Improved Signal Detection of Drug-Drug Interactions in the Post-Marketing Phase: Reference Sets, Quantitative Methods and Systems Pharmacology Aspects

Thesis submitted in accordance with the requirements of
the University of Liverpool for the degree of Doctor of Philosophy by

Elpida Kontsioti

June 2023

Declaration of Authorship

I, the undersigned, declare that the thesis has been composed by myself and that the work has not been submitted for any other degree or professional qualification. I confirm that the work submitted is my own, except where work which has formed part of jointly-authored publications has been included. My contribution and those of the other authors to this work have been explicitly indicated. I confirm that appropriate credit has been given within this thesis where reference has been made to the work of others.

Name: Elpida Kontsioti
Signature: 

Date: 25.06.2023

Abstract

In contemporary medicine, the continuous introduction of new medicines to the market has significantly contributed to disease prevention and improved patient outcomes. However, the administration of medicines is hindered by the emergence of unforeseen adverse effects, often observed during late-stage clinical studies or following authorisation. This has resulted in notable drug withdrawals due to these unexpected side effects. Of particular concern are the side effects resulting from drug-drug interactions (DDIs). The empirical study of DDIs before the drugs enter the market is challenging due to the limited number of co-prescribed drugs typically included in late-stage clinical trials. Also, computational methods for identifying potential DDIs during drug development are not capable of successfully capturing all adverse DDIs that can occur in clinical practice. Therefore, post-marketing surveillance plays a crucial role in detecting and monitoring DDIs, making pharmacovigilance an integral part of the drug lifecycle.

To address these challenges, this thesis proposes a comprehensive approach consisting of four stages to enhance signal detection activities for DDIs in the post-marketing setting. Firstly, the level of agreement on DDI information across major online drug information resources is assessed. The assessment results show considerable variation in the interacting drug pairs of the examined resources, together with variability in categorisation of severity and clinical management recommendations for the included DDIs. Such variability presents potentially deleterious consequences for patient safety and demonstrates a need for harmonisation and standardisation of the information available on drug information resources. In the second stage, a normalised reference set called CRESCENDDI (Clinically-relevant REference Set CENtred around Drug-Drug Interactions) is introduced. This publicly available dataset provides comprehensive information on DDIs and the individual behaviour of interacting drugs, facilitating research in signal detection methodologies and enabling quantitative performance evaluation. The third stage seeks to investigate the impact of confounding factors on existing signal detection methodologies. It is concluded that reference sets populated with some of the examined confounding factors can significantly impact the performance evaluation metrics, potentially altering the conclusions regarding which methodologies are perceived to perform best. The final stage proposes a novel signal detection method built upon a Bayesian hypothesis testing framework and combined with a systems pharmacology network to refine potential DDI signals and assess their biological plausibility. The results of this study showcase the potential of systems pharmacology to enhance signal detection in pharmacovigilance, with DDIs being an important and promising area of application.

In conclusion, this thesis presents a comprehensive framework that addresses the challenges of signal detection of DDIs. By focusing on data standardisation, reference set development, signal detection method building and signal refinement using biological plausibility aspects, the findings and tools developed in this thesis offer valuable insights for enhancing pharmacovigilance and ultimately promoting better healthcare outcomes.

Publications

The main chapters of this thesis are comprised of the following journal articles:

I **Kontsioti, E**, Maskell, S, Bensalem, A, Dutta, B, Pirmohamed, M. Similarity and consistency assessment of three major online drug–drug interaction resources. *Br J Clin Pharmacol.* **88**(9): 4067- 4079 (2022). https://doi.org/10.1111/bcp.15341

II **Kontsioti, E**, Maskell, S, Dutta, B, Pirmohamed, M. A reference set of clinically relevant adverse drug-drug interactions. *Scientific Data* **9**, 72 (2022). https://doi.org/10.1038/s41597-022-01159-y

III **Kontsioti, E**, Maskell, S, Pirmohamed, M. Exploring the impact of design criteria for reference sets on performance evaluation of signal detection algorithms: the case of drug-drug interactions. *Pharmacoepidemiology and Drug Safety* (2023). https://doi.org/10.1002/pds.5609

IV **Kontsioti, E**, Maskell, S, Pirmohamed, M. Identifying drug-drug interactions in spontaneous reports utilising signal detection and biological plausibility aspects. [Manuscript in preparation for submission to *Nature Communications*].

Acknowledgements

First and foremost, I am grateful for the invaluable opportunity to be supervised by Professor Simon Maskell and Professor Sir Munir Pirmohamed. Their support, encouragement, and insightful feedback have been instrumental in bringing this thesis to fruition.

Also, I express my appreciation to Antoni Wisniewski, Dr Bhaskar Dutta, and Dr Isobel Anderson at AstraZeneca for their indispensable guidance and contributions throughout this research endeavour. Their expertise has significantly enriched the quality and depth of this thesis.

Furthermore, I would like to extend my gratitude for the generous funding support provided by the Engineering and Physical Sciences Research Council (EPSRC) (award reference: 1959810) and AstraZeneca's Patient Safety Centre of Excellence. Their financial assistance has played a pivotal role in enabling the successful completion of this project.

I would like to offer my special thanks to my advisors Professor Jason Ralph and Dr Vincent Yip for their suggestions during the annual PhD progress interviews.

I would also like to acknowledge the organisations that have made their publicly available data and software accessible for this work. Their contributions have been crucial in facilitating the data collection and analysis presented in this thesis.

To my cherished friends from Liverpool, Alicia, Katerina, Marta and Tonya. Thank you for all the wonderful memories we have created together: yoga, secret concerts, running - everything has been filled with joy and countless laughs with you!

I would also like to deeply thank my office mates and friends, Marco, Alessandro, Dominik, Nestoras, Kira, Conor, Diego, Caroline, Noemie, Lykourgos, Manos, Matthew, Theofilos, Alexey, Vincent, Katerina. Our shared experiences have made the research journey more enjoyable and fulfilling.

A special and heartfelt thanks goes to my family. Firstly, I am indebted to my parents, Maria and Kostas, for their unconditional support throughout my life. You have always encouraged and enabled me to pursue my dreams. To my brother, Thomas, thank you for continuously inspiring me to persevere, even during the most challenging times. Also, my cousin Eugenia, for her support, encouragement, and the countless enjoyable moments we have shared throughout the years. Your presence has been a constant source of motivation and joy.

Last but not least, Nikos, my best friend and partner, thanks for your unconditional love that helped me keep going during difficult times and your endless patience. Without your support and encouragement, I would not have had the strength to complete this PhD. From venture to venture, from country to country, you have been there for me. Life is much more fun with you!

Contents

List of Figures

List of Tables

Nomenclature

ADR	Adverse Drug Reaction
AE	Adverse Event
AEOLUS	Adverse Event Open Learning through Universal Standardisation
AI	Artificial Intelligence
API	Application Programming Interface
ATC	Anatomical Therapeutic Chemical (Classification System)
AUC	Area Under the Receiver Operating Characteristic Curve
BR	Benefit-Risk
BNF	British National Formulary
CDF	Cumulative Distribution Function
CDS	Clinical Decision Support
CI	Confidence Interval
CIOMS	Council for International Organisations of Medical Sciences
CNN	Convolutional Neural Network
COX-2	Cyclooxygenase-2
CRESCENDDI	Clinically-relevant REference Set CENtred around Drug-Drug Interactions
CPRD	Clinical Practice Research Datalink
CSV	Comma Separated Values
CYP	Cytochrome P450
DC	Design Criterion
DDI	Drug-Drug Interaction
DIR	Drug Information Resource
EBGM	Empirical Bayes Geometric Mean
EHR	Electronic Health Record
EMA	European Medicines Agency
FAERS	Food and Drug Administration Adverse Event Reporting System
FDA	Food and Drug Administration
FN	False Negative
FP	False Positive

