University of Groningen

Signal detection in pharmacovigilance: time for a new era?

Scholl, Joep

DOI:
[10.33612/diss.204264948](https://doi.org/10.33612/diss.204264948)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2022

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Scholl, J. (2022). *Signal detection in pharmacovigilance: time for a new era?* [Thesis fully internal (DIV), University of Groningen]. University of Groningen. <https://doi.org/10.33612/diss.204264948>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Signal detection in pharmacovigilance: time for a new era?

Joep Scholl

Paranimfen

Hanneke Nugteren

Han Weerdesteyn

Signal detection in pharmacovigilance: time for a new era?

© 2022, Joep Scholl

No parts of this thesis may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system, without permission of the author.

The work presented in this thesis was performed at the Netherlands Pharmacovigilance Centre Lareb and the University of Groningen, Groningen Research Institute of Pharmacy, Unit Pharmacotherapy, - Epidemiology & -Economics

Het drukken van dit proefschrift werd mede mogelijk gemaakt met financiële steun van het Nederlands Bijwerkingen Fonds, de Rijksuniversiteit Groningen en Bijwerkingencentrum Lareb

Cover design: Joep Scholl & proefschriftmaken.nl

Lay-out design and printed by: Proefschriftmaken.nl, Utrecht

rijksuniversiteit
 groningen

Signal detection in pharmacovigilance: time for a new era?

Proefschrift

ter verkrijging van de graad van doctor aan de
Rijksuniversiteit Groningen
op gezag van de
rector magnificus prof. dr. C. Wijmenga
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

dinsdag 15 maart 2022 om 12.45 uur

door

Johannes Henricus Gerardus Scholl

geboren op 21 december 1975
te 's-Hertogenbosch

Promotores

Prof. dr. E.P. van Puijenbroek

Prof. dr. E. Hak

Copromotor

Dr. F.P.A.M. van Hunsel

Beoordelingscommissie

Prof. dr. P.M.L.A. van den Bemt

Prof. dr. O.H. Klungel

Prof. dr. P.G.M. Mol

Table of contents

CHAPTER 1	General introduction	7
CHAPTER 2	When More Is Less: An Exploratory Study of the Precautionary Reporting Bias and Its Impact on Safety Signal Detection <i>Clin Pharmacol Ther</i> 2018 Feb;103(2):296-303	19
CHAPTER 3	The Value of Time to Onset in Signal Detection	37
	3.1 Parametric time to onset models were developed to improve causality assessment of adverse drug reactions from antidiabetic drugs <i>J Clin Epidemiol</i> . 2015 Dec;68(12):1423-31	39
	3.2 The value of time to onset in statistical signal detection of adverse drug reactions: a comparison with disproportionality analysis in spontaneous reports from the Netherlands <i>Pharmacoepidemiol Drug Saf</i> . 2016 Dec;25(12):1361-1367	55
	3.3 Time to onset in statistical signal detection revisited: A follow-up study in long-term onset adverse drug reactions <i>Pharmacoepidemiol Drug Saf</i> . 2019 Oct;28(10):1283-1289	69
CHAPTER 4	A prediction model-based algorithm for computer-assisted database screening of adverse drug reactions in the Netherlands <i>Pharmacoepidemiol Drug Saf</i> . 2018 Feb;27(2):199-205	83
CHAPTER 5	General Discussion	99
	Summary	111
	Samenvatting	115
	Dankwoord	119
	List of publications	121
	About the author	125

CHAPTER 1

General introduction

PHARMACOVIGILANCE BASED ON SPONTANEOUS REPORTING OF ADVERSE DRUG REACTIONS

The role of pharmacovigilance in drug regulation

Pharmacovigilance (PV) is defined by the World Health Organization (WHO) as *the science and activities relating to the detection, assessment, understanding and prevention of adverse effects or any other drug-related problem* [1]. Since pharmacovigilance is based on clinical observations and experiences to gain knowledge about a particular topic, it belongs to the empirical sciences. Empirical research is often based on the empirical cycle, starting with the observation of a phenomenon and the inquiry regarding its causes (observation). This observation leads to the formulation of hypotheses (induction) and subsequently to the formulation of experiments to test the hypotheses (deduction), procedures used to test the hypotheses (testing) and interpretation of the data and formulation of a theory (evaluation) [2].

In pharmacovigilance, knowledge about the safety profile of a drug is gained throughout different stages of the drug's life cycle. In the clinical development phase (phase I-III), clinical trials are performed based on hypothesis testing and evaluation. Subsequently, new hypotheses are formulated which can be tested in the next phase. After registration of a drug, phase IV trials can be conducted to answer questions not provided in earlier phases. This may concern safety and/or efficacy related questions. This cascade of consecutive trials is based on the principles of the empirical cycle as described above. However, phase I-III or phase IV trial results alone do not provide all safety information relevant for the risk-benefit balance of a drug. In particular, clinical trials are often based on a more or less selective sample of the target population, excluding several relevant groups of patients (elderly, pregnant women, children etc.) who may need the drug. In addition, rare adverse drug reactions (ADRs), or those with a long time to onset, are unlikely to be detected in clinical trials due to the limited number of patients exposed to the drug [3]. This became painfully obvious during the thalidomide (Softanon®, Distaval®) tragedy in the 1960s. During this period, thousands of children with congenital deformities (deformity of the limbs or phocomelia) were born after their mothers had been using thalidomide as an anti-emetic and sedative. The drug had already been approved for market authorization for several years, before its role in this embryopathy came to light. The tragedy made society realize that continuous surveillance based on a more proactive (instead of the previously more reactive) approach was necessary both prior to and after drug approval. In fact, the letter from Dr. McBride to the Lancet in 1961, in which he described a case of phocomelia and asked the reader to report similar cases, can be considered to be one of the first signal detection initiatives in pharmacovigilance [4].

The role of spontaneous reporting within pharmacovigilance

Today, several data sources are used to monitor drug safety profiles after registration (also called post marketing phase). Two major sources are electronic healthcare records (EHR) and spontaneous reporting systems (SRS). EHRs use electronic data from patient medical records and can be used in both the detection of signals and in signal strengthening [5]. SRSs are mainly used for signal detection and were originally developed to identify rare ADRs [6]. They are based on the analysis of safety reports, voluntarily (spontaneously) reported by healthcare professionals and consumers. Each safety report is an observation of clinical phenomena (such as the use of a drug and the occurrence of a medical event) which may be causally related to each other. Whether or not there is a true relationship for each individual case, is usually not known.

At pharmacovigilance centres, a causality assessment is performed by pharmacovigilance assessors (PVA), who are clinically and pharmacologically trained for this purpose. To make a proper assessment of the causal relationship between the drugs the patient has been exposed to and the clinical observation under study, additional information on both intrinsic and extrinsic factors is needed. These may include the medical and pharmacological history, of the patient, but also characteristics of the drug, timing and background-incidence [7]. It is this clinical content of the report against the background of similar observations and clinical / pharmacological knowledge, which can make the observation valuable. However, different stakeholders within the reporting process have access to, and knowledge of, different aspects of the clinical content. Some aspects are known to both the PVA and the reporting healthcare professional (HCP), such as the suspected drug(s) and experienced ADR(s). Other aspects are known to only PVA or HCP, and some are known to neither of them (see also table 1). As a result, the information present in SRS is, in the majority of cases, limited to the information known to both PVA and HCP. This may lead to selective detection of ADRs because reporters may not recognize certain events in their daily practice as ADRs, e.g., because they are unaware of information that already exists on the matter. On the other hand, assessors may not value reported ADRs as true ADRs because they lack information known to the reporter. It should be taken into account that this selection of available information influences signal detection on SRS.

Table 1. Information related to different aspects of drug – ADR associations

	Known to the reporter (healthcare professional)	Less well known to the reporter (healthcare professional)
Known to the assessor	<ul style="list-style-type: none">• Suspected drug• Suspected ADR(s)• Concomitant medication• Concomitant disorders/diseases• Timing of ADR(s)• Outcome of ADR(s)• Risk factors• Part of medical history	<ul style="list-style-type: none">• Potential mechanism of action• Information from other databases• Pre-registration safety information• Risk Management Plans• ‘Pharmacovigilance memory’
Less well known to the assessor	<ul style="list-style-type: none">• Specialist knowledge• Detailed medical history• Experience with the course of diseases• Experience with patient• Access to lab/test results	<ul style="list-style-type: none">• Unknown mechanisms• Unrevealed disorders and diseases in this patient• Non-reported signs and symptoms• Adherence

Signal detection using spontaneous reporting systems

The main goal of using spontaneous reporting systems is to enable safety signal detection, where a safety signal can be seen as information on a new or known adverse event that may be caused by a medicine and requires further investigation [8]. Such a signal can initially be triggered by individual assessment of each reported case, by applying statistical signal detection methods, or a combination of both. In terms of the empirical cycle, both approaches are limited to the observation and induction phase, not deduction, testing or evaluation. The main argument for this is twofold. First, there is no specific research question to be answered. Theoretically, any combination of ADRs with any combination of drugs can be reported. As a result, spontaneous reporting databases are heterogeneous sets of observations, containing hundreds of thousands of unique drug – ADR associations. Second, spontaneous reporting systems were initially developed to find new drug – ADR associations. In empirical terms, these are observations (signals), which can lead to the hypothesis formulation (induction). Often, these signals should be followed by epidemiological studies for hypothesis testing and quantitative assessments, which is beyond the scope of pharmacovigilance using spontaneous reporting systems [9].

At pharmacovigilance centres, safety reports are assessed for causality and evaluated for signal value. Signal detection of spontaneously reported ADRs has its origin in case-by-case analysis where all case reports containing one or more ADRs are assessed by clinically qualified assessors. During assessment, assessors judge each report for signal value, often using a causality assessment algorithm, such as Naranjo [10] or the WHO causality scale [11]. When deemed necessary, a detailed analysis of associations looking at both intrinsic and extrinsic elements of information can be performed [12], which can ultimately lead to a signal. For many years, this has been the main method of signal detection. However, over the past decades the number of ADR reports received by pharmacovigilance centres and Marketing Authorization Holders (MAHs) has increased dramatically due to e.g., the introduction of patient reporting as a source of data. In addition, obligations for MAHs and competent authorities, arising from pharmacovigilance legislation that has become more strict over the years, have expanded and this has led to an increase of ADR reports. One example is the reporting of non-serious ADRs as a safety report to the authorities which was not mandatory a few years ago. This increase in reports has resulted in the situation where manual review of all reports is no longer feasible in many cases. Another issue to take into consideration is that signal detection based on case-by-case analysis is prone to a certain degree of subjectivity because assessors have different qualifications and competencies, based on their prior education and work experience. As a result, different signals may be detected, by different assessors, based on their knowledge and experience [13]. This implies that, due to differences in interpretation, signals also may remain undetected or deemed to have a low priority due to these differences among assessors.

To face these challenges, statistical methods based were developed as a complementary detection tool. If one considers case-by-case signal detection to be a bottom-up approach where individual cases are assessed (bottom) and the database and literature are searched (up) for additional evidence for a particular drug – ADR association, then statistical signal detection can be viewed as a top-down approach starting with a statistically significant

finding for a group of reports (top), followed by a review of the individual reports (down). This top-down approach allows for the detection of signals that may not have been detected using the case-by-case approach for reasons mentioned above [14, 15], and has proven to be a valuable complement to a bottom-up strategy.

Statistical signal detection: a historical perspective

Back in the 1980s, one of the first statistical signal detection methods were the observed vs. expected measures, also known as disproportionality analyses. These methods compare the observed and expected number of cases of a certain drug – ADR association, under the assumption that their relative numbers are equal. Commonly used methods include frequentist approaches (e.g., Reporting Odds Ratio (ROR), Proportional Reporting Ratio (PRR)). Later, Bayesian approaches (e.g., Bayesian Confidence Propagation Neural Network (BCPNN), Multi-item Gamma Poisson Shrinker (MGPS) [16-20] were developed, which allow for prior probabilities to be taken into account. Although different in their approach, all techniques are based on estimating an observed vs. expected measure with its confidence / credibility interval. When performing a disproportionality analysis, the point estimate (ratio observed/expected) is a measure of the strength of the association, whereas the associated confidence / credibility interval (most commonly 95%) indicates precision / statistical significance. When the results of a disproportionality analysis for one or more drug – ADR associations exceed a certain threshold, the individual reports are reviewed and a safety issue may be signalled.

A major limitation of this methodology is that these types of analyses are mainly suited for the detection of new, previously unknown ADRs. However, signal detection in pharmacovigilance deals with more than finding new ADRs, such as the detection of previously unknown aspects of known ADRs (e.g., risk factors, severity, time to onset) or the detection of drug-drug interactions. Since the (statistical) goal of these types of analyses is to identify those drug – ADR associations that have been reported more often than statistically expected, they basically create order in a chaotic database containing an abundance of drug – ADR combinations. However, as mentioned earlier, the value of a case describing an ADR is often determined by its clinical content and context. This constitutes another major limitation of traditional disproportionality analyses, which from a statistical point of view, weighs all cases equally, independent of their clinical value. As a result, reports of ADRs that could be of clinical interest in signalling safety issues may not be recognized as such. Furthermore it implies that clinical case analyses cover a larger part of the empirical cycle than disproportionality analyses, because of the different aspects taken into account in the former (see also figure 1).

Figure 1. The empirical research cycle applied to clinical case analysis and disproportionality analysis in spontaneous reporting systems (SRS).

Bias in signal detection from spontaneous reporting systems

Although the signal detection process described previously seems straight forward and can be easily applied to SRSs, there is an important phenomenon that should be taken into account, which is bias. When bias is not considered in signal detection, it may lead to false positive or false negative results, which may jeopardize drug surveillance. SRSs are sensitive to several (specific) types of bias. This includes reporting biases, such as (i) underreporting, indicating that only a proportion of the ADRs that actually occur are reported [21]; (ii) notoriety bias, where exposure to information on the existence of an ADR (e.g., due to media attention) trigger the reporting of that particular ADR [22, 23]; (iii) the Weber effect [24], characterized by the increased reporting of ADRs in the first 2 years after marketing authorization of a drug; (iv) the relative overreporting of serious ADRs compared to non-serious ADRs [25]; (v) recall bias, which favours the reporting of ADRs that can be remembered by the patient. In addition to these types of information bias, protopathic bias [26], confounding by indication [27] and channelling bias [28] should be taken into account in signal detection using spontaneous reports. In protopathic bias, a drug is started for symptoms of a yet undiagnosed disease. When the disease is diagnosed at a later stage, it may be suggested that the drug caused the disease. An example is the use of painkillers for general pain complaints, which are due to the presence of an undiagnosed tumour. When the tumour is diagnosed, a causal role may be unjustly

attributed to the painkiller. Protopathic bias is sometimes confused with confounding by indication, where the suspected ADR is caused by an underlying disease instead of the exposure to a drug. For instance, a patient with a history of diabetes mellitus who recently started a SSRI for depression and who experiences hypoglycaemia, may conclude that this was caused by the SSRI, where it was actually caused by the pre-existing diabetes mellitus. Channelling bias occurs when different drugs with similar indications are prescribed to different patient groups based on comorbidities and/or clinical status. The effect is similar to confounding by indication in the sense that the comorbidities may be incorrectly attributed to the drug. The types of bias described above may be relevant for both case-by-case signal detection and statistical signal detection, because they can result in misclassification (protopathic bias / confounding by indication) and may lead to an over- / underrepresentation of numbers used in observed vs. expected analyses.

In addition to the type of bias described above, precautionary reporting bias may occur. This type of bias can arise when selective ADRs are reported in so called patient support programs. This type of bias and the nature of patient support programs will be further discussed in chapter 2.

UNMET NEEDS IN STATISTICAL SIGNAL DETECTION

Over the past decades, disproportionality analysis has long been the method of choice, despite such analyses having several (severe) limitations as described earlier. Despite various efforts to renew methods for statistical signal detection in spontaneous reporting systems, implementation in daily pharmacovigilance practice is limited. This relative lack of search for novel methodologies inspired the design of the studies described in this thesis. The studied novel methodologies are based on either the investigation of a particular clinical parameter (univariable approach), in our case the time to onset of an event as a possible signal detection tool, or methodologies applying a multivariable approach.

AIM AND OUTLINE OF THE THESIS

Aim of the thesis

The aim of this thesis is to explore novel statistical methods of signal detection in pharmacovigilance and to evaluate their practical value in spontaneous reporting systems.

Outline of the thesis

This thesis consists of five studies, divided into three chapters, followed by a general discussion on the results and implications of our research.

Chapter 2 elaborates on the introduction of a new type of bias, called precautionary reporting bias. This type of bias represents a type of information bias and occurs mainly in reports from patient support programs. The objective of this study is to investigate to what extent this increase in reports has an effect on signal detection, in particular on the results of disproportionality analyses.

Chapter 3 provides insight into the possible role of the time to onset of an ADR in pharmacovigilance. Chapter 3.1 describes the role of the time to onset in causality assessment. In this study, the time to onset for ADRs of patients in a cohort study is modelled using common parametric distributions. The main objective is to investigate if the time to onset could be a valuable parameter in causality assessment. In addition, it is a prelude to studies investigating the time to onset in statistical signal detection. In chapter 3.2, a proof of concept study explores a time to onset based method for signal detection in spontaneous reporting. A comparison with disproportionality analysis is made. The study is performed using a limited number of drugs ranging from a single administration to chronic use. Chapter 3.3 expands on the possible role of time to onset in signal detection. Based on the results of the proof of concept study, a full database approach is investigated, and data are limited to reports with a relatively short time to onset.

Chapter 4 presents a multivariable regression model as a method to develop a signal detection screening tool. The rationale behind the study is the hypothesis that a multivariable prediction model could generate a priority list of associations to be analysed in further detail. This prioritization step could lead to a more efficient use of pharmacovigilance resources.

Chapter 5 contains a general discussion on the outcomes of the studies described in this thesis. In addition, it provides suggestions for further research in the field of statistical signal detection.

REFERENCES

1. The importance of pharmacovigilance: Safety monitoring of medicinal products [Available from: <https://apps.who.int/iris/rest/bitstreams/50360/retrieve>].
2. De Groot AD. Methodology: Principles of Research and Thinking in Behavioural Sciences. The Hague, The Netherlands: Mouton; 1961.
3. Stricker BH, Psaty BM. Detection, verification, and quantification of adverse drug reactions. *BMJ*. 2004;329(7456):44-7.
4. McBride WG. Thalidomide and congenital abnormalities. *Lancet*. 1961;278(7216):1358.
5. Pacurariu, AC, Straus SM, Trifiro 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*;38(12):1201-10
6. Temple R. Meta-analysis and epidemiologic studies in drug development and postmarketing surveillance. *JAMA*. 1999;281(9):841-4.
7. Ralph Edwards I. Causality Assessment in Pharmacovigilance: Still a Challenge. *Drug Saf*. 2017;40(5):365-72.
8. Agency EM. Guideline on good pharmacovigilance practices (GVP). Module IX – Signal management (Rev 1) [Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-good-pharmacovigilance-practices-gvp-module-ix-signal-management-rev-1_en.pdf].
9. Aronson JK, Hauben M. Anecdotes that provide definitive evidence. *BMJ*. 2006;333(7581):1267-9.
10. 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(2):239-45.
11. WHO-UMC. The use of the WHO-UMC system for standardised case causality assessment [Available from: https://www.who.int/medicines/areas/quality_safety/safety_efficacy/WHOcausality_assessment.pdf].
12. Begaud B, Evreux JC, Jouglard J, Lagier G. [Imputation of the unexpected or toxic effects of drugs. Actualization of the method used in France]. *Therapie*. 1985;40(2):111-8.
13. Rolfes L, Kolschoten J, van Hunsel F, van Puijenbroek E. The validity and reliability of a signal impact assessment tool. *Pharmacoepidemiol Drug Saf*. 2016;25(7):815-9.
14. van Hunsel F, van Erp A. Risico op hypokaliëmie bij gebruik van flucloxacilline. *Pharmaceutisch Weekblad*. 2020;155(45):16-9.
15. Lareb. Lamotrigine and photosensitivity [Available from: https://databankwslareb.nl/Downloads/Signals_2019_Lamotrigine_photosensitivity.pdf].
16. Stricker BH, Tijssen JG. Serum sickness-like reactions to cefaclor. *J Clin Epidemiol*. 1992;45(10):1177-84.
17. van Puijenbroek EP, Bate A, Leufkens HG, Lindquist M, Orre R, Egberts AC. 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.
18. 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;54(4):315-21.
19. Noren 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. *Stat Med*. 2006;25(21):3740-57.

20. 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.
21. Hazell L, Shakir SA. Under-reporting of adverse drug reactions : a systematic review. *Drug Saf*. 2006;29(5):385-96.
22. Pariente A, Gregoire F, Fourrier-Reglat A, Haramburu F, Moore N. Impact of safety alerts on measures of disproportionality in spontaneous reporting databases: the notoriety bias. *Drug Saf*. 2007;30(10):891-8.
23. de Boissieu P, Kanagaratnam L, Abou Taam M, Roux MP, Drame M, Trenque T. Notoriety bias in a database of spontaneous reports: the example of osteonecrosis of the jaw under bisphosphonate therapy in the French national pharmacovigilance database. *Pharmacoepidemiol Drug Saf*. 2014;23(9):989-92.
24. 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(6):743-9.
25. Moulis G, Sommet A, Durrieu G, Bagheri H, Lapeyre-Mestre M, Montastruc JL, et al. Trends of reporting of 'serious' vs. 'non-serious' adverse drug reactions over time: a study in the French Pharmacovigilance Database. *Br J Clin Pharmacol*. 2012;74(1):201-4.
26. Gerhard T. Bias: considerations for research practice. *Am J Health Syst Pharm*. 2008;65(22):2159-68.
27. Salas M, Hofman A, Stricker BH. Confounding by indication: an example of variation in the use of epidemiologic terminology. *Am J Epidemiol*. 1999;149(11):981-3.
28. Petri H, Urquhart J. Channeling bias in the interpretation of drug effects. *Stat Med*. 1991;10(4):577-81.

CHAPTER 2

2

When More Is Less: An Exploratory Study of the Precautionary Reporting Bias and Its Impact on Safety Signal Detection

Kevin Klein*, Joep Scholl*, Marieke de Bruin, Eugène van Puijenbroek, Bert Leufkens, Pieter Stolk

* Authors contributed equally to this work

ABSTRACT

Concerns have been expressed that large numbers of nonvalue-added reports have been accumulating in adverse drug reaction (ADR) databases, for example, via patient support programs. We performed an assessment of the impact of such reports, which we refer to as “precautionary reports,” on safety signal detection in the Netherlands. The case narratives of ADR reports of three case products were screened with text-mining algorithms to identify those reports that lack a causal relationship with the suspected medicinal product. We demonstrate that precautionary reports impede the optimal use of the pharmacovigilance system by, on the one hand, masking safety signals and, on the other hand, creating spurious signals. The precautionary reporting bias and its suppressing effect on statistical signal detection results in an altered adverse event safety profile. The findings from this study highlight the need for a better alignment between regulatory authorities and marketing authorization holders regarding pharmacovigilance guidelines.

INTRODUCTION

The reporting of an adverse drug reaction (ADR) by healthcare professionals or patients plays a key role in the post marketing surveillance of medicinal products in the European Union (EU). ADRs can be reported directly to national pharmacovigilance centres (NPCs) or to the marketing authorization holder (MAH) of the suspected medicinal product [1-4]. Whereas NPCs mostly receive spontaneous reports, which are voluntary reports from healthcare professionals and increasingly from patients themselves, that describe an ADR, MAHs also receive solicited ADR reports, e.g., from organized data collection systems, such as patient support programs (PSPs), noninterventional studies, or compassionate use programs [5]. ADR reports collected by NPCs and MAHs are both submitted to the EudraVigilance PostAuthorisation Module (EVPIM), the European ADR database that is managed by the European Medicines Agency (EMA). ADR data can be exchanged through EudraVigilance with NPCs to complement national ADR databases with ADR reports received by MAHs in the respective Member State [5,6].

Statistical signal detection based on disproportionality methods are frequently applied to European and national ADR databases to detect safety signals for medicinal products. These automated methods screen ADR databases for signals of disproportionate reporting (SDRs), which indicate any adverse event for a medicinal product that is disproportionately highly represented [7,8]. In the ADR database of the Netherlands Pharmacovigilance Centre Lareb, the reporting odds ratio (ROR) is the standard method for detecting SDRs. This statistical measure is part of the method used to screen the complete ADR database (combining spontaneous and solicited reports), to identify SDRs that exceed a predefined threshold, which are then considered for further assessment and evaluation [9].

Various studies have highlighted the existence of *competition biases* that may decrease the performance of statistical signal detection methods applied to ADR databases to detect safety signals. For example, Pariente *et al.* presented a “signal competition bias,” where the event of interest is significantly associated with other medicinal products, thus increasing the background reporting rate, and hence the threshold for the medicinal product of interest to create an SDR [10,11]. The subsequent reduction in SDR detection is referred to as the *masking effect* [12,13]. The same masking effect was also identified for a different type of competition bias reported by Hauben and Hochberg [14] and later described as the “event competition bias” by Salvo *et al.*, where the reporting of a large number of (often well-known) events for a given medicinal product may mask (previously unknown) associations of the same medicinal product with other events [15].

In this study we investigated the existence of another form of competition bias that could lead to the masking of SDRs. Some concerns have been expressed that the extensive patient contact by MAHs through organized data collection systems (in particular, PSPs) leads to the elevated reporting of (often unrelated) events [16,17]. One example of this type of reporting is the disproportionate reporting of patient deaths found in the US Food and Drug Administration's Adverse Event Reporting System (FAERS). These reports of patient deaths often concerned terminally ill patients enrolled in PSPs, for which the causality with the product was not confirmed [16,18]. It has been pointed out that

these reports are characterized by poor documentation of the medical context [16]. The rationale behind this reporting is not fully explained, but suggests a potential implication of the current pharmacovigilance requirements for MAHs. Current guidelines state that if the causal relationship between an event and the medicinal product cannot be ruled out, it should be considered an ADR. The nature of these guidelines implies that the biological plausibility as to whether the event is caused by the product is irrelevant [19]. Therefore, these guidelines oblige MAHs to also submit reports of (potentially unrelated) events for which the causal relationship cannot be ruled out (e.g., due to the limited information that is available), as these events still meet the regulatory definition of an ADR. These reports could be regarded as “precautionary reports” by MAHs to meet the regulatory requirements. It was argued that these precautionary reports, when submitted to ADR databases, could create spurious signals [16]. However, and more important, by increasing the background-reporting rate for the medicinal product of interest, these precautionary reports could increase the threshold for SDRs to be detected, thus potentially masking interesting (yet undetected) safety signals. Currently, limited information about such precautionary reports and their actual impact on statistical signal detection is available.

