims database (Kaiser Permanente Medical Care Plan) [13], and did not find an association (the exact estimate was not provided). However this study was not focused specifically on triptans as the authors investigated a wide range of diseases and drugs associated with IC. They also had limited power due to a low number of cases (16 triptans' users out of which 2 cases).

Our study also has several limitations. Firstly, the lack of power due to limited number of cases, which restricted us to further investigate exposure windows. As in any observational study, the

other limitations can be classified in: misclassification of outcome, misclassification of exposure and confounding.

Misclassification of outcome

The incidence of IC in the literature is 4.5 to 44 cases per 100,000 person-years and varies almost 100 fold with age, with one study reporting annual incidence rates ranging from 1.1 per 100,000 among those under age 40 years to 107 per 100,000 among those aged 80 years or older [11]. Our calculated incidence was lower than the estimates from the literature; (2.3 per 100,000 person-years) and this might be due to several reasons. The most logical reason is age, the migraine cohort is younger than the populations from other studies which investigated IC; whereas we have a mean age of 46.3, other studies have a mean age of 71.6 [11] or 69 years [15]. Secondly, IC is a diagnosis with unspecific symptoms (abdominal pain, diarrhea and rectal bleeding) and no specific diagnostic test, therefore it can be under-diagnosed by specialists. This is more likely for chronic IC, since acute IC is life-threatening and therefore much less likely to be underdiagnosed. Thirdly, to guard against case misclassification, we used a rather strict definition of IC and we might have missed some cases. However, we performed a sensitivity analysis with a broader diagnosis definition in an attempt to capture more cases. This led to a decrease in the estimate, suggesting that, if the association between triptans and ischemic colitis is real, the extended definition included false positive cases and should not be used. We did not validate the cases by chart review and we are not aware of any study which used a validated definition of IC.

Misclassification of exposure

It is known that the choice of the exposure risk window has a big impact on the risk estimate since risks vary over time. To account for prevalent users' bias, we excluded prevalent triptan users at cohort entry. We did not have any indication from literature about the exposure risk window for triptans and IC, therefore we used multiple risk windows. We assumed that an effect of triptans on gastrointestinal vasculature would have an acute or medium latency time (similar to the effects on other vasculature). However prescription records are inadequate to estimate accurately the actual timing of use since this class of drugs is used as needed, therefore we assumed there could be quite a lag time between prescription time and time of actual intake. In this case, misclassification of exposure is very likely, though non-differential between groups. To address this we investigated multiple risk windows: current, past and distant past use and then aggregated across the categories with similar risks.

Confounding

The potential confounders were identified from literature and other studies on ischemic colitis and they are quite diverse, including comorbidities, concomitant drugs and lifestyle-related factors such as smoking. With regards to the concomitant drugs, we decided not to adjust for NSAIDs since the evidence of them being a risk factor is weak and most NSAIDs are also prescribed over the counter and are therefore not captured in the database, so we would adjust for a highly misclassified

confounder [14]. We did adjust for smoking which is a known risk factor for the outcome, however we also know that smoking is misclassified in THIN, leaving some residual confounding.

In spite of potential misclassification of exposure, the finding of a significant increased risk of ischemic colitis following use of triptans would warrant further studies. Considering the wide confidence interval and the limited number of cases that we were able to include we consider that this finding should not be considered definitive and should be replicated and studied in a more powered study.

REFERENCES

1 Tepper SJ, Rapoport AM, Sheftell FD. Mechanisms of Action of the 5-HT1B/1D Receptor Agonists. Arch Neurol. 2002 Jul 1;59(7):1084–8.

2 MacIntyre PD, Bhargava B, Hogg KJ, Gemmill JD, Hillis WS. Effect of subcutaneous sumatriptan, a selective 5HT1 agonist, on the systemic, pulmonary, and coronary circulation. Circulation. 1993 Feb;87(2):401–5

3 Roberto G, Raschi E, Piccinni C, Conti V, Vignatelli L, D'Alessandro R, et al. Adverse cardiovascular events associated with triptans and ergotamines for treatment of migraine: systematic review of observational studies. Cephalalgia Int J Headache. 2015 Feb;35(2):118–31.

4 Dodick D, Lipton RB, Martin V, Papademetriou V, Rosamond W, MaassenVanDenBrink A, et al. Consensus Statement: Cardiovascular Safety Profile of Triptans (5-HT1B/1D Agonists) in the Acute Treatment of Migraine. Headache J Head Face Pain. 2004 May 1;44(5):414–25.

5 Mawe GM, Hoffman JM. Serotonin signalling in the gut--functions, dysfunctions and therapeutic targets. Nat Rev Gastroenterol Hepatol. 2013 Aug;10(8):473–86.

6 Nguyen TQ, Lewis JH. Sumatriptan-associated ischemic colitis: case report and review of the literature and FAERS. Drug Saf. 2014 Feb;37(2):109–21.

7 Alkhatib AA, Gangotena F, Peterson KA. Rizatriptan induced acute on top of chronic ischemic colitis. Am J Gastroenterol. 2009 Oct;104(10):2643–4.

8 Schwartz DC, Smith DJ. Colonic ischemia associated with naratriptan use. J Clin Gastroenterol. 2004 Oct;38(9):790–2.

9 Brandt LJ, Feuerstadt P, Blaszka MC. Anatomic patterns, patient characteristics, and clinical outcomes in ischemic colitis: a study of 313 cases supported by histology. Am J Gastroenterol. 2010 Oct;105(10):2245–2252; quiz 2253.

10 Higgins PDR, Davis KJ, Laine L. Systematic review: the epidemiology of ischaemic colitis. Aliment Pharmacol Ther. 2004 Apr 1;19(7):729–38.

11 Yadav S, Dave M, Edakkanambeth Varayil J, Harmsen WS, Tremaine WJ, Zinsmeister AR, et al. A population-based study of incidence, risk factors, clinical spectrum, and outcomes of ischemic colitis. Clin Gastroenterol Hepatol Off Clin Pract J Am Gastroenterol Assoc. 2015 Apr;13(4):731–738.e1–6; quiz e41.

12 Longstreth GF, Yao JF. Epidemiology, clinical features, high-risk factors, and outcome of acute large bowel ischemia. Clin Gastroenterol Hepatol Off Clin Pract J Am Gastroenterol Assoc. 2009 Oct;7(10):1075–1080.

13 Moszkowicz D, Mariani A, Trésallet C, Menegaux F. Ischemic colitis: the ABCs of diagnosis and surgical management. J Visc Surg. 2013 Feb;150(1):19–28.

14 Cappell MS. Colonic toxicity of administered drugs and chemicals. Am J Gastroenterol. 2004 Jun;99(6):1175–90.

15 Longstreth GF, Yao JF. Diseases and drugs that increase risk of acute large bowel ischemia. Clin Gastroenterol Hepatol Off Clin Pract J Am Gastroenterol Assoc. 2010 Jan;8(1):49–54.

16 Chang L, Chey WD, Harris L, Olden K, Surawicz C, Schoenfeld P. Incidence of ischemic colitis and serious complications of constipation among patients using alosetron: systematic review of clinical trials and post-marketing surveillance data. Am J Gastroenterol. 2006 May;101(5):1069–79.

17 Lo Re V, Haynes K, Forde KA, Localio AR, Schinnar R, Lewis JD. Validity of The Health Improvement Network (THIN) for epidemiologic studies of hepatitis C virus infection. Pharmacoepidemiol Drug Saf. 2009 Sep;18(9):807–14.

18 Khan NF, Harrison SE, Rose PW. Validity of diagnostic coding within the General Practice Research Database: a systematic review. Br J Gen Pract J R Coll Gen Pract. 2010 Mar;60(572):e128-136.

19 Richardson DB. An incidence density sampling program for nested case-control analyses. Occup Environ Med. 2004 Dec;61(12):e59.

20 Mosele M, Cardin F, Inelmen EM, Coin A, Perissinotto E, Sergi G, et al. Ischemic colitis in the elderly: predictors of the disease and prognostic factors to negative outcome. Scand J Gastroenterol. 2010 Apr;45(4):428–33.

21 Hegewald. A Backwards-Manual Selection Macro for Binary Logistic Regression in the SAS [Internet]. [cited 2017 Feb 23]. Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.553.7876&rep=rep1&type=pdf

Chapter 4

Exploring ways of improving signal detection methods

4.1

Drug safety monitoring in children: performance of signal detection algorithms and impact of age stratification

Osemeke U Osokogu
Caitlin Dodd
Alexandra Păcurariu
Florentia Kaguelidou
Daniel Weibel
Miriam CJM Sturkenboom

Drug Saf. 2016 Sep;39(9):873-81. doi: 10.1007/s40264-016-0433-x.

ABSTRACT

Background and Objectives: Spontaneous reports of suspected adverse drug reactions (ADRs) can be analysed to yield additional drug safety evidence for the pediatric population. Signal detection algorithms (SDAs) are required however the performance of SDAs in the pediatric population specifically is unknown.

Methods: We tested the performance of two established SDAs: Proportional Reporting Ratio (PRR) and Empirical Bayes Geometric Mean (EBGM) on a pediatric dataset from USA. We compared SDAs' performance to a published pediatric-specific reference set, by calculating diagnostic-test related statistics as the area under the Receiver Operating Characteristics curve (AUC). The impact of age stratification and age-adjustment SDAs' performance was assessed. Age adjustment was performed by pooling (Mantel-Hanszel) stratum-specific estimates.

Results: A total of 115,674 pediatric reports (patients aged 0–18 years) comprising 893,587 drug-event combinations were analysed. Crude values of the AUC were similar for both SDAs: 0.731 (PRR) and 0.745 (EBGM). Stratification unmasked four DECs, for example 'ibuprofen and thrombocytopenia', which prove to be real signals. Age-adjustment did not improve performance.

Conclusion: The performance of the two tested SDAs was similar in the pediatric population. Age adjustment does not improve performance and is therefore not recommended to be performed routinely. Stratification can reveal new associations, therefore is recommended when either drug use is age-specific or when an age-specific risk is suspected.

INTRODUCTION

Spontaneous reports of suspected adverse drug reactions (ADRs) can yield important information regarding the safety of drugs [1]. Usually, such reports are screened for emerging safety issues by applying statistical methods called signal detection algorithms (SDAs). Current SDAs compare the reporting rate of a drug-event combination (DEC) of interest with the expected count calculated from the overall reporting rate of that reaction in the entire database [1,2]. Although SDAs are routinely applied to reports pertaining to the general population, the performance of SDAs in the pediatric population specifically has not been investigated to date. Compared to adults, the pattern of drug use and occurrence of ADRs in pediatrics may differ [3-5] since the latter population comprises a heterogeneous group of subjects at various stages of development with age-dependent organ maturation and hormonal changes [6]. Several studies investigating ADR reporting in children identified different reporting patterns in this population compared to adults [3,5,7,8]. Since ADRs may be age-specific, adjustment for age seems to be a logical step when investigating pediatric ADRs and has been advocated by some researchers [4]. The major aim of stratification is verification of confounding and effect modification which otherwise may mask true signals [9]. Confounding by age can be dealt with by stratifying for age categories and pooling stratum-specific estimates. However if age specific estimates differ (in case of effect modification) pooling/adjustment should not be done, but instead, a verification of each individual stratum. While stratification has been investigated by some researchers [10], adjustment is routinely implemented in some Bayesian but not in frequentist SDAs [11-13]. Few studies have systematically addressed the impact of age stratification or adjustment and the results are contradictory [9,14,15].

Within the context of the Global Research in Pediatrics (GRiP) Network of excellence [16], we aimed to evaluate the performance of two well-established SDAs in the pediatric population and determine if age stratification or adjustment impacts signal detection in this population.

METHODS

Data source

Data was retrieved from the publicly available version of the US FDA Adverse Event Reporting System (FAERS), which comprises spontaneous reports of suspected ADRs submitted by manufacturers, healthcare professionals and patients. FAERS is one of the largest repositories of spontaneous reports in the world [17,18]. In this study, we analyzed reports received from the first quarter of 2004 through the third quarter of 2012.

For performance analysis, only reports of ADRs occurring in children and adolescents (<18 years of age) were retained. The ADRs in FAERS are coded according to the Medical Dictionary for Regulatory Activities (MedDRA®) [19].

To improve the quality of the dataset, we excluded reports with missing age, the main variable in our study. Also, reports with reported age equal to zero and with a MedDRA® preferred term indicating prenatal exposure were removed, as these imply *in-utero* drug exposure and were therefore not relevant for our study. We minimised the number of duplicates (i.e. the same report submitted by different reporters) by applying an algorithm based on case identifier, report identifier, drug and event names. For multiple reports (i.e. the same report is reported at a later time, with additional and updated information) [20], the most recent (and most updated) report was retained for analysis.

As drug names included in FAERS are not standardized, a harmonization procedure was implemented. Briefly, this consisted of removing superfluous characters and applying a generalized edit distance matching algorithm [21] to map free text drug names to synonyms and finally to the corresponding active substance and World Health Organization-Anatomic Therapeutic Chemical (WHO-ATC) code.

In this study, only those drugs reported as primary or secondary suspect in the FAERS database were retained for analysis. Analysis was performed at Drug-Event Combination (DEC) level, meaning that within each report, every suspect drug was combined with all reported ADRs. Thus, one report may comprise more than one DEC.

Signal detection algorithms (SDAs)

We tested two well established SDAs which are routinely used by various national and international regulatory and/or research institutions for signal detection: the proportional reporting ratio (PRR) [2] and the empirical Bayes geometric mean (EBGM) [13] (see Table 1). We also tested count of reports, as a positive control. In order to define a signal of disproportionate reporting (SDR) [22,23], we selected thresholds that are currently applied in routine practice. We applied the SDAs at the end of the study period, when the maximum number of reports had accrued.

Table 1: Signal detection algorithms and corresponding thresholds applied

Signal Detection Algorithm	Applied Threshold[a]	Institution where the method is currently used
Number of reports	n≥5	NA
PRR	PRR lower bound 95%CI≥1 & n≥5 reports	European Medicines Agency
EBGM	EB05 CI≥1.8 and n≥3 reports & EBGM ≥2.5	Medicines and Healthcare products Regulatory Agency (MHRA)

PRR=Proportional reporting ratio; EBGM=Empirical Bayes Geometric Mean; CI=confidence interval; NA=Not available; EB05=Lower bound of the 95% confidence interval; [a] Thresholds were obtained from Candore et al. [23]

Performance assessment measures

The performance of the SDAs was assessed by calculating diagnostic-test related statistics, namely specificity, sensitivity, positive predictive value (PPV), and negative predictive value (NPV) [24,25]. Sensitivity is the ability of the method to correctly identify true signals while specificity is the ability

to correctly exclude false signals. PPV and NPV are posterior probabilities, describing how many of the signals classified as positive or negative are correctly classified [24,25].

Since diagnostic-test related statistics are dependent on the threshold choice, their individual comparison has only limited, albeit practical value. Therefore, we also estimated the area under the curve (AUC) of receiver operating characteristics in order to compare the performance of the SDAs [32]; the AUC incorporates both sensitivity and specificity across all the possible values for a certain SDA. Calculation of AUCs was conducted by varying only the point estimate of each SDA and did not take into account the other components of the SDA.

For the purpose of performance evaluation, a previously constructed pediatric-specific reference set of positive and negative drug-event associations was used [26]. It consists of 37 positive and 90 negative DECs and includes drugs that are administered to children and events that are regarded as important for this population. The positive DECs are those that were confirmed to occur based on evidence from product information and the published literature, while the negative DECs are those that could not be confirmed at the time of literature review by neither the SmPC nor the published literature. For a full description of the reference set, see Osokogu et al. [26].

Stratification and adjustment for age

The impact of age stratification and adjustment on the performance of the SDAs was investigated. First, we checked for possible effect modification across age strata, by stratifying the data according to age categories defined according to International Conference on Harmonization (ICH) [27] and calculating stratum-specific measures for each SDA.

Secondly, we calculated age-adjusted estimates for PRR and EBGM by combining the stratum-specific estimates in an overall measure [28]. The performance of each SDA was reassessed after age adjustment.

Statistical analysis

Differences in the performance (AUC) of each SDA, crude versus age-adjusted and crude versus count of reports (positive control) were tested using paired chi-squared tests. Stratum-specific contingency tables were tested for homogeneity using the Breslow Day Tarone test [29]. The Mantel-Haenszel approach was used for pooling and calculating age-adjusted estimates [28]. The lower bound of the EBGM 95% confidence interval (EBGM05) was calculated using the EB05 for each stratum and then computing a Mantel-Haenszel average based upon Zeinoun [30]. Statistical significance was defined by p value <0.05.

Analysis was performed using SAS software version 9.2. Graphs were made in SAS software version 9.2 and R version 3.1.3.

RESULTS

Descriptive analysis

For the study period (first quarter of 2004 through the third quarter of 2012), a total of 4,285,088 reports were retrieved from FAERS. After eliminating duplicates (n=43,125), removal of adult reports (n=2,686,530) and reports with missing age (n=1,419,524) or reports indicating prenatal exposure (n=20,235), 115,674 reports corresponding to 893,587 individual DECs were retained for analysis of pediatric spontaneous reports (see Table 2).

Table 2: Description of pediatric reports by age categories

Age group	Number of reports, n (%)
Neonates: 0–27 days	5,091 (4.40%)
Infants: 28 days–23 months	12,566 (10.86%)
Children: 2–11 years	49,982 (43.21%)
Adolescents: 12–17 years	48,035 (41.53 %)
Total	**115,674 (100%)**

The total number of pediatric reports that included the investigated drugs and ADRs from the reference set can be observed in Figure 1, which also shows data regarding adults (for comparison purposes). The number of children exposed to the drugs of interest, for which any of the investigated ADRs was reported, varied from 26 patients (for praziquantel) to 7,535 patients (for ibuprofen) with a median of 781 patients exposed across all drugs. The number of events of interest in FAERS ranged from 164 reports (ventricular arrhythmia) to 14,777 (anaphylaxis), with a median of 1,004 reports across all events.

Overall performance of SDAs

Both SDAs showed high specificity and low sensitivity. They both had similar specificity values (PRR: 83.8% and EBGM: 91.9%), while sensitivity was lower for EBGM than for PRR (17.2% vs. 37.9%). The NPV and PPV were similar for both SDAs. When we applied the threshold-independent (AUC-based) approach, the tested SDAs showed similar performance in the pediatric population although the AUC value for EBGM (0.745) was slightly higher than for PRR (0.731). None of the SDAs performed better than the simple report count (AUC=0.634, p-values: PRR=0.27 and EBGM=0.14)

Stratification and adjustment for age and its impact on performance

Upon calculating SDA values per age stratum and testing for heterogeneity across strata, we observed effect modification for some associations. Some false negatives (positive DECs which failed to be highlighted as signals when analysing data pertaining to the entire pediatric population) were unmasked in some strata. Four DECs were unmasked: ibuprofen-thrombocytopenia and isoniazid-seizure (by PRR) and clarithromycin-erythema multiforme and ibuprofen-erythema multiforme (by EBGM). Conversely, 'ibuprofen-acute liver injury', also a positive DEC, was highlighted when we analyzed data pertaining to the entire pediatric population but after stratifying, it became clear that this DEC was highlighted only in older children (adolescents), and not highlighted

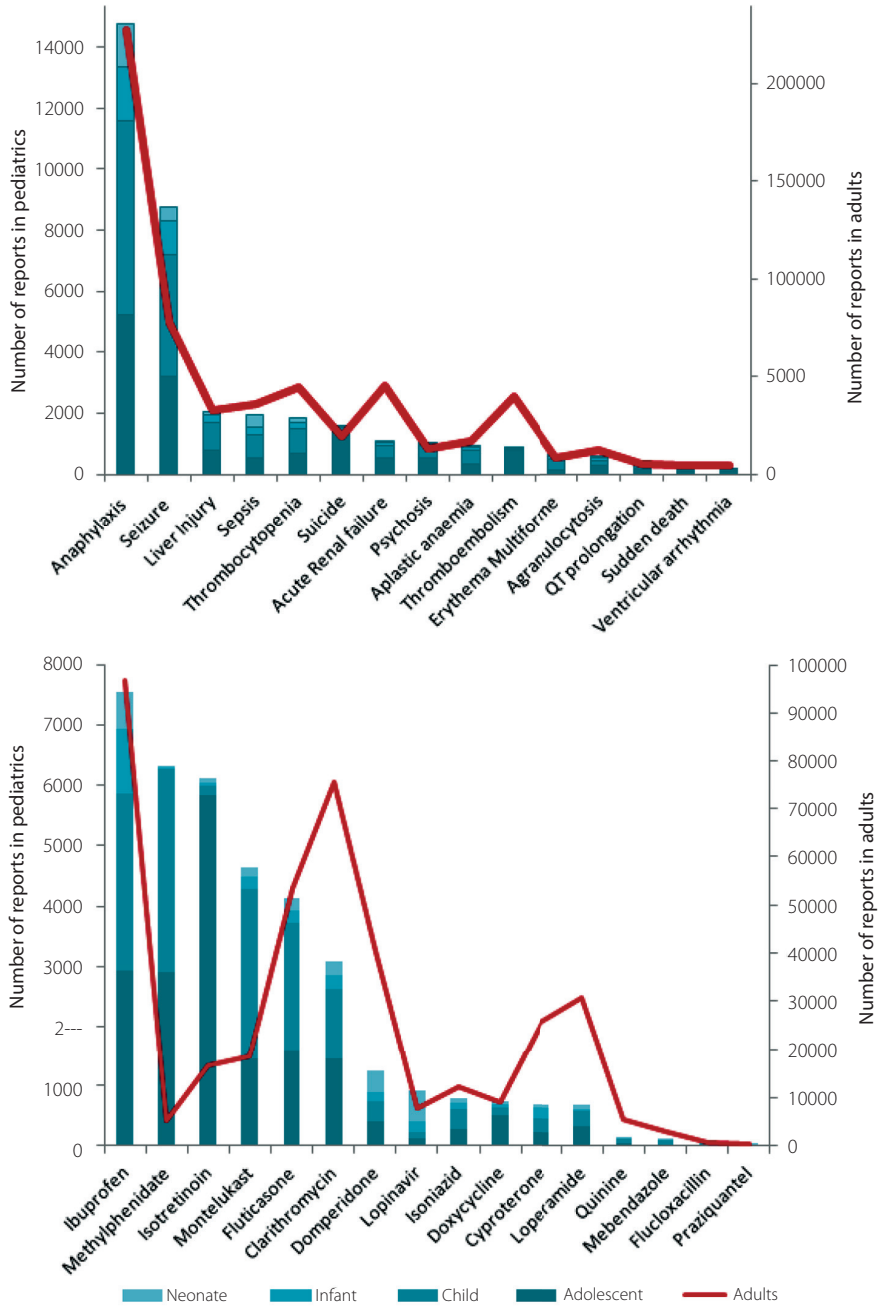


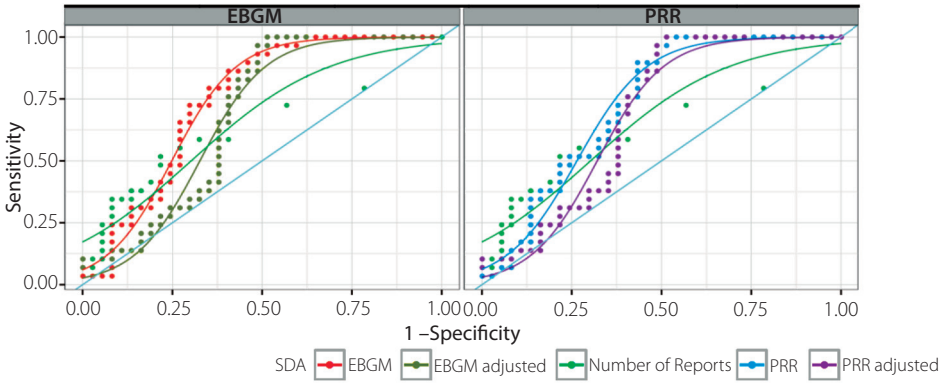
Figure 1: Count of reports in pediatric and adult population for the investigated ADRs and drugs, cumulatively for the period Q1 2004–Q3 2012[a]

[a] Number of reports in children is represented by bars and plotted on the left axis, while the number of reports in adults is represented by the red line and plotted on the right axis; Reports with missing age or age=0 were excluded. Only reports mentioning any of the drugs or events in the reference set were considered.

in younger children (see Figure 3). For an overview of SDA values across age strata and results of heterogeneity tests please refer to the Electronic Supplementary material Figures 1A and 1B.