GP	General Practice
GPT	Generative Pre-trained Transformer
HLGT	High Level Group Term
HLT	High Level Term
HPO	Human Phenotype Ontology
IC	Information Component
ICSR	Individual Case Safety Report
IME	Important Medical Event
IMI	Innovative Medicines Initiative
IntSS	Interaction Signal Score
MC	Medical Concept
MHRA	Medicines and Healthcare products Regulatory Agency
MedDRA	Medical Dictionary for Regulatory Activities
MGPS	Multi-Gamma Poisson Shrinker
LSTM	Long Short-Term Memory
NLP	Natural Language Processing
NSAID	Non-Steroidal Anti-Inflammatory Drug
OHDSI	Observational Health Data Sciences and Informatics
PD	Pharmacodynamic
PK	Pharmacokinetic
PMS	Post-Marketing Surveillance
PPI	Protein-Protein Interaction
PT	Preferred Term
PV	Pharmacovigilance
P-gp	P-glycoprotein
PPV	Positive Predictive Value
QSD	Quantitative Signal Detection
ROC	Receiver Operating Characteristic
RCT	Randomised Clinical Trial
SDA	Signal Detection Algorithm
SMQ	Standardised MedDRA Query
SNOMED-CT	Systematised Nomenclature of Medicine Clinical Terms
SRS	Spontaneous Reporting System
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
TF-IDF	Term Frequency - Inverse Document Frequency
THIN	The Health Improvement Network
TN	True Negative
TP	True Positive
UMC	The WHO Uppsala Monitoring Centre
WHO	World Health Organisation
WT	Wild Type

Chapter 1

Introduction

1.1 Motivation

In contemporary medicine, medicines are utilised as front-line tools intended to prevent, cure or mitigate diseases, with the aim of helping us live longer and better. With the global pharmaceutical industry's expenditure being more than 200 billion dollars per year for research and development, there has been a continuous increase in the number of new medicines that have been introduced to the market in the last few decades [1]. Although medicines are designed and administered to be beneficial, they can also turn out to be harmful to patients. One apparent reason is that they can be given in the incorrect dose, also according to Paracelsus' famous dictum [2]:

What is there that is not poison? All things are poison and nothing is without poison. Solely the dose determines that a thing is not a poison.

However, the fact that the dose of a medicine within the human body can end up causing unintended harm to a patient could arise from complex behaviour involving the drug molecule interacting with the metabolic and immunological pathways in unpredictable ways. Medicines, and in particular small molecules, can interact within the human body in multiple ways and get involved in various biological pathways that might not always be just the intended or the ones that are predictable based on current knowledge. Thus, apart from the desired effects, there might be other effects that arise from the interactions of the medicine with targets other than the intended therapeutic target. Any such unwanted and harmful response to a drug given at a therapeutic dose is described as an *adverse drug reaction* (ADR) [3]. The therapeutic effects are not always the same for every patient, such that not every drug has a beneficial effect for every patient; in a similar way, the secondary effects can also be subject to inter-individual variability.

This provides a rationale why so many candidate drugs fail during the development process, with 90% of those entering clinical trials never making it to the market [4]. Hence, we realise that, although the aim of pharmacotherapy is to provide benefits to patients, there are also potential risks. Nowadays, pharmacotherapy aims to strike a balance between benefit and risk. The benefit depends on multiple factors, including the

severity of the condition for which the medicine is used, the degree of beneficial impact on the patient's overall health and well-being, but also on the existence of alternative treatments and how their efficacy and safety compare to those for the drug in question. The risk, on the other hand, is calculated by taking into account the probability of the occurrence of the various potential side effects as well as how serious these are. This benefit-risk (BR) assessment for a drug is ongoing and extends beyond its marketing authorisation. This process involves doctors, regulators, and marketing authorisation holders who continuously collect evidence and evaluate the risk-benefit ratio of marketed products. The activity of gathering and analysing data in the post-marketing setting with an overarching aim of identifying previously unknown adverse drug effects or any other complications falls under the concept of *pharmacovigilance*.

The first and crucial point in pharmacovigilance activities following data collection is called *signal detection* and includes the statistical analysis of gathered data to spot any signals that could be indicative of a novel drug side effect. The collected data, traditionally in the form of reports submitted by health professionals or patients, contain information on patient demographics, concomitant medications, suspected drugs and adverse events (AEs) experienced following drug administration.

Increasing life expectancy rates worldwide contribute to rising multimorbidity rates, along with more medicines being approved for use. As a result, the issue of polypharmacy (i.e. regular simultaneous use of five or more medications) has become more prevalent and more urgent year by year. This can lead to a massive and ever-increasing number of potential interactions between administered drugs. There has lately been an increasing amount of computational work with the aim of identifying potential DDIs in the drug development stage, in an attempt to reduce rejection rates due to safety complications. However, post-marketing surveillance (PMS) still remains essential and critical due to the inherent complexity and the lack of complete understanding of drug behaviour and interactions within the human body in the presence of other drugs. At the same time, the statistical methods required to detect signals of potential DDIs in PMS are not as mature as the respective ones for single drugs.

ADRs have been identified as a major cause of hospitalisation, with the number of DDIs increasing from being accountable for one in six ADRs according to a study around 20 years ago [5] to being accountable for almost one in three in a more recent study [6]. More interestingly, a significant percentage of the DDIs identified in [6] were deemed preventable. Computerised systems that provide information at the time of prescribing aim to support decision-making and reduce the number of ADRs in clinical practice caused by DDIs. However, when it comes to DDI management in clinical practice, there is an issue of lack of standardisation and harmonisation among data resources. This comes as no surprise and might be explained by multiple factors, such as: different inclusion criteria that might affect the number and nature of the qualifying DDIs for consideration; decisions to be over-inclusive to limit legal liability; geographic variations

etc. Hence, developing an understanding of the level of discordance in the various types of information provided in different clinical resources is important.

The presence of a considerable degree of disagreement amongst existing resources for DDIs can also impede our ability to rely on a single source to gather sufficient evidence if we want to build a reference set with positive and negative controls to evaluate the performance of novel methods for signal detection of DDIs. This realisation, combined with the fact that the available evidence in drug safety is not static, means that definitive controls in the form of a "gold standard" cannot exist, making the task of building a reference set a challenging one. Traditionally, reference sets in pharmacovigilance have been manually curated, custom-made, and often limited in size [7, 8, 9], with only a few efforts having considered data standardisation practices so that the generated reference sets could be compatible across different sources of data [7, 10]. In other cases, large online databases containing potential DDIs (rather than clinically-relevant ones) have been used to evaluate novel methodologies. Hence, the possibility to automate data extraction, aggregate and standardise information from clinical resources on DDIs to build a reference set seems intriguing and, based on recent technologies that have been widely implemented in other scientific fields, also possible. At the same time, supporting this resource with data pertinent to the behaviour of interacting drugs in isolation of one another is an additional important consideration.

The use of small, custom-made reference sets that consider ad-hoc exclusion or inclusion criteria to define eligible controls complicates the comprehensive comparative evaluation and benchmarking of various methodologies in the literature. As the different choices that are made might affect the performance of each quantitative methodology for signal detection differently, they can act as potential confounders. By having access to a large and diversified reference set, it should be possible to assess the impact of those design criteria that we have seen driving choices when building reference sets in the literature on the relative performance evaluation of signal detection methodologies for DDI surveillance.

Some of the existing signal detection methodologies for DDI surveillance have utilised regression modelling [11], adapted shrinkage observed-to-expected ratios [12, 13], or association rule mining [14] to unveil signals that might indicate a novel DDI in the pharmacovigilance data. However, the inherent complexity of DDIs along with computational challenges have resulted in these efforts not being as mature as methodologies that have been developed for the identification of single-drug safety complications. At the same time, it is unfortunate that these efforts are not as developed considering the complexity of the problem. In the case of DDIs, the careful enumeration, estimation, and comparison of the different possible combinations of rates of occurrence of an AE that could arise when a drug combination is given, as opposed to the respective rates when the individual drugs are administered separately, could suggest the presence of a signal of disproportionate AE reporting in the data attributed solely to the drug combination, thus indicating the presence of a DDI. Furthermore, the potential of coupling

of statistical methodologies in pharmacovigilance with other types of data that could motivate the explanation of a predicted signal from a biological perspective has been suggested previously [15]. However, it has not been adequately explored in the case of DDIs. Multiple recent machine learning and deep learning approaches have attempted to predict novel DDIs using disparate data sources [16]. In pharmacovigilance, signal detection methods have been mainly disassociated from signal evaluation frameworks. Also, the availability of trustworthy data for drug targets and drug safety complications linked to target activity has not been leveraged so far for novel DDI identification. Open Targets [17] is a very good example and its potential to aid pharmacovigilance has been discussed in the literature [18, 19].

1.2 Thesis Aim & Objectives

The previous section has highlighted the landscape of opportunities and challenges related to the PMS of DDIs. The overarching aim of this thesis is to provide data, methods and research outputs that could overall lead to improved detection of adverse DDIs in the post-marketing setting. More specifically, the following objectives are proposed:

Objective I: Understand the existing evidence related to clinically relevant and observable DDIs by exploring the level of agreement on information listed in different drug information resources (DIRs).
This objective seeks to assess the concordance of leading clinical resources for DDIs from three different countries of origin in terms of: (1) inclusion of interacting drug pairs; (2) severity rating; (3) evidence rating and (4) clinical management recommendations.