Therefore, the objective of this study was to explore the effects of precautionary reports on the performance of statistical signal detection methods and the potential masking of safety signals with three real-world case studies in the Dutch ADR database.

RESULTS

Of the 157,833 reports collected in the Dutch ADR database since the start of data collection, 3,112, 2,555, and 689 reports were related to the erythropoietin (case product I), the bisphosphonate (case product II), and the endothelin receptor antagonist (case product III), respectively. These reports relate to 4,327, 6,531, and 1,555 medicinal product/event associations, respectively.

The assessment of the case narratives resulted in the identification of 2,757 (89%) precautionary reports for the erythropoietin, 2,023 (79%) for the bisphosphonate, and 409 (59%) for the endothelin receptor antagonist. These reports relate to 3,591 (83%), 4,922 (75%), and 716 (46%) medicinal product/event associations for the three respective case products. Furthermore, the assessment of the case narratives revealed that 59%, 98%, and 27% of the precautionary reports for the erythropoietin, bisphosphonate, and the endothelin receptor antagonist, respectively, were clearly attributable to organized data collection systems. A respective 81%, 51%, and 83% of the precautionary reports relate to spontaneous reports (**Figure 1**).

Figure 1. An overview of the reports for the erythropoietin ($n = 3,112$), the bisphosphonate ($n = 2,555$), and the endothelin receptor antagonist ($n = 689$), as a percentage of the total reports stratified by report source and report type. Report source: whether an ODCS was identified; report type: spontaneous, solicited, or other (if unclear); ODCS, organized data collection system.

An overview of the five most reported events based on the Medical Dictionary for Regulatory Activities (MedDRA) preferred term (PT) for the reports classified as precautionary reports for each case product is provided in Table 1. The event “death” was the most reported event in precautionary reports for all three case products, e.g., 36% of all medical product/event associations reported for the erythropoietin case are related to the event “death.” Looking at all reports with the event “death” for the erythropoietin, the bisphosphonate, and the endothelin receptor antagonist, our algorithm classified 99%, 95%, and 84% of the reports as precautionary reports, respectively (**Table 1**).

Table 1. Five most reported events for the precautionary reports of each of the three case products (in descending order)

Event (as MedDRA PT)	Medicinal product/events associations as a % of total for each case product	% of reports with the specific event classified as precautionary report
Erythropoietin		
Death	36%	99%
Hospitalization	15%	87%
Off label use	3%	78%
Dialysis	2%	90%
Terminal state	1%	68%
Bisphosphonate		
Death	9%	95%
Malignant neoplasm progression	5%	79%
Terminal state	5%	87%
General physical health deterioration	3%	81%
Neoplasm progression	3%	83%
Endothelin receptor antagonist		
Death	14%	84%
Right ventricular failure	2%	75%
Cardiac failure	1%	68%
Pulmonary arterial hypertension	1%	84%
Pulmonary hypertension	1%	62%

MedDRA, Medical Dictionary for Regulatory Activities; PT, Preferred Term.

Other events from the top 5 are also highly associated with precautionary reports. Initially, a total of 45, 175, and 64 SDRs were detected (based on the ROR) for the erythropoietin, the bisphosphonate, and the endothelin receptor antagonist, respectively (**Table 2, Figure 2**).

Table 2. Overview of SDR detection before and after exclusion of precautionary reports from the ADR database

Case product	Medicinal product/event associations [N]	Medicinal product/event associations from reports classified as precautionary reports [N (%)]	Number of SDRs detected [N]		Change in SDR detection	
			Before exclusion of precautionary reports	After exclusion of precautionary reports	SDRs no longer appearing [N (%)]	Unmasked SDRs [N (%)]
Erythropoietin	4,327	3,591 (83%)	54	43	27 (50%)	16 (37%)
Bisphosphonate	6,531	4,922 (75%)	175	82	101 (58%)	8 (10%)
Endothelin receptor antagonist	1,555	716 (46%)	64	43	26 (41%)	5 (12%)

SDR, signal of disproportionate reporting

Figure 2. Venn diagram of the number of SDRs detected before and after exclusion of the precautionary reports from the ADR database for the three case products. From left to right: Erythropoietin, bisphosphonate, and endothelin receptor antagonist; red circle: SDRs detected before exclusion of precautionary reports; green circle: SDRs detected after exclusion of precautionary reports; overlap: SDRs detected before and after exclusion of precautionary reports.

After the exclusion of the precautionary reports from the ADR database, 16 new SDRs were detected for the erythropoietin, 8 for the bisphosphonate, and 5 for the endothelin receptor antagonist. Of these unmasked SDRs, 10 (63%), 7 (88%), and 4 (80%) SDRs, respectively, are considered safety signals requiring further evaluation due to the absence of obvious confounders, i.e., these SDRs are not related to events associated with the natural course of underlying disease, the indication, (mis-) use, or a patient outcome (**Table 3**).

Table 3. An overview of the unmasked signals of disproportionate reporting (SDRs) after exclusion of precautionary reports from the ADR database.

Case product	Unmasked SDR (PT)
Erythropoietin	Deep vein thrombosis ^{a, b}
	Drug ineffective ^a
	Feeling abnormal ^a
	Fluid retention ^b
	Infection ^a
	Influenza ^a
	Influenza like illness ^a
	Injection site pain ^{a, b}
	Limb discomfort ^a
	Malaise
	Myocardial infarction
	Pain ^a
	Pallor
	Pyrexia ^a
	Renal failure
	Renal impairment
Bisphosphonate	Arthralgia ^b
	Arthritis ^{a, b}
	Blood glucose increased ^a
	Candida infection ^a
	Eye pain ^{a, b}
	Hypotension ^{a, b}
	Swelling ^{a, b}
Endothelin receptor antagonist	Tooth disorder ^{a, b}
	Abortion spontaneous ^a
	Anaemia ^{a, b}
	Deep vein thrombosis ^a
	Myocardial infarction
	Weight increased ^a

^a SDRs that represent events considered potential safety issues according to our definition, as they are not related to the underlying disease, the indication, the (mis-) use or a patient outcome; ^b SDRs that represent events that are listed in the Summary of Product Characteristic (SmPC) section on adverse reactions.

For all three case products, the exclusion of precautionary reports from the ADR database also resulted in SDRs that were no longer appearing. For the erythropoietin, bisphosphonate, and the endothelin receptor antagonist, 27, 101, and 26 SDRs, respectively, were not detected anymore after exclusion of the precautionary reports from the ADR database (**Table 2, Figure 2**).

DISCUSSION

The results from this study demonstrate for the first time the existence, and quantify the impact of, precautionary reports on statistical signal detection methods. The type of competition bias described in this article is hereafter referred to as the “precautionary reporting bias.” The exclusion of the large number of identified precautionary reports from the ADR database resulted in the unmasking of previously undetected SDRs. Moreover, the existence of precautionary reports in ADR databases can trigger the detection of spurious signals: SDRs that are generated by precautionary reports and do not appear any more after the exclusion of the precautionary reports from the database. The exclusion of precautionary reports from the database thus improved the quality of signal detection. Both effects result in an altered adverse event safety profile for all three cases, which signifies the public health relevance of this bias. For the erythropoietin and bisphosphonate cases, we identified that most of the precautionary reports were clearly attributable to organized data collection systems, all relating to PSPs (**Figure 1**). For the endothelin receptor antagonist, 27% of the precautionary reports were attributable to organized data collection systems, the majority relating to noninterventional studies in particular. For the erythropoietin and the endothelin receptor antagonist, however, we believe that the identification of organized data collection systems in precautionary reports is an underestimation, as for most of these reports the report source was not clearly identifiable. Precautionary reports negatively impact the performance of statistical signal detection based on disproportionality methods in two ways. First, precautionary reports increase the chance of the detection of false-positive signals (Type I errors), which we refer to as spurious signals. Second, precautionary reports increase the chance of not detecting signals that otherwise would have been detected, so-called false-negative signals (Type II errors), which is also described by the masking effect. Both Type I and Type II errors can be explained with the two-by-two contingency table (**Box 1**), whereby the C and D quadrants remain constant.

The presence of precautionary reports for a medicinal product of interest increases the background-reporting rate represented by the denominator (B) and thus the threshold for the event of interest (A) to be detected. In parallel, for events that are highly associated with precautionary reports, the numerator (A) artificially increases to such an extent that it produces spurious signals for these events. In addition to the masking effect described above, precautionary reports could delay the identification of safety signals, by increasing the time-to-detection due to an increased background-reporting rate (B), which requires more reports with the event of interest (A) over time to create an SDR. The majority of the spurious signals that are triggered by precautionary reports relate to extraneous events, for which the biological plausibility as to whether the event is caused by the medicinal product is questionable. Examples of such extraneous events that create spurious signals are “euthanasia” and “hospice care” (events that are not informative) or, for instance, events related to the indication (e.g., “haemoglobin increased” for erythropoietin) or the natural course of the underlying disease (e.g., “cancer pain” for bisphosphonate). The five most reported events for the three case products, for example, are all extraneous events (**Table 1**). There are also spurious signals that do not refer to extraneous events (e.g., “lung infection” for erythropoietin). Nonetheless, these events still lack a causal

relationship with the suspected medicinal product and could be related to comorbidities. There always remains a possibility that some of the events related to the indication or underlying disease could represent a paradoxical adverse event or point towards a lack of therapeutic effect [20,21]. However, as long as reporters of such an event do not provide any motive or rationale for a paradoxical adverse event or potential lack of effect, it hampers any meaningful interpretation of such safety signals. Therefore, their negative impact on signal detection that is demonstrated in this article; in our view, does not justify the potential (but limited) gains from these reports.

Box 1. Calculation of the reporting odds ratio (ROR) based on the two-by-two contingency table.

Two-by-two contingency table:

	Event of interest	All other events
Medicinal product of interest	a	b
All other medicinal products	c	d

The ROR is computed as: $\frac{(a/b)}{(c/d)}$

The standard error (se) of the natural logarithm of the ROR is as follows:

$$SE = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

The lower bound of the 95% confidence interval (CI) for the ROR is then calculated as:

$$95\%CI \text{ (lower bound)} = e^{\ln(ROR) - 1.96 \cdot se}$$

Limited information exists in the literature to explain the rationale behind precautionary reports, but it has been pointed out that current pharmacovigilance legislations and guidelines could play an important role. The guideline on Good Pharmacovigilance Practice (GVP) Module VI states that reports obtained from organized data collection systems should be considered as solicited reports and are subject to appropriate causality assessments to assess whether they meet the criteria for reporting [5]. According to the ICH guideline for Good Clinical Practice (GCP), all events for which the causal relationship between the medicinal product and the event cannot be ruled out should be considered an ADR [5,19]. Some documents found in the literature point out the challenges that MAHs encounter when trying to perform causality assessments and highlight the potential implications of these guidelines. These documents emphasize that reports from PSPs often lack the necessary information required for an appropriate causality assessment [16,17]. Moreover, efforts of MAHs to retrieve such important missing information during follow-up are extremely difficult or even impossible due to patient anonymity in these programs [22,23]. This inevitably prevents MAHs from being able to perform an appropriate causality assessment, and thus to rule out any causal relationship. Hence, these reports meet the definition of an ADR report from a regulatory perspective and have to be submitted by MAHs as part of their regulatory obligation. That this also occurs in daily practice can be observed in company statements in the EU and US commenting on

guideline proposals and a blog post commenting on the above-mentioned ISMP report, stating that MAHs acknowledge a positive causal relationship if the required information for such an assessment is not available [23–25].

Despite the fact that reports derived from organized data collection systems should generally be considered as solicited reports, the actual majority of the (precautionary) reports from organized data collection systems that we have identified in this study relate to spontaneous reports (**Figure 1**). For spontaneous reports, different guidelines are in place. In this case, the ICH E2D guideline states that if an event is spontaneously reported, even if the relationship is unknown or unstated, it implies causality and meets the definition of an ADR.²⁶ This was substantiated by company statements found in the case narratives of such reports, stating that it is company policy to consider every spontaneous report as *suspected*, including reports for which the causality was unsuspected. Regardless of the report type, we believe that three additional factors contribute to the precautionary reporting bias: 1) the large volume of reports that MAHs generally receive from PSPs due to extensive patient contact; 2) the 15-day time frame in which MAHs are required to submit serious events that they have been informed of to regulatory authorities; and 3) possible legal consequences and severe penalties that MAHs face if events are not submitted (in time) [27–30].

There are a number of limitations that apply to this study. The text-mining algorithm was prone to spelling/typing errors and changes in the textual structure and wording of case narratives, which could have resulted in an underestimation of the actual number of precautionary reports identified. Another limitation of this study is the selection of the three case products based on the MedDRA PT “death.” Due to this inclusion criterion, medicinal products were selected that were more likely related to severe diseases with high mortalities. Furthermore, this study was only performed in a single national ADR database, which means that the results should be extrapolated with caution. Nonetheless, since this study primarily served the purpose to demonstrate the existence of the precautionary reporting bias, this approach was regarded as sufficiently appropriate. We believe that the results of this study have EU-wide implications. Since the pharmacovigilance guidelines are effective at the EU level and organized data collection systems are widely integrated in EU healthcare systems, we expect the precautionary reporting bias to be also affecting other national ADR databases and aggregated ADR databases, such as EudraVigilance or the WHO pharmacovigilance database “VigiBase.” We recommend replicating this study in other ADR databases. A more robust algorithm or a case-by-case assessment could improve the identification of precautionary reports. Machine-learning techniques could help to develop and improve text-mining algorithms that can adapt to new case narrative structures (and changes over time). This not only allows for screening more complex narratives, but also for data stratification based on causality, which could improve statistical signal detection methods. Furthermore, we believe that a revision of the E2B-data elements could help identify reports from organized data collection systems, allowing for more stratified analysis. We also recommend an analysis of the potential increase of time-to-detection (of safety signals) caused by precautionary reports. The precautionary reporting bias is an example of the limited alignment between regulation and (clinical) practice. In general, regulators and MAHs act in a rational manner: regulators set low thresholds as

to what needs to be considered an ADR to minimize the chance that actual ADRs are not being reported and MAHs aim to be compliant with current regulations. However, as this study demonstrates, this rational behaviour is at the expense of safety signal detection and, consequently, has implications for the protection of public health. To solve this issue, we propose: 1) a more robust definition and clear understanding of a reportable event, and 2) better alignment between regulators and MAHs as to which reports need to be submitted to regulatory authorities. This study also tries to contribute to the discussion from a scientific perspective. In our opinion, the regulatory science community has an important role in supporting the continuous dialog between stakeholders and evaluating current practices to learn how we can optimally use the pharmacovigilance system. In conclusion, the results of this study demonstrate for the first time the existence of the precautionary reporting bias and its suppressing effect on statistical signal detection methods in ADR databases. The accumulation of nonvalue-added precautionary reports in ADR databases impede the optimal use of the pharmacovigilance system and thus the protection of public health. The exclusion of precautionary reports from the ADR database resulted in the unmasking of previously undetected safety signals. We therefore urge stakeholders to address this issue and improve the alignment between regulation and practice in this area. This will not only decrease burdens for regulators and MAHs but also increase the effectiveness and efficiency of pharmacovigilance systems and thus contribute to a better public health protection.

METHODS

The Netherlands has a long tradition of PSPs and other forms of organized data collections systems that are integrated in the national healthcare system. The Dutch ADR database is therefore a viable setting to conduct this study since the database is complemented with serious spontaneous and solicited reports received via the MAH. All ADR reports that were reported since the start of data collection were included in the analysis. Events were coded according to the MedDRA PT. Medicinal products were coded according to the Anatomical Therapeutic Chemical (ATC) classification system. PL/SQL Developer Version 11.0.6.1776 was used for the extraction of ADR data from the database. Statistical analyses were performed using R statistical software v. 3.2.2. We approached this assessment with three case studies: 1) an erythropoietin for the treatment of anaemia in patients with chronic renal failure and chemotherapy; 2) a bisphosphonate for the treatment of osteoporosis and bone diseases in patients with cancers; and 3) an endothelin receptor antagonist for the treatment of pulmonary artery hypertension (PAH). We selected our case products based on the largest number of ADR reports from MAHs with the MedDRA PT “death” in the ADR database, as this could potentially indicate the existence of precautionary reports [16,18]. We selected three case products from three different MAHs, which have different indications and an assessable case narrative structure. Case narratives of all reports were screened with a text-mining algorithm we developed for this study to determine if a report should be allocated to the “precautionary report” category or whether it is a “confirmed ADR report.” Reports were classified as precautionary reports if the information in the case narrative stated that the causal relationship between the medicinal product and the event was unknown, unsuspected, not related, not assessed, or not provided (e.g., due to insufficient information to establish a causal relationship). See **Box 2** for an example of a case narrative of a spontaneous report that was classified as a precautionary report.

Box 2. Example of a spontaneous report for the bisphosphonate that was classified as precautionary report.

This is an initial spontaneous report from a nurse for a patient support program received on <date> combined with follow-up from a patient’s daughter received on <date>. This report refers to a 73 -year-old male patient who received the *bisphosphonate* for the treatment of metastasized prostate carcinoma from an unspecified date. The last administration of the *bisphosphonate* occurred on <date>. At that time the patient’s condition was deteriorating. Seriousness, causality assessment and event outcome were not provided. The patient died <21 days after the last administration>. The cause of death and a causality assessment were not provided.

Confidential information has been removed/adjusted (italic).

In contrast, if the case narrative stated that the causal relationship between the suspected medicinal product and the event (or to one of the events) was likely, probable, possible, or unlikely (as there is still a possibility), ADR reports were classified as a confirmed ADR report. In line with our conservative approach, we classified reports that were not allocated to either category as a confirmed ADR report. This assessment was done at the level of the report: if a report was classified as both a precautionary report and a

confirmed ADR report, it was ultimately allocated to the confirmed ADR report category. Two scenarios may result in duplicate classification: 1) a report with two (or more) events, of which for one event a suspected causal relationship with the medicinal product exists, whereas this is not the case for the other event(s), or 2) a report with an event for which the causal relationship was initially not established (e.g., due to insufficient information), but was later confirmed during follow-up. The text-mining algorithm was refined until saturation, i.e., further adjustments did not lead to any substantial improvements in the identification of precautionary reports. Both report categories were further stratified by report type, to differentiate between spontaneous and solicited reports, and report source, to identify reports received from an organized data collections system. Two assessors (K.K., J.S.) reviewed the allocated report categories in the process of developing the text-mining algorithm by re-examining case narratives with a random sampling. For each of the three cases, SDR detection was performed before and after we excluded the precautionary reports from the ADR database. The ROR was used as the measure of disproportionality for the detection of SDRs, as this is the standard method applied by the Netherlands Pharmacovigilance Centre Lareb and the EMA, with a threshold for the lower bound of the 95% confidence interval (CI) of one and a minimum of three reported cases of medicinal product/event associations corresponding to the general standards of the EMA.⁹ SDR detection analysis was performed at the level of medicinal product/event associations. Three clinical experts from the Netherlands Pharmacovigilance Centre Lareb independently performed an assessment of all SDRs detected before and after exclusion of precautionary reports from the database, to identify SDRs that relate to events associated with the indication, the natural course of the underlying disease, other treatments for the disease, the (mis-) use of the medicinal product (e.g., “off-label use”), or events referring to patient outcomes (e.g., “terminal state”). The possibility that an SDR relating to such an event represents an actual ADR of the medicinal product can generally be eliminated, as the obvious cause is disease-related and not product-related. Inversely, this approach allowed us to consider every SDR that does not relate to any of the abovementioned event categories, as a safety signal requiring further evaluation due to the absence of obvious confounders. For all newly detected SDRs after exclusion of the precautionary reports from the ADR database (referred to as unmasked SDRs), the percentage of SDRs requiring further evaluation was calculated. The assessment was done in a blinded fashion; i.e., clinical experts did not know which SDRs were detected before and/or after the exclusion of the precautionary reports from the ADR database. Moreover, for each unmasked SDR, we assessed whether the event is mentioned in the SmPC undesirable effects section 4.8.

REFERENCES

1. Wise, L., Parkinson, J., Raine, J. & Breckenridge, A. New approaches to drug safety: a pharmacovigilance tool kit. *Nat. Rev. Drug Discov.* 8, 779–782 (2009).
2. European Commission. Directive 2001/83/EC of the European Parliament and of the Council of 6 November 2001 on the Community code relating to medicinal products for human use. <https://ec.europa.eu/health/sites/health/files/files/eudralex/vol1/dir_2001_83_consol_2012/dir_2001_83_cons_2012_en.pdf> (accessed 31 Aug. 2017).
3. European Commission. Regulation (EC) No 726/2004 of The European Parliament and of the Council of 31 March 2004 laying down Community procedures for the authorisation and supervision of medicinal products for human and veterinary use and establishing a European Medicines Agency. <<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2004:136:0001:0033:en:pdf>> (accessed 31 Aug. 2017).
4. European Medicines Agency. EU Individual Case Safety Report (ICSR) Implementation Guide. <http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2014/04/WC500165979.pdf> (accessed 27 Feb. 2017).
5. European Medicines Agency. Guideline on good pharmacovigilance practices (GVP) Module VI — Management and reporting of adverse reactions to medicinal products. <http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2012/02/WC500123203.pdf> (accessed 27 Feb. 2017).
6. Puijtenbroek, E. & Grootheest, K. Organization of Pharmacovigilance in the Netherlands. In *Mann's Pharmacovigilance* (eds. Andrews, E.B., Moore, N.) 213–216 (Oxford, UK, John Wiley & Sons; 2014).
7. Hauben, M. & Aronson, J.K. Defining 'signal' and its subtypes in pharmacovigilance based on a systematic review of previous definitions. *Drug Saf.* 32, 99–110 (2009).
8. Scholl, J.H.G. & van Puijtenbroek, E.P. The value of time-to-onset in statistical signal detection of adverse drug reactions: a comparison with disproportionality analysis in spontaneous reports from the Netherlands: Time-to-onset in signal detection of ADRs. *Pharmacoepidemiol. Drug Saf.* 25, 1361–1367 (2016).
9. European Medicines Agency. Screening for adverse reactions in EudraVigilance. <http://www.ema.europa.eu/docs/en_GB/document_library/Other/2016/12/WC500218606.pdf> (accessed 27 Feb. 2017).
10. Pariente, A. et al. A potential competition bias in the detection of safety signals from spontaneous reporting databases. *Pharmacoepidemiol. Drug Saf.* 19, 1166–1171 (2010).
11. Arnaud, M. et al. A method for the minimization of competition bias in signal detection from spontaneous reporting databases. *Drug Saf.* 39, 251–260 (2016).
12. Gould, A.L. Practical pharmacovigilance analysis strategies. *Pharmacoepidemiol. Drug Saf.* 12, 559–574 (2003).
13. Almenoff, J. et al. Perspectives on the use of data mining in pharmacovigilance. *Drug Saf.* 28, 981–1007 (2005).
14. Hauben, M. & Hochberg, A. The importance of reporting negative findings in data mining: the example of exenatide and pancreatitis. *Pharm. Med.* 22, 215–219 (2008).
15. Salvo, F. et al. A potential event-competition bias in safety signal detection: results from a spontaneous reporting research database in France. *Drug Saf.* 36, 565–572 (2013).
16. Sookoo, A. An inspector's perspective — considerations for patient support and reimbursement programmes (Stakeholder Meeting, 7th June 2013). <http://www.ema.europa.eu/docs/en_GB/document_library/Presentation/2013/06/WC500144661.pdf> (accessed 27 Feb. 2017).