Figure 2: Performance of signal detection algorithms within the entire pediatric population

SDA	Sensitivity	Specificity	PPV	NPV	AUC	p-value[b]
Number of reports	58.62	67.57	58.62	67.57	0.634	*reference*
PRR	37.93	83.78	64.71	63.27	0.731	0.266
EBGM	17.24	91.89	62.50	58.62	0.745	0.144
After age adjustment [a]				*(reference-crude PRR/EBGM)*		
PRR	34.48	86.49	66.67	62.75	0.688	0.267
EBGM	10.34	97.30	75.00	58.06	0.683	0.216



SDA-signal detection algorithm; PRR= Proportional reporting ratio; EBGM= Empirical Bayes Geometric Mean; AUC=area under the curve; PPV=positive predictive value; NPV-negative predictive value. [a] adjusted PRR/ROR values calculated by combining the individual estimates from each age stratum into one measure according to the Mantel-Haenszel approach. [b] paired chi-square test

We evaluated the performance of the methods within individual age strata (see Table 3). On average, performance of the SDAs was lower within age strata compared to the entire pediatric population and performance improved with increasing stratum size. For infants and neonates, the performance was very low, not better than chance (p-value>0.5 for both SDAs). The adolescent group exhibited the best performance, which was similar to the overall performance.

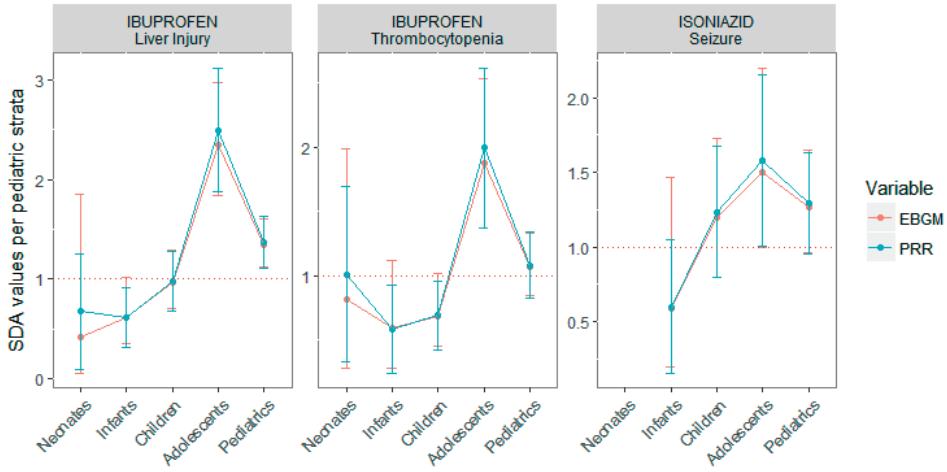


Figure 3: Variation of PRR and EBGM estimates across pediatric specific strata – selected examples

p-values were calculated with Breslow Day Tarone test for homogeneity

Table 3: Performance of signal detection algorithms across age strata

Age groups (number of reports)	Signal detection algorithms	AUC
Neonates (5,091)	Number of Reports	0.625
	EBGM	0.600
	PRR	0.65
Infants (12,566)	Number of Reports	0.667
	EBGM	0.548
	PRR	0.554
Children (49,982)	Number of Reports	0.654
	EBGM	0.698
	PRR	0.649
Adolescents (48,035)	Number of Reports	0.698
	EBGM	0.771
	PRR	0.718
Entire pediatric population (115,674)	Number of Reports	0.634
	EBGM	0.745
	PRR	0.731

PRR= Proportional reporting ratio; EBGM= Empirical Bayes Geometric Mean; AUC=area under the curve

After adjusting for age by pooling the stratum-specific estimates, the performance of the SDAs decreased, although not significantly (see Figure. 2; crude vs. adjusted AUC for PRR 0.731 vs. 0.688, p-value=0.267; crude vs. adjusted AUC for EBGM 0.745 vs. 0.683, p-value=0.216).

DISCUSSION

In this study, we have demonstrated that age stratification for detection of drug safety signals in children may unmask some signals that do not appear in neither crude nor adjusted analysis. Adjustment for age does not improve performance of the PRR or EBGM.

For the investigated events, similar reporting patterns were observed for children and adults while the investigated drugs appeared to have different reporting patterns (see Figure 1). Different drug-related reporting patterns in children vs adults were previously reported [5]. Consequently, reported DECs for children may differ from adults, [3,5] underlining the need for pediatric-specific approaches to signal detection especially when we consider that even within the pediatric population, reported drugs may vary by age group [3,31].

Overall, the PRR and EBGM showed good performance although results were slightly lower than results reported on other (not pediatric-specific) reference sets [32,33]. The similarity in performance between PRR and EBGM is in accordance with the recent results from the PROTECT project [23]. The fact that the performance (based on AUC) of PRR and EBGM was not statistically significantly better than simple report count may be due to the lack of power. Within age strata, performance seemed to correlate with stratum size: the poorest results were observed for infants and neonates (the smaller groups), slightly improving for children while the best performance was observed for adolescents, the age stratum with the highest number of tested DECs. Decrease in power due to fewer reports and therefore DECs may account for this observation. The fact that we used lower bounds of confidence intervals for signaling instead of point estimates might have exacerbated the influence of sample size on the results, since smaller strata will have higher variability. In neonates and infants for whom expected counts were difficult to calculate because of few reports, we observed that simple report counts performed similar or even better than the SDAs and might be an alternative to commonly used SDAs. The fact that simple report count performed better than SDAs may have been because the reference set comprised known DECs (which in turn may have influenced reporting) rather than emerging safety issues, a hypothesis proposed by Noren et al. [34].

Inspection of SDA values across child specific strata (age-stratification) revealed some heterogeneity in estimates pointing to some effect modification. For example, 'ibuprofen-thrombocytopenia', was found as a signal in the adolescents' group but not detected in the entire pediatric population or the younger age categories. This suggests that age-specific SDA calculations are sometimes needed, rather than age-adjusted SDA estimates. The age-adjusted estimates did not improve performance; in fact even PPV unexpectedly decreased. Simulation studies have shown that when adjusted for

strata, Bayesian methods such as EBGM tend to be underestimated when there are sparse strata [15]; this was also the case in our study. Previous studies in adults show contradictory results, with some showing a beneficial effect [9] while others did not [15]. The reason for our finding is not entirely clear; a possible explanation is that age is not a strong confounder for the investigated DECs. Also, the method of weighting (Mantel-Haenszel approach) may have played a role since more weight was assigned to age groups with more reports (adolescents and children). This may have masked signals occurring in age groups with fewer reports.

The limitations of data mining in FAERS include those inherent to spontaneous reporting databases: underreporting, lack of denominator data and control group, biases in reporting, as well as missing and poor quality data [35]. Missing information regarding age substantially reduced the study sample size since we could not determine whether these reports described patients aged less than 18 years old. While these biases are well acknowledged and have a definite impact, they cannot be completely avoided. Compared to adults, there are fewer reports and different reporting patterns for children [3,36,37] which may complicate signal detection in the pediatric population.

Evaluating performance of SDAs is a constant challenge due to lack of standard methodologies, imperfect reference standards and uncertainty regarding the best thresholds. Some of the drugs and events in the reference set are specific to one age group within pediatrics and this is obvious in Figure 1, even though the reference set was designed to be relevant for the entire pediatric population. We acknowledge that the reference set used, although specifically constructed for this purpose, does not include all the ADRs that are highly specific for pediatrics. This highlights the need for pediatric-specific approaches to signal detection; accounting for not just the entire pediatric population but also the different age strata within pediatrics. Still, the reference set captures various drug use and ADRs patterns [38] and is currently the only available pediatric-specific reference set. The thresholds applied to define a signal were obtained from previous publications and other cut-off points may generate better results; further research on pediatric-specific thresholds should be encouraged.

CONCLUSION

Our study revealed that age adjustment did not improve performance of the SDAs. However, stratification revealed some variation in SDAs' values across strata (effect modification) and inspection of stratum-specific estimates might sometimes yield useful information during routine surveillance.

REFERENCES

1 Bate A, Evans SJW. Quantitative signal detection using spontaneous ADR reporting. Pharmacoepidemiology and Drug Safety. 2009 2009;18:427–36.

2 Evans SJW, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. Pharmacoepidemiology and Drug Safety. 2001 10/2001;10:483-6.

3 Blake KV, Zaccaria C, Domergue F, La Mache E, Saint-Raymond A, Hidalgo-Simon A. Comparison between paediatric and adult suspected adverse drug reactions reported to the European medicines agency: implications for pharmacovigilance. Paediatric Drugs. 2014 Aug 2014;16:309-19.

4 Star K, Edwards IR. Pharmacovigilance for children's sake. Drug Safety. 2014 Feb 2014;37:91-8.

5 Star K, Norén GN, Nordin K, Edwards IR. Suspected adverse drug reactions reported for children worldwide: an exploratory study using VigiBase. Drug Safety. 2011 May 1, 2011;34:415-28.

6 Kearns GL, Abdel-Rahman SM, Alander SW, Blowey DL, Leeder JS, Kauffman RE. Developmental pharmacology - Drug disposition, action, and therapy in infants and children. New England Journal of Medicine. 2003 2003;349:1157-67.

7 Aagaard L, Weber CB, Hansen EH. Adverse drug reactions in the paediatric population in Denmark: a retrospective analysis of reports made to the Danish Medicines Agency from 1998 to 2007. Drug Safety. 2010 Apr 1, 2010;33:327-39.

8 Kimland E, Rane A, Ufer M, Panagiotidis G. Paediatric adverse drug reactions reported in Sweden from 1987 to 2001. Pharmacoepidemiology and Drug Safety. 2005 Jul 2005;14:493-9.

9 Woo EJ, Ball R, Burwen DR, Braun MM. Effects of stratification on data mining in the US Vaccine Adverse Event Reporting System (VAERS). Drug Safety. 2008 2008;31:667-74.

10 Orre R, Lansner, A., Bate, A., & Lindquist, M. Bayesian neural networks with confidence estimations applied to data mining. Computational Statistics & Data Analysis,. 2000 2000;34(4):473-93.

11 Almenoff JS, LaCroix KK, Yuen NA, Fram D, DuMouchel W. Comparative performance of two quantitative safety signalling methods. Drug Safety. 2006;29(10):875-87.

12 Norén GN, Bate A, Orre R, Edwards IR. Extending the methods used to screen the WHO drug safety database towards analysis of complex associations and improved accuracy for rare events. Statistics in Medicine. 2006 Nov 15, 2006;25:3740-57.

13 DuMouchel W. Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. The American Statistician. 1999;53(3):177-90.

14 Almenoff JS, Pattishall EN, Gibbs TG, DuMouchel W, Evans SJW, Yuen N. Novel statistical tools for monitoring the safety of marketed drugs. Clinical Pharmacology & Therapeutics. 2007;82(2):157-66.

15 Hopstadius J, Norén GN, Bate A, Edwards IR. Impact of stratification on adverse drug reaction surveillance. Drug Safety. 2008;31(11):1035-48.

16 Global Research in Paediatrics. Network of excellence. Global Research in Paediatrics. Network of excellence,. [cited 2016 28th February]; Available from: http://www.grip-network.org/index.php/cms/en/home

17 Moore TJ, Furberg CD. The Safety Risks of Innovation: The FDA's Expedited Drug Development Pathway. JAMA. 2012 2012-09-05;308:869.

18 Weiss-Smith S, Deshpande G, Chung S, Gogolak V. The FDA drug safety surveillance program: adverse event reporting trends. Archives of Internal Medicine. 2011 Mar 28, 2011;171:591-3.

19 MedDRA MSSO. Introductory Guide MedDRA Version 15.1. 2012 [cited 2016 28th February]; Available from: http://www.meddra.org/sites/default/files/guidance/file/intguide_15_1_English_0.pdf

20 Poluzzi E, Piccinni C, Raschi E, De Ponti F. Data Mining Techniques in Pharmacovigilance: Analysis of the Publicly Accessible FDA Adverse Event Reporting System (AERS): INTECH Open Access Publisher; 2012.

21 SAS. The power to know. SAS(R) 9.2 Language Reference: Dictionary, Fourth Edition. COMPGED Function. Returns the generalized edit distance between two strings. Computing the Generalized Edit Distance. [cited; Available from: http://support.sas.com/documentation/cdl/en/lrdict/64316/HTML/default/viewer.htm#a002206133.htm

22 Hauben M, Aronson JK. Defining 'signal' and its subtypes in pharmacovigilance based on a systematic review of previous definitions. Drug Safety. 2009;32(2):99-110.

23 Candore G, Juhlin K, Manlik K, Thakrar B, Quarcoo N, Seabroke S, et al. Comparison of Statistical Signal Detection Methods Within and Across Spontaneous Reporting Databases. Drug Safety. 2015 2015.

24 Zhou X-H, McClish DK, Obuchowski NA. Statistical methods in diagnostic medicine: John Wiley & Sons; 2009.

25 Strom BL KS, Hennessy S,. Pharmacoepidemiology. 5th ed. Chichester, West Sussex, UK: Wiley-Blackwell; 2012.

26 Osokogu OU, Fregonese F, Ferrajolo C, Verhamme K, de Bie S, Catapano M, et al. Pediatric Drug Safety Signal Detection: A New Drug–Event Reference Set for Performance Testing of Data-Mining Methods and Systems. Drug Safety. 2015;38(2):207-17.

27 international conference on harmonisation of technical requirements for registration of pharmaceuticals for human use ich harmonised tripartite guideline. Clinical investigation of medicinal products in the pediatric population e11. Current step 4 version 2000 [cited 2016 28th february]; available from: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E11/Step4/E11_Guideline.pdf

28 Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. Journal of the National Cancer Institute. 1959 Apr 1959;22:719-48.

29 Breslow NE, Day NE. Statistical methods in cancer research. Volume I - The analysis of case-control studies. IARC scientific publications. 1980 1980:5-338.

30 Zeinoun Z, Seifert H, Verstraeten T. Quantitative signal detection for vaccines: effects of stratification, background and masking on GlaxoSmithKline's spontaneous reports database. Human vaccines. 2009 Sep;5(9):599-607.

31 de Bie S, Ferrajolo C, Straus SMJM, Verhamme KMC, Bonhoeffer J, Wong ICK, et al. Pediatric Drug Safety Surveillance in FDA-AERS: A Description of Adverse Events from GRiP Project. PloS One. 2015 2015;10:e0130399.

32 van Puijenbroek EP, Bate A, Leufkens HGM, Lindquist M, Orre R, Egberts ACG. A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions. Pharmacoepidemiology and Drug Safety. 2002 2002;11:3–10.

33 Harpaz R, DuMouchel W, LePendu P, Bauer-Mehren A, Ryan P, Shah NH. Performance of Pharmacovigilance Signal-Detection Algorithms for the FDA Adverse Event Reporting System. Clin Pharmacol Ther. 2013 2013;93:539-46.

34 Norén GN, Caster O, Juhlin K, Lindquist M. Zoo or savannah? Choice of training ground for evidence-based pharmacovigilance. Drug Safety. 2014 Sep 2014;37:655-9.

35 Almenoff J, Tonning JM, Gould AL, Szarfman A, Hauben M, Ouellet-Hellstrom R, et al. Perspectives on the use of data mining in pharmacovigilance. Drug Safety. 2005;28(11):981-1007.

36 Aagaard L, Strandell J, Melskens L, Petersen PSG, Holme Hansen E. Global patterns of adverse drug reactions over a decade: analyses of spontaneous reports to VigiBase™. Drug Safety. 2012 Dec 1, 2012;35:1171-82.

37 Morales-Olivas FJ, Martínez-Mir I, Ferrer JM, Rubio E, Palop V. Adverse drug reactions in children reported by means of the yellow card in Spain. Journal of Clinical Epidemiology. 2000 Oct 2000;53:1076-80.

38 Impicciatore P, Choonara I, Clarkson A, Provasi D, Pandolfini C, Bonati M. Incidence of adverse drug reactions in paediatric in/out-patients: a systematic review and meta-analysis of prospective studies. British Journal of Clinical Pharmacology. 2001 Jul 2001;52:77-83.

Chapter 5

After signal detection: prioritization and triage

5.1

Decision making in signal management: A literature review of criteria used to prioritize newly detected safety issues

Alexandra C Păcurariu
Preciosa M Coloma
Liana Gross-Martirosyan
Miriam C Sturkenboom
Sabine M Straus

Pharmacoepidemiol Drug Saf. 2016 Nov 23. doi:10.1002/pds.4128.

ABSTRACT

Background and Objectives: In drug safety, there is a lack of guidance on how prioritization of safety issues should be performed. The aim of this literature review is to provide an overview of criteria used for signal prioritization and of the associated decision support frameworks.

Methods: A search strategy was constructed to identify relevant articles in Medline/Embase databases from the period from 1st January 1995–31st August 2015. The prioritization criteria were extracted and classified in relevant categories.

Results: From an initial set of 63 articles, 11 were retained for full review. The articles mentioned 48 criteria used in the prioritization process, with a median of 6 criteria per study (range: 1–16). More than half of the criteria (63%), referred to strength of evidence while 19% related to public health impact, 14% to general public and media attention and 4% to novelty of the drug event association. Fifteen criteria were tested for predictive value with 11 showing positive results, most of them from the strength of evidence category. Six decision making frameworks are presented, which incorporate criteria from various categories. Five of these frameworks were tested against expert decisions or by other means, but only in one database each and for a limited set of products.

Conclusions: There is a wide range of prioritization criteria described in the literature, however few of them demonstrated predictive value. Many criteria with predictive value were related to strength of evidence category and to novelty category. There were few attempts at integrating different criteria in decision support frameworks. Five of the frameworks were tested for validity and showed usefulness, while at least three are already in use for prioritization.

INTRODUCTION

Signal management is a key activity in pharmacovigilance, defined as 'a set of activities performed to determine whether, based on a detailed examination of individual case safety reports, aggregated data from active surveillance systems or studies, literature information or other data sources, there are new risks associated with an active substance or a medicinal product or whether risks have changed' [1].

The process starts with an exploratory phase, namely the detection of signals [2]. Hauben and Aronson define signal as 'information which arise from one or multiple sources (including observation and experiments), which suggest a new potentially causal association or a new aspect of a known association, between an intervention and a set of related events, either adverse or beneficial, which would command regulatory, societal or clinical attention, and is judged to be of sufficient likelihood to justify verificatory and, when necessary, remedial actions' [2]. Detection of signals is done in several ways, varying from employment statistical methods to a direct review of data by experienced professionals [3].

Signal detection is followed by several steps of additional data collection and analysis, aimed to increase the available evidence and to form the basis for a decision regarding the safety issue. Since the amount of detected signals usually surpasses the capacity for analysis, [4] prioritization should be done early in the process in order to focus resources on the most important signals, often from a public health perspective.

Similar to the triage applied in a clinical setting, [5] the objectives of prioritization in pharmacovigilance context are to ensure that the most important signals (i.e., with highest public health impact) are investigated first, to optimize the use of resources and to decrease time from signal detection to action. Prioritization is an ongoing process since priority may change in view of newly available information. In this study we focused on 'early prioritization' or 'first pass screening' which is applied right after signal detection and before thorough signal evaluation takes place.

Signal prioritization is a multifactorial decision making process, based on a combination of clinical, epidemiological, pharmacological and regulatory information. Weighting and integration of different information components are human tasks and therefore not always objective since they are influenced by prior knowledge and investigator bias [6]. Criteria for signal prioritization are suggested in various guidance documents, [3,7] however they are rarely standardized or validated [8]. The information available in the field of signal prioritization is scattered and there are no agreed criteria or guidance on how to combine them in decision support frameworks [9-11].

Therefore, the aim of this literature review is to provide an overview of the existing criteria and frameworks for signal prioritization.

METHODS

A broad literature search strategy was constructed to identify articles describing criteria for signal prioritization. The search was performed in Medline and Embase databases, from 1st of January 1995 – 31st of August 2015, and restricted to English language. The query was constructed based on MeSH terms and keywords extracted from an initial set of relevant publications. This was further optimized in an iterative process.

Articles were initially selected based on title and abstract screening, and subsequently the full text article was reviewed. Those that met the inclusion criteria were also checked for cross references of relevant studies ('snowballing'). Articles were included if they referred to one or more prioritization criteria used for the prioritization of previously detected signals.