Objective II: Build a clinically relevant reference set for DDIs.
This objective aims to develop a reference standard that can be used to facilitate research and allow common ground for comparing methodologies used for DDI surveillance in pharmacovigilance. Although a definitive reference standard including the complete set of DDIs cannot exist, a scalable approach that requires less manual effort for future updates, considering the dynamic nature of data and evidence availability, should be an adequate solution.

Objective III: Assess the impact of different choices on the nature of the controls included in a reference standard on the performance assessment of existing signal detection algorithms (SDAs) for DDI surveillance.
This objective aims to identify the relative impact of different factors that could be potential sources of confounding on the performance evaluation of existing methods for signal detection of DDIs. The utilisation of a large and diversified reference set enables the generation of smaller, custom-made reference sets considering multiple design criteria to assess any differences observed in the quantitative evaluation of SDAs for two-way DDIs.

Objective IV: Evaluate the ability of a novel Bayesian hypothesis testing framework to identify signals of disproportionate reporting indicative of DDIs in PMS data and assess the potential for signal refinement using biological plausibility aspects.
This objective is related to: (a) the development of a novel SDA for adverse DDIs that could produce a pharmacology-driven output by detecting increasing reporting rates in spontaneous reporting system (SRS) data while being able to distinguish signals that might arise from constituent drugs; (b) the assessment of the developed SDA in comparison to existing methodologies by utilising a large and diversified reference set; and (c) the generation of a systems pharmacology framework of established associations between biological nodes (i.e. drug targets, drug ingredients, and AEs) to refine the signals of potential DDIs and assess their biological plausibility.

1.3 Thesis Structure

The remaining chapters are organised as follows: first, the thesis continues with a Preliminaries chapter (**Chapter 2**); this is followed by four core chapters that are presented in a journal paper format (**Chapters 3-6**), and a final discussion chapter (**Chapter 7**).

Chapter 2 provides the theoretical background of this thesis. The first part presents the aims and importance of pharmacovigilance, along with current practices for collecting, analysing, and leveraging data to identify drug complications in the post-marketing phase. The second part introduces key concepts around DDIs from a pharmacological perspective, including classification and clinical implications, as well as methods and frameworks for their postmarketing surveillance.

Chapter 3, aiming to address **Objective I**, explores the level of agreement on DDI information listed in three major online drug information resources (DIRs) in terms of: (1) interacting drug pairs; (2) severity rating; (3) evidence rating and (4) clinical management recommendations. By extracting and normalising the data included in the different resources, the overlap of information pertinent to DDIs is assessed. Free text provided in the DIRs related to clinical management recommendations is also annotated, either manually, where possible, or through the application of a machine learning algorithm, for nine different clinical management categories.

Chapter 4 proposes a scalable approach to address **Objective II**. It presents CRESCENDDI (Clinically-relevant REference Set CENtred around Drug-Drug Interactions), a normalised reference set for DDIs coupled with information on the individual behaviour of interacting drugs. CRESCENDDI was built following the FAIR Data Principles (Findable, Accessible, Interoperable, and Reusable). The automatic extraction and aggregation of information from multiple clinical resources on DDIs and the individual behaviour of interacting drugs led to this publicly available data set that can be used to facilitate research in SDAs and allow common ground for comparing methodologies, requiring less manual effort for future updates.

Chapter 5 is linked to **Objective III**. This chapter investigates the impact of 14 criteria that could act as potential confounders (e.g. AE frequency) on three methods that have been developed to detect signals of potential adverse DDIs. Given that each method may be impacted to a different extent by those criteria, the assumption is that the relative composition of reference sets can significantly affect the evaluation metrics, potentially altering the conclusions regarding which methodologies are perceived to perform best. This is particularly relevant when using custom-made and small-in-size reference sets, which in many cases represent published work in the area of signal detection methods in pharmacovigilance, impeding a comprehensive and "fairer" evaluation.

Chapter 6 addresses **Objective IV**, by first introducing a novel signal detection method for adverse DDIs that could produce a pharmacology-driven output by detecting increasing reporting rates in spontaneous reporting system data while being able to distinguish signals that might arise from constituent drugs. It then utilises a systems pharmacology framework of established associations between biological nodes (i.e. drug targets, drug ingredients, and AEs) to refine the signals of potential DDIs and assess their biological plausibility.

Finally, **Chapter 7** provides a discussion of the thesis contributions and conclusions, also suggesting future directions for research and areas for improvement.

At the beginning of each chapter, there is a short introduction that tries to identify the connections between the material presented in the chapter and the remaining chapters, as well as connections to the overarching aim of this thesis.

1.4 Contributions

The four core chapters of this thesis (**Chapters 3-6**) are presented in the form of scientific journal publications and have been either published or are being prepared for submission.

Chapter 3 is a reproduction of a publication in the *British Journal of Clinical Pharmacology* [20]. I would like to thank Prof Simon Maskell and Prof Sir Munir Pirmohamed for their contribution to the conceptualisation of the original method and the analysis design. In this chapter, I performed all analyses and produced all writing and figures while taking into account comments and critical feedback from the co-authors. I would like to extend my appreciation to Dr Amina Bensalem for her contribution to the data analysis by independently annotating and validating French language free text from ANSM Thesaurus for clinical management recommendations.

Chapter 4 is a reproduction of published work in *Scientific Data* [21]. I conducted the data extraction, developed the analysis code, produced all data sets and wrote the manuscript while taking into account comments and critical feedback from the co-authors.

Chapter 5 is a reproduction of published work in the *Pharmacoepidemiology and Drug Safety* journal. I conceived the work and designed the analysis, with contributions

from the other two publication co-authors. I conducted the data extraction, developed the analysis code and produced all writing and figures while taking into account comments and critical feedback from the co-authors.

Chapter 6 is currently the draft version of a manuscript that is intended to be submitted to the *Nature Communications* journal. Prof Simon Maskell and I jointly developed the novel method for Bayesian signal detection of DDIs in PMS. I designed the biological plausibility framework, performed all analyses and produced all writing and figures while taking into account comments and critical feedback from the co-authors.

Chapter 2

Preliminaries

This chapter provides the theoretical foundations for this thesis. The first part presents an overview of current pharmacovigilance processes and discusses the concepts, goals, practices, and methodologies in the field. The second part of this chapter introduces the main concepts of DDIs from a pharmacological perspective and explains how DDI surveillance is performed.

2.1 Pharmacovigilance

2.1.1 Aim and Importance

In the 1950s, thalidomide was increasingly used in Europe to treat morning sickness in pregnant women. However, doctors started to notice limb malformations and other severe birth defects in newborn children associated with maternal exposure to thalidomide during pregnancy. These observations were reported to scientific journals [22] and, by early 1960s, thalidomide was withdrawn from multiple countries. The thalidomide tragedy indicated the urgent need for a data collection system of suspected side effects to medicines and the establishment of an orchestrated surveillance system that could capture future adverse drug reactions that are not understood before a drug reaches the market.

As drug development is a lengthy process that includes several steps across multiple years, it is often falsely perceived that a medicine's safety profile is fully understood by the time it gets approved for human use. Given the complexity of the human body, it is expected that a molecule can have multiple effects when administered by binding to multiple proteins and being involved in various biological pathways before it is excreted from the human body. Some of these effects are therapeutic and intended while others are called side effects and have the potential to cause harm. Hence, it is important to stress that, despite the popular public misconception that exists, no drug comes without side effects nor is it 100% safe. With the average time from drug discovery to drug approval being around 10 years nowadays [4], multiple pre-clinical as well as clinical assessments try to gather sufficient evidence that the candidate molecule meets the safety and efficacy requirements to be released onto the market. Safety is tested in multiple ways, such as laboratory (*in vitro*) experiments, animal models and computational models, followed by clinical studies in humans [23].

Regulatory bodies in the respective countries are responsible for interpreting the collected data to decide whether the perceived benefit of the drug outweighs its perceived risk [24]. Benefit-risk (BR) assessment is a complex procedure that takes into account multiple parameters to evaluate whether there is a balance between a drug's benefits and risks. Over the years, it has been considered a subjective process in its nature [25, 26]. It was only in the late 1990s when the need for more systematic and consistent frameworks and guidelines on how to perform BR assessment for regulatory decision-making was highlighted [27, 28]. The European Medicines Agency (EMA) has adopted a combination of qualitative and quantitative approaches to conducting BR evaluation. In the United States, the FDA has traditionally leaned towards the adoption of a qualitative approach, suggesting that this would be more appropriate and capable of including expert judgement, while it can accommodate quantitative elements if needed. In the past ten years, the Medicines and Healthcare products Regulatory Agency (MHRA) in the United Kingdom has emphasised the significance of making better-informed BR decisions. They have stressed the need to consider patient experiences, utilise real-time

data, enhance tools for continuous BR evaluation, and improve understanding of these decisions [29, 30]. But, inevitably, subjectivity is a component in BR evaluation that should be acknowledged and embraced [31].