17. Council for International Organizations of Medical Sciences. Current Challenges in Pharmacovigilance: Pragmatic Approaches (Report of CIOMS Working Group V). <https://cioms.ch/wp-content/uploads/2017/01/Group5_Pharmacovigilance.pdf> (accessed 31 Aug. 2017).
18. Institute for Safe Medication Practices. A critique of A Key Drug Safety Reporting System. <<http://www.ismp.org/QuarterWatch/pdf/s/2014Q1.pdf>> (accessed 27 Feb. 2017).
19. International Council for Harmonisation of technical requirements for pharmaceuticals for human use (ICH). Guideline for good clinical practice. <http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E6/E6_R2__Step_4.pdf> (accessed 1 Mar. 2017).
20. Windsor, A.C. et al. Paradoxical clinical deterioration despite nearcomplete pathological response to neoadjuvant chemotherapy for locally advanced gastro-oesophageal adenocarcinoma. *Surg. Oncol.* 4, 277–279 (1995).
21. Sfrikakis, P.P., Iliopoulos, A., Elezoglou, A., Kittas, C. & Stratigos, A. Psoriasis induced by anti-tumor necrosis factor therapy: A paradoxical adverse reaction. *Arthritis Rheum.* 52, 2513–2518 (2005).
22. Pharmaceutical Industry Associations. Management of Safety Data Originating from Patient Support and Market Research Programmes — Current Challenges from Pharmaceutical Industry and Proposals to Move Forward. <http://www.ema.europa.eu/docs/en_GB/document_library/Presentation/2013/06/WC500144670.pdf> (accessed 1 Mar. 2017).
23. Johnson & Johnson. Commentary on ‘Safety requirements for human drug and biological products.’ <<https://www.fda.gov/ohrms/dockets/dailys/03/oct03/101703/00N-1484-emc-000052-02.doc>> (accessed 1 Mar. 2017).
24. Balderson, D. Meeting Regulatory Agency Expectations on Reporting and Quality of ICSRs. <<http://www.sciformix.com/safety-riskmanagement-blog/meeting-regulatory-agency-expectations-wrt-reporting-and-quality-of-icsrs/>> (accessed 1 Mar. 2017).
25. European Federation of Pharmaceutical Industries and Associations. Submission of comments on legislative proposals to strengthen and rationalise the EU system of pharmacovigilance. <http://ec.europa.eu/health/sites/health/files/files/pharmacovigilance/docs/2007_02_26/48.pdf> (accessed 1 Mar. 2017).
26. International Council for Harmonisation of technical requirements for pharmaceuticals for human use (ICH). Post-approval safety data management: definitions and standards for expedited reporting. <https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E2D/Step4/E2D_Guideline.pdf> (accessed 1 Mar. 2017).
27. European Medicines Agency. European Medicines Agency acts on deficiencies in Roche medicines-safety reporting. <http://www.ema.europa.eu/docs/en_GB/document_library/Press_release/2012/06/WC500129047.pdf> (accessed 1 Mar. 2017).
28. European Medicines Agency. European Medicines Agency starts infringement procedure to investigate Roche’s alleged noncompliance with pharmacovigilance obligations. <http://www.ema.europa.eu/docs/en_GB/document_library/Press_release/2012/10/WC500134176.pdf> (accessed 1 Mar. 2017).
29. European Medicines Agency. Reporting requirements of Individual Case Safety Reports (ICSRs) applicable to marketing authorisation holders during the interim period. <http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2012/05/WC500127657.pdf> (accessed 1 Mar. 2017).
30. Medicines Evaluation Board. ADR reporting requirements based on pharmacovigilance legislation. <<https://english.cbg-meb.nl/human/for-marketing-authorisation-holders/contents/post-marketingauthorisation/reporting-adverse-events>> (accessed 1 Mar. 2017).

CHAPTER 3

3

The Value of Time to Onset in Signal Detection

3.1

Parametric time to onset models were developed to improve causality assessment of adverse drug reactions from antidiabetic drugs

Joep Scholl, Peter van de Ven, Eugène van Puijenbroek

ABSTRACT

Objectives: The aim of this study was to investigate whether the time to onset (TTO) of common adverse drug reactions (ADRs) of antidiabetic drugs could be modelled using parametric distributions and whether these TTO distributions were dependent on patient characteristics. Furthermore, information relevant for daily clinical practice was to be obtained.

Study Design and Setting: We performed an exploratory TTO modelling study, using a cohort of diabetes mellitus patients. Four parametric distributions (exponential, lognormal, gamma, and Weibull) were compared in terms of their goodness of fit. Covariates that could influence the TTO were investigated. In addition, TTO mean and median values were summarized for use in clinical practice.

Results: Overall, the gamma distribution provided the best goodness of fit, although differences with the Weibull distribution were negligible in some instances. No differences in TTO distributions between different antidiabetic drugs for a given ADR were found. The TTO was influenced by suspected concomitant medication for metformin-associated diarrhoea. Mean and median TTO values were similar for different drug – ADR combinations.

Conclusion: Our study shows that the TTO of common ADRs associated with antidiabetic drugs can be modelled using the gamma or Weibull distribution. Furthermore, clinically relevant information about these ADRs can be obtained.

INTRODUCTION

Every patient using a drug is at risk of developing adverse drug reactions (ADRs). These ADRs are a heterogeneous group of medical conditions with different pathophysiological mechanisms, which can be mild and transient in nature, but also life threatening or even fatal. The HARM (Hospital Admissions Related to Medication) study showed that, in the Netherlands, approximately 5.6% of unplanned hospital admissions were drug related. In addition, the results showed that 46.5% of these drug related admissions were possibly preventable [1].

Before registration of a drug, it has been tested in clinical trials for both efficacy and safety, resulting in, among others, a spectrum of ADRs that can be expected when using the drug. The information regarding ADRs is present in the summary of product characteristics (SPC), a regulatory document approved by the health authorities, containing the main characteristics of a drug, including ADRs that have been identified in research. Unfortunately, most of the information regarding ADRs that is present in the SPC of approved drugs is limited to the frequency of occurrence of the ADRs that can be expected. Although this provides the medical community with valuable information regarding the risk of experiencing a certain ADR on a population basis, it is less informative for causality assessment in an individual patient. In other words, when a physician sees a patient with a possible ADR, the frequency of occurrence in the population using the drug is of limited use to determine the actual causality in his/her patient, and additional information is required for causality assessment.

As mentioned in Hill's criteria for causation, temporality is one of the aspects that should be taken into account [2]. When this is applied to causality assessment of possible ADRs in individual patients, an important variable is the time to onset (TTO) of the event [3-7], which can be defined as the time from the start of the suspected drug(s) till the start of the patient's complaints. Intuitively, a long TTO may not always seem indicative of a causal relationship. However, depending on the drug, the type of ADR and the underlying mechanism of toxicity, the TTO can vary considerably in length. For instance, visual field defects due to vigabatrin use can manifest years after the start of the treatment [8], whereas varenicline-induced nausea usually occurs within several days [9]. During the period of drug use, the risk of developing the ADR is not necessarily a constant but can be seen as a function that changes over time.

Previous studies seem to indicate that parametric TTO modelling can be used as a signal detection tool in cohorts of patients starting treatment [10,11] and using registry data of spontaneous reports of adverse events [12]. The parametric methods either statistically test for a nonconstant hazard (rate of occurrence of the adverse event) [10,11] or allow for visual inspection of the hazard over time for the parametric model providing the best fit to the data [12]. In addition, nonparametric methods have been considered for comparing the TTO distributions of adverse events and different vaccines in both spontaneous reports and cohort data [13,14] and for assessing temporal associations between adverse events and medication use using electronic patient record data [15]. In pharmacovigilance, a signal detection tool can be used to pick up signals of ADRs

previously not recognized, whereas a prediction tool could help to determine causality in individual cases. The first occurrence of each type of ADR was included for analysis. Consequently, if a patient experienced the same type of ADR more than once, only the first occurrence after the start of the antidiabetic drug was included. aforementioned studies used the exponential, normal, lognormal, and Weibull distributions for modelling, but not the gamma distribution, which is commonly used in failure time models of continuous variables [16]. In addition, they did not investigate the possible effect of covariates that might influence the TTO of ADRs. This could, however, be relevant when assessing causality and determining optimal treatment in individual patients. In this exploratory study, we investigated whether the TTO distribution of common ADRs in a cohort of type 2 diabetes mellitus (T2DM) patients using antidiabetic drugs could be modelled into a cumulative survival function using different parametric distribution models. We investigated whether the TTO distribution depended on age, sex, and suspected concomitant medication. In addition, possible differences in TTO between different groups of antidiabetic drugs, given a certain common ADR, were investigated. Finally, descriptive information that could be relevant for prescribers and their patients regarding TTO of drug – ADR combinations was obtained.

This study provides insight into the distribution of TTO of some common ADRs in patients using antidiabetic drugs and will assist physicians and their patients to determine whether complaints experienced by patients are drug related or not.

METHODS

Patient selection

For this exploratory study, we used an existing cohort of patients with T2DM from the Lareb intensive monitoring (LIM) system. This is an observational Web-based intensive monitoring system developed to capture more detailed information on the time course of ADRs compared with spontaneous reporting. Patients were eligible if they had T2DM and started using an antidiabetic drug less than 2 weeks before inclusion. A period of 2 weeks was chosen to minimize the risk of recall bias. Data regarding the start of the drug and the start of the complaints were obtained through questionnaires that were sent 2 weeks, 6 weeks, 3 months, 6 months, 9 months, and 1 year after the patient started using the antidiabetic drug. More details regarding the LIM methodology have been described previously [17,18]. For our study, patients who started treatment with an antidiabetic drug that was a combination of more than one active substance were excluded from the analysis.

ADR selection

For each patient, all reported ADRs were screened. If a patient experienced more than one different type of ADR, the first occurrence of each type of ADR was included for analysis. Consequently, if a patient experienced the same type of ADR more than once, only the first occurrence after the start of the antidiabetic drug was included. ADRs were selected for analysis based on five criteria. First, the TTO of the ADR should be valid. Therefore, the exact start date of both the antidiabetic drug treatment and the ADR should be present,

and partial start dates were excluded. Second, the number of ADRs for a given drug – ADR combination should be at least 15. In a previous study, a minimum of five reports was deemed sufficient [12], but we preferred using a minimum number of 15 reports because this probably results in a more reliable estimation of the two-parameter TTO models. Third, the ADR should be present in the SPC of the specific drug to exclude events that have no causal relationship with the drug. Fourth, ADRs from patients who had one or more missing questionnaire before the first occurrence of the ADR were excluded from the analysis because this could result in biased estimates. To investigate the effect of this exclusion criterion, the mean and median values of the TTO of the analysed ADR-drug combinations (with and without exclusions) were compared. Fifth, because we were interested in comparing TTO distributions between different groups of antidiabetic drugs, our analysis was limited to ADRs that occurred in more than one group. For the purpose of uniformity, all reported ADRs were coded using the Medical Dictionary for Regulatory Activities (MedDRA).

TTO modelling

Four different continuous distributions for TTO modelling were used: exponential, lognormal, gamma, and Weibull. The choice for these distributions is based on their common use in failure time models of continuous variables [16]. Because it was already known that the data were right skewed and positive, the normal distribution was not used for modelling. To obtain a preliminary indication of the goodness of fit of the parametric models, two visual comparisons were made. First, cumulative survival was plotted for both the parametric models and the observed data. In addition, the probability density functions of the parametric models were plotted together with the histograms of the observed data. Statistically, the goodness of fit was determined using the Akaike Information Criterion (AIC) [19]. This statistic is based on the $-2\log$ likelihood, with a penalty for the number of parameters in the model, and allows for the comparison of models that are not nested. Subsequently, the differences in AIC values, which provide information about the ranking of the models, were used to calculate evidence ratios for a quantitative comparison of the different models. The evidence ratio is defined as the ratio of the relative likelihood of two models, which can be calculated from the AIC differences (see Equation 1) [20].

$$\text{Relative likelihood} = e^{-1/2*(AIC1-AIC2)} \quad (\text{Eq.1})$$

Finally, the hazard functions of the models with the best fit were plotted to allow a visual inspection of the developing hazard over time.

Analysis of covariates

The antidiabetic drugs were divided into three groups based on their mechanism of action: metformin, sulfonylurea (SU) derivatives, and dipeptidyl peptidase-4 (DPP-4) inhibitors. To investigate the possible effect of age on the TTO, patients were divided into three age groups: 18-44 years, 45-64 years, and 65 years or older. Because patients with T2DM often use several concomitant medications and one or more of these drugs could have an effect on the TTO of an ADR, each patient was screened for suspected concomitant medication. This was defined as a drug that was used during and/or after the start of the study drug (but before the start of the ADR) and that was known to possibly cause the ADR

to be analysed. Differences in TTO distribution related to sex, age group, and suspected concomitant medication, as well as differences between different groups of antidiabetic drugs were tested using two-sample Kolmogorov – Smirnov tests. Differences in median TTO were compared using Mann – Whitney U-test. To correct for multiple testing, a Bonferroni correction was performed. R statistics (Vienna, Austria) version 3.0.1 was used for parametric modelling. All other statistical analyses were performed using SPSS (IBM Corp: Armonk, NY) version 21.

RESULTS

Descriptive statistics

A total of 3,259 patients, who reported 2,751 ADRs, were included into the study. The most commonly used antidiabetic drug was metformin (64.2%), followed by SU derivatives (25.1%) and DPP-4 inhibitors (10.7%). More detailed information is shown in Table 1.

Table 1. Types of antidiabetic drugs used by patients in the study

Antidiabetic drug group	Antidiabetic drug	Number of patients
Metformin	Metformin	2,092
SU derivative	Glimepiride	419
	Tolbutamide	246
	Gliclazide	139
	Glibenclamide	13
DPP-4 inhibitors	Sitagliptin	272
	Vildagliptin	78
Total		3259

After the ADR selection process, 441 ADRs from 383 patients were included for analysis, consisting of three types of ADRs (diarrhoea, n = 255; nausea, n = 127; dizziness, n = 59). All three types of ADRs are type A ADRs, which are characterized as common (exaggerated) pharmacologic effects of a drug, and often have a suggestive time relationship [21,22]. The main reason for excluding ADRs from the analysis was a number of occurrence of less than 15. As a result of the low number of patients in the group of patients using DPP-4 inhibitors, no ADRs were selected for analysis for this group. The result of the selection process is shown in Figure 1.

Fig. 1. Schematic diagram of evaluation of ADRs for inclusion into the analysis. ADR, adverse drug reaction; SPC, summary of product characteristic; SU, sulfonylurea; DPP, dipeptidyl peptidase.

The summary statistics show that the TTO distribution is heavily right skewed and that the mean and median values are similar for the different drug – ADR combinations (see also Table 2).

Table 2. Mean and median time to onset (days) for evaluated ADRs. Mean values are displayed with 95% confidence intervals (95% CI) and median values with interquartile ranges (IQR).

	Diarrhoea		Nausea		Dizziness	
	Metformin (n=236)	SU (n=19)	Metformin (n=108)	SU (n=19)	Metformin (n=35)	SU (n=24)
Mean, days (95% CI)	16.2 (10.5 – 21.9)	11.0 (3.7 – 18.4)	10.9 (5.8 – 16.0)	17.5 (-3.0 – 38.0)	13.1 (5.6 – 20.5)	20.5 (0.4 – 40.6)
Median, days (IQR*)	2.0 (0.0 – 7.0)	5.0 (0.0 – 12.0)	1.5 (0.0 – 7.0)	5.0 (0.5 – 10.0)	2.0 (1.0 – 22.0)	2.0 (0.0 – 10.0)

* IQR, Interquartile range

No statistically significant differences in TTO distributions were observed between metformin and SU derivatives (diarrhoea, $p = 0.256$; nausea, $p = 0.770$; dizziness, $p = 0.952$). In general, the median TTO ranges from 2 to 5 days, however, for each drug – ADR combination a quarter of the patients reported a TTO of 1 week or more. An increase in mean and median TTO was observed for all drug – ADR combinations when cases that were excluded on the basis of one or more missing questionnaires before reporting of the ADR were added to the analysis. The increase was most pronounced in the mean values and less in the median values, which was due to an overrepresentation of high TTO values in the invalid cases (data not shown) and the fact that the median is less sensitive to outliers.

Parametric modelling

Based on the visual comparisons, the gamma and Weibull distributions seemed to have the best goodness of fit for all drug – ADR combinations. From the survival plots, it can be seen that the exponential distribution generally underestimates the number of events for the lower TTO values and overestimates for higher values, whereas the lognormal distribution fits the data rather well for lower TTO values but underestimates for higher values. Figure 2 shows the metformin – diarrhoea and SU derivatives – nausea associations as examples.

Fig. 2. (A + B) Cumulative survival plots of the time to onset of diarrhoea associated with the use of metformin (A) and nausea associated with SU derivatives (B). (C + D) Corresponding histograms and probability density plots of the time to onset of diarrhoea associated with the use of metformin (C) and nausea associated with SU derivatives (D). The histograms and black lines indicate the observed data, whereas the coloured lines represent the fitted parametric models (green = exponential, red = lognormal, blue = gamma, magenta = Weibull). SU, sulfonylurea.

The results of the preliminary visual inspections were confirmed by the AIC test results. For all drug – ADR combinations, the best fit was obtained using the gamma distribution, with the exception of metformin – diarrhoea, where the Weibull distribution showed the best fit (see also Table 3). The hazard functions of these distribution are decreasing over time, indicating that ADRs are more likely to occur shortly after the start of treatment (see Fig. 3).

Table 3. Goodness of fit of the parametric models for diarrhoea, nausea and dizziness associated with the use of metformin or SU derivatives, based on the Akaike Information Component (AIC). Lower values indicate a better fit.

	Diarrhoea		Nausea		Dizziness	
	Metformin	SU	Metformin	SU	Metformin	SU
Exponential	1,788	131	733	148	251	194
Gamma	991	91	365	93	187	97
Lognormal	1,026	102	387	100	204	101
Weibull	985	94	370	95	192	98

Fig. 3. Plots of the modelled hazard function of (A) diarrhoea associated with the use of metformin and (B) nausea associated with SU derivatives (blue = gamma, magenta = Weibull). SU, sulfonylurea.

In terms of quantitative evidence for the relative strength of the model, the evidence ratios in Table 4 show that it is highly unlikely that the exponential or lognormal models fit the data best. When the gamma and Weibull models are compared, the results are less obvious. Especially, for ADRs associated with SU derivatives, the evidence ratios are rather low, indicating that there is less evidence that the gamma model is actually the best model.

Table 4. Evidence ratios of the parametric models for diarrhoea, nausea and dizziness associated with the use of metformin or SU derivatives. Values are the ratios of the reference model (Ref) versus the model indicated. High values are an indication of better evidence that the reference model is superior to the indicated model.

	Diarrhoea		Nausea		Dizziness	
	Metformin	SU	Metformin	SU	Metformin	SU
Exponential	>1,000	>1,000	>1,000	>1,000	>1,000	>1,000
Gamma	20.1	Ref	Ref	Ref	Ref	Ref
Lognormal	>1,000	244.7	>1,000	33.1	>1,000	7.4
Weibull	Ref	4.5	12.2	2.7	12.2	1.7

Analysis of covariates

Based on the two-sample Kolmogorov – Smirnov test, there was no reason to assume that TTO distribution differed with sex, age group, or suspected use of concomitant medication (corrected for multiple testing $\alpha = 0.00278$). However, when comparing the median TTO using the Mann – Whitney test, we found the median TTO to be higher for patients using suspected concomitant medication compared with patients not using this type of medication for the metformin – diarrhoea association ($p = 0.002$, significant when corrected for multiple testing $\alpha = 0.00278$), suggesting a delay of occurrence of diarrhoea in the suspected concomitant medication group (median TTO = 1.0 day in suspected concomitant medication group, median TTO = 2.0 days in group not using suspected concomitant medication). When we compared the mean between the groups using a Gamma general linear model (GLM) with the inverse link, we did not find a significant difference ($p = 0.007$, not significant when corrected for multiple testing $\alpha = 0.00278$). More information is presented in Figure 4.

Fig. 4. Cumulative survival plots of the time to onset of diarrhoea associated with the use of metformin. The solid line indicates the patients not using suspected concomitant medication, whereas the dashed line indicates patients using suspected concomitant medication.

DISCUSSION

The purpose of this study was twofold: First, to investigate whether the TTO of ADRs associated with the use of antidiabetic drugs could be modelled using different parametric distributions and second, to provide health care professionals with practical information regarding the TTO for their daily practice. In terms of antidiabetic drug usage, our study population resembles the overall T2DM population physicians see in daily practice, with metformin being the most commonly used drug, followed by SU derivatives and DPP-4 inhibitors [23]. The limited use of DPP-4 inhibitors in the general T2DM population at

the time of inclusion into the study seems to be the main reason for the low numbers of reported ADRs in this group. In general, the results of our study show that the TTO of the investigated ADRs has a median value of several days, which corresponds to what is already known based on clinical experience. However, several patients reported TTO values that seemed extreme. Those might have been considered outliers and could have been excluded from the analysis in studies not aimed specifically at TTO research. For example, one patient reported diarrhoea 333 days after starting metformin. Although this could easily be considered an outlier because it is generally accepted that metformin-induced diarrhoea develops within several days, the TTO modelling may provide a different view. Because the Weibull and gamma models fitted the data rather well, but the TTO of 333 days seemed counterintuitive, a literature search concerning TTO of metformin-induced diarrhoea was performed. This resulted in one article describing that metformin-induced diarrhoea can occur several years after the start of the therapy [24]. This example shows how parametric modelling of the TTO can provide valuable additional information about known ADRs. Despite this finding, it should also be realized that diarrhoea experienced several months after the start of the metformin treatment can be due to other reasons, such as a gastrointestinal infection.

We investigated whether the exclusion of ADRs of patients for whom one or more questionnaires were missing before reporting of the ADR had an effect on the mean and median values of the TTO. The increase in both median and mean TTO that was observed was mainly due to an overrepresentation of high TTO values in the invalid cases. This could be the result of recall bias that occurs if patients already experienced an episode of symptoms earlier but did not remember it when they completed the questionnaire.

Parametric modelling of the TTO

Based on the AIC values, the gamma distribution generally fits the data best. Although this observation seems obvious for comparison with the exponential and lognormal distributions, the evidence ratios indicate that it is less clear for differences between the gamma and Weibull models, especially for the SU derivative-related ADRs. Because no formal cut-off values for evidence ratios are available, one could debate whether, for example, an evidence ratio of 1.7 (gamma vs. Weibull for dizziness associated with SU derivatives) is strong enough evidence to claim the gamma model as the best model. It is of importance, however, to realize that these low ratios may be due to the relatively small number of observations for the ADRs in the SU derivative group.

Beside the above, it should be noted that the ADRs analysed are all type A ADRs. In addition to this type of ADRs, there are also type B ADRs, which are often allergic or idiosyncratic reactions. Although these type B ADRs often have a sensitization period of approximately 10 days, a longer TTO is also possible [22]. It would be valuable to investigate type B ADRs (e.g., rash, pruritus) to see whether their TTO distribution can also be modelled using a Weibull or gamma distribution, but due to a lack of observations for these ADRs, this was unfortunately not possible with our currently available data. Furthermore, our study was performed on a cohort of patients, and it would be of interest to see whether similar studies on data received through spontaneous reporting will yield similar results.

Analysis of covariates

Although no statistically significant differences were found when the whole TTO distribution was compared between patients using suspected concomitant medication and patients not using this type of medication, for the association metformin – diarrhoea using the two-sample Kolmogorov – Smirnov test or when means were compared using a parametric Gamma GLM, a test for difference in location (median) showed the median TTO to be larger for patients using suspected concomitant medication that was known to cause diarrhoea. No explanation for this observation could be found in the literature, and the most likely hypothesis is that the differences are due to unmeasured variables that were not equally distributed between the groups, such as presence and type of concomitant diseases. For example, it is well known that renal insufficiency increases the half-life of metformin [25] and with that the plasma levels of the drug. The increased exposure would make patients more susceptible to the development of metformin-induced diarrhoea. If these patients were more present in the subgroup not using concomitant medication associated with diarrhoea, this could bias the results. In this example, the two-sample Kolmogorov – Smirnov test likely lacked power to detect the difference in medians because it is sensitive to many different aspects of the distributions, including the location, dispersion, and skewness. The Mann – Whitney test specifically focusses on the location, which would make it the preferred test when only differences in location (as shifts in median TTO) are of interest.

As the gamma distribution fitted our TTO data best, we also compared mean TTO between groups using and not using concomitant medication using a Gamma GLM. Although parametric methods typically have higher power when compared with nonparametric methods under the assumption that the parametric model is the true underlying model, we did not find a difference in means when using the Gamma GLM. This may be related to the high uncertainty in the group means due to their strong dependence on outliers corresponding to large TTOs. The medians are less sensitive to outliers, which may explain why the Mann – Whitney test showed a significant difference in location, whereas the Gamma GLM did not. In situations where covariates are identified that influence the distribution, separate details should be provided of distributions in specific subgroups based on covariates. Medians and confidence intervals for the subgroup could for instance be obtained using quantile regression.

Study limitations

A possible limitation of the study could be the introduction of selection bias because ADRs that occur later could be less frequently reported because they are less likely to be associated with the drug by the patient. This was partly dealt with by sending questionnaires at regular intervals up to 1 year after enrolment. Another limitation was the limited number of ADRs that were eligible for analysis, which was approximately 16% of the total number of ADRs reported. This was mainly because of the minimum number of 15 observations for each drug-ADR combination was often not reached. Because the number of 15 reports was rather arbitrary, future studies could investigate the possibility of a lower threshold but more likely larger data sets will be required. In addition, our direct modelling approach could be influenced by the presence of events that occurred coincidentally. Nonparametric methods that compare TTO distributions

between different subsets of data are not hampered by this issue as long as distribution of the coincident events does not differ between the subsets. For modelling, inclusion of coincident events may lead to biased estimates as ADRs reported later may be less likely to be causally related to the drug. The use of more advanced parametric models involving mixtures of, for instance, a gamma distribution for the ADRs causally related to the drug and an exponential distribution for the coincident ADRs may solve this issue. In addition, this may allow the estimation of an odds for an ADR observed after a specific time to be causally related to the drug. Such a modelling strategy will, however, require a large number of reported ADRs and long follow-up to be able to accurately estimate the uniform probability for the occurrence of coincident ADRs over time.

Study implications

The TTO is an important variable in the causality assessment of ADRs [3,4]. The results of our study show that the TTO of common type A ADRs in patients using antidiabetic drugs can be modelled using the gamma and Weibull distributions. These models summarized by information as the mean, median values, and quantiles (e.g., the 95th and 99th percentile) of the fitted distribution may be used to develop a structured approach for causality assessment in individual cases. In clinical practice, several patient characteristics are taken into account in diagnosis and treatment of a patient, and it is obvious that a causal relationship in a patient experiencing diarrhoea 6 months after starting metformin is less likely than diarrhoea after 1 week. However, although other aetiologies after several months of treatment cannot be ruled out, the possibility of a causal relationship should not be dismissed in advance, as underlying conditions of patients involved that may influence the aetiology of ADRs may also change over time. In this study, we only investigated several drug – ADR combinations, it seems plausible that in general, the gamma and Weibull distributions can be used for modelling of common type A ADRs. This should, however, be confirmed with additional research. This also holds for the analysis of data from spontaneous reports and electronic health record databases, which may be different from cohort data, and could be used to validate our method.

CONCLUSION

Our study shows that the TTO of common ADRs associated with antidiabetic drugs can be modelled using the gamma or Weibull distribution. Although the gamma models seemed to fit the data best in most cases, differences with the Weibull models were modest in some instances, and neither distributions can be considered as generally best suited for modelling the TTO of these types of ADRs. These models can be used as a concept for the development of a structured approach in the causality assessment of ADRs. In addition, this study shows that TTO modelling can be useful to ascertain more clinically relevant information about these ADRs, and although causality assessment in ADRs is challenging, a causal relationship should not be dismissed based on a long TTO alone.