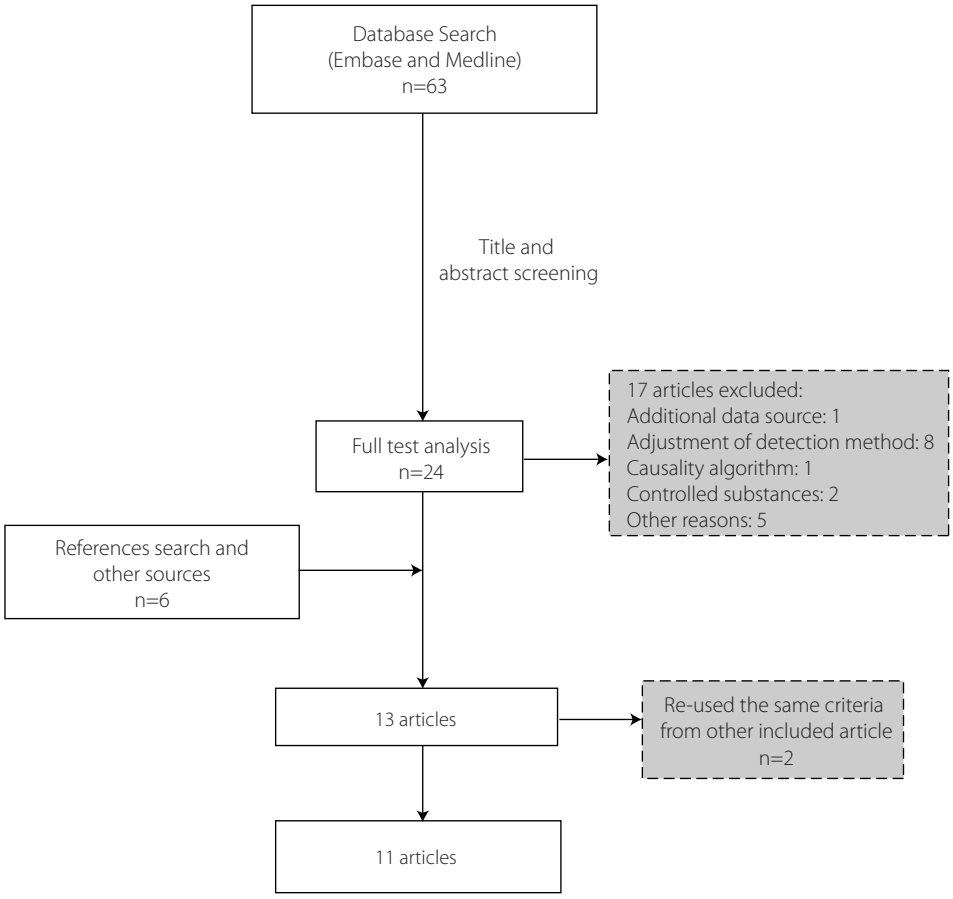


Figure 1: Flowchart of studies selection

Some of the variables can be used both in signal detection and prioritization steps. For example reporting rate of certain adverse events can be used to flag potential signals however, at the same time, an increased reporting rate could be a reason to prioritize a signal for further evaluation. For the purpose of this paper, a clear distinction was made and only criteria which were discussed in the context of prioritization of already detected signals were included in the analysis. Therefore, papers which reported updates of data mining algorithms with additional variables that are meant to be used for signal detection were excluded. Similarly, causality assessment methods meant to be applied at report level, as the Naranjo scale [12] or the WHO causality assessment system [13] were also excluded.

Data collection

From the included papers the following characteristics were extracted: publication title, year, prioritization criteria, its definition and mode of calculation, source and the predictive value (i.e. capacity to identify signals which later prove to be true).

All prioritization criteria were identified from the studies and then categorised independently by two authors (AP and LGM). Disagreements were arbitrated by a third reviewer (SS). Categorization was done into several pre-defined categories that were selected initially and then modified based on the available criteria described in the studies: novelty of drug event association, public health impact, strength of evidence, and general public or media attention. The definitions used for classifying the criteria into these categories are provided below:

- **Novelty of drug event association** referred to an association that was not previously recognised or that is not labelled in the product information of the drug [14].
- **Public health impact** was defined as the impact that a potential safety issue is likely to have on patients' health at population level, usually through the number of patients affected by an adverse reaction and their consequences [11,15]. The magnitude of the public health impact is usually considered in relation to the size of the general population, the population with the target disease and the treated population [7].
- **Strength of evidence** was defined as the degree of evidence supporting a causal relationship between the drug and the event [16].
- **General public and media attention** was defined as increased awareness and attention from the media or general public regarding a specific safety issue, due to either its serious consequences, difficulty to prevent/control or propensity to affect vulnerable populations as children or pregnant women [15].

Data analysis

This is a descriptive study therefore no hypothesis testing was performed. Summary measures for categorical data were used for the data description. SPSS version 21 was used for data analysis. The decision support frameworks were described individually, due to a large degree of heterogeneity.

RESULTS

A total of 63 papers were identified, out of which 11 papers were retained for analysis [4,9-11,15-21] (Figure 1). All studies described prioritization of signals arising from spontaneous reporting databases. One of the studies [21] focused on drug-drug interactions. A total of 48 different prioritization criteria were identified, with a median of 6 criteria per study (range: 1–16) (see Table 1).

Table 1: Criteria used for signal prioritization and number of occurrences in the literature

Strength of evidence			
Disproportionate reporting*	7	Confounded by indication	1
Alternative data sources	4	Mechanism potential	1
Quality/completeness of reports*	4	Narrative present*	1
Rechallenge positive	4	Nested case control studies	1
Multinational reporting*	3	Pharmacological expectedness	1
Biological plausibility	2	Recent reporting*	1
Class effects	2	Plausible CYP metabolism¥	1
Dechallenge positive	2	Presence of a similar association	1
Typical drug-related event*	2	Reporter qualification	1
Altered therapeutic effect at concomitant use¥	1	Reporting rate	1
Alternative cause	1	Specific, characteristic event	1
Rapid reporting increase*	1	Unexpected therapeutic response ¥	1
Background frequency	1	Suspected interaction by reporter¥	1
Fractional reporting ratio	1	Targeted comparisons	1
Causality	1	Temporality	1
Public health impact		**General public and media attention**	
Seriousness*	6	Factors likely to cause public anxiety	1
Number of reports	5	Media attention	1
Drug exposure	2	Other public concern	1
Frequency of ADR in users	2	Public misperceptions	1
Potential for prevention	2	External interest	1
Potential health consequences	2	Health authority concerns	1
Critical term*	1	Recent parliamentary questions	1
Targeted medical events	1		
Severity	1		
Novelty			
Novelty of the drug event association*	4		
Novelty of the drug *	2		

ADR=adverse drug reaction; * Demonstrated predictive value (i.e., capacity to identify signals which upon further analysis were proven to be true); ¥ Applicable only to drug-drug interactions

More than half of prioritization criteria referred to strength of evidence (63%), while 19% related to public health impact, 14% to general public and media attention and 4% to novelty of the drug event association (see Table 2). The most frequently used criteria were disproportionate reporting (7 studies), seriousness (6 studies) and number of reports (5 studies).

Six studies combined multiple criteria into decision support frameworks.

Table 2: Description of prioritization criteria

	Count	Percentage
Category		
Strength of evidence	30	63%
Public health impact	9	19%
Novelty of drug event association	2	4%
General public and media attention	7	8%
Source of the criteria		
Spontaneous reporting databases	23	48%
Drug utilisation data	3	6%
Product information	2	4%
Other data sources	27	56%
Predictive value tested (Yes)	15	31%

Percentages do not add to 100 since some variables can have more than one source

Criteria related to strength of evidence

Thirty different criteria related to strength of evidence were reported in 10 studies, with the most frequently reported ones being: disproportionate reporting, alternative data sources confirming the signal, quality/degree of completeness of reports and positive rechallenge. In addition to the disproportionality measures, rapid reporting increase [9,16] and recent reporting [10] were also categorised as related to strength of evidence.

An important criterion was the quality and completeness of reports. Two independent studies have shown a correlation between the quality of reports and true signals [10,17]. Caster et al. used two different criteria related to the quality/completeness of reports, one being a completeness score and the other a dichotomous criterion which indicates if the narrative is present or not in the report [10].

Another criterion is multi-national reporting (i.e., reports of a specific drug-event association originating from different countries; this increased the probability that the signal is true [9,10].

A third important criterion related to the strength of evidence was use of additional data sources. Some authors mentioned pre-clinical, clinical or epidemiological studies, [14,20] while others considered only randomized clinical trials or meta-analysis of clinical trials to be valuable sources of evidence [15].

Criteria related to novelty

Two criteria were identified that relate to novelty: novelty of drug event association and novelty of drug. Novelty of the drug event association was mentioned as a criterion to prioritize signals in four studies [14,16,17,20]. The novelty was usually assessed by checking if the ADR is mentioned in the product information or other sources of safety data, such as scientific literature, medical textbooks or Pharmacopoeias. One study [17] proved that true signals are correlated with absence of the ADR from the product information.

The novelty of drug (i.e., time on the market) was also used in prioritization [9,17] and showed predictive value. Instead of using it as a continuous variable, it was dichotomized in new vs. old drugs, according to an empirically chosen threshold: in one study a threshold of five years was chosen, while in another one, the threshold was 2 years.

Criteria related to public health impact

Nine criteria related to public health impact were identified in eight studies. The most frequent criteria were seriousness, drug exposure and number of reports.

Firstly, the seriousness of the reports was the most often used criterion (n=6), and also showed predictive value [9,17]. Studies that attempted to quantify the seriousness of the reports divided it into a fatal component and a non-fatal component (e.g., potential to cause major/permanent or minor disabilities).

Secondly, the extent of drug exposure was used to estimate health impact [15,18]. The pre-defined threshold of drug exposure that was used in the different studies to decide if an issue has a potential high public health impact varied from a threshold of 100,000 exposed patients/year (United Kingdom) [15] to over 1 million patients exposed cumulatively (United States) [18].

Thirdly, the absolute number of reports was another frequently used criterion for health impact [11,14-16]. Usually, the higher number of reports, the higher the health impact. The paper by Stahl et al. introduced an unusual threshold to the number of reports, which described a maximum instead of a minimum of reports as a prioritization criterion [9]. This criterion was developed for VigiBase®, the WHO international database of suspected adverse drug reactions maintained by UMC on behalf of the WHO Programme for International Drug Monitoring.

Except these three main factors, public health impact could have been assessed in other ways as well. Some organizations developed lists of 'drug related events' and 'targeted medical events', based on event seriousness and its likelihood to be drug induced. Examples are: targeted medical events developed by Food and Drug administration (FDA), [22] the critical terms list developed by UMC [23,24] and the important medical events list developed and maintained by EudraVigilance Expert Working Group in collaboration with MedDRA®. It was demonstrated that ADR being a critical term was predictive of a true signal [9,17].

The potential for prevention was also taken into account in two studies, [14,20] however none of these studies explained this any further nor described exactly how the concept can be evaluated or measured.

Criteria related to general public and media attention

Potential public misperceptions about the safety of the drug could cause harm through a behaviour change (e.g. decreased vaccine uptake, abrupt discontinuation of medicine leading to poor outcomes) [15] and therefore should be considered during signal evaluation. Factors likely to cause public anxiety either due to serious consequences of the reaction, difficulty to prevent/control the hazard or signals that are likely to affect vulnerable populations such as children/pregnant women were mentioned. In several studies the presence of external interest from media, health authorities, scientific community or patients was taken into account [15,16].

Predictive value of criteria

The predictive value of 15 prioritisation criteria was evaluated in five studies [9,10,17,19,21] and eleven criteria predicted real signals to a certain extent. Various approaches were used to evaluate the predictive value of criteria, for example logistic regression models or comparison with an alternative gold-standard method (e.g., expert review).

Many criteria with predictive value were related to strength of evidence category (rapid reporting increase, recent reporting, disproportionate reporting, multinational reporting, quality/degree of completeness reports, and presence of narrative in the reports). Both criteria categorised under novelty were shown to have predictive value (novelty of the drug and novelty of drug event association). From public health impact category, two criteria, namely seriousness and ADR being a critical term were shown to have predictive value.

Four criteria which failed to show predictive value were: positive dechallenge and positive rechallenge, number of reports and reporter qualification. The rest of them have not been tested.
Frameworks for signal prioritization

Six decision support frameworks were identified in the literature (see Table 3) and all of these contain a mix of the criteria presented above. Some introduced graded scales to limit subjectivity [11,15,16] and some included weighting schemes, [10,11,15,16] assigning more importance to some criteria than to others.

Five of the frameworks were tested to see how well they predict signals. This was done by calculation of agreement coefficients [11,15,16] or by regression models [10]. The gold standard against which evaluation was made consisted of reference sets of either known signals or expert judgment.

Table 3: Decision support frameworks in signal prioritization

	1	2	3	4	5	6
Publication year	**2004**	**2005**	**2013**	**2008**	**2012**	**2014**
Reference	Stahl et al. [9]	Waller et al [11]	Seabroke et al [15]	Levitan et al [16]	FDA guidance [18]	Caster et al [10]
Number of criteria incorporated	7	6	16	11	3	5
Key concepts						
Strength of evidence	x	x	x	x		x
Public health impact	x	x	x	x	x	
Novelty of drug event association	x		x	x		
General public and media attention			x			
Graded scales for qualitative variables	No	Yes	Yes	Yes	No	NA
Weighting of variables	No	Yes	Yes	Yes	No	Yes
Validity testing	Yes	Yes	Yes	Yes	No	Yes
Testing method (reference)	Method not mentioned	Agreement and/ or correlation coefficients (expert group opinion)			NA	Multiple logistic regression (set of known signals)
Currently in use	Partially, at UMC	Yes-at MHRA		Unknown	Unknown	Yes-at UMC

FDA- Food and Drug administration; MHRA= Medicines and Healthcare Products Regulatory Agency; UMC= Uppsala Monitoring Centre

The initial prioritization algorithm was developed by Stahl et al. [9] as early as 2001 at UMC, and took into account rapid reporting increase, seriousness, time on the market for the involved drug and if the ADR was of special interest. It was mentioned that these criteria proved to be successful in selecting true signals, however the exact testing method was not provided.

Another decision support framework [11] took into account the strength of evidence and the potential public health impact. The included components of strength of evidence were: disproportionality score, quality of reports and biological plausibility. The components of public health score were: number of reports, seriousness and reporting rate. Using an empirical cut-off point for both scales, four priority categories for a signal were obtained, each having a different course of regulatory action. All subjective variables were quantified using graded scales. This tool was validated [25] and is currently used for prioritization in a regulatory setting at Medicines and Healthcare Products Regulatory Agency (MHRA) in the UK.

Seabroke et al. [15] updated the framework mentioned above by adding two categories: agency regulatory obligations and public perceptions. This updated tool was designed to be used in a later stage of signal management. The tool was piloted and validated against expert group opinion and is also routinely used in the same organization.

A multi criteria decision analysis (MCDA) weighted framework was developed by Levitan et al. [16] based on 11 criteria related to novelty of event (10% weight), strength of evidence (40% weight) and medical impact (50% weight). In addition, two extra criteria were used for pre-selection: evaluation of novelty of the event and of confounding by indication. Each criterion had an assigned weight and there were graded scales for each attribute. The model was tested against expert group judgment and the agreement was found to be moderate.

FDA drafted a prioritization guidance aimed to classify post-marketing drug safety issues [18]. This guidance recommends to estimate the hazard posed by a safety issue, based on three criteria: (1) the seriousness of the issue; (2) the estimated size of the population exposed to the drug; and (3) the suspected frequency of harm for exposed patients. The combination of factors 2 and 3 provides an estimate of population risk, while the combination of factors 1 and 3 provides an estimate of individual risk.

The vigiRank predictive model developed by Caster et al. [10] is an algorithm for emerging safety signals that accounts mainly for reports quality and content. The following criteria were included: disproportionate reporting, number of informative and recent reports, number of reports with a narrative and multinational reporting. The advantages of this method are that it is automated in VigiBase® and was tested in a comprehensive manner, by means of multiple logistic regression, and against a reasonably large reference set. Public health impact was not considered within this algorithm.

Last but not least, Strandell et al. [21] propose two prioritization frameworks specific to drug-drug interactions. This was the first application of predictive regression models for first-pass screening of large collections of spontaneous reports, when looking for drug interactions. Due to its specificity for drug interaction, this was not described further.

DISCUSSION

Prioritization decisions are typically complex and resource intensive, as they blend the numerical information with scientific knowledge and judgment [8]. In this paper, we conducted a review of signal prioritization criteria and associated decision support frameworks that were built upon those, in order to increase awareness and facilitate the process.

A total of 48 criteria were identified in the literature for signal prioritization, and they were categorized according to the following key concepts: novelty, strength of evidence, public health impact and general public and media attention.

One important distinction should be made early on regarding the criteria found in the studies: some of them were used for prioritization based on their predictive value, while others were used independent of this property. Usually, criteria related to strength of evidence would fall in the first category, whereas criteria related to public health impact would be included in the second one.

The novelty concept is intrinsic to signal definition and six studies mentioned it, albeit this might be an underestimate of its actual use in signal prioritisation process. Novelty could related to the drug event association or just to the drug. Although 'Weber effect' [26] (i.e., AE reporting peaks at the end of the second year after approval) was not reproduced, [27,28] two studies [9,17] showed that new drugs are more likely to have more safety signals.

Another key concept, strength of evidence, was at the core of prioritization algorithms, being considered by 10 out of 11 studies and by all six decision support frameworks. This was expected, since it is logical to focus the resources, even from a very early stage, on those signals which have a high probability to be true. From the strength of evidence related criteria, the multi-national reporting was repeatedly demonstrated to have predictive value [9,10]. In addition, the quality of reports predicted a true signal and, therefore, it is worthwhile to consider this when prioritizing. The fact that the report quality/completeness is associated with true signals might seem counter-intuitive at first, since an increase in the amount of information should not necessarily mean an increase in likelihood of a causal association. A potential explanation might be that the reporter is more likely to provide more complete information about a report once he genuinely believes that the drug is the real culprit. An alternative possibility is that only complete reports can provide the necessary information for a causality assessment that can give rise to a true signal.

Besides focusing resources on true signals, a second purpose of prioritization process is to give precedence to signals with a higher public health impact [11]. Public health impact criteria can have a contribution as high as 50% [16] in decision support frameworks and it might well happen that weaker signals from an aetiological perspective will gain a higher priority, based on their potential impact on public health [11]. When estimating the public health impact, the number of reports and drug exposure data are deemed to be essential and are incorporated in various measures across studies. Waller et al. [11] highlights that in some situations the drug exposure variable might underestimate the importance of the signal when the drug use is limited to a particular sub-population, and suggest that in these situations a correction factor should be applied.

Criteria were rarely tested for predictive value (31%), mainly because it is particularly difficult to create a benchmark against which to perform testing. On the other hand, as previously mentioned, not all prioritization variables need to have predictive value. Most of the criteria are complementary and can be used in combination. As with any decision support system, an increase in the number of variables taken into account might enhance the accuracy of the decision, provided that the variables are fit for purpose and the data is of good quality. It appears that a combination of criteria from all four main key concepts (e.g. novelty, strength of evidence public health impact and general public/media attention) is necessary for a robust decision, however the specific choice of the criteria within these concepts is less straightforward.

CONCLUSIONS

There is a wide range of prioritization criteria described in the literature, however few of them demonstrated predictive value. Many criteria with predictive value were related to strength of evidence category (rapid reporting increase, recent reporting, disproportionate reporting, multinational reporting, quality/degree of completeness reports, presence of narrative in the reports) and to novelty (novelty of the drug and novelty of drug event association). Using these criteria is likely to increase the number of true signals in the post-prioritization set.

There were few attempts at integrating different criteria in decision support frameworks. Five of the frameworks were tested for validity and showed usefulness, while at least three are already in use for prioritization [10,11,15].

We recommend more testing of currently available prioritization criteria and frameworks as this would support creation of a robust evidence-based prioritization process. Testing should be done through comparison with existing prioritization procedures, in order to ensure that important signals are not missed by the updated process.

REFERENCES

1 Guideline on good pharmacovigilance practices (GVP) Annex I -Definitions. 2012.

2 Hauben M, Aronson JK. Defining "signal" and its subtypes in pharmacovigilance based on a systematic review of previous definitions. Drug Saf. Int. J. Med. Toxicol. Drug Exp. 2009;32:99–110.

3 CIOMS Working Group. Practical aspects of signal detection in pharmacovigilance . Geneva; 2010.

4 Meyboom RH, Egberts AC, Edwards IR, Hekster YA, de Koning FH, Gribnau FW. Principles of signal detection in pharmacovigilance. Drug Saf. Int. J. Med. Toxicol. Drug Exp. 1997;16:355–65.

5 Robertson-Steel I. Evolution of triage systems. Emerg. Med. J. EMJ. 2006;23:154–5.

6 Wynder EL. Investigator bias and interviewer bias: the problem of reporting systematic error in epidemiology. J. Clin. Epidemiol. 1994;47:825–7.

7 EMA. Guideline on good pharmacovigilance practices, Module IX – Signal management. 2012.

8 Hauben M, Bate A. Decision support methods for the detection of adverse events in post-marketing data. Drug Discov. Today. 2009;14:343–57.

9 Ståhl M, Lindquist M, Edwards IR, Brown EG. Introducing triage logic as a new strategy for the detection of signals in the WHO Drug Monitoring Database. Pharmacoepidemiol. Drug Saf. 2004;13:355–63.

10 Caster O, Juhlin K, Watson S, Norén GN. Improved Statistical Signal Detection in Pharmacovigilance by Combining Multiple Strength-of-Evidence Aspects in vigiRank : Retrospective Evaluation against Emerging Safety Signals. Drug Saf. Int. J. Med. Toxicol. Drug Exp. 2014;37:617–28.

11 Waller P, Heeley E, Moseley J. Impact analysis of signals detected from spontaneous adverse drug reaction reporting data. Drug Saf. Int. J. Med. Toxicol. Drug Exp. 2005;28:843–50.

12 Naranjo CA, Busto U, Sellers EM, Sandor P, Ruiz I, Roberts EA, et al. A method for estimating the probability of adverse drug reactions. Clin. Pharmacol. Ther. 1981;30:239–245.

13 Meyboom RH, Hekster YA, Egberts AC, Gribnau FW, Edwards IR. Causal or casual? The role of causality assessment in pharmacovigilance. Drug Saf. Int. J. Med. Toxicol. Drug Exp. 1997;17:374–89.

14 Meyboom RHB, Lindquist M, Egberts ACG, Edwards IR. Signal selection and follow-up in pharmacovigilance. Drug Saf. Int. J. Med. Toxicol. Drug Exp. 2002;25:459–65.

15 Seabroke S, Wise L, Waller P. Development of a novel regulatory pharmacovigilance prioritisation system: an evaluation of its performance at the UK Medicines and Healthcare products Regulatory Agency. Drug Saf. 2013;36:1025–32.

16 Levitan B, Yee CL, Russo L, Bayney R, Thomas AP, Klincewicz SL. A model for decision support in signal triage. Drug Saf. Int. J. Med. Toxicol. Drug Exp. 2008;31:727–35.

17 Puijenbroek EPV, Van Grootheest K, Diemont WL, Leufkens HGM, Egberts ACG. Determinants of signal selection in a spontaneous reporting system for adverse drug reactions. Br. J. Clin. Pharmacol. 2001;52:579–586.

18 FDA. Classifying Significant Postmarketing Drug Safety Issues (Internet). 2012. Available from: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM295211.pdf

19 Edwards IR, Lindquist M, Wiholm BE, Napke E. Quality criteria for early signals of possible adverse drug reactions. Lancet. 1990;336:156–8.

20 Waller PC, Lee EH. Responding to drug safety issues. Pharmacoepidemiol. Drug Saf. 1999;8:535–52.

21 Strandell J, Caster O, Hopstadius J, Edwards IR, Norén GN. The development and evaluation of triage algorithms for early discovery of adverse drug interactions. Drug Saf. 2013;36:371–88.

22 FDA. Safety reporting requirements for human drug and biological products: proposed rule. Federal Register 2003 Mar 14; 67 (50) 483-6.