Pre-approval clinical trials involve three different phases (Phase I-III), which aim to gather evidence that could inform the BR assessment of drugs. Provided that the BR profile of a candidate drug remains favourable throughout the multiple steps of pre-marketing clinical trials, marketing authorisation is granted. Subsequently, BR evaluation continues to ensure that there is no notable divergence from the original assessment back when the drug was approved [24, 31]. To be able to do that, the first step is to collect data that could enable the reevaluation of the drug's perceived BR ratio by analysing the collected data. These activities fall under the concept of *post-marketing surveillance* (PMS), or *pharmacovigilance* (PV). Manufacturers alongside national authorities that are responsible for marketing authorisations of the medicines are primarily responsible for PMS activities. Additionally, the World Health Organisation (WHO) has established a monitoring system called *Programme for International Drug Monitoring* since 1968, which today counts 151 member countries and aims to promote PV at the country level and enhance patient safety through cross-country collaboration.

PMS data are invaluable and essential, as we need to take into account that clinical trials are inevitably limited in size and also not fully representative of the population that utilises the drug in the real world. This happens for multiple reasons, including the inadequacy of some patient groups to participate in clinical trials due to reasons such as multimorbidity, lack of eligibility due to age (e.g. paediatric use of medicines), minority ethnic groups being under-represented in clinical trials, variations in drug responses due to genetic makeup (pharmacogenomic factors) and the inability of clinical trials to capture long-term effects due to their necessarily limited length. Hence, we need to recognise that patient recruitment processes, eligibility criteria, and the design of clinical trials might lead to BR assessments that require updating following drug approval.

In case the additional data support the change of the BR ratio, there are multiple possible outcomes. The decisions, apart from merely considering the incidence rates of the various ADRs that occur in the general population as well as in specific subgroups, also rely on other factors. Some of these include: the existence of other therapeutic alternatives in the market and how their safety profile compares to the medicine in question; the intended benefit from drug use, given that some ADRs might remain acceptable if the benefits from using the medicine are significant and cannot be achieved otherwise; and the exposure (i.e. how many people take the medicine). There are cases where there is evidence that specific subgroups should avoid taking a specific medicine, which leads to an update of the patient information leaflet. On the other hand, it is also possible that the medicine is withdrawn from the market, as it poses a significant risk to the patients. For example, rofecoxib, widely known by its brand name Vioxx, is a cyclooxygenase-2 (COX-2) selective nonsteroidal anti-inflammatory drug (NSAID) that was withdrawn from the market in early 2000. Similarly, rosiglitazone (an antidiabetic

agent) was withdrawn in 2010. In both cases, other alternative treatments existed from the same therapeutic class that supported the decision to withdraw.

Hence, PMS includes multiple activities and entails a number of possible implications for medicines in the post-marketing setting. Since these range from revealing safety issues in specific sub-populations that might lead to patient information leaflet updates to even drug withdrawal, it is important to ensure that the mechanisms in place are up to date and supported by the most recent scientific evidence so that patient safety is maximised.

2.1.2 Spontaneous Reporting System Databases

The reporting of suspected adverse drug reactions and other complications that might be linked to the administration of the approved drug that remains unidentified is called *spontaneous reporting* [32]. This process involves the submission of Individual Case Safety Reports (ICSRs) by health professionals, drug manufacturers, but also patients and consumers. ICSRs are maintained in spontaneous reporting system (SRS) databases.

The collection of spontaneous reports and maintenance of the SRS databases are performed by the national authorities who are responsible for the authorisation of the drugs in the respective countries. In the United Kingdom, the Yellow Card Scheme was established in 1964 [33]. ICSRs for suspected ADRs can be filed by health professionals (**Figure 2.1**) or patients (**Figure 2.2**). Today, online reporting through the Medicines and Healthcare products Regulatory Agency (MHRA) Yellow Card app or website (https://yellowcard.mhra.gov.uk/) are the most common ways to submit a report to the MHRA. In the United States, MedWatch is the FDA's PMS programme, with the FDA Adverse Event Reporting System (FAERS) database representing one of the largest SRS databases worldwide [34]. The FAERS database is available online via a public dashboard[1] that provides descriptive statistics dating back to 1968 or in downloadable data files, with quarterly updates available since 2004. As of a recent update (June 30th, 2022), the database contained a total of 24,809,611 reports.

OpenFDA[2] is an application programming interface (API) that functions as a platform for accessing open FDA data. It provides an endpoint to extract drug adverse events reported to FDA and stored in FAERS since 2004, allowing easy FAERS data extraction in JSON format. This API simplifies the extraction process by providing the data in JSON format, facilitating integration with other platforms. This allows for building applications that can also perform data mining tasks, such as predicting the seriousness of ADRs, by combining features from FAERS with drug molecule structure and pharmacological class information from other data sources [35].

Apart from national SRS databases, there exist efforts from multiple countries to aggregate their reports into a centralised database. VigiBase is the WHO Uppsala

[1] https://www.fda.gov/drugs/questions-and-answers-fdas-adverse-event-reporting-system-faers/fda-adverse-event-reporting-system-faers-public-dashboard

[2] https://open.fda.gov/

In Confidence

REPORT OF SUSPECTED ADVERSE DRUG REACTIONS Yellow Card

COMMISSION ON HUMAN MEDICINES (CHM)

It's easy to report online: mhra.gov.uk/yellowcard or via the app

If you suspect an adverse reaction may be related to one or more drugs/vaccines/complementary remedies, please complete this Yellow Card. See 'Adverse reactions to drugs' section in the British National Formulary (BNF), visit **mhra.gov.uk/yellowcard**, or see the back of this form for guidance. Do not be put off reporting because some details are not known.

PATIENT DETAILS Patient Initials:_____ Sex: M / F Ethnicity:_____ Pregnant? Y / N Weight (kg): _____

Age (at time of reaction):_____ Identification number (e.g. Your Practice or Hospital Ref):_____

SUSPECTED DRUG(S)/VACCINE(S)

Drug/Vaccine (Brand if known)	Batch	Route	Dosage	Date started	Date stopped	Prescribed for

SUSPECTED REACTION(S) Please describe the reaction(s) and any treatment given:
(Please attach additional pages if necessary)

Outcome
Recovered ☐
Recovering ☐
Continuing ☐
Other ☐

Date reaction(s) started:_____ Date reaction(s) stopped: _____

Do you consider the reaction(s) to be serious? Yes / No
If yes, please indicate why the reaction is considered to be serious (please tick all that apply):

☐ Patient died due to reaction
☐ Life threatening
☐ Congenital abnormality
☐ Involved or prolonged inpatient hospitalisation
☐ Involved persistent or significant disability or incapacity
☐ If medically significant, please give details:

Do you consider that the suspected reaction(s) resulted from a medication error? Yes/No

OTHER DRUG(S) (including self-medication and complementary remedies)

Did the patient take any other medicines/vaccines/complementary remedies in the last 3 months prior to the reaction? Yes / No
If yes, please give the following information if known:

Drug/Vaccine (Brand if known)	Batch	Route	Dosage	Date started	Date stopped	Prescribed for

Additional relevant information e.g. medical history, test results, known allergies, rechallenge (if performed), further details about any medication error (e.g. errors in prescription, dosing, dispensing or administration). For reactions relating to use of a medicine during pregnancy please state all other drugs taken during pregnancy, the last menstrual period, information on previous pregnancies, ultrasound scans, any delivery complications, birth defects or developmental concerns.