REFERENCES

1. Leendertse AJ, Egberts AC, Stoker LJ, van den Bemt PM. Frequency of and risk factors for preventable medication-related hospital admissions in the Netherlands. *Arch Intern Med* 2008;168:1890-6.
2. Hill AB. The environment and disease: association or causation? *Proc R Soc Med* 1965;58:295-300.
3. Kelly WN, Arellano FM, Barnes J, Bergman U, Edwards RI, Fernandez AM, et al. Guidelines for submitting adverse event reports for publication. *Drug Saf* 2007;30(5):367-73.
4. Aronson JK. Anecdotes as evidence. *BMJ* 2003;326:1346.
5. Arimone Y, Begaud B, Miremont-Salame G, Fourrier-Reglat A, Molimard M, Moore N, et al. A new method for assessing drug causation provided agreement with experts' judgment. *J Clin Epidemiol* 2006;59:308-14.
6. Theophile H, Andre M, Arimone Y, Haramburu F, Miremont-Salame G, Begaud B. An updated method improved the assessment of adverse drug reaction in routine pharmacovigilance. *J Clin Epidemiol* 2012;65:1069-77.
7. Agbabiaka TB, Savovic J, Ernst E. Methods for causality assessment of adverse drug reactions: a systematic review. *Drug Saf* 2008;31(1): 21-37.
8. Eke T, Talbot JF, Lawden MC. Severe persistent visual field constriction associated with vigabatrin. *BMJ* 1997;314:180-1.
9. Oosterhuis I, Harmark L, van Puijenbroek E. Experiences with the use of varenicline in daily practice in the Netherlands: a prospective, observational cohort study. *Drug Saf* 2014;37(6):449-57.
10. Cornelius VR, Sauzet O, Evans SJ. A signal detection method to detect adverse drug reactions using a parametric time-to-event model in simulated cohort data. *Drug Saf* 2012;35(7):599-610.
11. Sauzet O, Carvajal A, Escudero A, Molokhia M, Cornelius VR. Illustration of the weibull shape parameter signal detection tool using electronic healthcare record data. *Drug Saf* 2013;36(10):995-1006.
12. Maignen F, Hauben M, Tsintis P. Modelling the time to onset of adverse reactions with parametric survival distributions: a potential approach to signal detection and evaluation. *Drug Saf* 2010;33(5): 417-34.
13. van Holle L, Zeinoun Z, Bauchau V, Verstraeten T. Using time-to-onset for detecting safety signals in spontaneous reports of adverse events following immunization: a proof of concept study. *Pharmacoepidemiol Drug Saf* 2012;21(6):603-10.
14. van Holle L, Tavares Da Silva F, Bauchau V. Signal detection based on time-to-onset: extending a new method from spontaneous reports to observational studies. *Pharmacoepidemiol Drug Saf* 2014;23(8): 849-58.
15. Norén GN, Hopstadius J, Bate A, Star K, Edwards IR. Temporal pattern discovery in longitudinal electronic patient records. *Data Min Knowl Discov* 2010;20(3):361-87.
16. Kalbfleisch JD, Prentice RL, editors. *The statistical analysis of failure time data*. 2nd ed. Hoboken: Wiley and Sons; 2002.
17. Harmark L, van Puijenbroek E, Straus S, van Grootheest K. Intensive monitoring of pregabalin: results from an observational, web-based, prospective cohort study in the Netherlands using patients as a source of information. *Drug Saf* 2011;34(3):221-31.
18. Harmark L, Puijenbroek E, Grootheest K. Longitudinal monitoring of the safety of drugs by using a web-based system: the case of pregabalin. *Pharmacoepidemiol Drug Saf* 2011;20(6):591-7.

19. Akaike H. A new look at the statistical model identification. *IEEE Trans Automatic Control* 1974;19:716-23.
20. Burnham KP, Anderson DR. Model selection and multimodel inference. A practical information-theoretic approach. 2nd ed. New York, USA: Springer; 2002.
21. Meyboom RH, Egberts AC, Edwards IR, Hekster YA, De Koning FH, Gribnau FW. Principles of signal detection in pharmacovigilance. *Drug Saf* 1997;16(6):355-65.
22. Meyboom RH, Lindquist M, Egberts AC. An ABC of drug-related problems. *Drug Saf* 2000;22(6):415-23.
23. College voor Zorgverzekeringen. GIP Databank. College voor Zorgverzekeringen. GIP Databank. version date: 22-3-2011, Available at <http://www.gipdatabank.nl/>, Accessed August, 2014.
24. Lysy J, Israeli E, Goldin E. The prevalence of chronic diarrhea among diabetic patients. *Am J Gastroenterol* 1999;94(8):2165-70.
25. Farmacotherapeutisch Kompas. Farmacotherapeutisch Kompas. (version date: 30-3-2011), Available at <http://www.fk.cvz.nl/> Accessed March 30, 2011.

3.2

The value of time to onset in statistical signal detection of adverse drug reactions: a comparison with disproportionality analysis in spontaneous reports from the Netherlands

Joep Scholl, Eugène van Puijenbroek

ABSTRACT

Purpose In pharmacovigilance, the commonly used disproportionality analysis (DPA) in statistical signal detection is known to have its limitations. The aim of this study was to investigate the value of the time to onset (TTO) of ADRs in addition to DPA.

Methods We performed a pilot study using individual case safety reports (ICSRs) for three drugs (Cervarix®, nitrofurantoin and simvastatin) from the Lareb spontaneous reporting database. TTO distributions for drug – ADR associations were compared to other ADRs for the same drug and to other drugs for the same ADR using two-sample Anderson–Darling testing. Statistically significant associations were considered true positive (TP) signals if the association was present in the official product information of the drug. Sensitivity and specificity for the TTO method were compared with the DPA method. As a measure of disproportionality, the reporting odds ratio (ROR) was used.

Results In general, sensitivity was lower, and specificity was higher for the TTO method compared to DPA. The TTO method showed similar sensitivity for all three drugs, whereas specificity was lower for Cervarix®. Eight additional TP signals were found using the TTO method compared to DPA.

Conclusions Our study shows that statistical signal detection based on the TTO alone resulted in a limited number of additional signals compared to DPA. We therefore conclude that the TTO method is of limited value for full database statistical screening in our setting.

INTRODUCTION

Disproportionality analysis (DPA) is a common method for statistical signal detection (SSD) in spontaneous reporting systems (SRS) of adverse drug reactions (ADRs). Several methods of DPA exist, and their common goal is to detect statistically significant differences in observed-versus-expected ratios for different drug-ADR combinations [1]. Although DPA has demonstrated its value as a screening tool in recent years, it also has its limitations. One major issue is that it does not take the quality of individual case safety reports (ICSRs) into account, because it is solely based on the (relative) number of ICSRs for a certain association. SSD is subject to several types of bias, including selective reporting, notoriety bias [1], masking [2] and the Weber effect [3].

One of the trivial elements present in ICSRs is the time to onset (TTO) which can be defined as the duration from the start of the suspect drug until the start of the ADR. Although this information is usually applied in a qualitative manner in the analysis of ICSRs, it is quantitative in nature. For different drug-ADR combinations, the TTO distributions can vary considerably because of several factors, including the underlying mechanism of toxicity. This aspect can be used in both clinical causality assessment and SSD, as shown in previous studies [4-11]. In particular, several studies by Van Holle *et al.* have shown the additional value of TTO in SSD for events following immunization, using non-parametric Kolmogorov–Smirnov (KS) testing [7,8,12]. However, these studies were limited to vaccines and therefore concern a rather selective population of mainly healthy individuals. Additionally, TTO distributions for ADRs related to vaccines may differ from those related to non-vaccines because of several factors. Most importantly, vaccines are generally administered once whereas many non-vaccines are administered on a daily basis, both temporary and chronically. Therefore, ADRs with longer TTOs may be less likely to be reported for vaccines. Furthermore, the spectrum of ADRs related to vaccines, and therefore their TTO distributions, differs from that of non-vaccines. For instance, a large proportion of vaccine – related ADRs concern injection site disorders with a typical TTO whereas most ADRs related to non-vaccines in our database are reported for drugs that are taken orally. Taking these differences in TTO distributions into account, the Anderson–Darling (AD) test seems more appropriate for databases containing both vaccines and non-vaccine ICSRs, because it has more power than the KS test [13].

In this study, we used two-sample AD testing to investigate how TTO-based signal detection performs compared to DPA in a spontaneous reporting database containing ICSRs of both vaccines and non-vaccines. In order to capture the diversity of the database, we analysed three different types of drugs used in patient populations with different demographical and clinical features. The objective is to determine if TTO-based signal detection can be a useful addition to standard DPA in a database containing reports related to both vaccines and non-vaccines.

METHODS

In this proof of concept study, differences in TTO distributions for selected drug–ADR combinations were tested using two-sample AD testing. The performance, in terms of sensitivity and specificity, was compared with DPA based on the reporting odds ratio (ROR) [14].

Data source

The data for this study were derived from the database of the Netherlands Pharmacovigilance Centre Lareb. This database consists of spontaneous reports of suspected ADRs reported by both healthcare professionals and consumers from all parts of the Netherlands. All reports (both serious and non-serious) since the start of reporting (1986) until the cut-off date (31 July 2015) were eligible.

Drug and ADR selection

In order to cover the diversity of ICSRs in our database, three drugs with different indications and different patient populations were selected: bivalent human papillomavirus vaccine (Cervarix®), used for prophylaxis in a young and relatively healthy female population; nitrofurantoin, used for uncomplicated urinary tract infections in a diverse population in terms of age and co-morbidities; and simvastatin, used in a generally older population with multiple co-morbidities. Drugs were classified according to the WHO Anatomical Therapeutic Chemical (ATC) classification system, using the level of chemical substance (fifth level) [15]. ICSRs with suspected drugs containing more than one active substance (e.g., simvastatin/ezetimibe), and those with drugs reported as interacting were excluded. Additionally, duplicate reports were excluded, based on the duplicate detection procedure used at Lareb during assessment of individual ICSRs. All ADRs with an exact TTO were selected and coded using the preferred terms (PTs) from the Medical Dictionary for Regulatory Activities (MedDRA, version 18.0). The TTO was considered to be exact if both the start date of the suspect drug and ADR were full dates (i.e. yyyy-mm-dd). This was done because the majority of associations in the database has a TTO of several days to weeks, and using partial dates would introduce too much statistical imprecision.

Statistical testing

Statistical testing for differences in TTO distributions was performed using the two-sample AD test. The test statistic of this non-parametric test belongs to the quadratic class of empirical distribution function (EDF) statistics and determines if two samples come from the same continuous distribution. Compared to the KS test, it has in general more power and is more sensitive to differences in the tails of the distributions, shift, scale or symmetry [13,16]. For each drug – ADR combination, the TTO distribution was compared with two comparator distributions:

- The TTO distribution of all other ADRs for the same drug
- The TTO distribution of all other drugs for the same ADR

Each ICSR was used only once per statistical test. In other words, if an ICSR was used for the drug – ADR combination to be tested, it was not used for “all other ADR for same drug” or “all other drugs for same ADR”. The null hypothesis for each test was that both EDFs came from the same distribution and the alternative hypothesis was that both EDFs did not come from the same distribution (two-sided testing). The significance level was set at $\alpha = 0.05$. DPA was performed using the ROR, which is similar to the odds ratio in a case–control study [17] and can be calculated from a standard 2×2 table using the formula $ROR = ad/bc$ (see Table 1). Additionally, a 95% confidence interval (95% CI) is calculated and the ROR is considered statistically significant if the lower limit of the 95% CI exceeds 1.

Table 1. Two-by-two contingency table for calculation of the reporting odds ratio (ROR)

	ICSRs with event	ICSRs with other events
ICSRs with the suspect drug	a	b
ICSRs with other drugs	c	d

ICSR, Individual Case Safety Report; ROR = ad/bc

Sensitivity and specificity analyses

The performance of the TTO method in terms of sensitivity and specificity was compared with the ROR, which is the standard method for SSD at Lareb. Sensitivity and specificity were defined as described in Equations (1) and (2).

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (\text{Eq.1})$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (\text{Eq.2})$$

where TP is the number of true positive, TN is the number of true negative, FP is the number of false positive and FN is the number of false negative signals. The summary of product characteristics (SPC) was used as a reference for signal classification (see Table 2). For each ADR described in the SPC, the applicable MedDRA PTs were matched. Cases of reported symptoms not listed in the SPC that were considered highly indicative of a diagnosis mentioned in the SPC were discussed with a clinically qualified assessor. Based on the outcome of the discussion, it was decided whether the association was considered to be present in the SPC or not. Although there are issues and limitations regarding its use as a gold standard, the SPC has been used in several studies aimed at SSD [7,18]. For each of the drugs investigated, one SPC was used as the gold standard. In case of multiple Marketing Authorization Holders (MAHs) for one drug the SPC of the innovator was used. Because the AD test is sensitive to both dispersion and skewness, outliers may influence the results and performance. Therefore, sensitivity and specificity were determined for the full dataset without restrictions and for the dataset restricted to ICSRs with a TTO of 0 – 30 days. Because the TTO has no effect on disproportionality, separate sensitivity analysis for the 0 – 30 day time window for the DPA method was not applicable. Statistical analyses were performed with R statistics version 3.2.3.

Table 2. Definitions of true positive (TP), true negative (TN), false positive (FP) and false negative (FN) signals

	TTO	ROR
True positive	Both AD tests $p < 0.05$ and ADR present in SPC	$LL95\%CI > 1$ and ADR present in SPC
True negative	At least one AD test $p \geq 0.05$ and ADR not present in SPC	$LL95\%CI \leq 1$ and ADR not present in SPC
False positive	Both AD tests $p < 0.05$ and ADR not present in SPC	$LL95\%CI > 1$ and ADR not present in SPC
False negative	At least one AD test $p \geq 0.05$ and ADR present in SPC	$LL95\%CI \leq 1$ and ADR present in SPC

AD, Anderson-Darling; LL95%CI, lower limit of 95% confidence interval; ROR, reporting odds ratio; SPC, summary of product characteristics; TTO, time to onset

RESULTS

Descriptive statistics

A total of 3,313 ICSRs, containing 6,660 drug – ADR associations, were included into the analysis. The total number of unique drug–ADR associations was 252. For the analysis of the 0 – 30 days time window, the numbers are similar, with the exception of the relatively high percentage of drug – ADR associations with a TTO of more than 30 days for simvastatin (see also Table 3). This was mainly because of the high number of reports of simvastatin-associated musculotendinous complaints. The percentage of reports containing an exact TTO was 79.1% for Cervarix®, 85.2% for nitrofurantoin and 73.6% for simvastatin.

Table 3. Overview of number of reports included in the study

	Number of ICSRs	Total number of ADRs*	Number of unique ADRs*
Cervarix®			
All	966	2757	55
0-30 days	929	2664	54
Nitrofurantoin			
All	893	1592	76
0-30	848	1561	76
Simvastatin			
All	1454	2311	121
0-30	811	1319	78

* coded as MedDRA preferred term
ICSR, individual case safety report; ADR, adverse drug reaction

Performance

From the 252 unique associations, 42 TP signals (17%) were identified by the TTO method compared to 94 by DPA (37%). For the full dataset, sensitivity ranged from 0.27 to 0.32 for TTO and from 0.57 to 0.77 for DPA, where sensitivity was highest for Cervarix® in both methods. Sensitivity for TTO decreases when the time window is restricted to 0 – 30 days, particularly for simvastatin. The TTO method showed a higher specificity (TTO, 0.75 – 0.98; DPA, 0.42 – 0.79) (see Table 4 for more details). Eight associations were identified by the TTO method that were not identified by DPA. Conversely, 59 associations were identified by DPA and not by TTO. Because sensitivity was in general low for the TTO method, empirical cumulative distribution (ECD) plots were made for the TTO of several drug – ADR associations to provide a visual insight into the TTO distributions. For illustrative purposes, one TP signal (nitrofurantoin – interstitial lung disease) and one false negative signal (simvastatin – myopathy) were selected based on the well-known causal relationship of both associations [19,20].

Table 4. True positive signals, sensitivity and specificity for TTO and DPA signal detection for the total dataset and the 0 – 30 days time window

	Cervarix®		Nitrofurantoin		Simvastatin	
	DPA	TTO	DPA	TTO	DPA	TTO
True positive signals (%)						
All		10 (18)		15 (20)		17 (14)
0-30 day	24 (44)	9 (17)	34 (45)	6 (8)	36 (30)	6 (8)
Sensitivity						
All		0.32		0.29		0.27
0-30 days	0.77	0.29	0.63	0.12	0.57	0.15
Specificity						
All		0.75		0.88		0.98
0-30 days	0.42	0.87	0.79	1.00	0.79	0.97

* Percentage of the total number of association per drug
DPA, disproportionality analysis; TTO, time to onset

Figure 1. Empirical cumulative distribution plots of the TTO for the true positive signal nitrofurantoin—interstitial lung disease ($n = 23$) and the false negative signal simvastatin—myopathy ($n = 18$). a) nitrofurantoin—interstitial lung disease (green) versus nitrofurantoin with other ADRs (blue). b) nitrofurantoin—interstitial lung disease (green) versus interstitial lung disease for all other drugs (blue). c) simvastatin—myopathy (green) versus simvastatin with other ADRs (blue). d) simvastatin—myopathy (green) versus myopathy for all other drugs (blue). p = p -value of the two-sample AD test

DISCUSSION

In this study, we investigated the performance in terms of sensitivity and specificity of TTO-based signal detection versus DPA, to see if the TTO method would be a valuable addition to DPA. The quality of the TTO data in our database was satisfactory with an exact TTO reported in approximately 75% – 85% of the analysed associations. For simvastatin, the relatively high percentage of drug – ADR associations with a TTO of more than 30 days may be explained by the fact that it was the only drug in our study that is used chronically. The major difference between the current study and the initial study by Van Holle *et al* [7], is the use of the AD test instead of the KS test. Because of its higher power in case of,

among others, differences in the tails of the distributions, we consider the AD test to be more appropriate in databases containing both vaccines and non-vaccines, because of the (theoretical) diversity of TTO distributions in these databases.

In our study, we used three types of drugs with different indications and user populations as a proxy for the diversity of our database. It should be noted, however, that this selection was somewhat arbitrary, and a different set of drugs could have been used for our study. On the other hand, our results regarding sensitivity were similar for all three drugs and clearly show that the added value of the TTO method is limited and in our opinion a different set of drugs would not have made a substantial difference. One could debate whether pharmacokinetics (PK) should have been incorporated in the analyses, because there are several parameters that may influence the TTO, such as the half-life ($t_{1/2}$) of the drug or its time to maximum plasma concentrations ($t_{\max}$). We did not consider this appropriate for several reasons: First, PK is not fully informative about the presence of the drug in the body. Rather, it is a measure of the drug concentrations in the blood. For several drugs, however, concentration in different tissues can be substantially different from those in plasma. As a result, the tissue concentrations for each drug should be taken into account depending on the type of ADR (e.g., concentrations in skin tissue for dermatological ADRs, concentrations in cerebral spinal fluid for central nervous system ADRs etc.). In addition, it is not clear when an ADR may occur. This may be the case when the maximum concentration has been reached, but possibly also sooner or later. Second, the TTO of an ADR cannot be predicted by the PK of the drug alone. This typically applies to type B ADRs, but even type A ADRs can have a longer TTO than would be expected based on their PK. For instance metformin-induced diarrhoea can still occur several months or even years after initiation [21]. Third, several drugs have active metabolites with their own PK characteristics, and it would not be feasible to incorporate this into our study.

In general, the sensitivity for the TTO method was low compared to DPA, but showed similar results for the three drugs. The low sensitivity may be because of several causes. First, reporters tend to report the TTO in different units as its duration increases. Therefore, it is likely that actual TTOs of, for example, 8 or 10 days, respectively, will both be reported as one week. This phenomenon leads to TTO clustering, possibly resulting in decreased statistical precision. Second, most associations in our study have a relatively short TTO, which makes it more difficult to achieve statistical significance. It would be of interest to investigate the current method in a specific dataset containing drug – ADR associations with a longer TTO to see how this will affect performance. Third, our definition of a TP signal required both AD tests to be statistically significant for each association. This may have been an overly conservative approach, but enables us to make a proper comparison with the results from Van Holle *et al* [7]. Fourth, the use of the SPC as the gold standard has its limitations because not all ADRs listed in this document have a proven causality. Vice versa, if an ADR is not listed in the SPC, it does not necessarily mean that a causal relationship is absent. The appropriateness of use of the SPC as gold standard depends on the goal of signal detection. Because SRS are initially intended to detect hitherto unknown associations, the SPC is less suitable for validation purposes, because, ADRs are listed in this document with a more or less proven causal relationship. However, we expect reporters to submit ADRs with an established causal relationship more frequently. It

should be noted that the above also affects the performance of DPA, and can therefore not be the only explanation for the difference in performance between these two methods. Although specificity was in favour of the TTO method compared to DPA, we considered the added value of introducing TTO into the analysis of our full dataset to be limited. In SSD with spontaneous data, it is important not to miss a true signal, and as a result, a high sensitivity (albeit at the cost of a higher number of false positives) is considered more important than a high specificity.

In order to analyse the effect of outliers on performance, we decided to perform an additional analysis with a restricted TTO time window of 0 – 30 days. This analysis resulted in a decreased sensitivity, most likely because of the lower number of cases compared to the full data set, resulting in decreased statistical precision. The time to onset ECD plots did not reveal clues that might explain the low sensitivity for this method. As can be seen from Figure 1b, ECD curves that seem similar at visual inspection can still result in a statistically significant difference for the AD test ($p < 0.001$). Conversely, two visually distinct curves (Figure 1d) showed no statistically significant difference ($p = 0.352$). Although sensitivity was low for the TTO method, this method could still have been a valuable addition as a screening tool if the identified TP signals were different from those identified by DPA. However, because this was the case for only eight associations (where 59 associations identified by DPA were not identified by TTO), this strengthens our conclusion of the limited value of the TTO method for full database screening. Because the Lareb database has a limited number of reports compared to, for example, the EudraVigilance database, it would be of interest to investigate the performance of our method in a larger database. This was, however, beyond the scope of our study because the intention was to develop an additional SSD method specifically for the Lareb database. Additionally, using a larger dataset would have led to results that cannot automatically be extrapolated to our local database because different databases show different results for SSD methods [22]. However, the method we used can be tested in other databases as well, but in our opinion, the results acquired would only apply to that particular database.

Based on the current results, the TTO method does not provide additional value to DPA and does not seem suitable for general screening purposes. However, this does not imply that the method described may be of additional value for specific drugs or ADRs with a distinctive time of onset. Furthermore, as mentioned above, our approach may have been rather conservative because a TP signal required both AD test results to be statistically significant. For future research, it would be of interest to investigate how different definitions of a TP signal would influence sensitivity and specificity. We hypothesize that comparing the EDF for a drug – ADR association with that for all other ADRs for the same drug will lead to different results than the comparison with the same ADR for all other drugs. Because sensitivity for the TTO method was similar for the three drugs (range 0.27 – 0.32), the method seems robust enough for additional research at first sight. It should be emphasized that the results of this study should be seen in light of the full pharmacovigilance process. For the process at Lareb, this means that any result from any method of SSD is always followed by case-by-case review if a possible signal is identified. And in addition to SSD, all ICSRs received by healthcare professionals and consumers are assessed in a case-by-case manner, increasing the likelihood of finding new signals.

It is the opinion of the authors that using a single method only for signal detection is a suboptimal approach and increases the risk of missing signals.

CONCLUSIONS

Our study shows that TTO-based SSD was less sensitive than the DPA-based method we used for comparison, identifying only a small number of additional associations as possible signals. Therefore, we consider it of no additional value for full database screening purposes in its current form. Whether TTO is the method that can be useful with a different definition of a true positive signal, or in screening subsets of drug – ADR associations, remains to be investigated.

REFERENCES

1. Bate A, Evans SJ. Quantitative signal detection using spontaneous ADR reporting. *Pharmacoepidemiol Drug Saf* 2009; 18(6): 427–436.
2. Maignen F, Hauben M, Hung E, van Holle L, Dogne JM. Assessing the extent and impact of the masking effect of disproportionality analyses on two spontaneous reporting systems databases. *Pharmacoepidemiol Drug Saf* 2014; 23(2): 195–207.
3. 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(6): 743–749.
4. Cornelius VR, Sauzet O, Evans SJ. A signal detection method to detect adverse drug reactions using a parametric time-to-event model in simulated cohort data. *Drug Saf* 2012; 35(7): 599–610.
5. Sauzet O, Carvajal A, Escudero A, Molokhia M, Cornelius VR. Illustration of the weibull shape parameter signal detection tool using electronic healthcare record data. *Drug Saf* 2013; 36(10): 995–1006.
6. Maignen F, Hauben M, Tsintis P. Modelling the time to onset of adverse reactions with parametric survival distributions: a potential approach to signal detection and evaluation. *Drug Saf* 2010; 33(5): 417–434.
7. van Holle L, Zeinoun Z, Bauchau V, Verstraeten T. Using time-to-onset for detecting safety signals in spontaneous reports of adverse events following immunization: a proof of concept study. *Pharmacoepidemiol Drug Saf* 2012; 21(6): 603–610.
8. van Holle L, Tavares Da Silva F, Bauchau V. Signal detection based on time-to-onset: extending a new method from spontaneous reports to observational studies. *Pharmacoepidemiol Drug Saf* 2014; 23(8): 849–858.
9. van Holle L, Bauchau V. Signal detection on spontaneous reports of adverse events following immunisation: a comparison of the performance of a disproportionality-based algorithm and a time-to-onset-based algorithm. *Pharmacoepidemiol Drug Saf* 2014; 23(2): 178–185.
10. Norén GN, Hopstadius J, Bate A, Star K, Edwards IR. Temporal pattern discovery in longitudinal electronic patient records. *Data Min Knowl Discov* 2010; 20(3): 361–387.
11. Scholl JH, van de Ven PM, Van Puijenbroek EP. Parametric time-to-onset models were developed to improve causality assessment of adverse drug reactions from antidiabetic drugs. *J Clin Epidemiol* 2015; 68(12): 1423–1431.
12. van Holle L, Bauchau V. Use of logistic regression to combine two causality criteria for signal detection in vaccine spontaneous report data. *Drug Saf* 2014; 37(12): 1047–1057.
13. Engmann S, Cousineau C. Comparing distributions: the two-sample Anderson–Darling test as an alternative to the Kolmogorov–Smirnov test. *JAQM* 2011; 6(3): 1–17.
14. Stricker BH, Tijssen JG. Serum sickness-like reactions to cefaclor. *J Clin Epidemiol* 1992; 45(10): 1177–1184.
15. Structure and principles of the ATC classification system. Available at: http://www.whocc.no/atc/structure_and_principles/ (accessed 10 January 2016).
16. Razali NM, Yap BW. Power comparisons of Shapiro–Wilk, Kolmogorov–Smirnov, Lilliefors and Anderson–Darling tests. *J Stat Model Anal* 2011; 2(1): 21–33.
17. Waller P, van Puijenbroek E, Egberts A, Evans S. The reporting odds ratio versus the proportional reporting ratio: ‘deuce’. *Pharmacoepidemiol Drug Saf* 2004; 13(8): 525–526.