23 Lindquist M, Edwards IR, Bate A, Fucik H, Nunes AM, Ståhl M. From association to alert--a revised approach to international signal analysis. Pharmacoepidemiol. Drug Saf. 1999;8 Suppl 1:S15-25.

24 CIOMS V. Current Challenges in Pharmacovigilance: Pragmatic Approaches. 2001.

25 Heeley E, Waller P, Moseley J. Testing and implementing signal impact analysis in a regulatory setting: results of a pilot study. Drug Saf. Int. J. Med. Toxicol. Drug Exp. 2005;28:901–6.

26 Weber JCP. Epidemiology of adverse reactions to nonsteroidal anti-inflammatory drugs. Adv. Inflamm. Res. :1–7.

27 Hartnell NR, Wilson JP. Replication of the Weber effect using postmarketing adverse event reports voluntarily submitted to the United States Food and Drug Administration. Pharmacotherapy. 2004;24:743–9.

28 Wallenstein EJ, Fife D. Temporal patterns of NSAID spontaneous adverse event reports: the Weber effect revisited. Drug Saf. Int. J. Med. Toxicol. Drug Exp. 2001;24:233–7.

Chapter 6
Towards a risk-based monitoring

6.1

Is patient exposure pre and post-authorization a determinant of the timing and frequency of occurrence of safety issues?

Alexandra C Păcurariu
Christel Hoeve
Peter Arlet
Georgy Genov
Jim Slattery
Miriam CJM Sturkenboom
Sabine MJM Straus

Drug Saf. 2015; 38(12), doi: 10.1002/pds.4359.

ABSTRACT

Background and Objectives: The amount of drug exposure, pre and post approval, is considered to be a direct determinant of knowledge about safety of a drug. A larger pre-approval exposed population is supposed to reduce the risk of unanticipated safety issues post-approval. Post-approval population should also influence the number and occurrences of safety issues. We investigated how the amount of pre and post approval exposure influences the occurrence of safety issues post-approval.

Methods: Analysis was performed on a group of newly approved drugs in Europe, monitored for a median time of median time of 15.8 months. The outcome of interest was the first safety issue occurred in the period. We use a Cox model applied for analysis and we adjusted for drug related characteristics which were considered to be confounders.

Results: The amount of pre-approval exposure was not associated with the risk of safety issues when adjusting for anatomical therapeutical chemical (ATC) class, biological status and treatment duration. The post-approval exposure was associated with the risk of new safety issues (HR=2.44 (95%CI=1.12–5.31)) for drugs with more than 1,000 patient-years of cumulative exposure compared to drugs with less than 1,000 patient years of exposure.

Conclusion: Our results suggest that low pre-approval exposure does not lead to more post-approval safety issues while post-approval exposure influences to some extent the occurrence of safety issues.

INTRODUCTION

Traditionally the lifecycle of a drug is split into two main phases: the pre-approval phase when all of the exposure occurs during randomized clinical trials and the post-approval phase when most exposure occurs during so-called "real-world" use. Despite interest in more iterative approaches that might facilitate the access to innovative drugs for patients with unmet medical needs, [1] the traditional pre- to post dichotomy remains the norm.

Medicines regulators grant a marketing authorization for a new drug based on the assessment of the product's safety, quality and efficacy and the judgment that the benefits outweigh the risks for the target population and in the respective indication. However, at market entry, the knowledge of the product's safety profile is restricted due to the well-known limitations of pre-approval clinical trials [2]. Due to strict inclusion criteria, clinical trials often include a smaller, healthier and more homogenous population than the one for which the drug is intended post-approval. The total pre-approval exposure consists of a median of 1,700 patients [3]. In addition, the follow-up duration may be shorter than the intended drug use and consequently, adverse drug reactions (ADRs) with longer latency may be missed [3].

After approval, the initial safety profile is complemented with ADRs which occur and are detected post-approval, during the use of the product in the "real-world" setting. For this article, the terms occurrence and detection (of a safety issue) will be used interchangeably, the term "detection" being a proxy for "occurrence". Usually the post-approval population is larger and more heterogeneous than pre-approval population and its composition is influenced by various drug characteristics such as intended duration of treatment, drug legal status, target population and approved indication.

The amount of patients exposed to a drug is considered a direct and critical determinant for the knowledge on safety, [4,5] however there are only few studies which have investigated this relationship. One might expect that a larger pre-approval population will lead to a better established safety profile at market entry and consequently fewer safety issues detected post-approval. On the contrary, the post-approval population is supposed to directly influence the number and speed of occurrence of safety issues, mainly by increasing the sample size. Neither of these hypotheses is completely clarified at the moment.

One study conducted in Europe found that a larger pre-approval exposure leads to more serious safety issues post-approval and that the magnitude of post-approval exposure does not influence the occurrence of safety issues [6]. A second study [7] found that an important amount of safety signals occurs during the first 18 months post-approval, regardless of whether the product met a pre-defined exposure threshold of 10,000 patients or not; therefore the authors have concluded that post-approval exposure does not influence markedly the occurrence of safety issues. A third study tried to predict the amount of exposure needed to detect safety signals in electronic healthcare records and showed that the exposure is a function of the minimal detectable risk and background

incidence of the ADRs to be detected [4]. The study showed that the power to detect new safety issues is low for drugs with limited exposure [4].

In view of the limited existing research, the question of how pre and post-approval exposure to a drug influences the occurrence of safety issues post-approval is not yet completely answered. Clarification is relevant for regulators and pharmacovigilance specialists since this would inform policy-making, decisions on the authorization of new medicines, and would enable risk proportionate safety monitoring, with priority given to the drugs that are more likely to be associated with safety issues. This would further support risk management planning by targeting signal detection and post-authorization studies on situations where safety issues are most likely to occur.

This study was designed to address the question whether drug exposure is associated or not with the occurrence of new safety issues. We focused on newly approved drugs, expecting that any influence of exposure will be most pronounced during the first period on the market. Despite limited knowledge about their safety profile, new drugs might have a very rapid market uptake due to unmet medical need [6,8] and they might also be channeled to sicker patients who did not respond to existing alternatives.

Objectives

In this study we investigated if the amount of pre and post approval exposure is associated with the frequency and time to occurrence of safety issues post-approval.

METHODS

Study design and data collection

All innovative drugs, [12] approved in Europe by the European Commission between 1st January 2012–1st January 2016, were retrospectively studied. Both biological and small molecules were included. Scientific and regulatory information about drugs was obtained from the publicly available European Public Assessment Reports (EPARs), [13] which are summaries of drug related data compiled at the moment of authorization.

The main outcome of interest was whether the drug had a new safety issue in the period. We defined a new safety issue as occurrence of a new ADR or an increase of frequency or severity of a known ADR, which led to updates of the SmPCs, so-called "variations". Safety-related variations were identified from the website of the European Medicines Agency (EMA) [14]. For the scope of our research, we only included changes in the sections "Undesirable effects", and "Special warnings and precautions for use". Safety related changes that were the result of an extension of the indication or minor changes as a result of misspellings and correction of errors were not considered as outcome. The official date of completion of the variation was captured as the date of occurrence of the outcome. If mentioned, the source of information (spontaneous databases, clinical trials, other) substantiating the changes was recorded.

The drugs were followed until the earliest occurrence of the following: first new safety issue (addition of ADR to the SmPC), suspension/withdrawal of drug for safety reasons or end of study period.

Pre-approval exposure was defined as the cumulative number of patients and healthy volunteers exposed to the product in all clinical trials conducted as part of the clinical development program, as described in the scientific discussion of the EPARs [13].

Post-approval exposure was defined as the number of patient-years of exposure after approval and was calculated based on sales data. Sales data were obtained from periodic safety update reports provided by the marketing authorization holders. Post-approval exposure is a time dependent variable, and was recorded at fixed time points during a drug lifecycle: every 6 months for the first three years on the market. The values of post-approval exposure at intermediate time points were imputed by a simple linear interpolation between the two closest reported values.

Other characteristics collected for the included drugs were ATC class, biological status, orphan status and type of approval (e.g., regular approval vs. conditional approvals or approval under exceptional circumstances). In Europe there are two different types of approval, besides the regular one: exceptional circumstances (EC) [15] and conditional approval (CA), [16] both allowing drugs to be approved on more limited clinical datasets in order to fulfil unmet medical need.

Analysis

Descriptive statistics were used to characterize the drugs. Differences in baseline characteristics across exposure categories were tested with chi-square tests for categorical data and Mann Whitney-U test for continuous data.

Since the hazard of finding a new ADR is time dependent, [17,18] Kaplan-Meier survival analysis was used to estimate the probability of the occurrence of a first safety issue post-approval at 12 and 36 months respectively as well as the time from the approval date to a first safety issue.

We used Cox proportional hazards regression model to investigate the association between exposure variables and the hazard of having a first safety issue. Pre-approval exposure was incorporated as a fixed variable, collected at the beginning of follow up (e.g., time when the product was placed on the market), while post-approval exposure was incorporated as a time dependent variable, with the last value collected at time of the outcome or censoring. Both exposure variables were transformed from continuous into categorical, to make the results easier to interpret from a clinical perspective.

Sensitivity analysis was performed by including all variables which differed significantly between exposure categories and which were considered to be potential confounders based on clinical judgment and previous research [6,10,19]. SPSS version 21 and R version 3.3.1 were used for analysis.

RESULTS

A total of 233 drugs were centrally approved in European Union during the study period. After applying the exclusion criteria, 136 drugs were retained in the analysis and monitored for a median time of 15.8 months (interquartile range (IQR): 8–26) (see Figure 1).

The three most prevalent drug categories were Antineoplastic and immune-modulating agents (32%), Alimentary tract and metabolism (14%) and Anti-infectives for systemic use (12%). A third of the drugs were orphans (29%) and slightly more than one third (34%) were biological drugs. Other key characteristics of the included drugs are presented in Table 1.

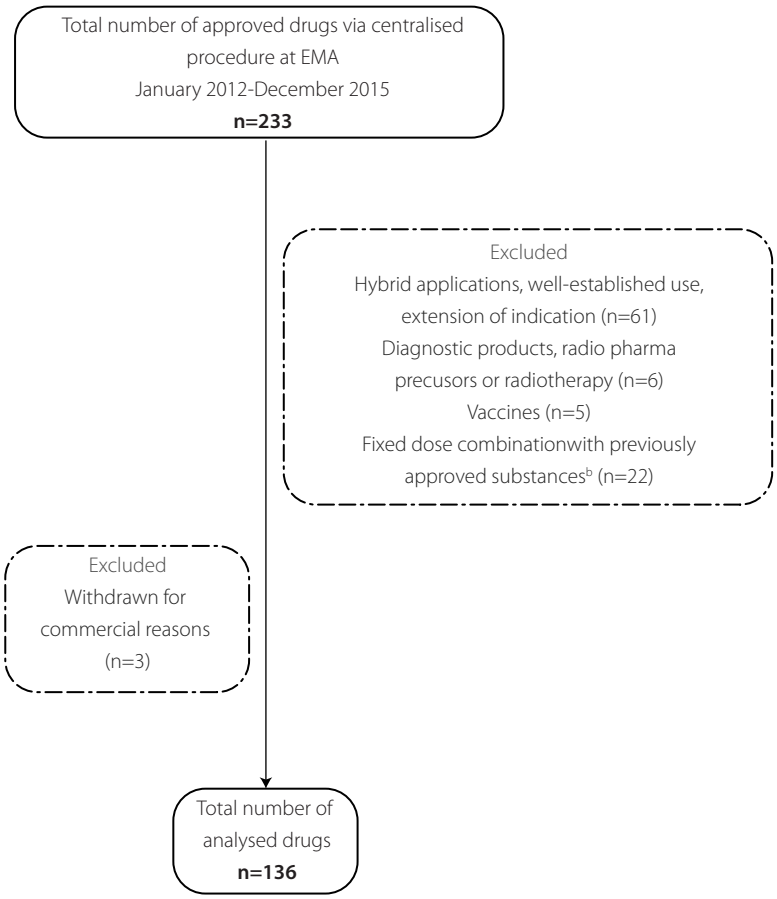


Figure 1: Attrition chart for study drugs

a- Approved under article 10(a), 10(c) or 10(3) of Directive 2001/83/EC

b- Approved under Article 10(b) of Directive 2001/83/EC

Table 1: Key characteristics of investigated drugs

Variable	All drugs (n=136)
ATC class [a]	
Antineoplastic and immune- modulating agents	44 (32%)
Alimentary tract and metabolism	19 (14%)
Anti-infectives for systemic use	17 (12%)
Other	56 (41%)
Indication [b]	
Type 2 diabetes mellitus	17 (12%)
Chronic obstructive pulmonary disease	8 (6%)
Chronic hepatitis C	6 (4%)
HIV infection	6 (4%)
Multiple sclerosis	6 (4%)
Prostate cancer	5 (4%)
Treatment duration	
Short term	14 (10%)
Medium term	49 (36%)
Long term	71 (52%)
Biological (Y)	46 (34%)
Orphan drugs (Y)	40 (29%)
Exposure related variables	
Pre-approval exposure (patients), median, (IQR)	1,111 (445–2,166)
0-500	39 (29%)
500-2,000	59 (43%)
>2,000	38 (28%)
Post-approval exposure (patient-years) median, (IQR)	1,076 (89–12,560)
<1,000	36 (50%)
1,000-10,000	15 (21%)
>10,000	21 (29%)
Procedural aspects	
EC and CA registration (Y), n (%)	20 (15%)
Total follow up time, months, median, IQR (range)[c]	15.8 (8–26)
Number of newly added ADRs per drug, median, range[d]	2 (0–11)

ATC=anatomic therapeutic chemical; IQR=interquartile range; EC=exceptional circumstances, CA=conditional approval; [a] Three classes (A, J, L) comprise more than 50% of the drugs and are therefore separately presented, the rest are grouped in the category "other"; [b] Only indications with 5 occurrences are presented; [c] Follow up time is measured until first outcome, study end or drug withdrawal; [d] Multiple ADRs added within the same regulatory procedure (grouped variations) were included as separate occurrences

The drugs had a median pre-approval exposure of 1,111 patients (IQR: 445–2,166) and a median cumulative post-approval exposure of 1,076 patient-years (IQR: 89–2,560) (see Figure 2).

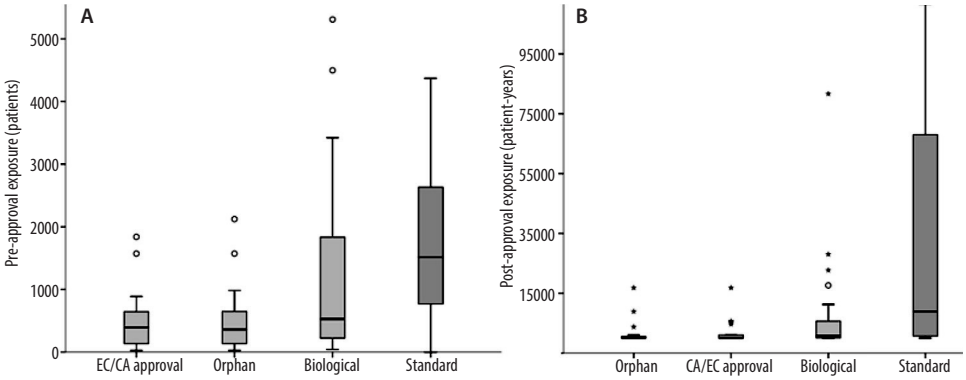


Figure 2: Pre-approval and post-approval exposure across different drug categories

A pre-approval exposure; B- post-approval exposure. Standard category is represented by drugs which do not belong to any of mentioned categories (i.e. non-orphan, non-biological and approved via a regular procedure). The other three categories are not mutually exclusive. The dots represent outliers.

During the period under review, 104 new ADRs were added to the SmPC of 35 drugs. Out of these, four (3.8%) were not new but known ADRs with increased frequency of occurrence. The most frequently identified ADRs referred to gastrointestinal disorders (15%), skin and subcutaneous tissue disorders (11.4%) and immune system disorders (11.4%). Three withdrawals for commercial reasons took place during the study period and no drugs were withdrawn because of safety reasons.

The cumulative hazard (Kaplan-Meier probability) of a drug having its first ADRs added to the SmPC was 10.4% (95%CI=4.6%–15.9%) in the first year after approval and increased to 47.7% (95%CI=31%–60.3%), three years after approval.

The pre-approval exposure appeared to be correlated with the risk of having an ADR added to the SmPC post-approval, in the first 2 years after approval; in the unadjusted model the risk being almost double (hazard rate (HR)=2.0 (95%CI =1.17–3.44)) for drugs with more than 2,000 patients exposed in clinical trials, compared with drugs with less exposure. However, the relationship was not maintained in the adjusted model (see Table 2), where the HR decreased to 1.28 (95%CI=0.55–2.46).

The amount of post-approval exposure was associated with the risk of having an ADR added to the SmPC in the first 2 years after approval; drugs with more than 1,000 patient-years of cumulative exposure had a higher risk of SmPC updates (HR=2.58 (95%CI=1.39–4.77)) as compared to drugs with less than 1,000 patient years of exposure. At very high levels of post-exposure (>10,000 patient-years) the risk starts to decrease but is still higher than for drugs in the lowest exposure category. The relationship was maintained after adjustment; HRs just decreased slightly compared to the non-adjusted model (see Table 2).

Table 2: Results of Cox proportional hazards regression analysis

Exposure related variables	HR crude model (95%CI)	HR adjusted model [a] (95%CI)
Pre-approval exposure (patients)		
≤500	*Reference*	*Reference*
>500 and <2,000	1.18 (0.69–2.01)	1.17 (0.55–2.46)
≥2,000	2.00 (1.17–3.43)	1.28 (0.54–3.04)
Post-approval exposure (patient-years)		
<1,000	*Reference*	*Reference*
≥1,000 and <10,000	2.58 (1.39–4.77)	2.44 (1.12–5.31)
≥10,000	0.86 (1.01–3.44)	1.29 (0.47–0.97)

HR=hazard rates; [a] adjusted for post/pre-exposure, biological status and ATC class

A secondary analysis using other variables potentially associated both with exposure and the safety issues, showed that drugs which are antineoplastic and immune-modulating agents, drugs indicated for acute treatment and non-biological drugs also had an increased risk of safety issues post-approval, independent of the exposure.

DISCUSSION

We initiated this study with the aim to test whether the hypothesis that the amount of patients exposed to a drug is associated with the occurrence of safety issues. So far, only one study [5] concluded that the amount of patients exposed is an important determinant of the statistical power for detection of safety signals post-marketing. Solely from a statistical perspective this might sound obvious. However, the post-approval setting is complex and therefore the amount of exposure might influence the safety in multiple and sometimes contradictory ways.

We were expecting that pre-approval exposure will be inversely correlated with post-approval safety issues due to the fact that a well-established safety profile at market entry will leave little room for new signals to be discovered post-approval. In our study, the initially increased risk for products with higher pre-approval exposure disappeared after adjustment for other drug related characteristics such as ATC class, treatment duration and biological status, suggesting that the amount of pre-approval exposure is not an independent risk factor for occurrence of post-approval safety issues. An explanation for this finding might be that the variation in pre-approval exposure is largely determined by the nature of the product and once we have accounted for this, the apparent relationship to post-approval safety disappears; therefore it is more likely that pre-approval exposure is a confounder or intermediate rather that an independent risk factor for safety issues.

Secondly, we hypothesized that post-approval exposure will be directly correlated with post-approval safety issues; the higher the exposure, the more safety issues will be found. Our results confirmed this second hypothesis; drugs with more than 1,000 patient-years of exposure had a 2.5 higher risk of having a safety related change in the SmPC compared to drugs with less exposure. This relationship was observed up to a certain exposure threshold (approximately 10,000 patient-years), after which the risk reached a plateau and then started to decrease, suggesting a non-monotonic relationship between post-approval exposure and occurrence of safety issues. The relationship remained similar after adjustment for other variables, increasing the robustness of our finding. Our results are opposing those from an earlier study which failed to find an association between post-approval exposure and occurrence of safety issues [6]. The discrepancy might be explained by use of different outcome definitions and different cut-off points for the exposure categories. Moreover, in the study by Mol et al., [6] the exposure data was collected only from one country (the Netherlands), while our exposure variable reflects worldwide exposure. Another study, [7] found that an important amount of safety signals occurs during the first 18-months post-approval, regardless of whether the product met a predefined threshold of 10,000-patient or not. This is partly in line with our results since for higher values of exposure we also found that the risk starts to decrease.

A secondary finding from our study referred to other drug related variables which might influence the hazard of having a safety issue. For example, some ATC classes might be associated with an increased risk of safety issues. In our study this ATC class was Antineoplastic and immune-modulating agents. The same result was reported before [6,9]. This finding must be interpreted with caution since a contributor factor might have been the over representation of this ATC class in the group of drugs even from the study start. Drugs indicated for acute treatment and non-biological drugs also had a slightly increased risk of safety issues, independent of exposure. On the contrary, the orphan drugs did not appear to have an increased risk, this also being in line with previous research [9]. The same lack of increased risk was found for drugs with exceptional or conditional approval [10].

Another secondary finding was that the most frequent ADRs added to the SmPC of new drugs in the first 2 years on the market were related to gastrointestinal disorders, skin and subcutaneous system disorders and immune system disorders. This is different from what was reported for all drugs (including both new and old drugs) for which the most common signals were related to general disorders and administration site conditions, nervous system disorders and gastrointestinal disorders [11]. Considering this difference, one might speculate that the type of reactions added to the SmPC change over time, and they also differ between new and old drugs, but we did not test formally this hypothesis in our study.

A few limitations of our study need to be mentioned. A first limitation is the relatively small sample size and few safety issues leading to lack of precision of estimates and potential false negative findings. A second limitation is the fact that we used the time of approval as a proxy for time of marketing, since the exact time of marketing is difficult to obtain. This led to a potential initial follow-up period during which no exposure occurred, however this is expected to be non-differential across groups and

therefore likely to affect just the precision and not the magnitude of the estimate. A third limitation pertains to the post approval variable; due to the fact that only values at specific time points were available we needed to impute missing values, which might have reduced the variability of data or have biased the variable in other ways. However we consider that the imputation method we used (simple linear interpolation between the closest existing values) is adequate, considering the short time period between the known exposure time points.

A strength of our study was the use of a statistical model which took the follow-up time into account, since both exposure and safety outcomes are known to be time dependent. To our knowledge this is the first study which investigated the relationship between exposure and safety outcome as its main analysis.

In conclusion, low pre-approval exposure does not lead to faster or more frequent post-approval safety issues, being rather an intermediate factor. The finding that pre-approval exposure does not influence the number of safety issues discovered post-approval confirms the idea that increasing the size of pre-approval population will not answer all the safety questions.

On the other hand, the amount of post-approval exposure is a clear determinant of the appearance of post-approval safety issues, at least in the first period on the market and until a certain exposure threshold is reached. Considering this finding, our recommendation is that the amount of patients likely to be exposed to a drug post-approval should be considered when planning the post-approval safety monitoring. Special consideration should be given to drugs with potential for high and rapid market uptake and to antineoplastic drugs. We would also recommend extending the study of the relationship between exposure and the timing, number and type of safety issues to a larger drug sample, as a better understanding of this relationship is likely to drive process improvement in risk management planning.