Please list any medicines obtained from the internet:

REPORTER DETAILS
Name and Professional Address: _____
_____
_____
Postcode: _____ Tel No:_____
Email: _____
Speciality: _____
Signature:_____ Date: _____

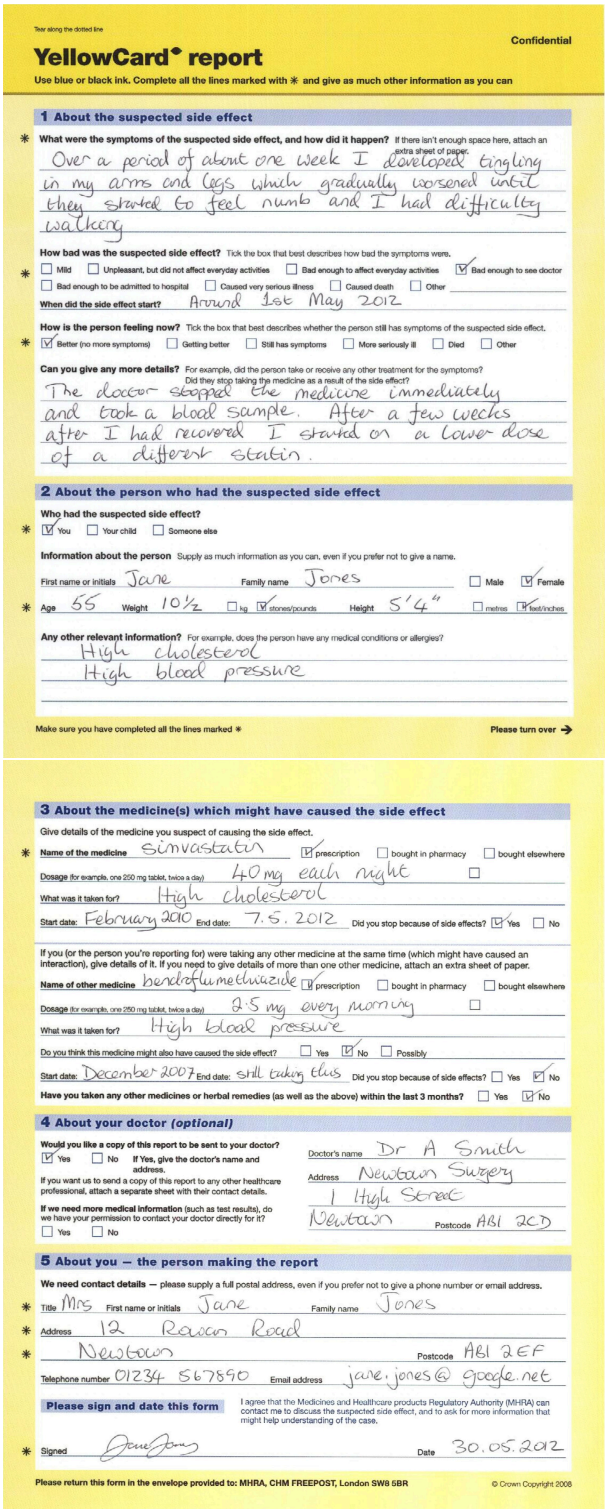
CLINICIAN (if not the reporter)
Name and Professional Address: _____
_____
_____
Postcode:_____ Tel No: _____
Email: _____
Speciality: _____
Date: _____

MHRA Regulating Medicines and Medical Devices

Interactive information on suspected adverse drug reactions received by the MHRA is available at **mhra.gov.uk/yellowcard** under Drug Analysis Profiles. Stay up-to-date on the latest advice for the safe use of medicines with our monthly bulletin *Drug Safety Update* at **gov.uk/drug-safety-update**

Please attach additional pages if necessary. Send to: FREEPOST YELLOW CARD (no other address details or postage required)

FIGURE 2.1: A healthcare professional Yellow Card reporting form from the Medicines and Healthcare products Regulatory Agency (MHRA).



Tear along the dotted line

Confidential

YellowCard report

Use blue or black ink. Complete all the lines marked with * and give as much other information as you can

1 About the suspected side effect

* **What were the symptoms of the suspected side effect, and how did it happen?** If there isn't enough space here, attach an extra sheet of paper.
Over a period of about one week I developed tingling in my arms and legs which gradually worsened until they started to feel numb and I had difficulty walking

How bad was the suspected side effect? Tick the box that best describes how bad the symptoms were.
* ☐ Mild ☐ Unpleasant, but did not affect everyday activities ☐ Bad enough to affect everyday activities ☑ Bad enough to see doctor
☐ Bad enough to be admitted to hospital ☐ Caused very serious illness ☐ Caused death ☐ Other

When did the side effect start? Around 1st May 2012

How is the person feeling now? Tick the box that best describes whether the person still has symptoms of the suspected side effect.
* ☑ Better (no more symptoms) ☐ Getting better ☐ Still has symptoms ☐ More seriously ill ☐ Died ☐ Other

Can you give any more details? For example, did the person take or receive any other treatment for the symptoms? Did they stop taking the medicine as a result of the side effect?
The doctor stopped the medicine immediately and took a blood sample. After a few weeks after I had recovered I started on a lower dose of a different statin.

2 About the person who had the suspected side effect

Who had the suspected side effect?
* ☑ You ☐ Your child ☐ Someone else

Information about the person Supply as much information as you can, even if you prefer not to give a name.
First name or initials Jane Family name Jones ☐ Male ☑ Female
* Age 55 Weight 10½ ☐ kg ☑ stones/pounds Height 5'4" ☐ metres ☑ feet/inches

Any other relevant information? For example, does the person have any medical conditions or allergies?
High cholesterol
High blood pressure

Make sure you have completed all the lines marked *

Please turn over →

3 About the medicine(s) which might have caused the side effect

Give details of the medicine you suspect of causing the side effect.
* **Name of the medicine** Simvastatin ☑ prescription ☐ bought in pharmacy ☐ bought elsewhere
Dosage (for example, one 250 mg tablet, twice a day) 40mg each night ☐
What was it taken for? High cholesterol
Start date: February 2010 End date: 7.5.2012 Did you stop because of side effects? ☑ Yes ☐ No

If you (or the person you're reporting for) were taking any other medicine at the same time (which might have caused an interaction), give details of it. If you need to give details of more than one other medicine, attach an extra sheet of paper.
Name of other medicine bendroflumethiazide ☑ prescription ☐ bought in pharmacy ☐ bought elsewhere
Dosage (for example, one 250 mg tablet, twice a day) 2.5 mg every morning ☐
What was it taken for? High blood pressure
Do you think this medicine might also have caused the side effect? ☐ Yes ☑ No ☐ Possibly
Start date: December 2007 End date: still taking this Did you stop because of side effects? ☐ Yes ☑ No
Have you taken any other medicines or herbal remedies (as well as the above) within the last 3 months? ☐ Yes ☑ No

4 About your doctor *(optional)*

Would you like a copy of this report to be sent to your doctor?
☑ Yes ☐ No **If Yes, give the doctor's name and address.**
Doctor's name Dr A Smith
Address Newtown Surgery
1 High Street
Newtown Postcode AB1 2CD

If you want us to send a copy of this report to any other healthcare professional, attach a separate sheet with their contact details.

If we need more medical information (such as test results), do we have your permission to contact your doctor directly for it?
☐ Yes ☐ No

5 About you – the person making the report

We need contact details – please supply a full postal address, even if you prefer not to give a phone number or email address.
* Title Mrs First name or initials Jane Family name Jones
* Address 12 Rowan Road
* Newtown Postcode AB1 2EF
Telephone number 01234 567890 Email address jane.jones@google.net

Please sign and date this form

I agree that the Medicines and Healthcare products Regulatory Authority (MHRA) can contact me to discuss the suspected side effect, and to ask for more information that might help understanding of the case.

* Signed Jane Jones Date 30.05.2012

Please return this form in the envelope provided to: MHRA, CHM FREEPOST, London SW8 5BR

FIGURE 2.2: A completed member of public Yellow Card reporting form from the Medicines and Healthcare products Regulatory Agency (MHRA).

Monitoring Centre (UMC) SRS database that collates ICSRs from 151 countries and has been in existence since 1968 [36]. As of January 2022, VigiBase contained over 30 million anonymised reports of suspected ADRs suffered by patients. The aim of pooling data into an international database is to identify signals that cannot be identified when screening national databases due to, for example, a limited number of reports regarding a suspected adverse reaction. Reporting policies for suspected ADRs vary in many cases between different countries (i.e. reporting might be either voluntary, encouraged or mandatory). Also, the types of information that are required from each responsible national authority might vary in terms of the level of detail, reporting of concomitant drugs, follow-up procedures etc. The types of eligible reporters can also be different. More specifically, patients are not allowed to directly report suspected side effects everywhere and this has only been introduced in the last twenty years in many countries [37, 38]. Hence, it is expected that national databases differ among countries in the data magnitude, completeness, quality, etc, unavoidably leading to data heterogeneity that should be taken into account when performing analyses using VigiBase.