18. Szarfman A, Machado SG, O'Neill RT. Use of screening algorithms and computer systems to efficiently signal higher-than-expected combinations of drugs and events in the US FDA's spontaneous reports database. *Drug Saf* 2002; 25(6): 381–392.
19. Mohassel P, Mammen AL. The spectrum of statin myopathy. *Curr Opin Rheumatol* 2013; 25(6): 747–752.
20. Kabbara WK, Kordahi MC. Nitrofurantoin-induced pulmonary toxicity: a case report and review of the literature. *J Infect Public Health* 2015; 8(4): 309–313.
21. Lysy J, Israeli E, Goldin E. The prevalence of chronic diarrhea among diabetic patients. *Am J Gastroenterol* 1999; 94(8): 2165–2170.
22. Candore G, Juhlin K, Manlik K, et al. Comparison of statistical signal detection methods within and across spontaneous reporting databases. *Drug Saf* 2015; 38(6): 577–587.

3.3

Time to onset in statistical signal detection revisited: A follow-up study in long-term onset adverse drug reactions

Joep Scholl, Florence van Hunsel, Eelko Hak, Eugène van Puijenbroek

ABSTRACT

Purpose: In a previous study, we developed a signal detection method using the time to onset (TTO) of adverse drug reactions (ADRs). The aim of the current study was to investigate this method in a subset of ADRs with a longer TTO and to compare its performance with disproportionality analysis.

Methods: Using The Netherlands's spontaneous reporting database, TTO distributions for drug—ADR associations with a median TTO of 7 days or more were compared with other drugs with the same ADR using the two-sample Anderson–Darling (AD) test. Presence in the Summary of Product Characteristics (SPC) was used as the gold standard for identification of a true ADR. Twelve combinations with different values for the number of reports and median TTO were tested. Performance in terms of sensitivity and positive predictive value (PPV) was compared with disproportionality analysis. A sensitivity analysis was performed to compare the results with those from the previous study.

Results: A total of 38,017 case reports, containing 32,478 unique drug—ADR associations. Sensitivity was lower for the TTO method (range 0.08–0.34) compared with disproportionality analysis (range 0.60–0.87), whereas PPV was similar for both methods (range 0.93–1.0). The results from the sensitivity analysis were similar to the original analysis.

Conclusions: Because of its low sensitivity, the developed TTO method cannot replace disproportionality analysis as a signal detection tool. It may be useful in combination with other methods.

INTRODUCTION

Statistical signal detection on databases containing spontaneous reports of adverse drug reactions (ADRs) has become a valuable addition to case-by-case assessment of individual case reports. Historically, these statistical methods are mainly based on observed versus expected ratios, using both Bayesian and frequentist approaches [1]. In recent years, additional methods have been investigated and developed as a way to generate potential signals, including the use of the time to onset (TTO) [2-10]. In general, the TTO studies are based on the hypothesis that the TTO distributions differ between true causally related ADRs and drug—event combinations without a causal relationship. This hypothesis makes sense from a pharmacological point of view since ADRs can have a different time course (rapid, first dose, early, intermediate, late, and delayed) [11]. However, contradictory results regarding the additional value of TTO analyses in signal detection have been reported in different studies [3,5,10].

Recently, we investigated the performance of a TTO-based method and compared it with disproportionality analysis based on the reporting odds ratio (ROR) used at our centre [10]. The main finding was that the sensitivity was too low to be useful for screening, most likely because of two main reasons: (a) most ADRs in our database have a median TTO of approximately 1 to 2 days, resulting in decreased discriminative power. This overrepresentation may have a pharmacological cause since the majority of ADRs are type A effects, and their TTO is related to the pharmacokinetic/pharmacodynamics properties of the suspect drug [12]. On the other hand, it may also be the result of selective reporting. Indeed, recall bias is likely to occur with an increasing TTO since the reporter does not associate the complaints with previous drug exposure. Additionally, coincidental events with a short latency may be reported unjustly and may therefore be misclassified as true ADRs. (b) TTO distributions of drug—ADR combinations were tested against all other ADRs for the same drug ($\text{drug—ADR}_{\text{other}}$) and against all other drugs for the same ADR ($\text{drug}_{\text{other}}\text{—ADR}$). For a drug—ADR combination to be a true positive signal, both test results had to be statistically significant. This may however, have been an overly conservative approach, and one could debate whether the comparison with $\text{drug}_{\text{other}}\text{—ADR}$ only would be more appropriate. After all, the hypothesis is that the TTO of a true ADR for a certain drug will have a different distribution compared with the same, possibly noncausally related symptoms for all other drugs, based on its pharmacology. The assumption behind this is that associations in the $\text{drug}_{\text{other}}\text{—ADR}$ subset will also contain reports where the ADR is not a true ADR but a suspected ADR and could reflect, e.g., background noise. Therefore, one may expect a more uniformly distributed TTO compared with a true ADR, and that is what is to be tested. However, when testing against $\text{drug—ADR}_{\text{other}}$, differences in TTO distributions are inevitable since different ADRs have different pharmacological mechanisms. The low sensitivity found in our previous study was somewhat unexpected and, based on the afore-mentioned overrepresentation of ADRs with a short TTO, led to the question if the method would perform better when applied to ADRs with a longer TTO. Additionally, our previous study was a proof of concept study investigating only three drugs, resulting in limited generalizability of the results. Therefore, we chose a full database approach in the current study.

The goal of this study was to compare the performance of TTO-based signal detection of ADRs with a longer TTO with disproportionality analysis in terms of sensitivity and positive predictive value (PPV). We expect that for longer TTOs, the misclassification is less outspoken, and thus, TTO analysis may yield a better performance. To our knowledge, this is the first full database approach investigating the TTO in statistical signal detection using a subset of ADRs with longer TTOs.

METHODS

In this study, we performed a retrospective analysis of reports of suspected ADRs to detect differences in TTO distributions using a subset of suspected ADRs with a longer TTO and compared the results with disproportionality analysis in terms of sensitivity and PPV. Presence in the Summary of Product Characteristics (SPC) was used as the gold standard to determine if a suspected ADR was a true ADR or not.

Data selection

Data from the spontaneous reporting system maintained by The Netherlands Pharmacovigilance Centre Lareb were used for the study. In the routine assessment of the reports received, presence or absence of the ADR in the SPC is logged by the assessor during assessment of the case report at the drug—ADR level, with the exception of vaccine-related reports, where presence in the SPC was not logged at all. Therefore, the latter were excluded from this study. Because the objective of the study was to investigate ADRs with a higher latency, only associations with a median TTO of 7 days or more were included. It should be noted that individual reports with a shorter TTO for a certain drug—ADR association were included if the median of the group was 7 days or more. To allow for a proper comparison of both methods, only associations included in the TTO analysis were used in the disproportionality analysis. However, the number of reports included for each association could differ between both methods since reports without a valid TTO were included in the disproportionality analysis but not the TTO analysis. All case reports since the start of reporting to Lareb (1986) until July 2017 were eligible for inclusion. Duplicate reports were excluded, based on the duplicate detection procedure used at Lareb during assessment of individual case reports. Reports from marketing authorization holders (MAH) were excluded. Reports from studies were not explicitly excluded, but because, at the time of data extraction, all reports from studies were received from MAHs only, the exclusion of those reports automatically implied an exclusion of reports from studies. Drugs were classified according to the WHO Anatomical Therapeutic Chemical (ATC) classification system [13], using the level of chemical substance (fifth level). ADRs were coded using the preferred terms (PTs) from the Medical Dictionary for Regulatory Activities [14] (MedDRA, version 19.0).

Disproportionality analysis

The ROR was used as the measure for disproportionality analysis as it is the standard method used at our centre [15,16]. It is based on a 2×2 contingency table as shown in Table 1. On the basis of Table 1, the ROR and its 95% confidence interval (95%CI) can be calculated using

$$\text{ROR} = \frac{A/B}{C/D} = \frac{AD}{BC} \quad (\text{Eq. 1})$$

$$95\% \text{CI} = e^{\ln(\text{ROR}) \pm 1.96 * \sqrt{\frac{1}{A} + \frac{1}{B} + \frac{1}{C} + \frac{1}{D}}} \quad (\text{Eq. 2})$$

The ROR was considered statistically significant if the lower limit of the 95%CI was greater than 1.

Table 1. A 2 x 2 contingency table used for ROR calculation

	Case reports with ADR of interest	Case reports with all other ADRs
Case reports with the drug of interest	A*	B*
Case reports with all other drugs	C*	D*

* Each case reports can attribute to only one value in the table. For instance, if the case report has the drug and ADR of interest but also an additional ADR, it is assigned to A, but not to B

Time to onset analysis

Differences in the TTO distributions of ADRs were tested using the two-sample non-parametric Anderson–Darling (AD) test. This test determines if two samples belong to the same continuous distribution, based on location, dispersion, and skewness [17]. To investigate the effect of the number of reports per association (N) and TTO values on the performance, several combinations were tested (see Table 2). For each of the combinations, differences in TTO distributions were tested using two-sample AD testing (drug—ADR vs drug_{other}—ADR). Our previous study was based on a double AD test for each drug—ADR:

- drug—ADR versus drug_{other}—ADR
- drug—ADR versus drug—ADR_{other}

In the interest of between-study validity, a sensitivity analysis was performed to investigate the effect of the current approach. The sensitivity analysis will be referred to as secondary analysis in order to avoid confusion with sensitivity as the measure of performance. Statistical testing was performed two sided with a significance level of $\alpha = 0.05$.

Table 2. Different combinations of number of reports and median TTO used in the analysis

Number of case reports	Median TTO (days)			
	≥ 7	≥ 14	≥ 30	≥ 60
≥ 5	x	x	x	x
≥ 10	x	x	x	x
≥ 15	x	x	x	x

TTO, time to onset

Performance

The performance of both methods was based on the sensitivity and PPV that were defined as described in Equations 3 and 4.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (\text{Eq. 3})$$

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (\text{Eq. 4})$$

Where TP is the number of true positive, FP is the number of false positive and FN is the number of false negative signals (see Table 3 for classification). For sensitivity and PPV, interpolated surface plots were generated for both methods to allow for a visual interpretation of the results. Interpolation was based on the Akima algorithm for scattered-data surface fitting [18]. Statistical analyses were performed with R statistics version 3.3.2. [19].

Table 3. Definitions of true positive, true negative, false positive and false negative signals

	TTO	ROR
True positive	AD test $p < 0.05$ and ADR present in SPC*	LL95%CI > 1 and ADR present in SPC
True negative	AD test $p \geq 0.05$ and ADR not present in SPC#	LL95%CI ≤ 1 and ADR not present in SPC
False positive	AD test $p < 0.05$ and ADR not present in SPC^	LL95%CI > 1 and ADR not present in SPC
False negative	AD test $p \geq 0.05$ and ADR present in SPC§	LL95%CI ≤ 1 and ADR present in SPC

AD test, Anderson-Darling test; LL95%CI, lower limit of the 95%CI of the ROR; SPC, summary of product characteristics; TTO, time to onset

For the secondary analysis the definitions are:

* Both AD tests (drug – ADR vs Drug_{other} – ADR and vs drug – ADR_{other}) $p < 0.05$ and ADR present in SPC

At least one AD test (drug – ADR vs Drug_{other} – ADR or vs drug – ADR_{other}) $p \geq 0.05$ and ADR not present in SPC

^ Both AD tests (drug – ADR vs Drug_{other} – ADR and vs drug – ADR_{other}) $p < 0.05$ and ADR not present in SPC

§ At least one AD test (drug – ADR vs Drug_{other} – ADR or vs drug – ADR_{other}) $p \geq 0.05$ and ADR present in SPC

RESULTS

Descriptive statistics

A total of 38,017 case reports, containing 3,247 unique drug—ADR associations, were included into the analysis. For the TTO analysis, less reports ($n = 29,876$) were included due to a lack of information on the TTO, which was missing in 26.5% of the associations. Additional descriptive information is presented in Table 4.

Table 4. Descriptive statistics of the case reports used in the study

	Number (n)
Number of case reports	
• TTO analysis	29,876
• ROR analysis	38,017
Number of associations	
• TTO analysis	45,904
• ROR analysis	62,440
Number of unique associations	3,247
Number of unique drugs*	338
Number of unique suspected ADRs#	484

ROR, reporting odds ratio; TTO, time to onset

* classified according to the Anatomical Therapeutic Chemical (ATC) classification system

coded as MedDRA Preferred Terms (PTs)

Performance

The sensitivity for the TTO method was low (range 0.08-0.34) compared with disproportionality analysis (range 0.60-0.87). In contrast, PPV was similar for both methods (range 0.93-1.00). Additional information can be found in Table 5. A more detailed analysis showed that both the number of TP and FP signals were three to five times higher for the disproportionality analysis method, whereas the number of FN signals was approximately two to three times lower. TN signals were similar between groups (data not shown). Interestingly, sensitivity increased with an increasing number of reports for the TTO method, whereas for disproportionality analysis, it increased with both increasing number of reports and increasing TTO (see Figure 1). For the subset with the highest TTO sensitivity (N15_TTO7), 96% of the associations detected by TTO were also detected by disproportionality analysis.

Table 5. The number of true positive signals, sensitivity and PPV for each of the combinations of the number of reports and TTO

	True positive signals (n)		Sensitivity		PPV	
	TTO	ROR	TTO	ROR	TTO	ROR
N5_TTO7	427	1,732	0.16	0.60	0,96	0,93
N10_TTO7	283	783	0.27	0.68	0,97	0,96
N15_TTO7	193	471	0.34	0.74	0,98	0,96
N5_TTO14	269	1,330	0.14	0.64	0,95	0,93
N10_TTO14	175	580	0.24	0.71	0,97	0,96
N15_TTO14	126	351	0.33	0.78	0,98	0,97
N5_TTO30	100	780	0.10	0.70	0,93	0,94
N10_TTO30	54	309	0.17	0.77	0,98	0,97
N15_TTO30	26	169	0.18	0.82	1,00	0,97
N5_TTO60	47	507	0.08	0.75	0,96	0,95
N10_TTO60	21	191	0.12	0.80	1,00	0,98
N15_TTO60	14	99	0.23	0.87	1,00	1,00

PPV, positive predictive value; ROR, reporting odds ratio; TTO, time to onset

Figure 1. Surface plots of sensitivity and PPV after interpolation. a) sensitivity for TTO. b) sensitivity for disproportionality analysis. c) PPV for TTO. d) PPV for disproportionality analysis. PPV, positive predictive value; TTO, time to onset

Secondary analysis

The secondary analysis showed a similar pattern for sensitivity and PPV as the original analysis (see Figure 2). However, absolute values for sensitivity were in general slightly lower for the secondary analysis (0.07 – 0.21).

Figure 2. Surface plots of sensitivity and PPV for the secondary analysis. a) sensitivity for original TTO analysis. b) sensitivity for secondary analysis. c) PPV for original TTO analysis. d) PPV for secondary analysis. PPV, positive predictive value; TTO, time to onset

DISCUSSION

In this follow-up study, we investigated the performance of a previously developed TTO signal detection method using the two-sample AD test applied to ADRs with a longer TTO and compared it with disproportionality analysis. The major reason for conducting this study was the previous finding that the majority of TTOs was 1 to 2 days, possibly limiting discriminative power of the statistical test. Reports from MAHs were excluded because it is our experience they are more often poorly documented than reports received directly by Lareb. In addition, presence of the ADR in the SPC is not logged since reports from MAHs are not assessed manually. There was a substantial number of case reports with missing

TTO values. However, we do not have any indication that this occurs on a selective basis (e.g., for certain types of drugs and/or ADRs).

The results show that sensitivity was lower for the TTO method, whereas PPV was similar and although there was a small number of associations that was detected by the TTO only, this did not apply to particular clinical entities. The similarity in PPV between the TTO method and disproportionality analysis can be explained by the three to five times higher FP value for the ROR method (thereby nullifying its three to five times TP advantage over the TTO method). The secondary analysis showed similar results although in general, sensitivity was slightly lower than in the original analysis. In signal detection using spontaneously reported data, it is most important not to miss a true signal as it is a timely detection of a signal. Therefore, we compared performance in terms of sensitivity and PPV, deliberately neglecting possible differences in specificity. The observation that sensitivity for the TTO method increases with increasing number of reports per association (n) cannot be explained unambiguously but may be a statistical artefact since more reports lead to a larger test sample and subsequently to more statistical power.

Previous similar studies by others were performed using the two-sample Kolmogorov–Smirnov test [3,4,6]. However, the AD test has generally more power and is more sensitive to differences in shift, scale or symmetry. In addition, it is better at detecting small differences, even when samples sizes are larger [17,20]. Given these facts, we considered the AD test to be more appropriate for our study. The major strength of this study is its full database approach covering a wide range of drug—ADR associations. Compared with methods where a subset of drugs is used in the analysis, this approach reduces selectivity in the results and increases generalizability. On the other hand, databases containing spontaneous reports show substantial differences, and a similar approach for a different database may result in a better performance [21]. Additionally, the comparison between the statistically stricter approach and the more tolerant approach (both tests significant versus only one test significant for a TP signal respectively) and the fact that the results were similar, increases the between-study validity.

As mentioned in our previous study, the use of the SPC as the gold standard (particularly in finding new, previously undocumented ADRs) has its drawbacks and could influence the results of this study by introducing misclassification because presence in the SPC does not necessarily imply a causal relationship between drug and ADR. The opposite is also true since absence in the SPC does not necessarily imply absence of causality. Moreover, the SPC lists both ADRs identified prior to and after market authorization, where the former may be preferentially reported immediately after authorization, artificially influencing some counts. Nevertheless, we consider the use of the SPC valid because (a) it is the only variable regarding causality that is consistently present for each drug—ADR association in our database since it is logged during manual assessment of the report; (b) if any misclassification occurs, it is most likely to be nondifferential and would therefore not account for any major differences between the two methods investigated. A second limitation of this study is the exclusion of vaccine related reports due to the lack of information about presence of the ADR in the SPC in our database. Since previous studies related to vaccine report by others [3-6] showed promising results,

it would be very valuable to investigate these reports in our database. Recently, Lareb has taken a new database into production providing the possibility to add information about presence in the SPC to vaccine related reports. In due time, this information can be used to perform similar studies for these type of reports. Finally, some selection bias may have been introduced. Since a positive ROR (or any other measure of disproportionality) may have been the trigger for a signal in the past, we cannot rule out that this trigger resulted in the inclusion of the ADR into the SPC. So theoretically, the ROR could have contributed directly to the outcome (SPC) introducing selection bias. We think however, that this will account for a limited number of drug—ADR associations and will not have had a serious influence on the results of our study, if any at all.

The results of the secondary analysis were in line with those of the original analysis, with similar patterns in sensitivity and PPV emerging from the surface plots. The lower absolute value for sensitivity compared with the original analysis can be explained by the fact that both AD tests had to be statistically significant to allow for a true positive signal in the secondary analysis, resulting in a decreased number of true positive signals. Additionally, the fact that only one of both tests had to be statistically significant to allow for a false negative signal resulted in an increase in false negative signals and therefore a corresponding decrease in sensitivity. One final issue to consider is the fact that databases containing spontaneous reports are cross-sectional. It is well-known that this type of data can be subject to various sorts of bias including recall bias. This type of bias may result in a less precise estimation of the TTO by the reporter and subsequently to a certain degree of randomness in the collected data, thereby decreasing statistical power. Additionally, TTO clustering may occur depending on the unit reported. For instance, both a TTO of 6 and 8 days may be reported as 1 week.

On the basis of the results of the current study, the TTO method does not perform well enough to replace disproportionality analysis as a screening tool. However, the results of the AD test may be useful as a parameter in a prediction model-based screening approach similar to the one we published recently [22].

CONCLUSIONS

The results of our study show that TTO-based signal detection, restricted to a dataset containing ADRs with a longer TTO only, cannot replace disproportionality analysis as a screening method. This may in part be due to the presence of several types of bias known to occur in spontaneous reporting.

REFERENCES

1. Bate A, Evans SJ. Quantitative signal detection using spontaneous ADR reporting. *Pharmacoepidemiol Drug Saf.* 2009;18(6):427-436.
2. Norén GN, Hopstadius J, Bate A, Star K, Edwards IR. Temporal pattern discovery in longitudinal electronic patient records. *Data Min Knowl Discov.* 2010;20(3):361-387.
3. van Holle L, Zeinoun Z, Bauchau V, Verstraeten T. Using time-to-onset for detecting safety signals in spontaneous reports of adverse events following immunization: a proof of concept study. *Pharmacoepidemiol Drug Saf.* 2012;21(6):603-610.
4. van Holle L, Tavares Da Silva F, Bauchau V. Signal detection based on time-to-onset: extending a new method from spontaneous reports to observational studies. *Pharmacoepidemiol Drug Saf.* 2014;23(8): 849-858.
5. van Holle L, Bauchau V. Signal detection on spontaneous reports of adverse events following immunisation: a comparison of the performance of a disproportionality-based algorithm and a time-to-onset-based algorithm. *Pharmacoepidemiol Drug Saf.* 2014;23(2):178-185.
6. van Holle L, Bauchau V. Use of logistic regression to combine two causality criteria for signal detection in vaccine spontaneous report data. *Drug Saf.* 2014;37(12):1047-1057.
7. Cornelius VR, Sauzet O, Evans SJ. A signal detection method to detect adverse drug reactions using a parametric time-to-event model in simulated cohort data. *Drug Saf.* 2012;35(7):599-610.
8. Sauzet O, Carvajal A, Escudero A, Molokhia M, Cornelius VR. Illustration of the weibull shape parameter signal detection tool using electronic healthcare record data. *Drug Saf.* 2013;36(10):995-1006.
9. Maignen F, Hauben M, Tsintis P. Modelling the time to onset of adverse reactions with parametric survival distributions: a potential approach to signal detection and evaluation. *Drug Saf.* 2010;33(5):417-434.
10. Scholl JH, Van Puijenbroek EP. The value of time-to-onset in statistical signal detection of adverse drug reactions: a comparison with disproportionality analysis in spontaneous reports from the Netherlands. *Pharmacoepidemiol Drug Saf.* 2016;25(12):1361-1367.
11. Aronson JK, Ferner RE. Joining the DoTS: new approach to classifying adverse drug reactions. *BMJ.* 2003;327(7425):1222-1225.
12. Meyboom RH, Lindquist M, Egberts AC. An ABC of drug-related problems. *Drug Saf.* 2000;22(6):415-423.
13. Structure and Principles of the ATC classification system. http://www.whoccno/atc/structure_and_principles/, 2011.
14. Brown EG, Wood L, Wood S. The medical dictionary for regulatory activities (MedDRA). *Drug Saf.* 1999;20(2):109-117.
15. Stricker BH, Tijssen JG. Serum sickness-like reactions to cefaclor. *J Clin Epidemiol.* 1992;45(10):1177-1184.
16. Van Puijenbroek EP, Bate A, Leufkens HG, Lindquist M, Orre R, Egberts AC. 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.
17. Engmann S, Cousineau C. Comparing distributions: the two-sample Anderson-Darling test as an alternative to the Kolmogorov-Smirnov test. *J Appl Quant Meth.* 2011;6(3):1-17.
18. Akima H. Algorithm 761: scattered-data surface fitting that has the accuracy of a cubic polynomial. *ACM Trans Math Software.* 1996;22(3):362-371.

19. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2016.
20. Razali NM, Yap BW. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *J Stat Model Anal.* 2011;2(1):21-33.
21. Candore G, Juhlin K, Manlik K, et al. Comparison of statistical signal detection methods within and across spontaneous reporting databases. *Drug Saf.* 2015;38(6):577-587.
22. Scholl JHG, van Hunsel FPAM, Hak E, Van Puijenbroek EP. A prediction model-based algorithm for computer-assisted database screening of adverse drug reactions in the Netherlands. *Pharmacoepidemiol Drug Saf.* 2018;27(2):199-205.

CHAPTER 4

A prediction model-based algorithm for computer-assisted database screening of adverse drug reactions in the Netherlands

Joep Scholl, Florence van Hunsel, Eelko Hak, Eugène van Puijenbroek

ABSTRACT

Purpose: The statistical screening of pharmacovigilance databases containing spontaneously reported adverse drug reactions (ADRs) is mainly based on disproportionality analysis. The aim of this study was to improve the efficiency of full database screening using a prediction model-based approach.

Methods: A logistic regression-based prediction model containing 5 candidate predictors was developed and internally validated using the Summary of Product Characteristics as the gold standard for the outcome. All drug-ADR associations, with the exception of those related to vaccines, with a minimum of 3 reports formed the training data for the model. Performance was based on the area under the receiver operating characteristic curve (AUC). Results were compared with the current method of database screening based on the number of previously analysed associations.