We consider that our study contributes to measuring the impact of pharmacovigilance, a key activity in driving process improvement.

REFERENCES

1 Eichler, H.-G. et al. Adaptive licensing: taking the next step in the evolution of drug approval. Clin. Pharmacol. Ther. 91, 426–437 (2012).

2 Vandenbroucke, J. P. Observational research, randomised trials, and two views of medical science. PLoS Med. 5, e67 (2008).

3 Duijnhoven, R. G. et al. Number of Patients Studied Prior to Approval of New Medicines: A Database Analysis. PLoS Med 10, e1001407 (2013).

4 Coloma, P. M. et al. Electronic healthcare databases for active drug safety surveillance: is there enough leverage? Pharmacoepidemiol. Drug Saf. 21, 611–621 (2012).

5 Hammond, I. W., Gibbs, T. G., Seifert, H. A. & Rich, D. S. Database size and power to detect safety signals in pharmacovigilance. Expert Opin. Drug Saf. 6, 713–721 (2007).

6 Mol, P. G. M. et al. Post-approval safety issues with innovative drugs: a European cohort study. Drug Saf. 36, 1105–1115 (2013).

7 Sekine, S. et al. Assessment of the impact of scheduled postmarketing safety summary analyses on regulatory actions. Clin. Pharmacol. Ther. 100, 102–108 (2016).

8 Lasser, K. E. et al. Timing of new black box warnings and withdrawals for prescription medications. JAMA 287, 2215–2220 (2002).

9 Heemstra, H. E., Giezen, T. J., Mantel-Teeuwisse, A. K., Vrueh, R. L. A. de & Leufkens, H. G. M. Safety-related regulatory actions for orphan drugs in the US and EU: a cohort study. Drug Saf. 33, 127–137 (2010).

10 Arnardottir, A. H. et al. Additional safety risk to exceptionally approved drugs in Europe?: Additional safety risk to exceptionally approved drugs in Europe? Br. J. Clin. Pharmacol. 72, 490–499 (2011).

11 Blake, K. V. et al. Enhanced Paediatric Pharmacovigilance at the European Medicines Agency: A Novel Query Applied to Adverse Drug Reaction Reports. Paediatr. Drugs 18, 55–63 (2016).

12 Eichler, H.-G., Aronsson, B., Abadie, E. & Salmonson, T. New drug approval success rate in Europe in 2009. Nat. Rev. Drug Discov. 9, 355–356 (2010).

13 European Medicines Agency European Public Assessment Report. (2016).at <http://www.ema.europa.eu/ema/index.jsp?curl=/pages/medicines/landing/epar_search.jsp&mid=WC0b01ac058001d124>

14 European Medicines Agency Procedural steps taken and scientific information after authorisation. (2016). at <http://www.ema.europa.eu/ema/index.jsp?curl=pages/medicines/human/medicines/000481/human_med_000822.jsp&mid=WC0b01ac058001d124>

15 Committee for Medicinal Products for Human Use (CHMP).Guideline on procedures for the granting of a marketing authorisation under exceptional circumstances, pursuant to article 14 (8) of regulation (EC) No 726/2004 (EMEA/357981/2005). 2005. (2005).

16 European Commission. Commission regulation (EC) No 507/2006 of 29 March 2006 on the conditional marketing authorisation for medicinal products for human use falling within the scope of Regulation (EC) No 726/2004 of the European Parliament and of the Council. 2006. (2006).

17 Norén, G. N., Caster, O., Juhlin, K. & Lindquist, M. Zoo or savannah? Choice of training ground for evidence-based pharmacovigilance. Drug Saf. 37, 655–659 (2014).

18 Hochberg, A. & Hauben, M. Time-to-Signal Comparison for Drug Safety Data-Mining Algorithms vs. Traditional Signaling Criteria. Clin. Pharmacol. 38 Ther. 85, 600–606 (2009).

19 Giezen, T. J. et al. Safety-related regulatory actions for biologicals approved in the United States and the European Union. JAMA J. Am. Med. Assoc. 300, 1887–1896 (2008).

Chapter 7

Summary, general discussion and future perspectives

SUMMARY OF FINDINGS

This thesis comprises studies that aim to answer questions related to sub-domains of signal management and regulatory science: data sources, detection methods, and the prioritisation process.

We started this thesis with a review of signals detected in the EU, as major legislative changes were adopted in 2010 and implemented in 2012. Based on the review of signals discussed in the PRAC we found that, similar to the USA, the most frequently used data source was spontaneous reports (72% of cases), see **Chapter 2** [1]. The most frequently discussed signals were related to skin and subcutaneous tissue disorders (12.8%), nervous system disorders (10.4%) and cardiac disorders and immune system disorders (6.4%). We observed that the mean time between a new signal occurrence and a decision by PRAC was 2.5 months, with 42.8% of all decisions taken during the first meeting. For 57.2% of the signals, additional information about the signal was requested after the first discussion in the PRAC. Several actions can be taken in relation to a newly validated signal. The decision most often taken was a change in the product information (54%). The decisions to start a referral and to send a direct healthcare professional communication, which are usually taken for more serious and urgent issues, were not that common (9.4 % and 7.3%); however they were decided more rapidly (1.8 months and 1.7 months, respectively).

Although spontaneous reports are still the mainstay for signal generation, their limitations are well-known and in view of these, additional data sources to augment signal detection are constantly being sought. After several large impact drug safety issues in the beginning of the new century the Institute of Medicine recognized that reliance on spontaneous reports alone may not be enough and suggested that potential electronic healthcare records (EHRs) may be used as alternative source [2]. To explore how EHRs can complement spontaneous reports in safety surveillance we investigated the role of European EHRs as a signal generation source and compared the ability to detect known associations between the databases participating in the EU-ADR project with Eudravigilance in **Chapter 3.1**. In this study, spontaneous reporting systems SRS outperformed EHRs in signal detection for events with a rare to very rare background incidence and ADRs which are easily attributable to drugs, e.g. bullous eruption and acute pancreatitis. On the other hand, there were events (e.g. hip fracture) for which detection in EHR led to better results.

An important issue to be taken into account when deciding which data source to use for signal generation is the noise-to-signal ratio. The false-positives associated with detection of signals in each data source (EU-ADR and EudraVigilance) depend both on the type of events and on the data source. Across the events, the lowest number of false positives was found for upper gastrointestinal bleeding and acute pancreatitis and the highest for bullous eruption. In terms of databases, for all events, it was more costly (e.g., more false positives need to be discarded) to detect safety signals in EHRs than in spontaneous reporting systems. In conclusion, an EHR-based system may have additional value for signal detection, alongside already established systems, especially for adverse

events with high background incidence. At this moment, spontaneous reporting systems appear to be more cost-effective than EHRs.

In **Chapter 3.2**, we described the use of EHR for signal evaluation by studying the association between triptans and ischaemic colitis. Due to the rarity of the outcome, only a few anecdotal reports have been reported to date. In the THIN primary care database, which covers ~6% of the UK population, we identified only 41 incident cases of ischemic colitis within a migraine cohort of approximately 300,000 patients. Use of triptans in the 12 months before the event, doubled the risk of ischemic colitis compared to no use (OR=2.29, 95%CI 1.02–5.15). After 12 months, the risk decreased to (OR=1.90, 95%CI 0.44–8.13). The lack of power restricted us to further investigate alternative exposure windows.

Apart from using different data sources as the basis for signal detection/evaluation, we investigated how to improve the performance of existing signal detection methods in paediatrics. In **Chapter 4.1** we tested the impact of age stratification or adjustment on signal detection performance. Two algorithms (PRR and EBGM) were applied to reports in children < 18 years of age, where the impact of age is expected to be higher due to organ maturation by age [3]. Performance was similar between methods (measured by area under the curve (AUC). Age adjustment decreased the performance of both methods compared to the unadjusted scenario. Stratification by age group, unmasked new signals in specific age strata, which were not discovered in the overall paediatric population analysis. We observed that age can be an effect modifier and therefore should be stratified for.

To increase efficiency and reduce the noise-to-signal ratio it may be useful to understand which type of drugs have a higher risk of new safety signals. Drugs for which a new signal was discussed at PRAC during the period September 2012 until December 2013, had a lower median post-marketing drug age (12.3 years), than other drugs existing on European market which did not have signals in the same period (19.7 years). However, it is not just the most recently marketed drugs that have safety signals: 58% of drugs with a signal discussed at PRAC had been marketed for more than 10 years, (range: 0.54–67.9), showing well-establish drugs might also generate signals.

Apart from drug age on the market, we explored other predictors for the frequency of safety signals in **Chapter 6.1**. ATC class Antineoplastic and immunomodulation drugs, drugs dedicated for acute treatment and non-biological drugs had a higher risk of safety issues post-approval. We found that the amount of pre-approval exposure time was not associated with the risk of detecting new safety issues. The opposite was true for the post-approval exposure. Drugs with higher post-approval exposure; namely more than 1,000 patient-years of cumulative exposure in the first two years on the market had a 2.4 fold increase in risk of having a new safety issue compared to drugs with less exposure in the same period. After a certain exposure threshold (approximated at 10,000 patient-years), the risk reaches a plateau and then decreases.

In order to inform regulators which criteria to consider when prioritizing safety signals, we reviewed different triage criteria (**Chapter 5.1**). We identified 48 different prioritisation criteria, which can be broadly categorized into four groups: novelty, public health impact, strength of evidence and public and media attention. More than half of the criteria (63%), referred to strength of evidence while 19% related to public health impact, 14% to general public and media attention and 4% to novelty of the drug event association. Fifteen criteria were tested for predictive value (e.g., they can predict true signals) with 11 showing positive results, most of them from the strength of evidence category. The multi-national reporting, quality and completeness of reports had predictive value. In contrast, the reporter qualification variable (whether the reporter is medically vs. non-medically qualified) was not associated with true signals. Six decision making frameworks were found in the literature, five of the frameworks were tested for validity and showed usefulness, while at least three are already in use for prioritization in various organizations.

Chapter 7 contains a general discussion about the implications of the findings from this thesis, as well as recommendations for current regulatory practice and future research.

GENERAL DISCUSSION

Regulatory decision making and science

The decisions of regulatory agencies should be grounded in science. Regulatory science, a newly emerging field, is *'the science of developing and validating new standards and tools to evaluate and assess the benefits and risks of medicinal products, facilitating sound and transparent decision making'.* [4] In an Editorial from 2011, the FDA commissioner stressed that it is vital for regulators to develop new tools, standards and approaches that efficiently and consistently assess the safety, efficacy and performance of products, and that this field has long been overlooked and underfunded. [5]

Regulatory decision making is a fragile balance between two choices that are not always easy to reconcile [6] and often need to be done in a realm of uncertainty due to limited data. The first choice is protection of public health, by approving only drugs whose benefits outweigh the risks and the other choice is to ensure timely access to therapeutic innovations for patients in need. The information available before a medicine is authorised is limited and 'real-world' use has an important contribution to the knowledge of the drug safety profile [7]. As Stricker and Psaty put it in an early editorial, 'widespread marketing of a new drug is, in fact, a large experiment on a population' [7]. Reflecting the need for a life time approach to monitor the benefit-risk profile of a drug, the Institute of Medicine in the USA states *'The approval decision does not represent a singular moment of clarity about the risks and benefits associated with preapproval clinical trials do not obviate continuing formal evaluations after approval'* [2].

Table 1: High profile drugs withdrawn/suspended in Europe during 2005-2015 for safety reasons

Substance name	Brand name	Time on the market	Indication in European Union	Reason for withdrawal	Main data source for detection of the safety issue
Thioridazine	(Mellaril®)*	1958-2005	Neuroleptic	Cardiac disorder, QT prolongation	RCTs and meta-analysis
Valdecoxib	(Bextra®)	2003-2005	NSAIDs	Cardiovascular and cutaneous disorders	RCTs and meta-analysis
Rosiglitazone	(Avandia®)*	2000-2010	Type II diabetes	Cardiovascular events, including congestive heart failure, myocardial infarction and stroke	RCTs, observational studies and meta-analyses
Rimonabant	(Acomplia®)	2006-2009	Treatment of obesity	Depression and increased suicide ideation/ attempts	RCTs
Benfluorex	(Mediator®, Modulator®)	1974-2009	Anorectic and hypolipidemic	Heart valve disease and pulmonary hypertension	Spontaneous reports, RCTs and case control studies
Sibutramine	(Reductil®)	1999-2010	Treatment of obesity	Increased risk of serious cardiovascular events as heart attack or stroke	RCTs
Aprotinine	(Trasylol®) **	1974-2007	To reduce perioperative blood loss during CABG	Increased risk of death in high risk cardiac surgery patients.	Interim analysis of clinical trials
Ximelagatran		2003-2006	Anticoagulant	Hepatotoxicity	RCTs

RCTs=randomized clinical trials; NSAID= non steroidal anti-inflammatory drug; CABG=coronary artery bypass graft;
Note: The basis for withdrawal and authorized indications were obtained from European Medicines Agency website and press releases, as well as other information available from the public domain.
* Still marketed in US, subject to black box warnings
**The product was re-authorized after analysis of final results from clinical trials

The safety related drug withdrawals in the last decades have reiterated the importance of a life-time approach to safety surveillance [8] (see Table 1). An example of a more flexible approach that is aligned with the need to monitor the effects over the entire drug lifetime is adaptive licensing a *'prospectively planned, flexible approach to regulation of drugs and biologics'* [9]. Adaptive licensing involves iterative phases of evidence gathering followed by regulatory evaluation and a drug-tailored approval [9]. Basically, by using adaptive licensing, the existence of uncertainty is acknowledged, and certain decisions can be postponed until more evidence is available. Not only regulators, but also patients, pharmaceutical industry and reimbursement bodies must balance uncertainties about the benefits and risks of drugs, each from their own perspective, while considering alternative treatment opportunities [10].

This thesis aimed to generate data tto support evidence-based decision making. As mentioned already in the introduction, we classified the areas in signal management addressed in this thesis in three categories: data sources, methods of detection and prioritisation process.

Data sources for signal management: spontaneous reports vs. electronic health care records

In this thesis we investigated different sources for signal detection: spontaneous reports and EHRs. Spontaneous reports are important for detection of safety signals post-marketing and they are, in many cases, the only source of evidence [11]. Spontaneous reports are efficient in identifying rare serious adverse events in a population. An example given by Stricker in a 2007 editorial is that in a country as UK, with 60 million people, a 1% cumulative exposure to a drug will lead to 600.000 people being exposed; even a rare event, with an incidence of 1:10 000 might be spontaneously reported and detected and in this case, especially if is recognized as drug-induced [7]. To identify such rare events in EHRs systems, very large populations (several hundred thousand or even several million people) and long follow up times would be needed [2]. Despite the efficiency of SRSs, there are many limitations to their use, as described in the **Introduction**. Not all events are equally detectable in SRSs, some of them might be more effectively monitored by using other data sources [12]. The rise of safety related warnings and product withdrawals as the notorious rofecoxib, led to investigation of alternative data sources for signal detection. It was suggested in the Institute of Medicine that alternative data sources such as EHRs might be used for detecting signals more rapidly [2].

Although initially not specifically designed for research, EHRs are used, beyond their primary administrative and clinical aim, for observational studies in the post-marketing setting. EHRs, either medical records or claims databases, have been evaluated as sources for signal detection and evaluation. Compared to spontaneous reports, these data sources provide a proper population and exposure base, more clinical information and have a longitudinal dimension [13]. The current advances in information technology offer the potential to access and integrate these data and generate evidence in a timely and meaningful way. Although they are widely used already for hypothesis testing, therefore for signal evaluation and assessment, the use of EHRs in signal detection

is still in its infancy. The potential utility of EHR for complementing signal detection was tested by multiple research groups during 2008-2016 [13–15] with different results.

The Observational Medical Outcomes Partnership (OMOP) project aimed to evaluate the performance of various analytical methods for signal detection, using a common data model on different data sources in the USA. With the aid of a specific reference set, the OMOP team evaluated a range of detection methods in different EHR databases [14]. They found that by holding the analysis constant, different data sources may yield different estimates. The opposite was also true by holding database constant, different analytic design choices may yield different estimates. A third finding was that the methods have quite a broad empirical performance, which can be improved with proper calibration [14]. The EU-ADR project tested 10 different signal detection methods in EU-ADR, an EHR network comprising over 20 million subjects from 7 databases across 3 European countries. A reference standard of known ADRs was used to test the performance of the methods. The highest performance was achieved by either longitudinal gamma poisson shrinker (LGPS) or case-control. [16] In addition, a method for removal of protopathic bias was tested – longitudinal evaluation of observational profiles of adverse events related to drugs (LEOPARD). LEOPARD increased the overall performance, but also flagged several known ADRs as caused by protopathic bias [16].

Trifiro et al. compared signal detection performed in the EHR datasets in EU-ADR with two spontaneous reporting databases (FAERS and Vigibase) [17]. They focused on six events of interest (bullous eruptions, acute renal failure, acute myocardial infarction, anaphylactic shock, rhabdomyolysis, and upper gastrointestinal bleeding) and concluded that the performance of signal detection in both EHR and SRSs varies across events. A similar conclusion was reached by Harpaz et al who compared the FAERS and OMOP [18], after performing a similar exercise in FAERS database. Harpaz et al compared MGPS, PRR, ROR and logistic regression. AUC was used as a performance metric. Examination of AUCs by event reveals that the methods applied in FAERS are most effective in identifying ADEs relating to gastrointestinal bleeding and acute renal failure, and least effective in signalling ADEs relating to acute myocardial infarction. For the outcome of myocardial infarction, detection in the OMOP network of claims databases had better performance.

The PROTECT initiative in Europe also explored signal detection methods tailored to EHRs [19], as described in the **Introduction**. Their applied method (temporal pattern discovery) was more conservative (more specific, less sensitive) identifying a lower number of drug-adverse event pairs than the reference method [20].

In our comparison of signal detection in EU-ADR versus EudraVigilance across five events of interest: bullous eruption, acute pancreatitis, upper gastrointestinal bleeding, hip fracture and acute myocardial infarction, we found out that the capacity of EU-ADR and EudraVigilance to detect signals differs depending on the nature and background incidence of the ADR investigated. EudraVigilance performed well in case of rare events such as bullous eruption and acute pancreatitis. This might not be unexpected since spontaneous reporting systems were specifically constructed

for efficient safety monitoring, while for EHR this is a secondary use. Additional explanations are the lower catchment population for EU-ADR compared to Eudravigilance (almost 10 times lower) and the fact that EudraVigilance has a worldwide coverage, whereas the source population in EU-ADR covers only 30 million persons. The EU-ADR network was most useful in identifying multi-factorial and more frequent events such as hip fractures. Hip fracture is multifactorial and occurs relatively common in the population, with an incidence around 117/100,000 person years [21]. These factors (common occurrence and multi-factorial aetiology) might contribute to a decreased recognition of the event as being an ADR and therefore to a decreased spontaneous reporting. Consequently, this type of events will not be easily recognized as drug-induced and therefore not reported. An illustration of this was the lack of recognition of the myocardial infarction associated with rofecoxib (see **Introduction**). Both systems detected gastrointestinal bleeding reasonably well and myocardial infarction rather poorly. If we extrapolate the findings, this means that there is a correlation between the background incidence of the ADRs and the capacity of each database to generate signals. We tested this and the correlation was statistically significant for EudraVigilance, but not for EU-ADR, which may be due to the low number of events tested. Applicability of our findings to a broader range of events requires further investigation. If this finding is reproduced, it might serve as a basis for tailored made signal detection: spontaneous reporting systems will be used only for certain events, and complemented by EHRs for the others.

Signal detection methods

Given the existing data source, another important question in signal detection is how we can improve the performance of detection methods by simultaneously increasing sensitivity and specificity. Especially in the area of paediatrics improvement of signal detection methods has been lagging behind, as initially children were not looked at separately [22]. Paediatric patients may differ from adults regarding the pattern of drug use and ADRs occurrence [22-24]. Applying the same methods as for the adult population may lead to masking of signals [25,26].

We investigated the impact of age on signal detection methods in paediatrics. Stratification is the correct method to inspect confounding and/or effect-measure modification across various strata [27]. Through stratification we create categories in which the confounding factor does not vary or varies very slightly. Afterwards, there are two ways to summarize information across strata- either by reporting the estimate in each stratum (subgrouping or stratification) or reporting an adjusted overall estimate (adjustment) [27]. Effect modification differs from confounding in several ways: whereas confounding is a bias that needs to be identified and removed from the effect estimate, effect modification is based on biology (e.g, higher or less susceptibility by age due to differences in organ maturation) and should be observed and described but not adjusted [27]. Our study in paediatrics showed that adjustment for age decreases overall the performance of signal detection, both in sensitivity and specificity. On the other hand, stratification by age led to unmasking of new signals in certain age categories. This supports the hypothesis that beyond a confounder, age may be an effect modifier in signal detection in children.

Our finding was in line with that of Seabroke et al. [28], who also found that age-subgroup analysis performs better than crude analysis in both sensitivity and precision. Other studies have also found that stratification may perform better than adjustment [25,29]. In one study conducted within the PROTECT consortium, signal detection performance (evaluated by sensitivity and precision) was compared for stratified, subgroup and unadjusted analyses within five spontaneous report databases (VigiBase, EudraVigilance, UK regulatory agency database, GlaxoSmithKline's and Astra Zeneca's safety databases). Analyses were repeated for a range of covariates: age, sex, country/region of origin, calendar time period, event seriousness, vaccine/non-vaccine, reporter qualification and report source. The most relevant variables were age and region/country of origin [28]. Subgroup analysis performed better than crude analysis in both sensitivity and precision, while stratified analysis did not. Subgrouping by type of drug (vaccines/non-vaccines) led to mixed results: it has resulted in a decrease in both precision and sensitivity [28]. One study performed on a large international spontaneous reports database (Vigibase), found that vaccines have a large and mathematically predictable impact on signal detection in the paediatric population [29]. Another study performed on a smaller company-owned vaccine-specific database [30] found a rather modest masking effect. In conclusion, with regards to vaccines stratification, the evidence is mixed.

The risk of over-stratification if too many variables are adjusted for was highlighted by Hopstadius et al. [31] in a simulation study performed in the WHO database where they compared crude and adjusted signal detection methods based on random allocation of reports to a set of strata with a realistic distribution of stratum sizes. The study demonstrated that adjustment led to underestimation of effects relative to the crude analysis, in the presence of very small strata. If strata are large enough, this tendency can be avoided and stratification might be useful.

Based on our results, and the sum of existing evidence, we conclude that routine use of age adjustment in paediatric signal detection is not recommended while stratification can be used on top of unadjusted analysis, to avoid masking of signals.

Drug utilisation data and its role in pharmacovigilance

Drug utilisation data are an important aspect for pharmacovigilance as clinical trials have limitations with regards to safety in terms of patient exposure and length of follow-up [32]. Knowledge of how a drug is used in real life in terms of actual usage, dosages, polypharmacy, helps estimating the absolute risks and public health impact and also informs risk minimisation strategies.