In the European Union, the EMA is also responsible for collecting reports from the member countries in a centralised database called EudraVigilance. As opposed to FAERS, EudraVigilance is not publicly accessible and its access is restricted to national competent authorities within the European Union and authorised personnel involved in pharmacovigilance activities. However, some aggregated safety information and reports are made available to the public through the EMA's website[3].

In addition, marketing authorisation holders also maintain their own proprietary databases to collect safety reports related to their marketed products. These databases are used to capture and store information on adverse events (either spontaneously reported or collected from clinical trials), product complaints, and other safety-related data. The purpose of these proprietary databases is to enable marketing authorisation holders to monitor the safety of their products, detect emerging safety signals, and take appropriate actions to ensure patient safety.

Spontaneous reporting enables everyone who wishes to report a suspicion regarding the safety of a medication to do so. In this way, it becomes possible to gather data from patient groups that are under-represented during clinical trials, including people with multimorbidity, people who take multiple medications, ethnic minorities, etc. The data collection process is also simple, non-interventional, and inexpensive. On the other hand, spontaneous reporting has some clear drawbacks, as the process of data collection is passive. First, the lack of a control group, as opposed to a randomised clinical trial (RCT) impedes the data analysis. Furthermore, the reported information might be of insufficient detail or incomplete, with the option of follow-up being difficult in some cases. Other data quality issues include the presence of duplicate reports, where the same occurrence of one or more AEs in the same patient is reported multiple times, probably by different sources, with the reports not being linked to the same case [39, 40, 41]. The

[3] https://www.adrreports.eu/en/index.html

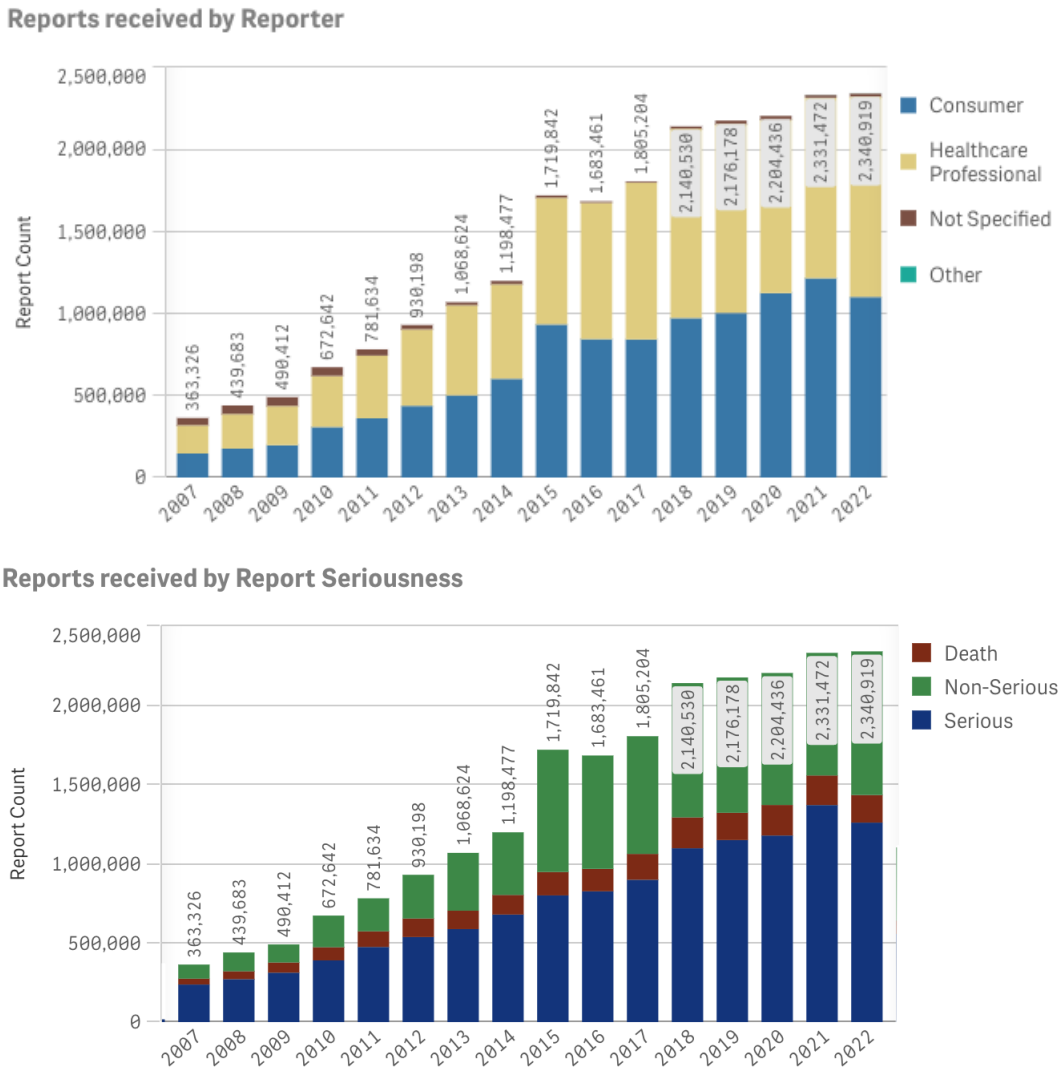


FIGURE 2.3: Number of FAERS Reports received by: (a) Reporter Type and (b) Seriousness (downloaded from the FAERS Public Dashboard).

issue of stimulated reporting is also important, as reporting processes might be driven by public disclosure of a drug safety issue and publicity [42]. Generally, under-reporting is problematic and might delay the detection of a novel signal [43].

SRS data volume has seen a sheer increase over the past 10 years (for example, FAERS report counts by year are shown in **Figure 2.3**). This phenomenon could be attributed to multiple factors, including: the ageing population; more people suffering from multiple conditions, hence being treated with more medicines; the widening phenomenon of polypharmacy; more new medicines reaching the market; clinicians and patients being more aware of the opportunity to report suspected adverse drug reactions and complications, and the switch from paper-based reporting methods to electronic methods. The increasing amount of data presents both a challenge and an opportunity to come up with methods that could provide us with better answers, maximise the value of the collected data, and promote patient safety.

2.1.3 Other Data Streams

Apart from the traditional SRS databases that have been used for pharmacovigilance purposes in the last few decades, alternative data streams have been considered to support the aim of timely and more accurate detection of safety issues related to medicines. These categories include: (i) electronic health record (EHR) databases; (ii) administrative claims data; (iii) scientific literature; (iv) social media data; and (v) patient registries.

2.1.3.1 Electronic health record databases

An EHR provides a comprehensive, up-to-date, and easily accessible collection of a patient's health information in a digital format. EHRs are routinely used in the clinical setting to help healthcare providers make informed decisions about a patient's care. EHR databases comprise a novel data stream that is increasingly used in pharmacoepidemiology and pharmacovigilance. This data stream can provide real-world evidence for further support of the BR assessment and the evaluation of drug effectiveness. Some of the types of information contained in EHRs can provide important context and insights into the patient's health status and are valuable for pharmacovigilance, including patient demographics, medical history (i.e. any prior diagnoses, treatments, and prescribed medications), laboratory test results, clinical notes (i.e. observations and assessments made by healthcare providers), detailed medication information (e.g. dose, frequency, and duration of use), and genomic data (e.g. targeted genotyping, next-generation sequencing) [44, 45].

EHR data offer the capability of estimating drug exposure and background incidence rates of AEs, as opposed to SRS databases, by knowing how many people took the drug but did not experience an AE. EHRs also enable the analysis of data over time (*longitudinal analysis*), allowing the identification of changes in patterns or trends that may indicate a safety signal in the context of pharmacovigilance. Changes in the incidence or severity of ADRs can be used to determine whether there is a temporal relationship between medication use and the manifestation of an AE. In addition to tracking the incidence of AEs, longitudinal analysis can also be used to evaluate the potential effects of interventions, such as changes in dose or the addition of concomitant medication, on the occurrence of AEs. This information can inform decisions about changes in treatment strategies and help to optimise patient safety. Time-to-onset is another important aspect of pharmacovigilance that can inform causality and explain links between medications and AEs, with time-stamped patient records being suitable to provide information on that end [46]. Especially in the context of drug-drug interactions (DDIs), analyzing the sequence of medication introduction to a treatment plan and the subsequent occurrence of AEs in a patient can offer valuable insights into identifying the drugs that may be associated with those AEs potentially via a DDI. Long-term or delayed complications from exposure to medicines can also be studied using routinely collected EHRs.