Results: A total of 25,026 unique drug-ADR associations formed the training data for the model. The final model contained all 5 candidate predictors (number of reports, disproportionality, reports from healthcare professionals, reports from marketing authorization holders, Naranjo score). The AUC for the full model was 0.740 (95% CI; 0.734–0.747). The internal validity was good based on the calibration curve and bootstrapping analysis (AUC after bootstrapping = 0.739). Compared with the old method, the AUC increased from 0.649 to 0.740, and the proportion of potential signals increased by approximately 50% (from 12.3% to 19.4%).

Conclusions: A prediction model-based approach can be a useful tool to create priority-based listings for signal detection in databases consisting of spontaneous ADRs.

INTRODUCTION

Spontaneous reporting systems have been the cornerstone of pharmacovigilance since their introduction in the 1960s. The main aim of spontaneous reporting is the early detection of previously unrecognized adverse drug reactions (ADRs). In addition, spontaneous reporting can also be useful for obtaining information on new aspects of known associations between drugs and ADRs [1]. Although spontaneous reporting has its methodological shortcomings from an epidemiological perspective, it is a valuable method for the early detection of ADRs [2,3]. An effective signal detection process is a key element of a spontaneous reporting system in pharmacovigilance. For the detection of signals, the Netherlands Pharmacovigilance Centre Lareb has historically relied on a case-by-case clinical review of incoming reports, directly submitted by health care professionals (HCP) and consumers. This review is performed by trained pharmacovigilance assessors, the majority of them being medical doctors and pharmacists. Reports that may represent a potential signal in the view of the assessor are discussed in a weekly scientific meeting. Potential signals undergo a more detailed analysis [4]. Lareb has criteria in place for assessors to determine which reports should be discussed at the weekly scientific meeting. However, because multiple assessors are involved in this process and the selection of reports for the weekly scientific is prone to some level of subjectivity, a computer-assisted database screening tool is in place as an additional approach to reduce the risk for missing potential signals [5]. The screening tool is even more important for ADRs reported by marketing authorization holders (MAHs) that may be indicative of potential signals, as these are not assessed on a case-by-case basis at Lareb.

The computer-assisted database screening tool used in the Netherlands relies on the number of reports of drug-ADR associations and disproportionality based on the reporting odds ratio (ROR). With the disproportionality analyses, the observed rate of a drug and ADR together is compared with an expected value based on their relative frequencies reported individually in the spontaneous reporting database [5,6]. In the approach applied at our centre, the lower limit of the 2-sided 95% confidence interval (CI) is used combined with a number of at least 3 reports per association. Associations can be automatically selected by the screening tool based on 1 or more of the following predefined criteria; Anatomical Therapeutic Chemical code (allowing the assessor to screen more efficiently), ADR being unlabelled in the Summary of Product Characteristics (SPC), number of reports (≥ 3), threshold of the lower limit of the 2-sided 95% CI of the ROR (ROR025) (>1), pre-specified calendar date, set during previous analysis. Associations highlighted by the screening tool undergo a short analysis by trained pharmacovigilance assessors. Based on the decision of the assessor, subsequent new thresholds can be specified (ADR, unlabelled, number of reports or lower limit 95% CI, new date) or the association can undergo further detailed analyses. The association will be highlighted again as soon as one of the aforementioned criteria is met [7]. Although the current approach facilitates the selection of potential signals, the downside of this approach is that it yields a high number of associations that need an initial, short analysis, which is a time-consuming process. With the current methods, associations can be ranked on the basis of number of reports or the level of disproportionality. However, it is not possible to prioritize based on other, possibly relevant features of the reported association. Prioritization based on associations that

would theoretically yield the number of highest potential signals is seen by Lareb as a way to improve timelines of the signal detection process.

The Uppsala Monitoring Centre, WHO Collaborating Centre for International Drug Monitoring, has developed a data-driven screening algorithm for emerging drug safety signals that accounts for report quality and content, called *vigiRank* [8]. *VigiRank* is a model which uses several predictive values as determined in the WHO Global ICSR database; *VigiBase*®. Some of the predictors that were found for this model are not applicable for a national database, such as geographical spread. The Lareb database contains a high number of reports with free text and have a relatively high documentation grade, as represented by the *vigiGrade*® completeness score of the Lareb reports in *VigiBase*® [9]. Because Lareb does a case-by-case analyses of all reports, except those received through the MAH, it is known for each association whether the ADR is labelled in the Dutch SPC. Also, for each report (except those received through the MAH), a causality score (Naranjo) is calculated [10]. Based on this, and other, additional information that is available for ICSRs, a more elaborate set of predictors would probably be suited for a screening tool on the Dutch national spontaneous database.

The primary aim of this study was to develop a new prediction model-based screening tool in order to improve statistical signal detection. Secondary aim was to compare this new model to the old screening tool, which is based on the number of reports and the ROR025.

METHODS

Setting

In this study, we developed a logistic regression-based prediction model for drug-ADR associations present in the Lareb spontaneous reporting database. Using the linear predictor of this model, a prioritized list of associations not present in the SPC was made for comparison with the current method. The data for this study were derived from the database of the Netherlands Pharmacovigilance Centre Lareb. This database consists of spontaneous reports of suspected ADRs reported to Lareb directly by both HCP and consumers. Additionally, reports from MAHs regarding events that occurred in The Netherlands are imported into our database from the European Medicines Agency database *EudraVigilance*. Each report contains 1 or more drug-ADR associations. For the development of the prediction model, all drug-ADR associations were extracted from each report. ADRs were coded using the preferred terms from the Medical Dictionary for Regulatory Activities [11]. Drugs were classified according to the WHO Anatomical Therapeutic Chemical classification system [12].

Outcome

The outcome of the model was defined as the presence in the SPC of each unique drug-ADR association at the time of the analysis. Although the use of the SPC to determine if an association is actually an ADR (implying causality) has its limitations, it has been used in several studies aimed at statistical signal detection [13,14]. At Lareb, for each association

present in an ICSR received directly from a HCP or consumer, a causality assessment using the Naranjo score is performed [10]. For Naranjo question 1: "Are there previous conclusive reports on this reaction?", 3 options are available at our centre: 1) "Yes, listed in SPC", 2) "Yes, described in other literature", 3) "No/unknown". In order to be in line with the original Naranjo scale, a score of +1 is assigned if option 1) or 2) is chosen and a score of 0 if option 3) is chosen in routine practice at Lareb. For the purpose of this study, associations with the outcome were defined as option 1), and associations without the outcome as option 2) or 3). Associations consisting of MAH reports only were manually assessed for the presence in the SPC.

Inclusion / exclusion criteria

All reports received until 12-May-2016 were considered eligible for inclusion with the exception of reports related to vaccines. For these reports, a method other than Naranjo is used to determine causality. Because that particular method lacks information about the presence in the SPC, reports related to vaccines were excluded. For statistical considerations, only associations with a minimum of 3 reports were selected, because this was deemed to be the minimum number of reports needed for a reliable ROR estimation.

Selection of candidate predictors

For each association, the following variables were selected as candidate predictors in the model:

1. The number of ICSRs.
2. The lower limit of the 2-sided 95%CI of the ROR (ROR025).
3. The percentage of ICSRs derived from health care professionals (HCP). This variable was selected because HCP reports differ from consumer reports [15].
4. The percentage of ICSRs derived from MAHs. This variable was selected because it is our experience that MAH reports differ from reports received directly from HCPs / consumers, most likely due to regulatory obligations for MAHs. This is confirmed by the fact that only 0.2% of the reports from MAHs are included in signals published by Lareb (for HCPs and consumers this is 3.9% and 2.0%, respectively [16]).
5. The mean Naranjo score across all reports containing the association. The answer to question 1 was excluded in the calculation of the Naranjo score because this question is the basis for the outcome variable of the prediction model. Additionally, because no Naranjo scores are present for MAH reports, scores for those reports were set at +2 because we considered it valid to assume that the ADR occurred after the suspect drug was given (Naranjo question 2).

Development of the model

A multivariable logistic regression model was developed using backward step-wise selection. The training data consisted of all drug-ADR associations with a minimum of 3 reports. The candidate predictors with $P < 0.05$ (Wald test) were fitted into the model.

Due to non-linearity of the predictors, they were converted to categorical variables with equally sized categories. For all candidate predictors, the generalized variance inflation factor, a measure of variance for categorical variables, was calculated to investigate collinearity, using 4 as a conservative cut-off value [17,18].

Evaluation of the model

Internal validation of the model was performed by means of a calibration curve of the observed versus predicted probabilities, Hosmer-Lemeshow goodness of fit testing, and bootstrap resampling. For the latter, the validate function from the R package “rms” was used (number of bootstrap samples = 1000, no backward step-down variable deletion). The performance of the model was based on the area under the receiver operating characteristic curve (AUC). Because our prediction model was developed to be used specifically for the Lareb database, external validation was not considered to be relevant. Additionally, previous signal detection research showed that there are substantial differences among different pharmacovigilance databases [19], making validation using an external data source (eg, a database from another country) of limited value.

Comparison with current method of signal detection

Although the current method of screening at Lareb is not based on a prediction model, we considered it of interest to compare the current model with a model containing only the number of reports and the ROR025, which are the basis for the current screening method. The comparison consisted of 2 elements: First, the performance in terms of the AUC was compared between both methods. Second, we constructed priority lists of associations not present in the SPC, containing the first 10% and 20% of the associations, respectively. The priority list was based on the value of the linear predictor from the old and new model, respectively. With the list, we determined the (relative) number of associations that had been previously analysed in depth (mainly triggered by case-by-case assessment), as a proxy for a possible signal, for both methods. This second method was performed to investigate whether a theoretical difference in performance, based on the AUC, also resulted in a different number of potentially interesting signals in practice.

RESULTS

Descriptive statistics and model development

A total of 151,033 ICSRs, containing 120,171 unique drug-ADR associations were extracted from the database. After the selection of associations with a minimum of 3 ICSRs, 25,026 associations remained as the training data for fitting of the model. Of these, 17,071 associations (68.2%) were present in the SPC. Additional information is presented in Table 1.

Table 1. Descriptive statistics of the ICSRs used for analysis

	Number (n)
Number of ICSRs	151,033
Number of associations	341,478
Number of unique associations (total)	120,171
Number of unique associations (n ≥ 3)	25,026
present in SPC	17,071 (68.2%)
Number of unique drugs*	1,745
Number of unique suspected ADRs#	5,726

* classified according to the Anatomical Therapeutic Chemical (ATC) classification system

coded as MedDRA Preferred Terms (PTs)

Median values with interquartile ranges (IQR) and proportions of the candidate predictors are shown in Table 2. Prior to the regression analysis, all candidate predictors were divided into 4 categories of equal sizes due to non-linearity. After the backward step-wise selection procedure, all candidate predictors were included into the final model. VIF values for the assessment of multicollinearity were below the pre-defined threshold (4) for all predictors in the model. The proportion of associations present in the SPC increased with increasing numbers of reports and an increasing proportion of HCP reports. For ROR025, Naranjo score and proportion of MAH reports the direction of the effect was less consistent. For ROR025, an increase in category was associated with increasing coefficients except for the highest category. For both Naranjo and MAH reports, the lower categories were associated with a higher coefficient (see also Table 3).

Table 2. Descriptive statistics of the candidate predictors used in the analysis

Candidate predictor	Associations listed in SPC (n=17,071)	Associations not listed in SPC (n=7,955)
Number of ICSRs per association (median / IQR)	5 (7)	4 (3)
ROR025 per association (median / IQR)	1.3 (3.1)	1.3 (5.3)
Naranjo score per association (median / IQR)	1.8 (0.7)	2.0 (0.4)
Associations from MAH reports (%)	57.4	42.6
Associations from HCP reports (%)	86.8	13.2

IQR, interquartile range; ROR025, lower limit of the two-sided 95%CI of the ROR

Table 3. Full multivariable model with model parameters and measure for multicollinearity (VIF)

Predictor	Number of Observations	Regression Coefficient	Standard Error	p-value	VIF
Intercept		-0.09	0.07	0.23	
Number of ICSRs per association (n)					1.50
3	7,859		Reference category		
4 – 5	7,008	0.34	0.04	<0.0001	
6 – 8	4,197	0.72	0.05	<0.0001	
> 8	5,962	1.22	0.05	<0.0001	
ROR025 per association					1.63
≤ 0.54	6,365		Reference category		
0.55 – 1.29	6,181	0.33	0.04	<0.0001	
1.30 – 4.18	6,224	0.52	0.04	<0.0001	
> 4.18	6,256	0.39	0.04	<0.0001	
Naranjo score per association					2.12
0 – 1.33	6,709		Reference category		
1.34 – 1.94	5,805	0.07	0.05	0.16	
1.95 – 2.00	6,784	-0.39	0.05	<0.0001	
> 2.00	5,728	-0.11	0.04	0.01	
Percentage of MAH reports					2.36
0%	10,071		Reference category		
0.1 – 20.0%	2,711	0.64	0.07	<0.0001	
20.1 – 75.0%	6,323	0.07	0.04	0.14	
> 75.0%	5,921	-0.79	0.07	<0.0001	
Percentage of HCP reports					2.47
0 – 12.5%	6,305		Reference category		
12.6 – 50.0%	7,130	0.41	0.06	<0.0001	
50.1 – 75.0%	5,654	0.48	0.06	<0.0001	
> 75.0%	5,937	0.70	0.07	<0.0001	

ROR025, lower limit of the 2-sided 95%CI of the ROR; VIF, variance inflation factor.

Performance and validation

The performance of the model was satisfactory, based on the area under the receiver operating characteristic curve (AUC = 0.740; 95%CI 0.734–0.747; see Figure 1). The 3 strongest predictors in the model were more than 8 reports per association, followed by a percentage of HCP reports of 75% or higher, and a percentage of MAH reports between 0% and 20% (see Table 3). The calibration curve of the model shows good calibration based on the observed versus predicted probabilities (see Figure 2). Additionally, bootstrap resampling showed only a marginal difference in the AUC of the model (AUC after bootstrapping = 0.739), indicating no overfitting of the model.

Fig 1. Receiver operating characteristics (ROC) curves for the various models. AUC, area under the curve; NUMBER, number of reports; ROR025, lower limit of the 2-sided 95% CI of the ROR; MAH, percentage of marketing authorization holder reports; HCP, percentage of reports by health care professionals; NARANJO, Naranjo score; FULL, full model

Fig 2. Calibration curve of the final model. The dotted line represents a perfect calibration

Comparison with current method

A comparison of the new model with the current method based on a model with only the number of reports and ROR025 as predictors showed an increased performance ($AUC_{new} = 0.740$; $AUC_{old} = 0.649$; see Figure 3). As mentioned previously, this is a theoretical comparison, because the current method used at Lareb is not based on a prediction model. Therefore, the models were not compared in terms of their AUCs, but priority lists of associations were made for both models, and these were compared in terms of possible signals. The results of these analyses show that the proportion of possible signals increased by 58.2% (from 12.3% to 19.4%) and 44.2% (from 9.6% to 13.9%) depending on the number of associations used (800 vs 1,600, respectively). Additional information is present in Table 4.

Fig 3. Receiver operating characteristics (ROC) curves for the new model (FULL) and the old model (NUMBER+ ROR025). AUC, area under the curve; NUMBER, number of reports; ROR025, lower limit of the 2-sided 95% CI of the ROR

Table 4. Performance of the new prediction model in terms of possible signals

	Number of possible signals (n/%)		Relative increase (%)
	Old model	New model	
Top-800*	98 (12.3)	155 (19.4)	58.2
Top-1,600*	154 (9.6)	222 (13.9)	44.2

* Refers to the first 800 (10%) and 1,600 (20%) associations not listed in the SPC of the linear predictor-based priority lists for each model respectively

DISCUSSION

In this study, we developed a prediction model-based screening tool aimed at improving statistical signal detection of our spontaneous ADR reports. Five relevant characteristics (number of reports, disproportionality, Naranjo score, proportion MAH reports, proportion HCP reports) were chosen as potential predictors in the model. For Naranjo, we considered to use the scoring (doubtful, possible, probable, definite) for categorization. However, this would result in a skewed distribution of observations due to the (very) low numbers of associations with outcome “definite”. Our choice to set the Naranjo score for all MAH reports to +2 was based on the assumption that the ADR occurred after the suspect drug was given, which gives a score of +2. Because no case-by-case assessment is performed for MAH reports at our centre, other items from the Naranjo algorithm could not be answered. This approach could lead to less diversity in the Naranjo scores, and as a result impact the performance of the model. However, a recent pilot study performed at our centre investigating case-by-case analysis of MAH reports showed that the documentation level of the reports was often poor [20]. Therefore, availability of information on additional Naranjo items would most likely be limited. Although important in statistical signal detection [13,21], the time to onset was not included as a predictor in our model. In a previous study, we investigated the values of the time to onset in statistical signal detection and found that it was of limited value in the investigated setting [22].

Overall, the model performed well (AUC = 0.740) and showed good calibration. The highly significant result of the Hosmer-Lemeshow test ($P < 0.001$), indicating a poor goodness of fit, is most likely explained by the large sample size and should be interpreted with caution [23]. Based on the calibration curve and the bootstrap resampling, the internal validation was considered satisfactory. The latter showed a negligible difference in AUC (0.740 versus 0.739). We found little differences in AUC values among individual predictors, although disproportionality (ROR025) seems to have the lowest predictive value. This may be explained by the fact that disproportionality is sensitive to selective reporting and other types of bias [24–26]. Within subgroups of predictors, we found some noteworthy results regarding the regression coefficients. The predictors “number of reports” and “percentage of HCP reports” showed a consistent increase in coefficients with increasing categories, which was as expected. For ROR025 and Naranjo, we anticipated similar results, which was not the case though (see Table 3). This may be explained by the fact that the outcome used in our model (presence of the association in the SPC) does not by definition imply causality because true ADRs are not necessarily present in this document, and on the other hand, events that are not true ADRs may be present. Additionally, the categories used for Naranjo in the model limit the conclusions that can be drawn from the regression coefficients. For fitting of the model, 4 more or less equally sized categories were made due to non-linearity of the parameter. As a result, the first 3 categories range from 0 to 2, and the fourth category contains all values above 2, whereas the actual Naranjo score has a range from –4 till 13.

For the percentage of MAH reports, there seems to be an inverse relationship with presence in the SPC. One striking difference between MAH reports and reports received directly from HCPs or consumers that we found in our previously described pilot study [20] was

the presence of ADRs in MAH reports that are not truly ADRs but, for example, outcomes (e.g., death, hospitalization, lack of efficacy) or events related to the indication of the drug (e.g., terminal state in cancer patients with metastases). The fact that these types of events are not likely to be present in the SPC and have a low percentage of reports directly reported to Lareb may be an explanation for the inverse relationship. Although one can debate the validity of the use of the SPC as the gold standard for causality, we considered it to be the most comprehensive and up-to-date data source with information regarding ADRs that is publicly available. A different approach could have been the method used for *vigiRank*, where a set of historical safety signals was used as a reference set of positive controls [8]. However, the issue regarding causality and presence in the SPC as mentioned earlier remains, because they used the SPC to define the negative controls. We considered it more appropriate to use the same gold standard for both the positive and negative controls and therefore decided to use the SPC for both. Additionally, the use of a reference set would most likely result in a less heterogeneous set of ADRs because it would probably not contain more common ADRs (eg, headache, dizziness, nausea, etc.), although these types of ADRs are among the most frequently present in spontaneous reports.

The linear predictor-based priority lists comparing the old and new model showed a substantial increase in potential signals among the most highly ranked drug-ADR combinations not present in the SPC. In this context, the increase in potential signals should be seen in terms of earlier detection due to prioritization and not in terms of signals that would, or would not be picked up by either method. Previous research suggests that results obtained from signal detection algorithms depend on the database the algorithm is applied to [8,19]. The same will hold for our algorithm. For example, in the Netherlands, we receive a substantial number of ICSRs reported by patients, but this is not necessarily the case in other countries. Therefore, the use of the number of HCP reports as a candidate predictor may not be a logical choice for other databases. Consequently, the development of such a model should be based on the reporting and database characteristics of the country or region it is applied to. Nevertheless, the method of generating a prediction model-based priority list of signals could be useful in other (spontaneous reporting) databases. One of the limitations of our study is the risk of bias due to selective reporting. Because the database contains well-established associations, it is reasonable to assume that these associations are reported more frequently than unknown associations, therewith influencing the predictors in the model. In an alternative approach, the values of the predictors immediately prior to the recognition of the association could be used in the model. However, recovering the date of recognition for several thousand associations may prove to be infeasible.

In conclusion, this study shows that a prediction model-based screening tool can be used to generate priority-based listings of drug-ADR associations for signal detection. Additionally, as seen in other studies [8,27], the introduction of variables other than the number of reports and disproportionality can increase screening efficiency due to priority-based assessment of drug-ADR associations.

REFERENCES

1. Practical Aspects of Signal Detection in Pharmacovigilance: Report of CIOMS Working Group VIII. Geneva: 2010.
2. Raine JM. Risk Management: A European Regulatory View. New York: Wiley; 2007.
3. Lester J, Neyarapally GA, Lipowski E, Graham CF, Hall M, Dal PG. Evaluation of FDA safety-related drug label changes in 2010. *Pharmacoepidemiol Drug Saf.* 2013;22(3):302-305.
4. van Hunsel F, Talsma A, van Puijenbroek E, de Jong-van den Berg L, van Grootheest K. The proportion of patient reports of suspected ADRs to signal detection in the Netherlands: case-control study. *Pharmacoepidemiol Drug Saf.* 2011;20(3):286-291.
5. Wisniewski AF, Bate A, Bousquet C, et al. Good signal detection practices: evidence from IMI PROTECT. *Drug Saf.* 2016;39(6):469-490.
6. Stricker BH, Tijssen JG. Serum sickness-like reactions to cefaclor. *J Clin Epidemiol.* 1992;45(10):1177-1184.
7. van Hunsel F, Ekhardt C. Experiences with a computer-assisted database screening tool at The Netherlands Pharmacovigilance Centre Lareb. *Pharmacoepidemiol Drug Saf.* 2015;24(S1):442
8. Caster O, Juhlin K, Watson S, Noren GN. Improved statistical signal detection in pharmacovigilance by combining multiple strength-of-evidence aspects in *vigiRank*. *Drug Saf.* 2014;37(8):617-628.
9. Bergvall T, Noren GN, Lindquist M. *vigiGrade*: a tool to identify well-documented individual case reports and highlight systematic data quality issues. *Drug Saf.* 2014;37(1):65-77.
10. Naranjo CA, Busto U, Sellers EM, et al. A method for estimating the probability of adverse drug reactions. *Clin Pharmacol Ther.* 1981;30(2):239-245.
11. Brown EG, Wood L, Wood S. The medical dictionary for regulatory activities (MedDRA). *Drug Saf.* 1999;20(2):109-117.
12. Structure and Principles of the ATC classification system. http://www.whocc.no/atc/structure_and_principles/, 2011 (Accessed April 3, 2013).
13. van Holle L, Zeinoun Z, Bauchau V, Verstraeten T. Using time-to-onset for detecting safety signals in spontaneous reports of adverse events following immunization: a proof of concept study. *Pharmacoepidemiol Drug Saf.* 2012;21(6):603-610.
14. Szarfman A, Machado SG, O'Neill RT. Use of screening algorithms and computer systems to efficiently signal higher-than-expected combinations of drugs and events in the US FDA's spontaneous reports database. *Drug Saf.* 2002;25(6):381-392.
15. Rolfes L, van Hunsel F, Wilkes S, van Grootheest K, Van Puijenbroek EP. Adverse drug reaction reports of patients and healthcare professionals-differences in reported information. *Pharmacoepidemiol Drug Saf.* 2015;24(2):152-158.
16. van Hunsel F, de Waal S, Harmark L. The contribution of direct patient reported ADRs to drug safety signals in the Netherlands from 2010 to 2015. *Pharmacoepidemiol Drug Saf.* 2017;26(8):977-983.
17. Fox J, Monette G. Generalized collinearity diagnostics. *J Am Stat Assoc.* 1992;87(417):178-183.
18. Neter J, Kutner MH, Nachtsheim CJ, Wasserman W. Applied linear statistical models. Irwin; 1996.
19. Candore G, Juhlin K, Manlik K, et al. Comparison of statistical signal detection methods within and across spontaneous reporting databases. *Drug Saf.* 2015;38(6):577-587.
20. Scholl JH, van Hunsel F, Harmark L. Case-by-case assessment of ICSRs from marketing authorization holders: a pilot study. *Drug Saf.* 2015;38(10):986 Ref Type: Abstract
21. Norén GN, Hopstadius J, Bate A, Star K, Edwards IR. Temporal pattern discovery in longitudinal electronic patient records. *Data Min Knowl Discov.* 2010;20(3):361-387.

22. Scholl JH, Van Puijenbroek EP. The value of time-to-onset in statistical signal detection of adverse drug reactions: a comparison with disproportionality analysis in spontaneous reports from the Netherlands. *Pharmacoepidemiol Drug Saf.* 2016;25(12):1361-1367.
23. Hosmer DW, Hosmer T, Le CS, Lemeshow S. A comparison of goodness-of-fit tests for the logistic regression model. *Stat Med.* 1997;16(9):965-980.
24. Bate A, Evans SJ. Quantitative signal detection using spontaneous ADR reporting. *Pharmacoepidemiol Drug Saf.* 2009;18(6):427-436.
25. Maignen F, Hauben M, Hung E, van Holle L, Dogne JM. Assessing the extent and impact of the masking effect of disproportionality analyses on two spontaneous reporting systems databases. *Pharmacoepidemiol Drug Saf.* 2014;23(2):195-207.
26. 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(6):743-749.
27. van Holle L, Bauchau V. Use of logistic regression to combine two causality criteria for signal detection in vaccine spontaneous report data. *Drug Saf.* 2014;37(12):1047-1057.