A study which investigated medicines approved between 2000 and 2010 found that the median total number of patients studied before approval was 1,708 ([IQR] 968-3,195) for standard medicines and 438 (IQR 132-915) for orphan medicines [32]. In addition, the same study found that for medicines intended for chronic use, the number of patients studied before marketing is insufficient [32]. The study concluded that both safety and efficacy require continued study after approval.

At the moment, information about drug exposure post-marketing is not very easy to collect, especially when multiple countries or regions are involved. Firstly, drug utilisation data can be collected at different points in the drug use chain: sales from manufacturers, dispensing, prescription or reimbursement. Secondly, the data might be collected by different parties: governmental agencies, insurance companies or pharmacies and can be recorded in different units. The PROTECT project, started to address this problem by building an inventory on drug consumption databases across Europe, which was finalized in February 2015 and serve as a comprehensive source of information [33].

Drugs differ quite a lot in terms of drug exposure, both qualitatively and quantitatively. This depends on indication(s), patterns of use and healthcare policies in the countries where those drugs are marketed. We showed that for drugs with a very rapid market uptake (e.g. more than 1,000 patient-years of exposure in the first two years after approval), we detect more new safety issues in this period. This doesn't necessarily mean the drugs are intrinsically riskier but rather that they reach faster a sample size large enough to allow identification of signals. It was also observed that after a certain exposure threshold, the chance of detecting a new ADR, decreases. We approximated this threshold to be around 10,000 patient-years. Since it was an exploratory categorical analysis, this finding should be reproduced by other researchers in different settings. Although this finding is not consistent with statistical reasoning, which states that more exposure should lead to more power and more issues detected, it may be considered a 'saturation' phenomenon; after most of the ADRs are known, less and less new ADRs will be reported and discovered. Additional explanations for this phenomenon could be better risk management; medical professionals becoming familiar with the product and decreased reporting. Our finding is in contrast to the one of Mol et al. [34] who found that the level of use in clinical practice does not affect the number of post-approval safety issues. Several methodological differences might have accounted for that: Mol et al selected only innovative drugs (in contrast to all new drugs in our case) and only hospital based Dutch exposure data was used (while we used worldwide sales data). Furthermore, they adjusted for different confounders and the study period was different.

In contrast to post-authorisation exposure, the amount of pre-authorisation exposure did not seem to influence the number of detected signals in our study, at least during the initial years on the market. Another study had a similar finding [35] while Mol et al found the opposite: a larger pre-approval study population led to an increased risk of a serious safety issue [34].

Our studies did not focus on the qualitative aspect of drug exposure: not only does the number of patients exposed change post-approval, but, more importantly, the type of patients may change. A literature review which compared RCTs population in the fields of cardiology, mental health, and oncology, with real-world populations for which the drugs were indicated, demonstrated that the real-world populations are very different from RCTs in terms of demographics, clinical characteristics and polypharmacy, up to the point where they might jeopardize the external validity of RCTs [36].

However it is different to study the impact of these changes on a global scale, since aggregated high level sales data often do not reveal these details.

In view of our findings on pre- and post-marketing exposure time regarding new safety issues, we recommend that drug exposure data are considered when planning the drug surveillance process, and particular attention should be given to drugs with a steep increase in exposure early after marketing. Also considering the conflictual findings existing in literature, more research in this area is needed. It appears that drug exposure is a variable that is correlated with drug safety and it was neglected over the last years.

Where to focus in signal detection; the need to prioritise

With the constant addition of data sources and automatization of the signal detection, there is a legitimate concern that the number of generated signals might surpass the current capacity for analysis [37]. Therefore, is important that signals are prioritised during the entire signal management process. The real challenge in drug surveillance is to detect serious ADRs as early as possible, before too many patients are exposed to harm. There are multiple factors that may influence occurrence of ADRs: the drugs' pharmacological and toxicological profile, the patterns of drug use, the population for which the drug is indicated, time on the market and prescriber and consumers' experience with it [38].

Novelty

In Chapter 2.1 we have shown that while most signals are identified for recently approved drugs also older drugs [39] still have new safety issues. Novel drugs being more likely to have safety signals is not a surprising, nor a new finding. A study from the US which included drugs approved between 1975-2000, found that new drugs have a higher likelihood to have safety issues (defined as 'black box warnings' and drug withdrawals). Half of 'black box warnings' occurred in the first 7 years of drug introduction and half of withdrawals occurred in the first 2 years [40]. Although the famous 'Weber effect' [41] (i.e., adverse events reporting peaks at the end of the second year after approval) was not reproduced in recent studies [42,43], there are studies which showed that novelty influences the risk of finding new safety issues [44,45]. Since we used age as a continuous variable, we cannot recommend a threshold for separating drugs into 'old' vs. 'new'. Other studies proposed a threshold of two and respectively five years for this classification. In conclusion, 'novelty' in regulatory and risk management field equates more uncertainty and therefore higher risk, leading to higher priority.

The fact that more than half of drugs with a signal discussed at PRAC have been marketed for more than 10 years might seem counter-intuitive. Still, the signals identified for well-established drugs might be explained by several factors such as: change in drug utilisation patterns, change in morbidity of underlying population or increased awareness about ADRs. An illustrative example is the one of codeine and life-threatening toxicity through overdose in CYP_{2D6} ultra-rapid metabolizers, [46] a signal discovered only recently, despite the fact that codeine has been used for more than 50 years. Codeine is a prodrug and needs to be converted to morphine by the cytochrome P-450

isoenzyme 2D6 (CYP_{2D6}); which later is inactivated through glucuronidation. Patients with a normal range of CYP_{2D6} activity represent 75-92% of the population, while 5-10% are ultra-rapid metabolizers, who can quickly convert codeine into large amounts of morphine. In children, the rapid conversion of codeine into morphine can be dangerous since the glucuronidation systems are immature and morphine inactivation is slower. In April 2012, a case series was published reporting two deaths and one case of respiratory depression in children 3 to 5 years of age who had received typical doses of codeine after tonsillectomy or adenoidectomy. In response to these cases, regulators in Europe and US started an evaluation of the safety of codeine in children and identified even more cases of morphine overdose. The regulatory actions included a 'black box warning' in US and restriction of codeine use in paediatric patients both in US and Europe. This case illustrates not only the value of anecdotal reports in pharmacovigilance but also the need for continuous monitoring during the entire lifetime of a product. In paediatrics, such surprises may occur with older drugs, as for their licensure, children were not included in clinical trials, a rule which now has been changed, with the updated paediatric legislation from 2007 [47].

Strength of evidence

The findings that the multi-national reporting (or geographic spread) has predictive value for true signals [44,48] needs to be emphasized more. EMA recently changed its practice based on new results [15] and now considers subgrouping by country or continent of origin in its signal detection practice. WHO also applies it in VigiBase [49] with promising results. Overall it highlights the need to create networks of multinational collaborations to augment signal detection.

Using the information on quality and completeness of spontaneous reports during prioritisation phase, seems to be useful as well, as Caster et al demonstrated during testing of Vigirank system [48]. Vigirank was designed as a data-driven screening algorithm for drug safety signals that accounts or report quality and content [48]. For the quality of the reports, two parameters were included: reports were scored on completeness (if they contain sufficient information to allow a causality assessment) and number of reports with free text narrative available were counted. Using Vigirank led to a statistically significant improvement in detection performance, assessed by AUC and compared with the classical disproportionality method and with raw number of reports. In a new study, involving real world use in WHO database, VigiRank significantly outperformed disproportionality analysis alone in signal detection [49]. Surprisingly, the reporter qualification (medical vs. non-medical) failed to show predictive value for signals [45]. This is interesting to note since, before the legislative changes in 2012, medically confirmed reports were perceived as being of higher quality and were prioritized in the reporting process (expedited reporting to authorities) as well as in signal detection procedures [50], which use to give less weight to patient reports. Although the USA, have given patients the possibility to report since the start of their pharmacovigilance systems, in Europe the role of patients as a source of ADRs has not been fully accepted until recently [51]. The Netherlands and Sweden were among the first countries to implement patient reporting and now this is an integral part of the SRS all over EU. there is evidence that reporting by patients can be an important complement to the reporting by healthcare professionals [52]. Examples of signals where consumer reports have

been of crucial importance for signal detection are electric shock-like sensations associated with the use of duloxetine, and persistent sexual dysfunction after discontinuation of selective serotonin reuptake [51].

Public health impact

The public health impact is the impact of the ADR at population level, expressed through the number of patients affected by an ADR and its consequences [53]. Variables related to public health impact are: the seriousness of the event, fatality rate or potential for serious consequences, size of the population with the target disease and size of the treated population, potential to affect vulnerable populations and potential for prevention. More recently, the individual health impact is also taken into account both by FDA and EMA authorities. FDA estimates the hazard posed by a significant safety issue, based on three variables: (1) the relative seriousness of the issue; (2) the estimated size of the population exposed to the risk of the drug; and (3) the suspected frequency of harm to patients exposed to the drug. The combination of factors 2 and 3 provides an estimate of population risk; the combination of factors 1 and 3 provides an estimate of personal risk to the patient [54]. Similarly, in Europe, in the newly updated guidance on signal management [55] both individual and public health impact are separately mentioned. When the public health impact is very high, it can even take precedence over strength of evidence, and regulatory actions can be taken before definitive evidence is gathered [53]. We illustrate this with examples related to two intensely mediatized signals; adjuvanted influenza A(H1N1) vaccine and narcolepsy and tiotropium and cardiovascular side-effects.

A safety signal around Pandemrix®, an adjuvanted influenza A (H1N1) vaccine was identified in August 2010, suggesting that the vaccine causes narcolepsy in children and adolescents. The signal originated from Finland and Sweden, two countries with very high vaccine coverage rate in the entire population (60-70%) [56]. The issue received extensive media attention and the Nordic countries as well as other EU countries conducted rapid risk assessment studies to substantiate the signal [57]. The results differed across countries and the biological mechanism is still unclear to date. At the same time, the extensive media awareness and public unrest urged for regulatory and political actions; an EU referral procedure, product information update, restriction of indication in children and compensation of the victims occurred. A global research project (SOMNIA) was set up to study this and demonstrated that the effects of adjuvanted pdm A (H1N1) vaccines were not observed in other countries [58].

Another example of a signal with a high public health impact is the increased rate of death associated with soft-mist Respimat® inhaler. Respimat® inhaler is a new formulation of the active substance tiotropium aimed to increase bioavailability and efficacy. A risk of increased mortality from cardiovascular disease and all-cause mortality in association with tiotropium Respimat® was identified from a large meta-analysis, back in 2008 [59]. The signal was analysed by the European Pharmacovigilance Working Party which decided to update the product information to reflect the increased mortality. In 2012, an editorial in the BMJ journal called for drug withdrawal based on the

current evidence [60]. The signal was re-discussed at EU level, the evidence was carefully assessed, and it was concluded that all the evidence presented in the editorial was already considered in the previous evaluation. Consequently, the previous conclusion was considered valid and it was decided to wait until the finalization of an ongoing large clinical trial, which could offer more insight. Meanwhile, a case-control study finalised in 2012 [61] identified an 27% increased risk in overall mortality and cardiovascular death. The authors stated that is unclear if this association is causal or due to residual confounding. In 2013, the awaited RCT (involving more than 17,000 patients with COPD for a median duration of 835 days), TIOSPIR trial [62] found no difference in the overall or cardiovascular mortality between tiotropium Respimat® and Handihaler® (the original formulation) in patients with and without baseline cardiac disorders. In view of these new results, the signal was reopened and discussed at PRAC. It was decided that no action is needed besides an update of product information to fully reflect the study results. Considering the high prevalence of COPD (7,6%, in Europe) [63], the widespread use of the drug, recommended as first line treatment and the most prescribed COPD treatment worldwide, summing more than 31 million patient-years of use [64] and the severity of side effects, the public health impact of this issue was considered to be very high. Therefore, the signal had top priority for evaluation and re-opened each time new evidence was available.

Although not as important as the strength of evidence, the factors related to public or media attention are important to consider, to prevent panic and unintended consequences as switching therapies or abrupt interruption of treatment.

Signal validation–a case study

Once a signal of a potential drug safety issue is identified, signal validation and evaluation have to follow rapidly to confirm or refute the association and to describe it in terms of frequency, seriousness and additional risk factors. In **Chapter 3.2**, we performed a signal validation through a pharmacoepidemiological study. We evaluated the association between triptans and ischemic colitis, by using a nested case-control study design in a primary care database from the UK (THIN). This association is another example where the precautionary approach took precedence over strength of evidence. Based on a few well-documented case reports, regulators decided to update the product information of triptans with the adverse event of ischemic colitis. At the moment the association was not confirmed in any hypothesis-testing study. The only evidence existing in the literature came from a case control study performed in an US claims database (Kaiser Permanente Medical Care Plan) [13], which included a very limited number of triptans users (n=16) and did not find an association between triptans use and ischemic colitis. More evidence was needed. We found an increased risk of ischemic colitis associated with triptans' treatment. The most important study limitation was the lack of power which led to borderline significant results and restricted us to further investigate exposure windows. We think the finding should be replicated in a more powered study, ideally a multi-national one.

Methodological considerations of the research contained in this thesis

The challenge of an evaluation exercise in absence of optimal reference standards

Before embarking on a mission to change the current practice of signal management we need to know what the actual performance is. The usual method used to assess performance of signal detection algorithms is by using diagnostic test-related parameters [65] as sensitivity, specificity, positive predictive value and negative predictive value or AUC. These metrics assess the algorithms' capacity to discriminate between true signals and non-causal associations. To be able to use these metrics we need a reference standard for comparison, composed of 'true positive' and 'true negative' signals, classified according to the best currently available evidence. The sources of evidence for true drug-disease associations may be: published scientific literature, product information leaflets or expert opinion.

The absence of a robust reference standard represents a major obstacle in evaluating the performance of signal detection methods. Even when they do exist, reference standards are mostly sub-optimal. Firstly, most reference standards are limited in size, due to time-constraints. They contain a limited number of drug event-associations and it is customary to focus on a small set of drugs or outcomes of interest. This is also the case for our paediatric reference standard used in **Chapter 4.1**, which focuses on of 16 paediatric drugs and 16 ADRs. Secondly, many reference standards also lack verified true negatives (controls), and their focus is on positive test cases only. This is a major limitation since in absence of true negative associations we cannot assess specificity of the method or the AUC. Only a partial performance can be calculated. We have avoided this in our research by using only reference standards with both positive and negative cases. A third limitation, which is also very difficult to avoid, is the possible correlation between the constructed reference standard and the database where the method is applied. Even if not directly consulted in the creation of the reference standard, information from spontaneous reporting often contributes to product labelling and the patients' perception of ADRs and might influence the classification [66]. We could not avoid this completely in either since we used information from product information leaflets for verifying the true positive signals. Information contained in product information leaflets might influence the reporting behaviour.

Ultimately, constructing a universally valid reference standard to test signal detection methods for challenging since causality assessment is not a black and white decision and is also fluctuating with time. Knowledge accrues over time as supplementary data like new studies, better conducted and in larger populations substantiated with biological evidence, or simply more cases become available. This is one reason why many research groups construct their own reference standards at the time of the study. The most common approach has been to use historical ('time-frozen') safety signals as positive controls. However, as mentioned before, the signals might change over time and therefore lead to misclassification. Noren et al. [67] argue that evaluation should be done against emerging and not established adverse events, and a time stamped reference database of ADRs would be the best way forward.

Earlier efforts to develop reference standards were usually not systematic or transparent about their decision process, were limited in the size and diversity of drug-outcome pairs included, or lacked negative controls. This eventually got better as various research groups attempted to create reference standards for the purpose of testing signal detection methods:

EU-ADR reference standard was based on existing scientific literature and expert opinion and included 44 positive associations and 50 negative controls for the ten outcomes 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 [68]. PROTECT reference standard was compiled based on information contained in the product information of 220 drugs approved in Europe [15]. Date when the ADR appeared in product information is also captured. It contains only positive test cases.

Harpaz et al. constructed a reference set based on drug labelling revisions, such as new warnings, which were issued and communicated by the US Food and Drug Administration in 2013. The reference standard includes 44 drugs and 38 events, both positive and negative cases and is time indexed, containing the date when an association (positive test case) became known according to product labels [66].

For the purpose of methods testing OMOP built a reference set of 399 test cases: 165 'positive controls' that represent medical product exposures for which there is evidence to suspect an association with the outcome, and 234 'negative controls' that are drugs for which there is no evidence that they are associated with the outcome, for four health outcomes of interest: acute myocardial infarction, acute liver injury, acute renal failure, and gastrointestinal bleeding. The reference standard spans 181 unique drugs, including nonsteroidal anti-inflammatory drugs, antibiotics, antidepressants, angiotensin-converting enzyme inhibitors, β-blockers, antiepileptics, and glucose-lowering drugs. The work is continued by OhDSI who tries to develop an impressive reference set of 1,000 active ingredients across 100 HOIs [69]. They want to capitalize on previusly contructed reference sets and use a wide range of information sources as: literature, product information and observational healthcare data.

Since none of the existing reference standards was fit for purpose (being restricted either in number of products or outcomes covered), in this thesis we used two reference standards tailored to our research. One was constructed based on published scientific literature and expert opinion (**Chapter 3.1**) and consists of both positive and negative reference drug events pairs, focused on selected outcomes of interest. We included both positive and negative controls, and scientific literature was used as source of information. In contrast to the approach used in previous studies, verification was performed for all drug-event associations associated with the events of interest, irrespective if they were highlighted as signals or not.

The second reference set used in **Chapter 4.1** was paediatric-specific and based on evidence from product information and the published literature. This is also a new addition in the research community, being the first paediatric specific reference set. This one is smaller in size, a combination of 16 paediatric drugs and 16 ADRs of interest for paediatric population, which resulted in 37 positive and 90 negative controls [70].

Statistical Power

Another problem frequently encountered in either signal detection or evaluation when we usually deal with rare events is lack of sufficient power. We encountered this in **Chapter 3.2** when trying to investigate the association between triptans and ischemic colitis. In signal detection, the overall size of a database is not the main determinant of the statistical power but rather the drug exposure data [71,72]. Coloma and colleagues estimated for which type of drugs we have enough power to detect signals in a network of European EHR databases, EU-ADR [72]. Their findings showed that combining EHRs for active drug safety surveillance does increase power, but it may still be not high enough for rare and very rare events and for drugs that are infrequently used. Multi-national and multi-database networks that offer access to large and heterogeneous populations might be of help.

CONCLUSION AND FUTURE PERSPECTIVES

The motivation behind exploring signal detection and evaluation processes is the timely detection of safety issues, and ultimately a better protection of public health. This thesis aimed and to provide informative evidence for regulatory decision making in the areas of signal detection and evaluation. In terms of additional data sources, we explored the use of EHRs, which proved to be challenging but their use might have added value for detection, especially in assessing frequent events or events which are not likely to be recognized as ADRs.

There is a shift lately in pharmacovigilance towards an increased usage of real-world data. Large networks of EHRs increase heterogeneity and size of available populations for analysis of drug safety, at a level that cannot be attained by individual databases alone [73]. The problem in Europe is that these resources are fragmented and highly heterogeneous in terms of structure, coding, and content. The heterogeneity problem can be addressed in different ways, summarized below.

Multi-database and multi-centre studies are essential for increasing power and generalizability of the results and several initiatives started in parallel in US, Canada and Europe to support them. Many of current studies use data from multi-database or multi-national networks. Some of these networks were formed ad-hoc for a particular study while other are more permanent/broad in scope and can be reused for different studies [74]. In Europe, multinational studies were further supported since 2007 through public and public/private EC initiatives such as the European Commission's Seventh Framework Programme (FP7) [73], Horizon 2020 [75] programmes and the Innovative Medicines Initiative (IMI) [76].

Examples of networks created to support multi-database and multi-national studies are ENCePP [13] which is network of centres and all the project based networks in Europe (e.g. EU-ADR, SOS, ARITMO, SAFEGUARD, CARING, PROTECT, VAESCO, ADVANCE, EMIF). Successful examples of multi-databases networks used in regulatory agencies are Sentinel and Canadian Network for Observational Drug Effect Studies (CNODES), both distributed data networks [77,78]. Sentinel is a FDA mandated and highly funded distributed data network that allows secure access and analysis of very large populations (more than 223 million members as of September 2017) [77]. In Canada, CNODES is a multi-provincial distributed network funded by Health Canada which started in 2011. An even more ambitious project, OMOP created a common data model that was implemented in over 50 databases, leading to a truly global network [79]. The work is now continued by OhDSI, whose focus is on the development of open source solutions for observational data [79]. (see a more complete description of these in **Chapter 1**).

In Europe, the ENCePP network focuses on gathering expertise and increase collaboration and does not have databases in-house. It has a Working Group dedicated to the initiation and conduct of observational research using multiple data sources and to develop guidance on conceptual models for multi-national and multi-database studies [80].

There are several methods to conduct multi-database or multi-national studies, ranging from less to more centralized [74,81], illustrated in Table 2. The traditional' and well accepted approach way of integrating results from different databases was by meta-analysis of results of individual studies designed by the local investigators. It requires no collaboration, however the lack of harmonization of study design makes heterogeneity of the results hard to attribute either to methodological or clinical heterogeneity [82,83]. The newest approaches are:

- *Common protocol, local data management and analysis.*
  Data are extracted and analysed locally on the basis of a common protocol. Definitions of exposure, outcomes and covariates, analytical programs and reporting formats are all standardised. Analyses are conducted locally and afterwards, the estimates are pooled together. This model ensures more harmonization of study design, and removes some potential sources of variability. It is a distributed approach where data partners maintain physical and operational control over electronic data in their existing environment. This approach was employed by PROTECT project [19], and TEDDY project in paediatrics [84]. Outside Europe, the Canadian Network for Observational Drug Effect Studies (CNODES) a multi-provincial distributed network started in 2011, is another successful example [78]. Similar to Sentinel, data are extracted and analysed locally and afterwards combined in a summary estimate [78].

- *Common protocol, project specific common data model and common analytics*
  In this distributed model study-specific data are extracted locally and transformed into a common data model. A standardized script can then be run across all sites, which generates aggregated output that can be shared and pooled. This is a very successful and efficient model in Europe. It has been employed by many studies as EU-ADR, SOS, GRIP, ARITMO, SAFEGUARD, EMIF, ADVANCE and many commercial post authorisation safety studies. A characteristic of this model is that the creation of study variables is local and requires close collaboration with the data access provider. This model has been much developed an many tools have been made available to support it: from code mapping tools (Codemapper [85]) to standardized analytics (in Jerboa, SAS, R), and fit for purpose assessments [86].

- *Common protocol, general common data model and common analytics*
  Local data are converted in a generalised full common data model which stays local, but can be analysed with standardized analytics. This is the most 'general' of all systems, which allow flexibility to the analysers to define study variables from different components of data in the common data model. Widely known examples of this approach are VSD, Sentinel, PCORnet and OMOP. The common data model differs between these approaches. The Sentinel common data model and OMOP one differ with regard to design and the availability of derived variables. Sentinel has the original variables and leaves it to the study to derive variables, whereas OMOP maps the original variables in several derived variables beforehand.