On the other hand, challenges in utilising EHR data for pharmacovigilance purposes might include the difficulty of performing data validation due to the anonymised nature of the system, the inability to follow up patients that change practices, the inability to retrieve data from linked external resources, or terminology inconsistencies among different data streams [47]. Also, free-text clinical narratives might contain valuable information which is not coded, thus requiring text mining methods [48]. Many reported medical events that are mentioned in an EHR can be linked to drugs but are irrelevant associations rather than AEs that represent potential ADRs.

In the United Kingdom, general practice (GP) electronic data resources are available, such as the Clinical Practice Research Datalink (CPRD) or The Health Improvement Network (THIN). Other examples in the US include the Sentinel Initiative and the Veteran's Affairs (VA) system [49, 50]. These resources have been used for answering research questions in pharmacovigilance and can be utilised alongside SRS data [51, 52]. For example, combining EHR with SRS data can support signal validation, with EHR data having supported the early detection of signals that were later validated using SRS data [53].

2.1.3.2 Administrative claims data

Administrative claims data refer to information that is collected by healthcare payers (e.g. insurance companies) and is used to reimburse healthcare providers for the services they provide to patients. This data stream has also been considered for surveillance activities [47].

The advantage of administrative claims databases is that they are capable of providing longitudinal follow-up of patients across more than one health system (i.e. hospitals, pharmacies, outpatient clinics, etc). However, they usually do not provide enough clinical details, at least not to the extent that EHR databases do. At the same time, several limitations should be taken into consideration when using this data stream for signal detection in pharmacovigilance. An important challenge is the presence of incomplete data that may not reflect the full extent of a patient's medical history or the complete list of medications taken. Additionally, administrative claims data are often generated by healthcare providers who may not have the necessary training or expertise to accurately code medical conditions and adverse events. This can result in coding errors that can impact data accuracy. Being primarily used for billing purposes, administrative claims data are also not designed to capture detailed clinical information, such as patient symptoms or the severity of AEs. For this reason, information on dose, frequency, or duration of medication use is missing in many cases.

Despite these limitations, administrative claims data remains a valuable resource for pharmacovigilance. However, it is important to use this data in combination with other sources of information, such as SRS, EHR or clinical trial data.

2.1.3.3 Biomedical literature

The biomedical literature is an important source of information for pharmacovigilance [48]. At the same time, it can be seen as challenging to use due to the unstructured nature of the data. In the last 20 years, there is an increasing availability of methodologies in the fields of text mining and Natural Language Processing (NLP) with astonishing capabilities that enable the utilisation of this novel data stream. Initial efforts to mine the literature for signal detection of novel ADRs mainly focused on biomedical text annotation for biomedical vocabulary concepts (e.g. UMLS concept text annotation with MetaMap [54, 55]), combined with rule-based approaches to defining relationships between the detected concepts (e.g. SemRep [56, 57]). Other approaches considered suitable Medical Subject Headings (MeSH) terms from the biomedical search engine PubMed (http://pubmed.gov) (e.g. 'chemically-induced'; 'adverse effects') to filter scientific papers containing drug-adverse event associations [58]. Other approaches have considered the extraction of causal relationships between drug entities and AE concepts from the title and abstract free-text using named entity recognition or syntactic pattern extraction [59].

Different machine learning (e.g. support vector machines), and, more recently, deep learning approaches (e.g. convolutional neural networks) have been considered for detecting sentences relevant to potential ADRs [60, 61]. The development of annotated benchmarks, such as the *ADE corpus* [62], can support the development and evaluation of these methods aiming to improve the extraction of drug-related adverse events from medical text. More recently, the development of more advanced methods to encode free text from the biomedical literature (e.g. word embeddings) have further contributed towards this direction [63, 64]. In the past few years, there has been significant progress in language models, with pre-trained language models (e.g. BioBERT [65]) playing a pivotal role and already having numerous applications in mining biomedical literature. Additionally, generative pre-trained transformers (GPT) [66] have primarily been utilised for language generation but show promise in other areas such as document categorisation and relation extraction specifically for biomedical text mining [67]. These advancements in language models offer exciting possibilities for various tasks in the field of NLP specifically in the context of biomedical literature mining for pharmacovigilance.

There are several key challenges of using biomedical literature for signal detection in pharmacovigilance [68]. First, identifying causal relationships (i.e. not just co-occurrence) of medications and AEs in free-text is an important limitation [69]. Drug names and AEs can be represented by multiple terms and appear uncoded in this data stream. The large volume of input data for processing introduces computational challenges, making it hard to consider full-text articles, with many methods relying on the extraction of terms from the title, abstract and related annotations. Also, the data quality in the biomedical literature can be variable and may not always reflect the most recent or accurate information in drug safety, hindering the support of signal detection efforts. Biased reporting of AEs can impact the accuracy of the information contained in

the biomedical literature. For example, studies with negative results may be less likely to be published. Finally, the availability of relevant data in the biomedical literature can be limited, particularly for rare or uncommon AEs.

Overall, it is clear that the biomedical literature in pharmacovigilance can be used to identify new or unexpected ADRs associated with medications, as well as to provide insights into the ADR mechanisms and risk factors. The biomedical literature can also be used to evaluate the evidence supporting the use of specific medications, including the strength of the evidence for their safety and effectiveness.

2.1.3.4 Social media and patient forum data

Social media and patient forum data have been identified as promising sources of information for pharmacovigilance [70, 71]. Social media platforms, such as Twitter, can provide real-time information on patient experiences and opinions related to medications, including AEs and other drug-related problems. Patient forums (e.g. PatientsLikeMe, Ask a Patient, HealthUnlocked) are online platforms where patients can discuss their health experiences, including the use of drugs and vaccines, and share information with others. These platforms can be used by patients to discuss their experiences with medicines and vaccines, including information on AEs and other drug- or vaccine-related problems. Hence, analysing these data sources has the potential to add value in the context of pharmacovigilance activities, as they may contain information that has not been reported through traditional SRSs or other data sources.

The Innovative Medicines Initiative (IMI) WEB-RADR project was launched in 2014 and aimed to explore the potential of using data from general social media (i.e. Twitter, Reddit, Facebook, Instagram) in pharmacovigilance [72]. The resulting recommendations of this project suggested not utilising this data stream for broad statistical signal detection activities [73]. This conclusion was based on a comparative analysis of SRS data and social media data and their relative ability to identify previously validated signals. The level of fairness of this analysis could be questioned, considering that it favoured the already established data stream based on the selected benchmark (i.e. validated signals that most likely arose from SRS data) and the poor performance of the applied methodology for AE recognition [74]. Also, the combination of both data streams (i.e. SRS and social media) was not considered in the analysis framework that was used for providing these recommendations. However, the project indicated the possibility of social media data adding value in specific niche areas (i.e. drug abuse and pregnancy-related outcomes).

There are multiple areas of opportunity in social media and patient forum data for pharmacovigilance purposes. This novel data stream is growing over time, considering the increasing number of Internet and social media users and the amount of information that they share on the Web. Also, there are areas that could be explored using these data streams, which could not be adequately captured and are often poorly reported in traditional SRS databases. These include, but are not limited to: more detailed

patient experiences following medication use; the off-label use of drugs; AEs experienced following the use of recreational drugs and herbal medicines (alone or in combination with prescribed medications); and the ability to capture data related to patient groups that would be less likely to report to traditional pharmacovigilance systems [74].

A clear challenge with this data source is again the need for text mining approaches that could also manage to infer causality and not simply capture the co-occurrence of terms [69]. The volume of this data stream entails computational efforts that also need to be considered, while extracting patient-reported AEs from unstructured text is challenging [74]. Furthermore, the quality of the data in these streams may not always reflect accurate or reliable information. For example, social media users may misinterpret the cause of an AE, or even provide false and misleading information. The scarcity of appropriate benchmarks that can be used for performance assessment can also impede the comprehensive evaluation and enhancements of the methods that consider social media and patient forum data. However, some recent efforts have resulted in manually generated reference sets that can be used to tackle this challenge [75]. Last but not least, ethical challenges related to the absence of explicit informed consent and the presence of potentially personally identifiable information, such as username, and geolocation, should be taken into account and properly addressed if possible [70].

Overall, both social media and patient forums are promising data streams for the future. By utilising appropriate text mining approaches and addressing the technical and ethical challenges, it will become possible to better understand the areas where they can provide added value and enable their real-world implementation.

2.1.3.5 Patient registries

Patient registries can be utilised in PMS for specific drugs, groups of drugs or diseases [76]. For example, the British Society of Rheumatology Biologics Register has been used to monitor long-term safety complications from the treatment of rheumatoid arthritis with biological agents, such as tumour necrosis factor (TNF) inhibitors. Furthermore, patient registries can be suitable for drug safety surveillance related to pharmacotherapy in orphan diseases.