CHAPTER 5

5

General Discussion

STATISTICAL SIGNAL DETECTION IN PHARMACOVIGILANCE

This thesis describes statistical methods that may aid the overall signal detection process in pharmacovigilance, in particular in spontaneous reporting systems (SRS). In the past decades a lot of experience has been gained with the use of disproportionality analyses to identify possible safety signals [1-5]. Even though these methods have proven to be valuable additions to the signal detection process, they have their limitations. First, they do not take clinically relevant parameters present in the individual reports into account. Although disproportionality analysis is just the first step, followed by clinical review of the corresponding cases, this initial step leads to a suboptimal use of data available in the reports, reducing each report to a number. Second, spontaneous reporting is sensitive to several types of bias [6-12], likewise leading to the possibility of biased results of disproportionality analysis. In particular, relevant associations may remain initially undetected because there are not enough reports to exceed the statistical threshold, though they may be of interest for clinical or pharmacological reasons. For the purpose of this thesis it should be noted that the goal of disproportionality analysis is to get an impression of the disproportional presence of drug – ADR combinations and it should not be regarded as a formal epidemiological study.

This thesis describes the research on additional methodology that may increase the arsenal of already available tools in signal detection. Prior to the actual methodology research, chapter 2 presents a previously unpublished type of bias and assessed its effects on disproportionality analysis based signal detection. Chapter 3 contains three articles investigating the value of the time to onset (TTO) of an adverse drug reaction (ADR) as a clinical parameter in signal detection methodology. The research in chapter 4 describes the use of a multivariable regression model as a tool for signal prioritization.

In this final chapter, the main findings from our studies, and lessons learned from them, are discussed and placed in broader perspective. Subsequently, a future perspective regarding the developments in statistical signal detection will be given.

MAIN FINDINGS AND LESSONS LEARNED

Precautionary reporting bias and its effect on database quality

Knowledge of factors influencing the degree of disproportionality is necessary to interpret disproportionality analysis outcomes correctly. Over the past years, many of these factors, such as notoriety bias (increased reporting due to e.g., media attention) and the Weber effect (increased reporting in the first 2 years after marketing authorization of a drug) have been described. However, the knowledge about the influence of including reports from specific data collection initiatives on disproportionality analysis was limited.

Chapter 2 shows that the bias resulting from these initiatives (precautionary reporting bias) can disrupt the results of DPA substantially. This type of bias is due to the selective reporting of certain drug – ADR associations. It is well known that Marketing Authorization Holders develop and implement so-called patient support programs. In these programs, a drug is administered to the patient in his own home by a third party, and because the Marketing Authorization Holders are responsible for the programs, they are also responsible for ADR reporting. Pharmacovigilance legislation dictates that any untoward medical event that occurs after the administration of a drug should be reported when a causal relationship with the administered drug cannot be ruled out. Marketing Authorization Holders take this legislation very seriously and often act on the saying: “It’s better to report too much than too little”. As a result, there is a selective (over)reporting of certain suspected ADRs for specific drugs. The consequences of this, in terms of disproportionality analysis, are twofold: first, statistically significant associations that are detected turn out to be irrelevant in terms of signal detection. Second, clinically relevant associations may not be detected. This leads to the question to what extent the fact that signal detection is influenced by these patient support programs, benefits the pharmacovigilance process. An example that is described in the study are reports of deaths occurring in patients in a patient support programs using a bisphosphonate in the oncological setting. These are often terminally ill patients who receive a bisphosphonate intravenously every 3 – 4 weeks for symptomatic bone metastases. When the nurse administering the bisphosphonate calls to make a new appointment and learns that the patient has died, he/she will report this death as an ADR to the Marketing Authorization Holder as required by pharmacovigilance legislation. However, a terminally ill patient dying from the underlying disease is more likely than dying from a bisphosphonate infusion. This procedure led to a large number of reports of death for the concerning bisphosphonate in the database, which does not contribute to a better surveillance. After all, as the study shows, other possibly relevant signal were masked because they could not exceed the statistical threshold for detection.

This leads to the question whether or not the Marketing Authorization Holder should be obligated to report these events as suspected ADRs and that’s where the problem is. From a legal point of view, each event should be reported if a causal relationship with the drug cannot be ruled out. However, in practice this results in a substantial number of poorly usable reports as shown in the study. There are a few solutions for this problem: i) Reports from patient support programs should not be included into the SRSs anymore. This does not seem to be the most desirable solution because it requires a modification of the current legislation and possible safety signals could be missed. Furthermore even though,

in our experience, the majority of patient support program reports have limited value, it cannot be ruled out that they contain reports with signal potential. ii) Reports from patient support programs should only be included in SRSs if the causal relationship with the drug is more likely than “not related”. Current legislation dictates that a medical event is to be reported if a causal relationship cannot be ruled out. In daily practice however, causality assessment is often complex. Because in most reports not all information necessary for a proper causality assessment is present, a causal relationship can rarely be ruled out with certainty. Another issue is the fact that the causality assessment would have to be performed by the Marketing Authorization Holders because they are responsible for the patient support programs and the registration of the ADRs occurring in these programs. The question arises if they are the most appropriate stakeholder to make well-considered causality decisions, taking commercial interests into account. After all, the past has shown cases where Marketing Authorization Holders failed to report crucial safety information in time, thereby jeopardizing the public health (Vioxx / Avandia cases) [13, 14]. iii) A final solution would be to make reports from patient support programs more easily identifiable in SRSs, making it easy to exclude them from statistical analyses. The main advantage would be that they would no longer disrupt the statistical signal detection, but would still be available for individual case review, when necessary. This solution would require an update of the current ICH-E2B(R3) format [15], as well as adjustments to all systems based on this format. Of these options, the latter seems to be the most adequate because it combines the completeness of data available with the option to selectively ignore reports during signal detection.

Time to onset as a clinical parameter in signal detection

Chapter 3 elaborates on the possible role of the TTO as a clinical parameter in statistical signal detection. It is well known that certain ADRs have a typical time to onset, including vigabatrin-induced visual field defects [16] and hypersensitivity reactions [17]. This aspect can be used in statistical signal detection.

Chapter 3.1 described a cohort study showing that the TTO is a variable that can be modelled using well known parametric distributions, and could be a relevant parameter in causality assessment of ADRs. Although the study did not focus on statistical signal detection, the results suggested that the TTO could play a role there. The hypothesis here is that when the TTOs of all reports concerning a particular drug – ADR association follow a certain (specific) statistical distribution, this distribution can be tested against the distribution of TTOs of other drug – ADR combinations. This could be an additional method for signal detection although it could be less useful to detect differences between similar drugs and / or ADRs.

In chapter 3.2 a pilot study investigating the TTO in statistical signal detection is described. The general idea behind the study is the hypothesis that the TTO of most associations follow a more or less uniform distribution when the suspected ADR is not a true ADR (comparable with background noise). In the event of a true ADR however, the TTO will cluster around a certain value characteristic for that drug – ADR combination (and the TTO distribution is thus no longer uniform). Subsequently the TTO distribution of each drug – ADR association was statistically tested against the TTO distribution of other drug – ADR

associations, where a significant result may indicate a true ADR. This principle resembles disproportionality analysis with the exception that disproportionality analysis is based on the (relative) number of reports whereas the TTO analysis is based on the statistical distribution of the TTOs. Our pilot study was aimed at the confirmation of previously published research by van Holle et al. [18-20] but also contained an extension. The studies by van Holle were limited to the value of TTO for vaccine signal detection only, and since vaccines concern a minority of reports in the Lareb database, two other type of drug were included in our study: a drug for short-term use and one for chronic use. This led to the inclusion of a more heterogeneous group of patients in the study. Furthermore, the Anderson-Darling test was used instead of the Kolmogorov-Smirnov test used by van Holle, because the former has more statistical power and is more sensitive to differences in the tails of the distributions [21, 22]. The results from the TTO analyses were compared with disproportionality analysis showing that the TTO approach was of limited additional value for these case drug associations. More specifically, the sensitivity was low, due to a low number of true positives. When reports were limited to those with a TTO of 0-30 days, specificity improved slightly, but sensitivity did not. The most important lesson learned from this study was that the TTO was not as robust as previously assumed. This became evident from the fact that: i) the TTO is not always reported, since this is not mandatory. ii) there seemed to be a clustering of TTOs around certain 'logical' values. For example, ADRs with an actual TTO of 6 days and 8 days respectively, may both be reported having a TTO of 1 week. This results in a decreased statistical precision. iii) There was an overrepresentation of ADRs with a short TTO. This may be partially due to recall bias where a reporter is less likely to recognize an event as being an ADR when the TTO increases.

This final finding led to the study described in chapter 3.3, which was aimed at signal detection using the same TTO method but with ADRs having a TTO of at least 7 days. Additionally, we investigated whether the number of reports for the concerning association would influence the performance of the method. Results were again compared with those from disproportionality analysis. Similar to the previous study, the current study showed no additional value for the TTO method compared to disproportionality analysis, again due to a low sensitivity for the TTO method. Also, the number of reports as a variable for performance showed little improvement of the performance.

In summary, the three studies described above show that although the TTO seems to be a promising candidate for signal detection on theoretical grounds, the method did not live up to its expectation in practice. One possible explanation for this could be that other variables that may affect the TTO were not taken into account. For instance, it is plausible that concomitant medication or the presence of comorbidities may influence the TTO of an ADR. Using certain types of concomitant medication may have an effect on the pharmacokinetic profile of the suspect drug resulting in a different TTO, thereby decreasing the discriminative power of the statistical test. Similarly, an impaired liver function may also influence pharmacokinetics. A suggestion for future research is to perform a pilot study on data from registration trials. These should contain time to onset data but information regarding the TTO is poorly documented in most product information documents such as the Summary of Product Characteristics (SPC). The above leads to the question whether the development of signal detection methods based on just one variable is the most ideal

approach, or that a model with more variables is desirable. This thinking has led to the research described in chapter 4.

Multivariable models

In the past, Lareb used disproportionality analysis as signal detection method besides case-by-case analysis. For disproportionality analysis, the Reporting Odds Ratio (ROR) was used in combination with a minimum number of reports per association. We investigated if using a model with more (clinical) variables would lead to a better performance than disproportionality analysis alone.

Chapter 4 describes the development and application of this model on Lareb's SRS. The model contained both quantitative variables (number of reports, disproportionality) and qualitative (clinically relevant) variables, including the type of reporter (consumer or healthcare professional) and the Naranjo score [23] as a proxy for causality. The resulting model showed a good performance and was able to generate a priority list of potential signals. This prioritization step can increase signal screening efficiency compared to screening with disproportionality analysis only. Currently, the model is used in routine pharmacovigilance activities at Lareb [24]. During model development we deliberately selected a limited number of variables as potential candidates, despite the fact that the ICH-E2B(R3) format, which is the basis of our SRS, has plenty of variables to choose from. The major reason for this is because there are very few variables in the format that are mandatory. Pharmacovigilance legislation dictates that pharmacovigilance centers (and Marketing Authorization Holders) are obligated to accept each report that has the four following items present: i) an identifiable patient, ii) an identifiable reporter, iii) an identifiable suspect drug, iv) an identifiable event [25]. Because of this, the information density differs considerably among reports of ADRs [26, 27] and in order to develop a model with a good performance, it is necessary to have sufficient data points per variable. If this is not the case, imputation could be a solution if the number of missing values is limited (and if they are random), but imputation becomes unreliable when the proportion of missing values becomes too high. Although no strict rules for the cut-off value of this proportion exist, approximately 5% is generally accepted [28]. However, for a lot of variables the proportion of missing values is much higher than this threshold. Another issue to take into account is that not all variables are relevant for all types of associations. For instance, the presence / absence of a positive dechallenge could be a promising variable for the model, but is not relevant for several drugs (e.g., vaccines) because you cannot perform a dechallenge for a single administration drug. As a result, you would introduce a parameter into your model that is not clinically applicable for a substantial proportion of your drugs. Moreover, since the goal of our research was to develop a full database screening tool, we decided to limit the variables to those that apply to all drugs / ADRs.

SUGGESTIONS FOR FUTURE RESEARCH IN SPONTANEOUS REPORTING

The development of statistical methods for signal detection of spontaneously reported ADRs has taken a great leap in the past decades, both out of necessity (because of increasing numbers of reports) and due to the increased availability of suitable techniques

and related software. Today, more than ever, the trigger for signal detection seems as much a statistical as well as a clinical/pharmacological discipline and the question is what innovations may occur in the coming years.

Our research on a prediction model to aid with signal prioritization showed that a multivariable model can be a useful tool in pharmacovigilance of spontaneous data. Conceptually, one can easily imagine to broaden this idea to be integrated into more complex models. Such models can include parameters that are relevant in causality assessment, such as positive de-/re-challenges, interacting concomitant medication and relevant comorbidities. As mentioned previously, one should bear in mind that these parameters should actually be present in substantial proportions in reports.

Besides the additions of more parameters to the model, there are other ways to investigate parameters and their interrelations. One way to realize this could be neural network analyses. Such network analyses can reveal complex relationships between relevant parameters (drugs, ADRs) within a system (the SRS). Analyses of these type of networks can help to visualize and interpret certain patterns that may be hidden in voluminous data. One of the advantages of such an approach is that no additional variables are necessary and that they are always present in each report because they are mandatory to be reported (suspect drug and ADR). Ball et al. showed that a neural network of reports containing suspect vaccines and ADRs was able to visualize clinically relevant patterns (intussusception with Rotavirus vaccination; syncope in mainly adolescents being vaccinated) in a large dataset [29]. The only variables needed were the suspect vaccine and reported ADRs (and age for stratification). Another way to detect signals using a limited set of variables is to use external sources of information. DrugBank for example, is a database containing pharmacological information on drugs (receptor affinity, PK/PD data) [30] that can be integrated into the SRS. This allows for the prediction of ADR profiles of drugs based on their pharmacology. Finally, Natural Language Processing (NLP) could be a valuable addition to signal detection methodology in spontaneous reporting. NLP is a type of artificial intelligence (AI) that allows for computers to interpret (among others) written text. Although the current reporting format used in spontaneous reporting contains a lot of coded information, there is still a substantial amount of uncoded (free) text that is only analysed manually. NLP could be helpful here by interpreting these text and (using text mining) analyse it in a quantitative manner. A feasibility study with electronic healthcare records showed that this method has potential [31], but to the best of our knowledge, no studies for SRSs have been published yet.

FINAL REMARKS

The studies described in this thesis show that statistical methods for signal detection in spontaneous reporting can be of value, but results vary. It turned out that the time to onset was not sensitive enough to detect enough signals, even though previous research by others showed substantially better results. On the other hand, a multivariable prediction model had an acceptable predictive performance and is currently used as a screening tool and to prioritize signals in the signal detection process at Lareb. How statistical signal detection will develop in the coming years remains to be seen, but regardless of the applied methodology, results will have to be supported by proper case analyses where a clinical judgment will have to be made in terms of causality.

REFERENCES

1. Stricker BH, Tijssen JG. Serum sickness-like reactions to cefaclor. *J Clin Epidemiol*. 1992;45(10):1177-84.
2. van Puijenbroek EP, Bate A, Leufkens HG, Lindquist M, Orre R, Egberts AC. 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.
3. 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;54(4):315-21.
4. Noren 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. *Stat Med*. 2006;25(21):3740-57.
5. 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.
6. Hazell L, Shakir SA. Under-reporting of adverse drug reactions : a systematic review. *Drug Saf*. 2006;29(5):385-96.
7. Pariente A, Gregoire F, Fourier-Reglat A, Haramburu F, Moore N. Impact of safety alerts on measures of disproportionality in spontaneous reporting databases: the notoriety bias. *Drug Saf*. 2007;30(10):891-8.
8. de Boissieu P, Kanagaratnam L, Abou Taam M, Roux MP, Drame M, Trenque T. Notoriety bias in a database of spontaneous reports: the example of osteonecrosis of the jaw under bisphosphonate therapy in the French national pharmacovigilance database. *Pharmacoepidemiol Drug Saf*. 2014;23(9):989-92.
9. 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(6):743-9.
10. Moulis G, Sommet A, Durrieu G, Bagheri H, Lapeyre-Mestre M, Montastruc JL, et al. Trends of reporting of 'serious' vs. 'non-serious' adverse drug reactions over time: a study in the French Pharmacovigilance Database. *Br J Clin Pharmacol*. 2012;74(1):201-4.
11. Gerhard T. Bias: considerations for research practice. *Am J Health Syst Pharm*. 2008;65(22):2159-68.
12. Salas M, Hofman A, Stricker BH. Confounding by indication: an example of variation in the use of epidemiologic terminology. *Am J Epidemiol*. 1999;149(11):981-3.
13. Krumholz HM, Ross JS, Presler AH, Egilman DS. What have we learnt from Vioxx? *BMJ*. 2007;334(7585):120-3.
14. Nissen SE. The rise and fall of rosiglitazone. *Eur Heart J*. 2010;31(7):773-6.
15. Agency EM. ICH guideline E2B (R3) on electronic transmission of individual case safety reports (ICSRs) - data elements and message specification - implementation guide [Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/international-conference-harmonisation-technical-requirements-registration-pharmaceuticals-human-use_en-4.pdf].
16. Wild JM, Martinez C, Reinshagen G, Harding GF. Characteristics of a unique visual field defect attributed to vigabatrin. *Epilepsia*. 1999;40(12):1784-94.
17. Demoly P, Adkinson NF, Brockow K, Castells M, Chiriac AM, Greenberger PA, et al. International Consensus on drug allergy. *Allergy*. 2014;69(4):420-37.
18. van Holle L, Bauchau V. Signal detection on spontaneous reports of adverse events following immunisation: a comparison of the performance of a disproportionality-based algorithm and a time-to-onset-based algorithm. *Pharmacoepidemiol Drug Saf*. 2014;23(2):178-85.

19. Van Holle L, Tavares Da Silva F, Bauchau V. Signal detection based on time-to-onset: extending a new method from spontaneous reports to observational studies. *Pharmacoepidemiol Drug Saf.* 2014;23(8):849-58.
20. Van Holle L, Zeinoun Z, Bauchau V, Verstraeten T. Using time-to-onset for detecting safety signals in spontaneous reports of adverse events following immunization: a proof of concept study. *Pharmacoepidemiol Drug Saf.* 2012;21(6):603-10.
21. Engmann S, Cousineau C. Comparing distributions: the two-sample Anderson–Darling test as an alternative to the Kolmogorov–Smirnov test. *JAQM.* 2011;6(3):1-17.
22. Razali NM, Yap BW. Power comparisons of Shapiro–Wilk, Kolmogorov–Smirnov, Lilliefors and Anderson–Darling tests. *J Stat Model Anal.* 2011;2(1):21-33.
23. 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(2):239-45.
24. van Hunsel F, Scholl J, van Puijenbroek E. First year results of using a new prediction model-based algorithm for computer-assisted database screening of adverse drug reactions in the Netherlands. *Pharmacoepidem & Drug Safety.* 2018;27(S2):258-9.
25. Agency EM. Guideline on good pharmacovigilance practices (GVP). Module VI – Collection, management and submission of reports of suspected adverse reactions to medicinal products (Rev 2) 2017 [Available from: https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/guideline-good-pharmacovigilance-practices-gvp-module-vi-collection-management-submission-reports_en.pdf].
26. Rolfes L, van Hunsel F, van der Linden L, Taxis K, van Puijenbroek E. The Quality of Clinical Information in Adverse Drug Reaction Reports by Patients and Healthcare Professionals: A Retrospective Comparative Analysis. *Drug Saf.* 2017;40(7):607-14.
27. Rolfes L, van Hunsel F, Wilkes S, van Grootheest K, van Puijenbroek E. Adverse drug reaction reports of patients and healthcare professionals-differences in reported information. *Pharmacoepidemiol Drug Saf.* 2015;24(2):152-8.
28. Schafer JL. Multiple imputation: a primer. *Stat Methods Med Res.* 1999;8(1):3-15.
29. Ball R, Botsis T. Can network analysis improve pattern recognition among adverse events following immunization reported to VAERS? *Clin Pharmacol Ther.* 2011;90(2):271-8.
30. Drugbank. [Available from: <https://go.drugbank.com/>].
31. Wang X, Hripcsak G, Markatou M, Friedman C. Active computerized pharmacovigilance using natural language processing, statistics, and electronic health records: a feasibility study. *J Am Med Inform Assoc.* 2009;16(3):328-37.

Summary

Drug safety, also called pharmacovigilance, is defined by the World Health Organisation (WHO) as *the science and activities relating to the detection, assessment, understanding and prevention of adverse effects or any other medicine/vaccine related problem*. One of the major objectives of pharmacovigilance is the timely detection of drug related safety hazards. This process starts during the developmental stages of a drug, even before it is approved for market authorization and before it can be prescribed. It is however essential to keep monitoring a drug's safety profile after approval because new insights may arise on its adverse drug reactions (ADRs). This is because the studies that were the basis for approval of the drug are performed with a limited study population. This often concerns both a limited number of participants as well as the exclusion of participants based on medical criteria. Examples are the exclusion of elderly participants, children, pregnant women or people with serious diseases, such as cancer.

As soon as a drug has obtained marketing authorization, its safety profile is monitored by, among others, national pharmacovigilance centres like Lareb. At these centres, a 'spontaneous reporting system' is used to collect and analyse reports of suspected ADRs reported by both healthcare professionals and consumers. All reported ADRs are entered into the database using a predefined format, which allows for the analysis for possible safety signals. Here, a safety signal can be a new ADR or a hitherto unknown aspect of a known ADR. This process of analysing data for signals is called signal detection and it can be performed in several ways. Traditionally, all received reports were manually assessed by assessors with a medical / pharmacological background. They assess each report to see if it may constitute a safety signal and if dissemination of this knowledge is required. Over the course of time, statistical methods have been developed to identify signals. This is referred to as statistical signal detection. These methods are based on the hypothesis that certain ADRs may occur more often in the database than is to be expected based on probability (the number of reports of the ADR for a particular drug is compared with the same ADR for all other drugs). If this is the case, this may constitute a safety signal, which may be a sign that it is a true ADR and that additional measures or research is warranted.

The objective of the studies described in this thesis is the development of additional statistical methods that may improve signal detection. This is because the currently used methods, the so-called disproportionality analyses have their limitations and only take the number of reports received into account. They only determine if a certain ADR is present more often than statistically expected. All other information in the report that may be relevant for the detection of a signal, such as pre-existing conditions or the use of concomitant medication, is not included. So, a well-documented report containing a lot of (clinical) information contributes as much as a less well-documented one. The development of novel methods including other information could contribute to the signal detection process and, as a result, to a safer use of drugs in daily practice.

Chapter 1 starts with an introduction on pharmacovigilance as a science and the role of spontaneous reporting systems therein. Subsequently the concept of signal detection is discussed, and in more detail, statistical signal detection. Although several new signal detection methods have been investigated, most of them did not pass the research stage and have not been implemented in daily pharmacovigilance practice. Finally, the concept of bias is introduced.

This thesis describes three main subjects playing a role in the improvement of statistical signal detection: i) the role of bias, ii) the additional value of the time to onset, iii) the additional value of a multivariable statistical model.

In **chapter 2** a new type of bias is described, called precautionary reporting bias. Pharmaceutical companies regularly implement *patient support programs*, where nurses are hired to administer medication to patients at home, where they would normally be required to go to the hospital. For example, this refers to patients with cancer who are very ill and for whom going to the hospital for a few hours is a heavy burden. These programs require the pharmaceutical company to report each medical event that the nurse hears of, as a possible ADR. Subsequently, the company is required to report these events to Lareb via the EU EudraVigilance database. The problem that arises is that a substantial number of events is reported, which are most likely not actual ADRs but a result of e.g., the underlying disease. Terminally ill cancer patients for example, will eventually die because of their disease, but when a nurse in a patient support program learns of a patient's death, she has to report this as a possible ADR. Our study shows that these type of reports interfere with statistical signal detection, leading to a masking effect of certain signals due to the large number of reports from patient support programs. As a result, true signals may be detected later or not at all because they do not exceed the statistical threshold

Chapter 3 discusses the possible role of the time to onset (TTO) within signal detection. The TTO is defined as the time window from the start of a suspect drug till the start of the event. This time window varies between ADRs and differences can be tested statistically, allowing the parameter to be used as a signal detection method. Prior to the actual research investigating the added value of the TTO in signal detection, **chapter 3.1** describes the effort to model the TTO using known statistical distributions. In other words, the study explored whether the TTO behaved as was to be expected based on well-known models. The results show that this was indeed the case, indicating that TTO might be a good parameter for signal detection. **Chapter 3.2** mentions a proof-of-concept study exploring the actual added value of the TTO in statistical signal detection. To investigate this, the TTO method was compared to disproportionality analysis for 3 drugs. These drugs were selected based on their use in different populations and concerned: a vaccine (young and relatively healthy population), an antibiotic (mixed population in terms of age and health status) and a statin (older population with often multiple comorbidities). In this study the TTO method performed worse than disproportionality analysis, which was mainly due to a low sensitivity of the method (true signals were poorly detected). One of the possible causes for this result, was that the majority of TTO values reported were around 1-2 days, while it is important for statistical methods to have a variation in values. This finding led to the follow-up study described in **chapter 3.3**. Here, the previously used

method was investigated using only ADRs with a TTO of at least 7 days. Additionally, the role of the number of reports was taken into account. The idea behind this is the fact that a statistically significant result is reached more easily with increasing numbers of reports. Finally, all drugs in the database were used instead of just 3 in the previous study. Despite these changes to the study, the results were similar to the proof of concept study, showing a sensitivity for the TTO method that was too low to have additional value over disproportionality analysis.

Chapter 4 explores if a multivariable model can be helpful in signal prioritization. In such a model, multiple parameters (hence multivariable) are used to predict a certain outcome. The model described here contained the following 5 parameters to predict whether a suspected ADR was actually an ADR: 1) the number of reports for the ADR; 2) the magnitude of the disproportionality (how many times more often than expected was it reported); 3) the type of reporter (healthcare professional or consumer); 4) the way of reporting (directly to Lareb or via the manufacturer); 5) the strength of the causal relationship (for each report of an ADR the likelihood of causality was assessed). The developed method was compared with the hitherto used model at Lareb containing only parameter 1) and 2). It was concluded that the new model showed a better performance over the old one and it could be useful for signal prioritization. This allows for the suspected ADRs that are most likely to be a signal to be analysed first, resulting in increased efficacy and efficiency of the signal detection process.