Table 2. Differences among the multi-databases strategies with respect to responsibility in the data management and analysis and expected output

Model	Data extraction	Data management	Data analysis	Output shared with partners	Examples
1. Common protocol, local data management and analysis	Local*	Local	Local	Final estimates	PROTECT TEDDY CNODES
2. Common protocol, project specific common data model and common analytics	Local	Local, study-specific	Central	Raw data in a common data model	EU-ADR,SOS, GRIP, ARITMO, SAFEGUARD, EMIF, ADVANCE
3. Common protocol, general common data model and common analytics	Local	Initially local, then central	Central	Patient level data, aggregated data or final estimates	VSD, Sentinel, PCORnet, OMOP

PROTECT= Pharmacoepidemiological Research on Outcomes of Therapeutics by a European ConsorTium; TEDDY = Task-force in Europe for Drug Development for the Young; CNODES= Canadian Network for Observational Drug Effect Studies; EU-ADR= Exploring and Understanding adverse drug reactions; SOS= Safety of Non-Steroidal Anti-Inflammatory Drugs; GRIP= Global research in Paediatrics; ARITMO= Arrhythmogenic Potential of Drugs; EMIF= European Medical Information Framework; ADVANCE= Accelerated development of vaccine benefit-risk collaboration in Europe; VSD= Vaccine safety datalink; *PCORnet= National Patient-Centered Clinical Research Network; OMOP= Observational Medical Outcomes Partnership*

* 'Local' refers to the data custodian, 'central' refers to the coordinating centre of the study;

The existence of multiple networks has transformed the way that we conduct pharmacoepidemiological studies and will likely continue to do so. The advantages of such networks, once created, are statistical power, high external validity and speed of delivery of results (e.g., Sentinel can deliver in days or weeks depending on the research question) [77]. What would be needed is such a sustainable system in the EU rather than project based networks that have a limited time-span and limited generalizability.

The role of regulators and researchers –recommendations for practice and areas of future research

- Which data source to use: spontaneous reports vs electronic healthcare records.
  spontaneous reporting systemss are still the mainstay of pharmacovigilance and should continue to be screened. Their value is greater in the area of rare and very rare reactions, easily recognized as ADRs, where they perform the best. Electronic healthcare records might have additional value, for more frequent reactions and those which are hard to be identified as ADRs by reporters. In Europe the problems of fragmentation and lack of compatibility of data source as a result of the significant heterogeneity among European data sources should be solved before being able to use EHRs for detection. We recommend that multi-national and multi-

database studies are encouraged by regulators and guidance, platforms and tools to facilitate this are built.

- Signal detection methods- age as a confounder or effect modifier
  Age appears to be an effect modifier rather than a confounder. Age adjustment was systematically demonstrated to decrease signal detection performance and should be avoided. Age stratification can increase sensitivity (especially in paediatrics) and lead to discovery of new signals therefore can be used complementary to standard methods.

- Predictors for new safety issues
  Newly approved drugs should be monitored with greater caution since the knowledge of their benefit-risk profile is still less mature. Post-approval exposure seems to be a determinant of safety issues, at least in the initial period on the market. Special attention during signal detection should be given to drugs with potential for high and rapid market uptake, at least until they achieve a certain exposure threshold. The exact threshold, estimated in our study at approximately 10,000 patient-years should be investigated in further research. Since the studies investigating the relation between drug exposure and frequency of safety issues have different results, more research in this area is recommended.
  Multi-national reporting and report quality should be considered when prioritizing signals. In contrast, reporter qualification should not be considered as a prioritization criteria since it was not proven to be associated with true signals.

- More testing of currently available prioritization criteria and frameworks should be done, as this would support creation of a robust evidence-based prioritization process.

- Drug exposure data
  Drug utilisation data have an increasingly important role in the review of benefit-risk of medicinal products post-marketing. Signal detection is no exception. To ensure optimal signal management, efforts should be made to improve collection and accessibility of drug exposure information, since exposure is needed to estimate the public health impact.

CONCLUSION

In conclusion, the dynamic nature of the drug safety field, both in the scientific and in the regulatory aspects, drives the continuous update of existing methods and exploration of other sources for investigating drug safety. There is a need to create big networks of EHR, to support signal detection and evaluation processes, to increase access to drug utilisation data and to invest in prioritisation systems.

REFERENCES

1 Ishiguro C, Hall M, Neyarapally GA, Dal Pan G. Post-market drug safety evidence sources: an analysis of FDA drug safety communications. Pharmacoepidemiol Drug Saf. 2012 Oct 1;21(10):1134–6.

2 Institute of Medicine. The Future of Drug Safety: Promoting and Protecting the Health of the Public [Internet]. 2006 [cited 2018 Jan 12]. Available from: https://www.nap.edu/catalog/11750/the-future-of-drug-safety-promoting-and-protecting-the-health

3 Kearns GL, Abdel-Rahman SM, Alander SW, Blowey DL, Leeder JS, Kauffman RE. Developmental pharmacology - Drug disposition, action, and therapy in infants and children. N Engl J Med. 2003;349(12):1157–67.

4 Gispen-de Wied CC, Leufkens HGM. From molecule to market access: drug regulatory science as an upcoming discipline. Eur J Pharmacol. 2013 Nov 5;719(1–3):9–15.

5 Advancing regulatory science [Internet]. Available from: https://www.fda.gov/ScienceResearch/SpecialTopics/RegulatoryScience/default.htm

6 Meagher E. Efficient drug approval and monitoring must rely on sound regulatory science : Nature Medicine : Nature Research [Internet]. [cited 2017 Feb 12]. Available from: http://www.nature.com/nm/journal/v17/n12/abs/nm1211-1535.html

7 Stricker BHC, Psaty BM. Detection, verification, and quantification of adverse drug reactions. BMJ. 2004 Jul 3;329(7456):44–7.

8 Krumholz HM, Ross JS, Presler AH, Egilman DS. What have we learnt from Vioxx? BMJ. 2007 Jan 20;334(7585):120–3.

9 Eichler H-G, Oye K, Baird LG, Abadie E, Brown J, Drum CL, et al. Adaptive licensing: taking the next step in the evolution of drug approval. Clin Pharmacol Ther. 2012 Mar;91(3):426–37.

10 Eichler H-G, Baird LG, Barker R, Bloechl-Daum B, Børlum-Kristensen F, Brown J, et al. From adaptive licensing to adaptive pathways: delivering a flexible life-span approach to bring new drugs to patients. Clin Pharmacol Ther. 2015 Mar;97(3):234–46.

11 Vandenbroucke JP. In defense of case reports and case series. Ann Intern Med. 2001 Feb 20;134(4):330–4.

12 Harpaz R, DuMouchel W, LePendu P, Bauer-Mehren A, Ryan P, Shah NH. Performance of pharmacovigilance signal-detection algorithms for the FDA adverse event reporting system. Clin Pharmacol Ther. 2013 Jun;93(6):539–46.

13 Coloma PM, Schuemie MJ, Trifirò G, Gini R, Herings R, Hippisley-Cox J, et al. Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: the EU-ADR Project. Pharmacoepidemiol Drug Saf. 2011 Jan;20(1):1–11.

14 Ryan PB, Madigan D, Stang PE, Overhage JM, Racoosin JA, Hartzema AG. Empirical assessment of methods for risk identification in healthcare data: results from the experiments of the Observational Medical Outcomes Partnership. Stat Med. 2012 Dec 30;31(30):4401–15.

15 Wisniewski AFZ, Bate A, Bousquet C, Brueckner A, Candore G, Juhlin K, et al. Good Signal Detection Practices: Evidence from IMI PROTECT. Drug Saf. 2016;39:469–90.

16 Schuemie MJ, Coloma PM, Straatman H, Herings RMC, Trifirò G, Matthews JN, et al. Using electronic health care records for drug safety signal detection: a comparative evaluation of statistical methods. Med Care. 2012 Oct;50(10):890–7.

17 Trifirò G, Patadia V, Schuemie MJ, Coloma PM, Gini R, Herings R, 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 Harpaz R, Vilar S, Dumouchel W, Salmasian H, Haerian K, Shah NH, et al. Combing signals from spontaneous reports and electronic health records for detection of adverse drug reactions. J Am Med Inform Assoc JAMIA. 2013 May 1;20(3):413–9.

19 PROTECT Home [Internet]. [cited 2018 Jan 12]. Available from: http://www.imi-protect.eu/symposium.shtml

20 Edwards I, Karimi G, Bergvall T, Caster O, Asiimwe A, Star K, Bhayat F, Soriano-Gabarro M, Thakrar B, Thompson M, Bate A, Norén G. Risk identification in healthcare records: comparison to epidemiological studies. Pharmacoepidemiol Drug Saf. 2013;22(S1):311–312.

21 Dhanwal DK, Dennison EM, Harvey NC, Cooper C. Epidemiology of hip fracture: Worldwide geographic variation. Indian J Orthop. 2011;45(1):15–22.

22 Star K, Edwards IR. Pharmacovigilance for children's sake. Drug Saf. 2014 Feb;37(2):91–8.

23 Star K, Norén GN, Nordin K, Edwards IR. Suspected adverse drug reactions reported for children worldwide: an exploratory study using VigiBase. Drug Saf. 2011 May 1;34(5):415–28.

24 Blake KV, Zaccaria C, Domergue F, La Mache E, Saint-Raymond A, Hidalgo-Simon A. Comparison between paediatric and adult suspected adverse drug reactions reported to the European medicines agency: implications for pharmacovigilance. Paediatr Drugs. 2014 Aug;16(4):309–19.

25 Zeinoun Z. Quantitative signal detection for vaccines: effects of stratification, background and masking on GlaxoSmithKline's spontaneous reports database. - PubMed - NCBI [Internet]. 2016 [cited 2016 Jan 19]. Available from: https://vpn-gw.erasmusmc.nl/+C SCO+10756767633A2F2F6A6A6A2E6 1706F762E61797A2E6176752 E746269++/pubmed/?term=Quantitative+signal+detection+for+vaccines%3A +effects+of+stratification+ GlaxoSmithKline%27s+spontaneous+reports+database

26 de Bie S, Verhamme KMC, Straus SMJM, Stricker BHC, Sturkenboom MCJM. Vaccine-based subgroup analysis in VigiBase: effect on sensitivity in paediatric signal detection. Drug Saf. 2012 Apr 1;35(4):335–46.

27 Rothman KJ. Epidemiology: an introduction. 2nd ed. New York, NY: Oxford University Press; 2012. 268 p.

28 Seabroke S, Candore G, Juhlin K, Quarcoo N, Wisniewski A, Arani R, et al. Performance of Stratified and Subgrouped Disproportionality Analyses in Spontaneous Databases. Drug Saf. 2016 Jan 9;

29 de Bie S, Verhamme KMC, Straus SMJM, Stricker BHC, Sturkenboom MCJM. Vaccine-Based Subgroup Analysis in VigiBase. Drug Saf. 2012 Apr;35(4):335–46.

30 Zeinoun Z, Seifert H, Verstraeten T. Quantitative signal detection for vaccines: effects of stratification, background and masking on GlaxoSmithKline's spontaneous reports database. Hum Vaccin. 2009 Sep;5(9):599–607.

31 Hopstadius J, Norén GN, Bate A, Edwards IR. Impact of stratification on adverse drug reaction surveillance. Drug Saf Int J Med Toxicol Drug Exp. 2008;31(11):1035–48.

32 Duijnhoven RG, Straus SMJM, Raine JM, de Boer A, Hoes AW, De Bruin ML. Number of Patients Studied Prior to Approval of New Medicines: A Database Analysis. PLoS Med. 2013 Mar 19;10(3):e1001407.

33 PROTECT Drug Consumption Database in Europe [Internet]. [cited 2018 Jan 12]. Available from: http://www.imi-protect.eu/drugConsumption.shtml

34 Mol PGM, Arnardottir AH, Motola D, Vrijlandt PJ, Duijnhoven RG, Haaijer-Ruskamp FM, et al. Post-approval safety issues with innovative drugs: a European cohort study. Drug Saf. 2013 Nov;36(11):1105–15.

35 Arnardottir AH, Haaijer-Ruskamp FM, Straus SMJ, Eichler H-G, de Graeff PA, Mol PGM. Additional safety risk to exceptionally approved drugs in Europe?: Additional safety risk to exceptionally approved drugs in Europe? Br J Clin Pharmacol. 2011 Sep;72(3):490–9.

36 Kennedy-Martin T, Curtis S, Faries D, Robinson S, Johnston J. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. Trials. 2015 Nov 3;16:495.

37 Coloma PM, Schuemie MJ, Trifirò G, Furlong L, van Mulligen E, Bauer-Mehren A, et al. Drug-induced acute myocardial infarction: identifying "prime suspects" from electronic healthcare records-based surveillance system. PloS One. 2013;8(8):e72148.

38 Practical Aspects of Signal Detection in Pharmacovigilance: Report of CIOMS Working Group VIII • COUNCIL FOR INTERNATIONAL ORGANIZATIONS OF MEDICAL SCIENCES [Internet]. COUNCIL FOR INTERNATIONAL ORGANIZATIONS OF MEDICAL SCIENCES. [cited 2018 Jan 13]. Available from: https://cioms.ch/shop/product/practical-aspects-of-signal-detection-in-pharmacovigilance-report-of-cioms-working-group-viii/

39 European Medicines Agency pre-authorisation procedural advice for users of the centralised procedure - WC500004069.pdf [Internet]. [cited 2017 Apr 2]. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2009/10/WC500004069.pdf

40 Lasser KE, Allen PD, Woolhandler SJ, Himmelstein DU, Wolfe SM, Bor DH. Timing of new black box warnings and withdrawals for prescription medications. JAMA. 2002 May 1;287(17):2215–20.

41 Weber JCP. Epidemiology of adverse reactions to nonsteroidal anti-inflammatory drugs. Adv Inflamm Res. (1984):1–7.

42 Hartnell NR, Wilson JP. Replication of the Weber effect using postmarketing adverse event reports voluntarily submitted to the United States Food and Drug Administration. Pharmacotherapy. 2004 Jun;24(6):743–9.

43 Wallenstein EJ, Fife D. Temporal patterns of NSAID spontaneous adverse event reports: the Weber effect revisited. Drug Saf Int J Med Toxicol Drug Exp. 2001;24(3):233–7.

44 Ståhl M, Lindquist M, Edwards IR, Brown EG. Introducing triage logic as a new strategy for the detection of signals in the WHO Drug Monitoring Database. Pharmacoepidemiol Drug Saf. 2004 Jun;13(6):355–63.

45 Puijenbroek EPV, Van Grootheest K, Diemont WL, Leufkens HGM, Egberts ACG. Determinants of signal selection in a spontaneous reporting system for adverse drug reactions. Br J Clin Pharmacol. 2001;52(5):579–586.

46 Racoosin JA, Roberson DW, Pacanowski MA, Nielsen DR. New Evidence about an Old Drug — Risk with Codeine after Adenotonsillectomy. N Engl J Med. 2013 Jun 6;368(23):2155–7.

47 Pediatric Regulation (EC) No 1901/2006 of the European Parliament and of the council. 2006.

48 Caster O, Juhlin K, Watson S, Norén GN. Improved Statistical Signal Detection in Pharmacovigilance by Combining Multiple Strength-of-Evidence Aspects in vigiRank : Retrospective Evaluation against Emerging Safety Signals. Drug Saf Int J Med Toxicol Drug Exp. 2014 Aug;37(8):617–28.

49 Caster O, Sandberg L, Bergvall T, Watson S, Norén GN. vigiRank for statistical signal detection in pharmacovigilance: First results from prospective real-world use. Pharmacoepidemiol Drug Saf. 2017 Aug;26(8):1006–10.

50 Volume 9A of The Rules Governing Medicinal Products in the European Union. 2008.

51 Härmark L, van Hunsel F, Grundmark B. ADR Reporting by the General Public: Lessons Learnt from the Dutch and Swedish Systems. Drug Saf. 2015 Jan 28;

52 Hunsel F. The contribution of direct patient reporting to pharmacovigilance. 2011.

53 Seabroke S, Wise L, Waller P. Development of a novel regulatory pharmacovigilance prioritisation system: an evaluation of its performance at the UK Medicines and Healthcare products Regulatory Agency. Drug Saf. 2013 Oct;36(10):1025–32.

54 FDA. Classifying Significant Postmarketing Drug Safety Issues [Internet]. 2012. Available from: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM295211.pdf

55 EMA. Guideline on good pharmacovigilance practices, Module IX – Signal management. 2012.

56 Björkman I, Sanner MA. The Swedish A(H1N1) vaccination campaign--why did not all Swedes take the vaccination? Health Policy Amst Neth. 2013 Jan;109(1):63–70.

57 Sturkenboom MCJM. The narcolepsy-pandemic influenza story: can the truth ever be unraveled? Vaccine. 2015 Jun 8;33 Suppl 2:B6–13.

58 Black S, Sturkenboom M, Ridder M, Dodd C, Weibel D, Bonhoeffer J, et al. The SOMNIA Study: Assessment of the Risk of Narcolepsy following Adjuvanted 2009 H1N1 Pandemic Vaccines. 2016.

59 Singh S, Loke YK, Enright PL, Furberg CD. Mortality associated with tiotropium mist inhaler in patients with chronic obstructive pulmonary disease: systematic review and meta-analysis of randomised controlled trials. BMJ. 2011 Jun 14;342:d3215.

60 Beasley R, Singh S, Loke YK, Enright P, Furberg CD. Call for worldwide withdrawal of tiotropium Respimat mist inhaler. BMJ. 2012 Nov 9;345:e7390.

61 Verhamme KMC, Afonso A, Romio S, Stricker BC, Brusselle GGO, Sturkenboom MCJM. Use of tiotropium Respimat Soft Mist Inhaler versus HandiHaler and mortality in patients with COPD. Eur Respir J. 2013 Sep;42(3):606–15.

62 Dahl R, Calverley PMA, Anzueto A, Metzdorf N, Fowler A, Mueller A, et al. Safety and efficacy of tiotropium in patients switching from HandiHaler to Respimat in the TIOSPIR trial. BMJ Open. 2015 Dec 1;5(12):e009015.

63 Raherison C, Girodet P-O. Epidemiology of COPD. Eur Respir Rev. 2009 Dec 1;18(114):213–21.

64 Halpin DM, Dahl R, Hallmann C, Mueller A, Tashkin D. Tiotropium HandiHaler(®) and Respimat(®) in COPD: a pooled safety analysis. Int J Chron Obstruct Pulmon Dis. 2015;10:239–59.

65 Zhou X-H, Obuchowski NA, McClish DK, Wiley InterScience (Online service). Statistical methods in diagnostic medicine [Internet]. New York: Wiley-Interscience; 2002 [cited 2013 Jun 13]. Available from: http://public.eblib.com/EBLPublic/PublicView.do?ptiID=468721

66 Harpaz R, Odgers D, Gaskin G, DuMouchel W, Winnenburg R, Bodenreider O, et al. A time-indexed reference standard of adverse drug reactions. Sci Data. 2014 11;1:140043.

67 Norén GN, Caster O, Juhlin K, Lindquist M. Zoo or savannah? Choice of training ground for evidence-based pharmacovigilance. Drug Saf. 2014 Sep;37(9):655–9.

68 Coloma PM, Avillach P, Salvo F, Schuemie MJ, Ferrajolo C, Pariente A, et al. A reference standard for evaluation of methods for drug safety signal detection using electronic healthcare record databases. Drug Saf. 2013 Jan;36(1):13–23.

69 Boyce RD, Ryan PB, Norén GN, Schuemie MJ, Reich C, Duke J, et al. Bridging islands of information to establish an integrated knowledge base of drugs and health outcomes of interest. Drug Saf. 2014 Aug;37(8):557–67.

70 Osokogu O, Fregonese F, Ferrajolo C, Verhamme K, de Bie S, Jong G 't, et al. Pediatric Drug Safety Signal Detection: A New Drug–Event Reference Set for Performance Testing of Data-Mining Methods and Systems. Drug Saf. 2015 Feb 1;38(2):207–17.

71 Hammond IW, Gibbs TG, Seifert HA, Rich DS. Database size and power to detect safety signals in pharmacovigilance. Expert Opin Drug Saf. 2007 Nov;6(6):713–21.

72 Coloma PM, Trifirò G, Schuemie MJ, Gini R, Herings R, Hippisley-Cox J, et al. Electronic healthcare databases for active drug safety surveillance: is there enough leverage? Pharmacoepidemiol Drug Saf. 2012;21(6):611–621.

73 Public Workshop: A Framework for Regulatory Use of Real-World Evidence | Margolis Center for Health Policy [Internet]. [cited 2017 Nov 30]. Available from: https://healthpolicy.duke.edu/events/public-workshop-framework-regulatory-use-real-world-evidence

74 Gini R, Schuemie M, Brown J, Ryan P, Vacchi E, Coppola M, et al. Data Extraction and Management in Networks of Observational Health Care Databases for Scientific Research: A Comparison of EU-ADR, OMOP, Mini-Sentinel and MATRICE Strategies. EGEMS Wash DC. 2016;4(1):1189.

75 Horizon 2020 - European Commission [Internet]. Horizon 2020. [cited 2018 Jan 12]. Available from: https://ec.europa.eu/programmes/horizon2020/

76 The Innovative Medicines Initiative [Internet]. IMI - Innovative Medicines Initiative. [cited 2018 Jan 12]. Available from: http://www.imi.europa.eu/

77 The FDA's Sentinel Initiative [Internet]. [cited 2017 Nov 30]. Available from: https://www.fda.gov/Drugs/NewsEvents/ucm586259.htm

78 Suissa S, Henry D, Caetano P, Dormuth CR, Ernst P, Hemmelgarn B, et al. CNODES: the Canadian Network for Observational Drug Effect Studies. Open Med Peer-Rev Indep Open-Access J. 2012;6(4):e134-140.

79 Observational Medical Outcomes Partnership [Internet]. 2013 [cited 2013 Mar 22]. Available from: http://omop.fnih.org/

80 ENcEPP database [Internet]. Available from: http://www.encepp.eu/

81 Trifirò G, Coloma PM, Rijnbeek PR, Romio S, Mosseveld B, Weibel D, et al. Combining multiple healthcare databases for postmarketing drug and vaccine safety surveillance: why and how? J Intern Med. 2014 Jun 1;275(6):551–61.

82 Hernández-Díaz S, Varas-Lorenzo C, García Rodríguez LA. Non-steroidal antiinflammatory drugs and the risk of acute myocardial infarction. Basic Clin Pharmacol Toxicol. 2006 Mar;98(3):266–74.

83 Pladevall-Vila M, Delclos GL, Varas C, Guyer H, Brugués-Tarradellas J, Anglada-Arisa A. Controversy of oral contraceptives and risk of rheumatoid arthritis: meta-analysis of conflicting studies and review of conflicting meta-analyses with special emphasis on analysis of heterogeneity. Am J Epidemiol. 1996 Jul 1;144(1):1–14.

84 TEDDY – European Network of Excellence for Paediatric Clinical Research [Internet]. [cited 2018 Jan 31]. Available from: https://www.teddynetwork.net/

85 Becker BFH, Avillach P, Romio S, van Mulligen EM, Weibel D, Sturkenboom MCJM, et al. CodeMapper: semiautomatic coding of case definitions. A contribution from the ADVANCE project. Pharmacoepidemiol Drug Saf. 2017 Aug;26(8):998–1005.

86 Sturkenboom M, Caitlin Dodd, Ana Correa, Hanne Dorthe-Emborg, Talita Duarte-Salles, Elisa Martin, Consuelo Huerta, Silvia Lucchi, Gino Picelli, Lara Tramontan, Lieke van der Aa, Daniel Weibel, Benedikt Becker, Guiseppe Roberto, Marius Gheorge,... Methodology to assess eligibility of electronic healthcare databases to conduct vaccine coverage, risk and benefits studies in Europe. In 2017.

Nederlandse samenvatting

Dit proefschrift omvat studies die beogen vragen te beantwoorden met betrekking tot de sub-domeinen signaal management en regulatory science: databronnen, detectie-methoden en het proces van prioritering.

We begonnen dit proefschrift met een overzicht van signalen in de EU, aangezien er grootschalige wetswijzigingen zijn goedgekeurd in 2010 en van kracht zijn geworden in 2012. Op basis van het overzicht van signalen die werden besproken in de Pharmacovigilance Risk Assessment Committee (PRAC), concludeerden we dat de meest gebruikte databron spontane meldingen waren (72% van de gevallen), net als in de VS , zie **Hoofdstuk 2** (1). De signalen die besproken werden, waren het vaakst gerelateerd aan "Huid en onderhuidse aandoeningen" (12,8%), "Zenuwstelstel aandoeningen" (10,4%), "Cardiale aandoeningen" en "Immuunsysteemaandoeningen" (6,4%). We zagen dat de gemiddelde tijd tussen het opkomen van een nieuw signaal en een beslissing van PRAC 2,5 maanden was, waarbij in 42,8% van de gevallen een beslissing genomen werd tijdens de eerste vergadering. Voor 57,2% van de signalen werd additionele informatie opgevraagd na de eerste bespreking in de PRAC. Er kunnen meerdere acties genomen worden na validatie van een nieuw signaal. De actie die het meest vaak werd genomen was een wijziging in de productinformatie (54%). Er werd niet vaak besloten tot het starten van een referral (9,4%) of het versturen van een Direct Healthcare Professional Communication (7,3%), maatregelen die voornamelijk genomen worden voor ernstige en urgente problemen. Echter, het besluit om over te gaan tot één van beide voorgenoemde acties werd wel sneller genomen (respectievelijk 1,8 maanden en 1,7 maanden).

Spontane meldingen vormen nog altijd de hoeksteen voor het opsporen van signalen, maar hun beperkingen zijn algemeen bekend en om deze reden is men continue op zoek naar nieuwe gegevensbronnen om het proces van signaaldetectie mee te verrijken. Door meerdere problemen omtrent de veiligheid van geneesmiddelen in het begin van de 21ste eeuw, die veel aandacht hebben gekregen, gaf het Institute of Medicine aan dat spontane meldingen alleen niet afdoende zijn. Elektronische gezondheidszorg databases (EHRs) werden toen voorgesteld als alternatieve gegevensbron (2). Om te achterhalen hoe EHRs spontane meldingen zouden kunnen aanvullen bij geneesmiddelentoezicht, hebben wij onderzoek gedaan naar Europese EHRs als bron voor het oppikken van signalen. In **Hoofdstuk 3.1** hebben we de databases vergeleken die meededen met het EU-ADR project met Eudravigilance op hun vermogen om bekende associaties op te sporen. In deze studie waren we beter in staat bijwerkingen met zeldzame tot zeer zeldzame achtergrondincidentie op te sporen in databases met spontane meldingen (SRS), evenals bijwerkingen die makkelijk aan geneesmiddelen toe schrijven zijn, zoals blaarziekten en acute pancreatitis. Daarentegen waren er bijwerkingen (bijv. heupfracturen) waar EHRs tot betere resultaten leidden.

Een belangrijke overweging bij het selecteren van gegevensbronnen voor het opsporen van signalen is de ratio ruis-op-signaal. Het aantal vals positieve signalen in elke gegevensbron (EU-ADR

en Eudravigilance) is afhankelijk van zowel het type signaal als de gegevensbron. Het laagste aantal vals positieven vonden we voor bovenste tractus digestivus bloedingen en acute pancreatitis, het hoogste aantal voor blaarziekten. Het was voor alle bijwerkingen duurder om signalen te detecteren in EHRs dan in SRS (bijv. omdat meer vals positieven moeten worden uitgesloten). Concluderend, een systeem gebaseerd op EHRs kan van toegevoegde waarde zijn voor signaaldetectie mits gebruikt naast de huidige systemen, vooral voor bijwerkingen met een hoge achtergrondincidentie. Op het moment lijken SRS meer kosteneffectief te zijn dan EHRs.

In **Hoofdstuk 3.2** hebben we het gebruik van EHRs voor evaluatie van signalen beschreven door onderzoek te doen naar de associatie van triptanen en ischemische colitis. Vanwege de zeldzaamheid van de uitkomst, zijn er tot nu toe maar een handvol meldingen geweest. In de THIN database, die eerstelijnszorggegevens van ongeveer 6% van de bevolking van het Verenigd Koninkrijk bevat, konden we maar 41 nieuwe gevallen van ischemische colitis identificeren in een cohort van ongeveer 300,000 migraine patiënten. Het gebruik van triptanen in de 12 maanden voor het event verdubbelde het risico op ischemische colitis vergeleken met geen gebruik (OR=2,29, 95%CI 1,02–5,15). Na 12 maanden werd het risico kleiner (OR=1,90, 95%CI 0,44–8,13). Door gebrek aan statistische kracht konden alternatieve risicovensters niet verder exploreren.

Naast het gebruik van verschillende datasoorten voor het opsporen en evalueren van signalen, hebben we ook onderzocht hoe de huidige methoden om signalen op te sporen bij kinderen zouden kunnen worden verbeterd. In **Hoofdstuk 4.1** hebben we het effect getest van stratificatie op leeftijd of het adjusteren voor leeftijd. We hebben twee algoritmes (PRR en EBGM) toegepast op meldingen in kinderen jonger dan 18 jaar oud, omdat te verwachten valt dat leeftijd een groter effect sorteert door de relatie met verschillende stadia van orgaanrijping. Beide methoden leverden vergelijkbare prestaties (zoals gemeten met een area under the curve (AUC)). De prestatie van beide methoden was minder wanneer er voor leeftijd geadjusteerd werd, vergeleken met wanneer niet voor leeftijd geadjusteerd werd. Stratificatie voor leeftijd leidde tot het ontdekken van signalen in specifieke leeftijdsstrata, die niet ontdekt waren bij de algemene analyse. We concluderen dat leeftijd een effect modifier kan zijn en dat er daarom gestratificeerd moet worden op leeftijd.

Het begrijpen welke soorten geneesmiddelen een hoger risico hebben op veiligheidssignalen kan de efficiëntie verbeteren en de ratio ruis-op-signaal verminderen. Geneesmiddelen, waarvoor een nieuw signaal werd besproken bij de PRAC in de periode tussen september 2012 en december 2013, hadden een lagere mediane post-marketingleeftijd (12,3 jaar) dan andere geneesmiddelen op de Europese markt waarvoor geen signalen waren in dezelfde periode (19,7 jaar). Het zijn echter niet alleen de meest recent toegelaten geneesmiddelen waar signalen voor zijn: 58% van de geneesmiddelen waarvoor een signaal besproken werd bij PRAC waren al langer dan 10 jaar op de markt (leeftijdsrange 0,54–67,9 jaar), dus ook voor gevestigde geneesmiddelen kunnen signalen worden gevonden.

In **Hoofdstuk 6.1** hebben we naast hoe lang een geneesmiddel al op de markt verkrijgbaar is, ook onderzoek gedaan naar andere voorspellers van de frequentie van veiligheidssignalen. Geneesmiddelen behorende tot de ATC klasse "Antineoplastische en immuunmodulerende middelen", geneesmiddelen die gebruikt worden voor behandeling van acute aandoeningen en niet-biologische geneesmiddelen hadden een hogere kans op veiligheidsproblemen na toelating op de markt. Daar waar de totale duur van blootstelling voor toelating niet geassocieerd was met de kans op detectie van veiligheidsproblemen, gold het tegenovergestelde voor de totale duur van blootstelling na toelating. Voor geneesmiddelen met meer dan 1.000 patiëntjaren van behandeling in de eerste twee jaar na toelating was de kans op veiligheidsproblemen 2,4 keer zo hoog als voor geneesmiddelen met minder blootstelling in dezelfde periode. Na een zekere drempel (ongeveer 10.000 patiëntjaren) bereikte deze verhoogde kans een plateau en nam daarna af.

Om regulatoire instanties te adviseren over welke criteria in beschouwing moeten worden genomen bij prioritering van veiligheidssignalen, hebben we onderzoek gedaan naar verschillende triage criteria (zz). We hebben 48 verschillende criteria geïdentificeerd, die in 4 groepen kunnen worden ingedeeld: noviteit, impact op de volksgezondheid, bewijslast en publieke- en media-aandacht. Meer dan de helft van de criteria (63%) had betrekking op de bewijslast, terwijl 19% betrekking had op de impact voor de volksgezondheid, 14% betrekking had op publieke- en media-aandacht en 4% op noviteit van de associatie tussen het geneesmiddel en de bijwerking. Van deze criteria hebben we er 15 getest op de voorspellende waarde (het vermogen om een daadwerkelijk signaal te voorspellen), waarbij 11 positieve resultaten opleverden. Hiervan behoorden de meesten tot de bewijslastcategorie, zoals meldingen uit meerdere landen, kwaliteit en volledigheid van de meldingen. Daarentegen was hoedanigheid van de melder (of de melder wel of niet medisch onderlegd was) niet geassocieerd met echte signalen. We hebben 6 besliskaders gevonden in de literatuur, waarvan we er 5 getest hebben op validiteit. Ze bleken alle 5 redelijk bruikbaar, en tenminste 3 ervan worden reeds gebruikt door verschillende instanties voor prioritering.

Hoofdstuk 7 bevat een algemene discussie over de gevolgen van de uitkomsten van dit proefschrift, evenals aanbevelingen voor de huidige regulatoire praktijk en toekomstig onderzoek.

List of Publications

Manuscripts related to this thesis

Păcurariu AC, Coloma PM, Gross-Martirosyan L, Sturkenboom MC, Straus SM
Decision making in drug safety-a literature review of criteria used to prioritize newly detected safety issues.
Pharmacoepidemiol Drug Saf. 2016 Nov 23. doi: 10.1002/pds.4128.

Păcurariu AC, Straus SM, Trifirò G, Schuemie MJ, Gini R, Herings R, Mazzaglia G, Picelli G, Scotti L, Pedersen L, Arlett P, van der Lei J, Sturkenboom MC, Coloma PM
Useful Interplay between Spontaneous ADR Reports and Electronic Healthcare Records in Signal Detection.
Drug Saf. 2015 Dec;38(12):1201-10. doi:10.1007/s40264-015-0341-5.

Păcurariu AC, Coloma PM, van Haren A, Genov G, Sturkenboom MC, Straus SM
A description of signals during the first 18 months of the EMA Pharmacovigilance risk assessment committee.
Drug Saf. 2014 Dec;37(12):1059-66. doi: 10.1007/s40264-014-0240-1.

Osokogu OU, Dodd C, Păcurariu A, Kaguelidou F, Weibel D, Sturkenboom MC
Drug Safety Monitoring in Children: Performance of Signal Detection Algorithms and Impact of Age Stratification.
Drug Saf. 2016 Sep;39(9):873-81. doi: 10.1007/s40264-016-0433-x.

Păcurariu AC, Hoeve C, Arlett P, Genov G, Slattery J, Sturkenboom M, Straus S
Is patient exposure pre and post-authorization a determinant of the timing and frequency of occurrence of safety issues?
Pharmacoepidemiol Drug Saf [in press]

Other Publications

Gadroen K, Straus SM, Păcurariu AC, Weibel D, Kurz X, Sturkenboom MC
Patterns of spontaneous reports on narcolepsy following administration of pandemic influenza vaccine; a case series of individual case safety reports in Eudravigilance.
Vaccine. 2016 Sep 22;34(41):4892-7. doi: 10.1016/j.vaccine.2016.08.062.

Osokogu OU, Dukanovic J, Ferrajolo C, Dodd C, Păcurariu AC, Bramer WM, 'tJong G, Weibel D, Sturkenboom MC, Kaguelidou F
Pharmacoepidemiological safety studies in children: a systematic review.
Pharmacoepidemiol Drug Saf. 2016 Aug;25(8):861-70. doi: 10.1002/pds.4041. Review.

Osemeke U Osokogu, Alexandra Păcurariu, Mees Mosseveld, Peter Rijnbeek, Daniel Weibel, Katia Verhamme, Miriam C J M Sturkenboom

Impact of different assumptions on estimates of pediatric disease occurrence from health care data: A retrospective cohort study

Submitted

Thomas Goedecke, Daniel Morales, Alexandra Pacurariu, Xavier Kurz

Measuring the impact of medicines regulatory interventions -systematic review and methodological considerations

British Journal of Clinical Pharmacology, 2017 November, epub ahead of print

List of Co-Authors (alphabetical order)

Anja van Haren
Medicines Evaluation Board, Utrecht, the Netherlands

Caitlin Dodd
Department of Medical Informatics, Erasmus Medical Center, Rotterdam, the Netherlands

Christel Hoeve
Medicines Evaluation Board, Utrecht, the Netherlands
Department of Medical Informatics, Erasmus Medical Center, Rotterdam, the Netherlands

Daniel Weibel
Department of Medical Informatics, Erasmus Medical Center, Rotterdam, the Netherlands

Florentia Kaguelidou
Hôpital Universitaire Robert Debré, Paris, France

Georgy Genov
Pharmacovigilance Division, European Medicines Agency, London United Kingdom

Giampiero Mazzaglia
European Medicines Agency, London United Kingdom

Gianluca Trifirò
Department of Clinical and Experimental Medicine, University of Messina, Messina, Italy

Gino Picelli
Pedianet-Società Servizi Telematici SRL, Padova, Italy

Jim Slattery
Surveillance and Epidemiology Division, European Medicines Agency, London United Kingdom

Johan van der Lei
Department of Medical Informatics, Erasmus Medical Center, Rotterdam, the Netherlands

Lars Pedersen
Department of Clinical Epidemiology, Aarhus University Hospital, Aarhus, Denmark

Liana Gross-Martirosyan
Medicines Evaluation Board, Utrecht, the Netherlands

Lorenza Scotti
Department of Statistics, Università di Milano-Bicocca, Milan, Italy

Maria de Ridder
Department of Medical Informatics, Erasmus Medical Center, Rotterdam, the Netherlands

Martijn Schuemie
Janssen Research and Development

Miriam C Sturkenboom
Department of Medical Informatics, Erasmus Medical Center, Rotterdam, the Netherlands

Osemeke Osokogu
Department of Medical Informatics, Erasmus Medical Center, Rotterdam, the Netherlands

Peter Arlett
Pharmacovigilance Division, European Medicines Agency, London United Kingdom

Preciosa M Coloma
Department of Medical Informatics, Erasmus Medical Center, Rotterdam, the Netherlands

Remy D C Francisca
Medicines Evaluation Board, Utrecht, the Netherlands
Department of Medical Informatics, Erasmus Medical Center, Rotterdam, the Netherlands

Ron Herings
PHARMO Institute for Drug Outcomes Research, Utrecht, the Netherlands

Rosa Gini
Agenzia Regionale di Sanità della Toscana, Florence, Italy

Sabine MJM Straus
Medicines Evaluation Board, Utrecht, the Netherlands
Department of Medical Informatics, Erasmus Medical Center, Rotterdam, the Netherlands

Swabra Nakato
Department of Medical Informatics, Erasmus Medical Center, Rotterdam, the Netherlands

PhD Portfolio

Name: Alexandra Cristina Păcurariu
Promotores: Prof. M.C.J.M. Sturkenboom
Copromotor: Dr. S.M.J.M. Straus

Affiliation: Erasmus University Medical Center
Departments: Medical Informatics

PhD period: 2013–2017

PhD Training

Research Skills

2011–2013 Master of Science in Pharmacovigilance and Pharmacoepidemiology, Bordeaux University
2013 Scientific English writing course, Erasmus University, Rotterdam
2014 Biostatistical Methods course, Erasmus University, Rotterdam

Oral Presentations

2015 Spontaneous reports and electronic healthcare records for safety signal detection – yin and yang
15th Annual Meeting of the International Society of Pharmacovigilance (ISoP 2015),
27–30 October 2015, Prague, Czech Republic

2016 Exposure related variables –their influence on the occurrence of safety issues post approval
Pharmacovigilance's role in rapid access to safer drugs
30–31 May 2016, Uppsala, Sweden

Poster Presentations

2014 A Comparative Analysis of Signal Detection in Electronic Healthcare Database vs Spontaneous Reporting Databases
30th International Conference on Pharmacoepidemiology & Therapeutic Risk Management
Oct 24–27, 2014, Taipei, Taiwan

A Comparative Analysis of Signal Detection in Electronic Healthcare Database vs. Spontaneous Reporting Databases
EURODURG
August 27–29, 2014, Groningen, the Netherlands

2015 Spontaneous reports and electronic health records for safety signal detection
Drug Information Association Annual Meeting
June 15–17, 2015, Washington, DC

Memberships

International Society of Pharmacoepidemiology, 2013

Peer reviewer

Drug safety
Pharmacoepidemiology and Drug safety

Teaching activities

2013–2016 Teaching assistant for the online Master European Programme in Pharma-coigilance and Pharmacoepidemiology

2015–2016 Teaching assistant for Master of Science in Paediatric Medicines Development and Evaluation

Acknowledgements

Firstly, to one of my dearest mentors, prof. Wolfgang Graier, who made me fall in love with science, even if I 'betrayed' you by leaving the world of fluorescent mitochondria, your contagious enthusiasm for science has never left me.

To Prof. Miriam Sturkenboom – dear Miriam, thank you taking a chance on me, and giving me this truly life-changing opportunity. I learned a lot from you, both professionally and personally. None of this would have happened were it not for you. To my co-promotors who were both true mentors; dear Sabine, we went through so many things together but you always supported and understood, while constantly trying to save me from myself. Thank you for finding the perfect balance between giving me the freedom to explore whatever topic I was interested in and offering a helping hand whenever I was struggling. To Preci –thank you for being such a guiding, comforting and inspiring presence in my life, no matter the distance between us. I hope I made you proud! To prof.dr. Bruno Stricker, prof.dr. H.G.M. Leufkens and prof.dr. E.P. van Puijenbroek, whose previous work inspired and guided me and formed a solid base of this thesis.

Thank you to all dear colleagues and friends in IPCI and Pharmacoepi group: Katia, Inge, Marjolein, Gwen, Daniel, Ann, Carmen, Osemeke, Nico, Maarten, Esme, Christel, Ruben and Peter, for sharing the frustration of research work. To Swabra, Caitlin and Ingrid, for sharing the room on 27th floor and going through my bipolar moods every week and regularly supplying the keys whenever I lost them. To the colleagues in BIOS and to other colleagues in Medical Informatics, and to the Secretariat: Desiree, Tineke, Carmen, Petra, and Sander – thank you for your support. To Mees, Kris, Marius and Marcel - thank you for all the technical assistance.

Thanks from heart to my colleagues from CBG: Negar, Florianne, Laurens, Ineke, Menno, Quirine, Remy, Liana, Anouk, Fakhredin, Evelyn, Sara, Marcel, Maarten, Maria, Stephany, Anita, Ursula, Rianne, Esther, Inge, Daniel. Ik heb ontzettend veel van jullie mogen leren en ik heb het naar mijn zin met jullie, leuke samenwerking en de gezelligheid. You are all special! I especially miss the FT2 team, thank you for adopting me and for all the nice things we did together, at work and outside. To Fakhredin, thank you for the guidance and training and for taking the anxiety away with your calm tone. Menno, for always being there for my sometimes stupid questions and for the interesting discussions we always had.

To Joan and Joris, for the great support all these years, thank you for believing in me and giving me the freedom to do it my own way. You both took a great risk, I hope it was worth it!

Acknowledgements

To Anja, my dearest friend in the Netherlands, for being always around and relentlessly trying to improve my organizational skills and failing gloriously. Probably the only thing you ever failed to accomplish beautifully. To Remy, for the wonderful talks, I love that we can speak about interrupted time series analysis and poetry in the same coffee break. You are one of a kind! And you always enlighten my day with your smile.

To Rosa, for the laughter and precious advice and for reminding me to enjoy life more.

To Oana, Leo, Laura, Gertjan and Robert- for the sleepless nights discussing politics, morality and other not so respectable subjects, for making me laugh and allowing me to cry next to you. And special thanks to Leo for constantly rescuing me from any cold station in the Netherlands I might happen to get lost at!

To my friends from Romania, we do not see each other often enough but you will always have a special place in my heart. To Marius, thank you for having been next to me for so long, for supporting me all throughout, for the unwavering love and strong character,

To my parents and the rest of my family Buna, Andreea, Radu, for always supporting me, even when you disagreed with my decisions, and for loving me every single moment. Special thanks to my parents for constantly sending Romanian food all over Europe. I am sorry for leaving you! I miss you every moment!

To Scott, for sharing with me the frustrations and joys of finalizing this book and this road and for daring to walk with me on new roads lying ahead.

About the author

Alexandra Cristina Pacurariu was born on January 5th, 1985, in Bucharest, Romania. She attended the Mihai Viteazul Gymnasium Bucharest, where she graduated in 2003 with a profile in mathematics and informatics. In the same year, she started the pharmaceutical studies at Carol Davila University, Bucharest. Consequently, she obtained her Master of Science degree in Pharmaceutical Sciences in 2008 with the thesis 'Effects of imipramine on an animal model of depression'. Afterwards, she became interested in scientific research and in the autumn of 2009 she followed an internship in endocannabinoid research, at the Molecular Biology department from Karl-Franzens-Universität Graz, Austria.

In 2013, she completed a Master of Science in Pharmacovigilance and Pharmacoepidemiology, at the University of Bordeaux, while at the same time working in the pharmaceutical industry, in the field of Drug safety. This is how she became interested in the drug safety area. Pharmacoepidemiology was still far away. In 2013, she started her PhD project in pharmacovigilance as described in this thesis, with a focus on signal detection methods. At the same time she worked as a pharmacovigilance assessor at the Dutch Medicines Evaluation Board, Utrecht.

Alexandra was tutored by Prof. Miriam Sturkenboom, dr. Sabine Straus and dr. Preciosa Coloma, to whom she is perennially grateful for guidance and support.

During her PhD training, she has collaborated with the Global research in Pediatrics project and was a tutor for EU2P and GRiP Master program. When she is not in front of a computer or at conferences, she likes the great outdoors, sports and making people laugh.