Chapter 5 contains a general discussion and focuses on the lessons that can be learned from the study results and the implications thereof for future research. The most important finding was that statistical methods can provide added value within the signal detection process, although results vary by method. The TTO as a parameter for signal detection performed less than expected based on previous research. This was mainly due to the low sensitivity of the method. The multivariable model, on the other hand, showed promising results and is currently used in routine signal detection at Lareb.

Future research could focus both on models containing more and / or different parameters than described in this thesis, and explore different types of models such as neural networks. Another possibility is the implementation of external sources, such as databases containing extensive information on drugs to signal detection methodology. Finally, natural language processing (NLP, a method to interpret text automatically) could be a useful addition. The current format used to register ADRs contains substantial amounts of free text (text that is not coded) and this is difficult to analyse using statistical methods. Using NLP, free text could be interpreted allowing it to be analysed statistically.

In summary, this thesis shows that statistical methods can have added value in the signal detection process within pharmacovigilance.

Samenvatting

Geneesmiddelenbewaking, ook wel farmacovigilantie genoemd, wordt door de Wereldgezondheidsorganisatie (WHO) gedefinieerd als *'de wetenschap en activiteiten met betrekking tot de opsporing, beoordeling, kennis en preventie van mogelijke bijwerkingen of andere geneesmiddel-gerelateerde problemen'*. Eén van de belangrijkste doelen van farmacovigilantie is het tijdig signaleren van geneesmiddel-gerelateerde veiligheidsrisico's. Dit proces start al bij de ontwikkeling van een geneesmiddel, nog voordat het goedgekeurd is voor toelating op de markt en kan worden voorgeschreven. Het is echter van wezenlijk belang om na de goedkeuring de veiligheid te blijven bewaken omdat er nieuwe inzichten kunnen ontstaan over de bijwerkingen van een geneesmiddel. Dit komt onder andere doordat het geneesmiddel in de studies die hebben geleid tot goedkeuring, in een beperkte populatie onderzocht is. Zo betreft het vaak een beperkt aantal deelnemers en worden diverse mensen uitgesloten van deelname op basis van medische criteria. Voorbeelden hiervan zijn oudere mensen, kinderen, zwangere vrouwen of mensen met een ernstige ziekte zoals kanker.

Zodra een geneesmiddel eenmaal op de markt is wordt de veiligheid bewaakt door o.a. nationale bijwerkingencentra zoals Lareb. Hiervoor wordt een 'spontaan rapportage systeem' gebruikt waarin meldingen van vermoede bijwerkingen worden opgenomen die gemeld zijn door zorgverleners en / of consumenten. Alle gemelde bijwerkingen worden volgens een vast format vastgelegd in een database en in deze database wordt gezocht naar mogelijke signalen, waarbij een signaal een nieuw gevonden bijwerking kan zijn of een tot dan toe onbekend aspect van een bekende bijwerking. Dit proces van signalen zoeken wordt signaaldetectie genoemd en kan op verschillende manieren gedaan worden. Van oudsher worden alle meldingen die worden ontvangen handmatig beoordeeld door beoordelaars met een klinische / farmacologische achtergrond. Zij kijken bij iedere beoordeelde melding of er mogelijk sprake is van een signaal en of er aanleiding is om kennis hierover te delen. In de loop van de tijd zijn er echter ook statistische methoden ontwikkeld om signalen in de database op te sporen en dit wordt statistische signaaldetectie genoemd. De gedachte achter deze methoden is dat bepaalde bijwerkingen vaker voor kunnen komen in de database dan je op basis van toeval zou verwachten (hierbij wordt het aantal meldingen van de bijwerking bij een bepaald geneesmiddel vergeleken met dezelfde bijwerking voor alle andere geneesmiddelen). Wanneer dit het geval is, is er mogelijk sprake van een veiligheidssignaal. Zo'n signaal is een teken dat er mogelijk daadwerkelijk sprake is van een bijwerking en dat aanvullende maatregelen of aanvullend onderzoek nodig is.

Het doel van de onderzoeken in dit proefschrift is het ontwikkelen van aanvullende statistische methoden die de signaaldetectie kunnen verbeteren. De tot nu toe gebruikte methoden, de zogenaamde disproportionaliteitsanalyses, hebben namelijk hun beperkingen en kijken alleen naar de aantallen meldingen die ontvangen zijn. Er wordt hierbij alleen gekeken of een bepaalde bijwerking vaker voorkomt dan je statistisch gezien zou verwachten. Alle inhoudelijke informatie die in de melding staat en die relevant

kan zijn voor het detecteren van een signaal, zoals reeds bestaande ziektebeelden of het gebruik van andere medicatie, wordt hierbij echter niet meegenomen. Een goed gedocumenteerde melding met veel informatie telt dus even zwaar mee als een minder goed gedocumenteerde. Het ontwikkelen van nieuwe methoden waarbij ook naar andere informatie gekeken wordt, zou een bijdrage kunnen leveren aan de signaaldetectie en daarmee aan een veiliger gebruik van geneesmiddelen in de dagelijkse praktijk.

Hoofdstuk 1 geeft een inleiding over farmacovigilantie als wetenschap en de rol van spontane rapportagesystemen hierin. Vervolgens wordt besproken wat we precies onder signaaldetectie verstaan en kijken we meer specifiek naar de statische signaaldetectie. Hoewel er in de laatste jaren veel nieuwe signaaldetectiemethoden onderzocht zijn, zijn de meesten niet verder gekomen dan het onderzoekstadium en worden de meesten niet routinematig in de dagelijkse praktijk toegepast. Als laatste komt het concept '*bias*' (vertekening) aan de orde.

Dit proefschrift beschrijft drie hoofdthema's die een rol spelen bij de verbetering van de statistische signaaldetectie, namelijk i) de rol van bias, ii) de toegevoegde waarde van de latentietijd, iii) de toegevoegde waarde van een multivariabel statistisch model.

In **hoofdstuk 2** wordt een nieuwe vorm van bias beschreven, namelijk precautionary reporting bias. Farmaceutische bedrijven maken regelmatig gebruik van zogenaamde *patient support programs* waarbij verpleegkundigen ingehuurd worden om geneesmiddelen thuis bij de patiënt toe te dienen waarvoor hij/zij anders naar het ziekenhuis zou moeten gaan. Dit gebeurt o.a. bij patiënten met kanker die vaak erg ziek zijn en voor wie een paar uur naar een ziekenhuis voor een dagbehandeling een zware belasting kan zijn. Deze programma's brengen voor de fabrikant de verplichting met zich mee om ieder medisch voorval dat de verpleegkundige ter ore komt te melden als mogelijke bijwerking. De fabrikant is vervolgens verplicht om deze voorvallen te melden bij Lareb via de Europese EudraVigilance database. Het probleem hierbij is dat er heel veel meldingen gedaan worden die hoogstwaarschijnlijk geen bijwerking zijn maar een gevolg van b.v. het onderliggende ziektebeeld. Zo zullen terminale kankerpatiënten uiteindelijk overlijden door hun ziekte, maar wanneer een verpleegkundige hoort dat de patiënt overleden is, moet dit gemeld worden als mogelijke bijwerking. In ons onderzoek tonen we aan dat dit soort meldingen een verstoring heeft op de statistische signaaldetectie en dat bepaalde signalen onderdrukt worden door de grote aantallen meldingen uit patient support programs. Het gevolg hiervan kan zijn dat echte signalen niet of later ontdekt worden doordat ze (in eerste instantie) de statistische drempel niet overschrijden.

Hoofdstuk 3 beschrijft de mogelijke rol van de latentietijd (time to onset) binnen de signaaldetectie. De latentietijd is de tijdsduur vanaf de start van een verdacht geneesmiddel tot het moment dat de klachten optreden. Deze tijdsduur is anders voor verschillende bijwerkingen en dit verschil kan statistisch getoetst worden. Op die manier kan het als signaaldetectiemethode gebruikt worden. Voordat daadwerkelijk onderzocht is of er sprake is van een toegevoegde waarde binnen de signaaldetectie wordt in **hoofdstuk 3.1** eerst gekeken of de latentietijd zich goed laat modelleren middels bestaande statistische

verdelingen. Concreet betekent dit dat onderzocht werd of de latentietijd zich gedraagt zoals we zouden verwachten op basis van bekende modellen. Het onderzoek laat zien dat dit inderdaad het geval is waardoor het aannemelijk is dat het een goede parameter zou kunnen zijn voor de signaaldetectie. In **hoofdstuk 3.2** wordt een eerste studie beschreven waarbij gekeken wordt of de latentietijd daadwerkelijk een toegevoegde waarde heeft binnen de signaaldetectie. Om dit te onderzoeken werd voor drie geneesmiddelen de latentietijd methode vergeleken met disproportionaliteitsanalyse. Er werd gekozen voor geneesmiddelen die door verschillende patiëntengroepen gebruikt worden: een vaccin (jonge en relatief gezonde meisjes), een antibioticum (gemengde populatie zowel qua leeftijd als gezondheid), en een statine (wat oudere populatie met vaak meerdere aandoeningen). De resultaten laten zien dat de latentietijd slechter scoort dan de disproportionaliteitsanalyse en dat komt vooral omdat de gevoeligheid laag was (echte signalen werden niet goed gedetecteerd). Eén van de mogelijke oorzaken hiervoor was dat er vooral veel latentietijden van 1-2 dagen gemeld werden, terwijl het voor een statistische methode van belang is dat er variatie is in de waarden. Dit gegeven leidde tot een vervolgonderzoek dat beschreven wordt in **hoofdstuk 3.3**. Hier wordt onderzocht of de eerder gebruikte methode beter zou werken wanneer alleen gekeken werd naar bijwerkingen met een latentietijd van minimaal 7 dagen. Verder werd nog gekeken of het aantal meldingen van een bepaalde bijwerking een rol speelt. De gedachte hierachter is dat je met meer meldingen makkelijker een statistisch resultaat behaalt dan met minder meldingen. Verder werden nu alle geneesmiddelen in de database gebruikt en werd het niet beperkt tot slechts 3. Ondanks deze aanpassingen waren de resultaten vergelijkbaar met het eerder uitgevoerde onderzoek en bleek de gevoeligheid van de latentietijd methoden te laag om een toegevoegde waarde te hebben ten opzichte van disproportionaliteitsanalyse.

In het onderzoek beschreven in **hoofdstuk 4** werd onderzocht of een multivariabel model zou kunnen helpen om signalen te prioriteren. In zo'n model worden meerdere parameters (multivariabel) gebruikt om een bepaalde uitkomst te voorspellen. In het beschreven model werden gekeken of de volgende 5 parameters konden voorspellen of een vermoede bijwerking ook daadwerkelijk een bijwerking was: 1) het aantal meldingen voor die bijwerking; 2) de mate van disproportionaliteit (hoeveel vaker kwam het voor dan je zou verwachten); 3) het type melder (zorgverlener of consument); 4) de wijze van ontvangst van de melding (direct aan Lareb of via de fabrikant); 5) de sterkte van het causale verband (voor iedere melding van een bepaalde bijwerking werd bepaald hoe waarschijnlijk een causaal verband is). Het ontwikkelde model werd vergeleken met het tot dan toe gebruikte model bij Lareb waarin slechts parameter 1) en 2) zaten. De conclusie was dat het nieuwe model beter presteerde dan het oude model en dat het gebruikt kan worden om signalen te prioriteren. Op deze manier kunnen vermoede bijwerkingen die de hoogste kans hebben om ook daadwerkelijk een signaal te zijn als eerste geanalyseerd worden wat de effectiviteit en efficiëntie van de signaaldetectie verhoogt.

Hoofdstuk 5 bevat een algemene discussie over het proefschrift en richt zich met name op wat geleerd kan worden uit de resultaten van de beschreven onderzoeken en de implicaties daarvan voor toekomstig onderzoek. De belangrijkste bevindingen zijn dat statistische methoden toegevoegde waarde kunnen hebben binnen de signaaldetectie,

maar dat de resultaten wisselend zijn. De latentietijd als parameter voor signaaldetectie presteerde minder goed dan verwacht op basis van eerder uitgevoerd onderzoek. Dit kwam vooral doordat de gevoeligheid van de methode te laag was. Het multivariabele model daarentegen, liet goede resultaten zien en de methode wordt momenteel gebruikt binnen de routine signaaldetectie binnen Lareb.

Toekomstig onderzoek zou zich kunnen richten op zowel modellen met meer en/of andere parameters dan in dit proefschrift beschreven, maar er zou ook andere soorten modellen gebruikt kunnen worden, zoals b.v. neurale netwerken. Een andere mogelijkheid is het toevoegen van externe bronnen aan de signaaldetectie, zoals databases met uitgebreide geneesmiddelinformatie. Als laatste zou natural language processing (NLP, een methode om geautomatiseerd tekst te interpreteren) van toegevoegde waarde kunnen zijn. In het huidige format dat gebruikt wordt om bijwerkingen te registreren wordt nog redelijk veel vrije tekst gebruikt (tekst die niet gecodeerd wordt) en deze is lastig te analyseren met statistische methoden. Middels NLP zou een interpretatieslag uitgevoerd kunnen worden waarna statistische analyse mogelijk zou zijn.

Samenvattend laat dit proefschrift zien dat statistische methoden toegevoegde waarde kunnen hebben binnen de signaaldetectie van bijwerkingen.

Dankwoord

Na mijn afstuderen als medisch bioloog leek een promotie traject niet iets dat op mijn pad zou komen. Vijf dagen per week laboratorium onderzoek doen past niet bij me en binnen de functies die ik in de farmaceutische industrie had waren op dat moment ook geen mogelijkheden. Maar een balletje kan raar rollen, en toen ik bij Lareb binnen kwam kreeg ik de mogelijkheid om een epidemiologie master te volgen, gecombineerd met een promotie onderzoek. Niet bepaald een standaard carrière-route, maar om op je 46^e nog als 'jonge' doctor door het leven te gaan spreekt me dan wel weer aan! Graag wil ik iedereen bedanken die het mogelijk heeft gemaakt dit promotietraject tot een goed einde te kunnen brengen.

In eerste instantie wil ik mijn beide promotoren Eugène van Puijenbroek en Eelko Hak bedanken. Beste Eugène, ik ben er tot op de dag van vandaag nog niet uit of je een generalistische specialist of een specialistische generalist bent. Laten we het erop houden dat je van veel dingen veel weet. De discussies die we de afgelopen jaren over mijn onderzoek hadden gingen over farmacovigilantie, geneeskunde, farmacologie, epidemiologie, statistiek en machine learning, regelmatig voorzien van een ietwat filosofische inslag (*"In der Beschränkung zeigt sich erst der Meister"*) maar ook het nodige pragmatisme (*"Laten we dit eens op z'n Brabants aanpakken"*). Bedankt voor alles wat je me geleerd hebt in de afgelopen jaren!

Beste Eelko, hoewel je pas wat later in mijn promotietraject in beeld kwam, heb je toch duidelijk je stempel gedrukt op de inhoud van dit proefschrift. Met name je vermogen om de epidemiologische puntjes op de i te zetten heb ik als zeer waardevol ervaren.

Ook mijn copromotor Florence van Hunsel wil ik danken voor haar bijdrage aan dit proefschrift. Beste Florence, altijd was je bereid om te discussiëren over een onderzoeksoptzet of om een kritische blik op mijn teksten te werpen, ook in tijden van extreme drukte zoals we die de afgelopen COVID periode ervaren hebben. Bedankt daarvoor! Ik hoop dat we binnenkort weer tijd hebben om een aantal nieuwe ideeën voor onderzoek vorm te geven.

Ook dank aan de (oud-)directeuren van Lareb. Beste Agnes, hoewel een promotie traject een academische exercitie is, heb ik mijn onderzoeken vooral binnen Lareb uitgevoerd en ik ben je erg dankbaar dat je me die mogelijkheid gegeven hebt en dat je het doen van onderzoek binnen Lareb zo hoog in het vaandel hebt staan.

Beste Kees, hoewel je niet betrokken bent geweest bij mijn daadwerkelijk promotie onderzoek, ben je wel degene geweest die mij in dienst heeft genomen bij Lareb, en heb je me de mogelijkheid gegeven een opleiding tot epidemioloog te volgen. Zonder die voorwaarden was dit proefschrift er niet geweest en ik ben je er dan ook erg dankbaar voor.

Verder wil ik de leden van de beoordelingscommissie, Prof. dr. P.M.L.A. van de Bemt, prof. dr. P.G.M. Mol en prof. dr. O.H. Klungel, bedanken voor de beoordeling van dit proefschrift en voor de adviezen tot verbetering ervan.

Ook wil ik mijn coauteurs bedanken voor hun bijdrage en in het bijzonder Kevin Klein en Peter van de Ven. Beste Kevin, diverse middagen hebben we bij Lareb narratives door zitten spitten om te kijken hoe we meldingen uit patient support programs konden identificeren. En het resultaat daarvan mag er zijn met ons artikel over precautionary reporting bias. Nu mijn promotietraject afgesloten is, lijkt het me leuk om weer een keer samen een onderzoek op te zetten.

Beste Peter, wat was het ontzettend fijn om een statisticus als stagebegeleider te hebben tijdens de epidemiologie master die de opmaat was voor deze promotie. Los van de inzichten in statistiek die je me hebt gegeven, heb je me kennis laten maken met R, waar ik tot op de dag van vandaag de vruchten van pluk.

Zonder goede collega's is het lastig onderzoek doen (dat zou zomaar een stelling bij een proefschrift kunnen zijn) en ik wil alle (oud-)collega's van Lareb dan ook bedanken voor de collegialiteit en fijne werkomgeving. Hopelijk hebben we tegen de tijd dat dit proefschrift daadwerkelijk gedrukt is de COVID pandemie grotendeels achter ons gelaten en zijn we weer "gewoon" op kantoor aan het werk. In het bijzonder wil ik Annet van Erp – van Boekel en Annemarie Muller bedanken voor hun hulp bij de klinische interpretatie van de resultaten uit de studie over precautionary reporting bias.

Dank aan mijn paranimfen Hanneke en Han. Ik ben erg blij dat jullie me bij willen staan tijdens de verdediging. Ruim 21 jaar geleden stonden we in een vergelijkbare opstelling en zoals je ziet is leeftijd geen beperking voor een promotie, dus als jullie nog plannen hebben.....

Lieve familie en vrienden, bij werk en wetenschappelijke inspanning hoort sociale ontspanning. Bedankt voor alle ontspanning die jullie me hebben gegeven in de afgelopen jaren!

Lieve pa en ma, onderzoek doen en promoveren begint met (wetenschappelijke) nieuwsgierigheid en de basis daarvoor wordt al in de kinderjaren gelegd. Jullie hebben me altijd gestimuleerd om het beste uit mezelf te halen (al deed ik daar halverwege de middelbare school niet zo veel mee) en gezorgd voor een fijne thuissituatie die dat mogelijk maakte. Hiervoor ben ik jullie ontzettend dankbaar!

Lieve Inge en Luuk, als laatste genoemd maar uiteindelijk het belangrijkste. Onderzoek doen en daarover "een boekje" schrijven is allemaal leuk en aardig. Maar het belangrijkste is dat je een thuis hebt met degenen van wie je houdt, en ik kan me geen beter thuis wensen. Ik ben zo blij dat ik jullie heb. Bedankt voor alles!

List of publications

Publications marked with an asterix (*) are related to the work presented in this thesis

PUBLICATIONS PRESENTED IN THIS THESIS

Klein K, Scholl JHG, De Bruin ML, van Puijenbroek EP, Leufkens HGM, Stolk P. When More Is Less: An Exploratory Study of the Precautionary Reporting Bias and Its Impact on Safety Signal Detection. *Clin Pharmacol Ther.* 2018;103(2):296-303.

Scholl JH, van de Ven PM, van Puijenbroek EP. Parametric time-to-onset models were developed to improve causality assessment of adverse drug reactions from antidiabetic drugs. *J Clin Epidemiol.* 2015;68(12):1423-31.

Scholl JH, van Puijenbroek EP. The value of time-to-onset in statistical signal detection of adverse drug reactions: a comparison with disproportionality analysis in spontaneous reports from the Netherlands. *Pharmacoepidemiol Drug Saf.* 2016;25(12):1361-7.

Scholl JHG, van Hunsel F, Hak E, van Puijenbroek EP. Time to onset in statistical signal detection revisited: A follow-up study in long-term onset adverse drug reactions. *Pharmacoepidemiol Drug Saf.* 2019;28(10):1283-9.

Scholl JHG, van Hunsel F, Hak E, van Puijenbroek EP. A prediction model-based algorithm for computer-assisted database screening of adverse drug reactions in the Netherlands. *Pharmacoepidemiol Drug Saf.* 2018;27(2):199-205

OTHER PUBLICATIONS

Glerum PJ, Maliepaard M, de Valk V, Scholl JHG, van Hunsel F, van Puijenbroek EP, et al. Quantification of Adverse Drug Reactions Related to Drug Switches in The Netherlands. *Clin Transl Sci.* 2020;13(3):599-607.

Brand W, Noorlander CW, Giannakou C, De Jong WH, Kooi MW, Park MV, et al. Nanomedicinal products: a survey on specific toxicity and side effects. *Int J Nanomedicine.* 2017;12:6107-29.

Klein K, Scholl JH, Vermeer NS, Broekmans AW, Van Puijenbroek EP, De Bruin ML, et al. Traceability of Biologics in The Netherlands: An Analysis of Information-Recording Systems in Clinical Practice and Spontaneous ADR Reports. *Drug Saf.* 2016;39(2):185-92.

Venhuis BJ, Scholl JH, Keizers PH. Naar schatting ten minste 150.000 gebruikers van illegale erectiemiddelen in Nederland. *Ned Tijdschr Geneesk.* 2016;160:D302.

Scholl JH, van Eekeren R, van Puijenbroek EP. Six cases of (severe) hypoglycaemia associated with gabapentin use in both diabetic and non-diabetic patients. *Br J Clin Pharmacol*. 2015;79(5):870-1.

Scholl J, Dahan A, Hansma A. Afwijkende leverwaarden en behandeling met esketamine in verband met pijn; mogelijk een causaal verband. *PW Wetenschappelijk Platform* 2015;9(6):136-138.

Scholl J. Perifeer oedeem bij gebruik van metronidazol. *GeBu* 2013;47(5):57-58.

Scholl JH, van Puijenbroek EP. Hearing impairment associated with oral terbinafine use: a case series and case/non-case analysis in the Netherlands Pharmacovigilance Centre Lareb database and VigiBase. *Drug Saf*. 2012;35(8):685-91.

Scholl J. Sitagliptine en dyspneu. *GeBu* 2012;46(6):69-70.

Scholl J. Latanoprost en reukstoornissen. *GeBu* 2012;46(3):32-33.

Scholl J. Fluorchinolonen en stomatitis. *GeBu* 2011;45(93).

ABSTRACTS

Nguyen M, Velthuis K, Scholl J, Jansen J, Kosse L, ten Klooster P, Jessurun N, Vonkeman H. Survival analysis of time to first adverse drug reaction and drug survival in rheumatoid arthritis patients treated with adalimumab and etanercept. *Annals of the Rheumatic Diseases*. 2021;80:1122

Scholl J, van Hunsel F, Hilgersom W, van Puijenbroek E. Vaxbatch: a stratified disproportionality analysis tool for batch-related AEFI monitoring in vaccines. *Pharmacoepidemiol Drug Saf*. 2020;29(suppl 3):501-502.

Glerum PJ, Maliepaard M, de Valk V, Scholl JHG, van Hunsel F, van Puijenbroek EP, et al. Quantification of Adverse Drug Reactions Related to Drug Switches in The Netherlands. *Pharmacoepidemiol Drug Saf*. 2020;29(suppl 3):492.

*van Hunsel F, Scholl J, van Puijenbroek E. First year results of using a new prediction model-based algorithm for computer-assisted database screening of adverse drug reactions in the Netherlands. *Pharmacoepidemiol Drug Saf*. 2018;27(suppl 2):258-259.

*Scholl J, van Hunsel F, Hak E., van Puijenbroek E. Time-to-onset in statistical signal detection of ADRs: a follow-up study. *Pharmacoepidemiol Drug Saf*. 2018;27(suppl 2):348.

Scholl J, van Hunsel F, van Puijenbroek E. Changepoint analysis in pharmacovigilance signal detection. *Pharmacoepidemiol Drug Saf*. 2017;26(suppl 2):534-535.

*Scholl J, van Hunsel F, van Puijenbroek E. Prediction model based algorithm for a computer-assisted database screening tool. *Pharmacoepidemiol Drug Saf.* 2016;25(suppl 3):501-502.

Scholl J, van Hunsel F, Härmark L. Case-by-case assessment of ICSRs from Marketing Authorization Holders: a pilot study. *Drug Saf.* 2015;38(10):986.

*Scholl J, van de Ven P, van Puijenbroek E. Parametric time-to-onset (TTO) modelling of common ADRs in patients using antidiabetic drugs. *Pharmacoepidemiol Drug Saf.* 2015;24(9):656.

Scholl J, van Puijenbroek E, van Grootheest K. Hearing impairment associated with oral terbinafine use. *Drug Saf.* 2012;21(8)Supplement:90.

About the author

Joep Scholl was born on December 21st 1975 in 's-Hertogenbosch, The Netherlands. He completed his pre-university education at the Jacob-Roelandslyceum in Boxtel in 1995. Subsequently, he started his study medical biology at the Catholic University Nijmegen (now Radboud University) in Nijmegen, from which he graduated (Msc.) in 2001. In the same year he started his career in the pharmaceutical industry, where he held several roles in clinical drug development and medical affairs, working for contract research organisations (CROs) and a pharmaceutical company. In October 2010, he started working at the Netherlands Pharmacovigilance Centre Lareb as a scientific assessor. During his employment at Lareb he completed his master in epidemiology from the EpidM institute, Amsterdam and started the research described in this thesis.