2.1.3.6 Other novel data streams

The use of wearable devices, such as smartwatches, fitness trackers, and mobile health applications, can potentially play a role in PMS. Wearables can provide a continuous and objective source of real-time data on patient health status and drug exposure, which can be used to detect potential ADRs. Biomarker monitoring has been already applied in the clinical trial setting, but could also be considered for PMS considering the volume and variety of collected data on physical activity, heart rate, sleep patterns, and other health-related metrics [77, 78]. For example, a change in physical activity levels or heart rate could indicate, for example, a myocardial infarction that might be related to a drug. Additionally, wearables can also collect patient-reported data on drug exposure,

including information on drug dosing and timing. However, it is important to note that wearable data also have limitations, such as the potential for measurement error and bias, and they require advanced methods that can extract relevant information reliably.

2.1.3.7 Challenges and opportunities

The data streams that are covered in this section can be broadly categorised into three groups: (i) *established* (i.e. SRS databases, biomedical literature); (ii) *emerging* (i.e. EHRs, administrative claims data, patient registries), and (iii) *experimental* (i.e. data from social media, patient forums, wearable devices and apps) [74]. The concept of multimodality in pharmacovigilance describes the use of these diverse data sources to enhance the detection of novel ADRs and other drug-related problems [79]. In view of recent advances in the storage, retrieval, and analysis of distributed data that can enable federated analyses across high-dimensional and multi-source data, using these technologies in pharmacovigilance to improve patient safety is promising [80].

An increasing number of initiatives, such as the Observational Health Data Sciences and Informatics (OHDSI) programme, the IMI PROTECT project and EU-ADR, have attempted to tackle data heterogeneity. During the COVID-19 pandemic, the vaccine roll-out was inevitably followed by a large amount of generated social media data with relevant content, ranging from public sentiment [81] and vaccine hesitancy [82] to AE experiences following vaccination. However, vaccine surveillance has almost entirely been performed on traditional data streams [83]. Information from alternative streams could also be harvested by building appropriate frameworks and methodologies that can not only allow data harmonisation across disparate sources but also enable multimodal analysis while taking into account the challenges related to the different data streams.

The integration of multiple disparate data systems is apparently not straightforward. The increasing amount and complexity of collected data that further complicate the pharmacovigilance ecosystem also requires being mindful of the principle of parsimony [84]. Also, it is important to acknowledge that utilising the above-mentioned data sources for PMS purposes is considered secondary use of data (i.e. use for purposes other than the ones they were collected for), thus it entails apparent challenges given that these data sources might not be well suited to answer the study question. At the same time, while SRS databases remain an important source of information for PMS, complementary data sources are also needed. Understanding how these novel data streams could be utilised and combined with the established ones in the best possible way remains an open challenge.

2.1.4 Signal Detection

Signal detection in pharmacovigilance is only the first step following data collection towards a complex, time-consuming process that may end up in a drug label change or, in rare cases, even drug withdrawal. The term *signal detection* arises from electronic engineering, where there is a need to distinguish *real signals* (i.e. information-bearing

patterns) from accompanying random patterns in the background (called *noise*) that distract from the information.

The first definition [85] of a signal in the context of pharmacovigilance by WHO in 2002 is the following:

> *Reported information on a possible causal relationship between an adverse event and a drug, the relationship being previously unknown or incompletely documented. Usually, more than a single case report is required to generate a signal, depending on the seriousness of the event and the quality of the information.*

An updated definition [86] was provided by the Council for International Organisations of Medical Sciences (CIOMS) in 2010:

> *Information that arises from one or multiple sources (including observations and experiments), which suggests a new potentially causal association, or a new aspect of a known association, between an intervention and an event or set of related events, either adverse or beneficial, that is judged to be of sufficient likelihood to justify verificatory action.*

The evolution of the original definition is clear in several aspects: first, multiple sources are mentioned; beneficial events (apart from adverse ones) might also be considered in scope for signal detection activities; and, last but not least, moving from "*usually more than a single case report*" to "*association [...] judged to be of sufficient likelihood*" in light of the increasing magnitude and complexity of data. This last change also reflects the evolution of methods that have been used for signal detection from manual, case-by-case clinical expert review to statistical tools and quantitative signal detection (QSD). These methods emerged in the late 1990s and have turned into an indispensable part of the pharmacovigilance life cycle [87].

It is important to note that safety signals do not necessarily indicate a causal relationship between drugs and the occurrence of AEs. They instead involve cases that need further investigation to either be validated or rejected. Quantitative methodologies in signal detection aim to classify drug-AE combinations based on cases of co-reporting as either *signals* (which require further evaluation to be validated or refuted) or *noise* (i.e. combinations that should not trigger any further evaluation), with the aim of reducing the number of classifications characterised as false positives (i.e. calling noise a signal) and false negatives (i.e. calling a true signal noise). A perfect signal detection methodology would be able to perform a 100% accurate classification of the drug-AE combinations between the two possible classes. However, this is never the case and this challenging task is tackled by considering an adequate trade-off between sensitivity (i.e. the ability to correctly identify signals) and specificity (i.e. the ability to correctly identify non-signals) [88]. In the context of PMS, reduced specificity would result in more cases needed for manual clinical review, while reduced sensitivity would keep real signals unrevealed, thus prolonging their impact on patient safety.

		AE		
		Mentioned	**Not mentioned**	**Total**
Drug	**Mentioned**	a	b	$a+b$
	Not mentioned	c	d	$c+d$
	Total	$a+c$	$b+d$	$a+b+c+d$

TABLE 2.1: A 2-by-2 contingency table in the case of one drug and one AE.

2.1.4.1 Disproportionality analysis in SRS Data

Disproportionality analysis is the cornerstone of QSD in pharmacovigilance and measures the degree of *unexpectedness* in the reporting of a drug-AE association. In the simplest case, the observed reporting rate of one single drug with an AE is compared to the reporting of the same AE with other drugs in the same database. This *observed to expected* approach is suitable given the inability to readily calculate from the SRS data the rates of exposure to the drug (i.e. how many people took the drug) or the background incidence of the AE in the general population.

By considering one single drug and an AE, reports from an SRS database could be classified into four distinct categories using a case/non-case analysis: (i) reports that mention both the drug and the AE (a); (ii) reports that mention the drug but not the AE (b); (iii) reports that mention the AE but not the drug (c); and (iv) reports that mention neither the drug nor the AE (d). These report counts are usually placed in a 2-by-2 contingency table (**Table** 2.1).

There are various disproportionality measures (**Table** 2.2), which are broadly categorised based on their statistical approach into frequentist and Bayesian methods. Frequentist methods, such as the Proportional Reporting Ratio (PRR) [89] and the Reporting Odds Ratio (ROR) [90], were the first to appear in the literature and are still extensively used. For example, the PRR is used for routine screening of the Eudravigilance database. In terms of Bayesian methods, the Information Component (IC) was developed and is used to date as the preferred method for routine screening of the VigiBase at UMC. IC is a measure of the observed-to-expected ratio that is based on the Bayesian Confidence Propagation Neural Network (BCPNN) method [91]. Also, the Multi-Item Gamma Poisson Shrinker (MGPS) is an empirical Bayesian method that calculates the Empirical Bayes Geometric Mean (EBGM) measure [92, 93]. It has been implemented and is still used by the FDA.

Frequentist methods assume repeated random sampling of data and fixed model parameters, while Bayesian methods consider the data as fixed data and non-fixed (i.e. uncertain) parameters that are described by a probability distribution [94]. Frequentist methods can be computationally cheap for simple models and may be preferred in some cases because they provide a straightforward way to estimate statistical significance and confidence intervals. However, frequentist methods often require assumptions about the underlying distribution of the data, which can be difficult to specify correctly in practice. Bayesian methods, on the other hand, can be more computationally intensive. At the

Disproportionality measure	**Estimator**	**Probabilistic interpretation**
Relative Reporting (RR)	$\frac{a/(a+b)}{(a+c)/(a+b+c+d)}$	$\frac{P(AE\|Drug)}{P(AE)}$
Proportional Reporting Ratio (PRR)	$\frac{a/(a+b)}{c/(c+d)}$	$\frac{P(AE\|Drug)}{P(AE\|-Drug)}$
Reporting Odds Ratio (ROR)	$\frac{a/c}{b/d}$	$\frac{P(AE\|Drug)/P(-AE\|Drug)}{P(AE\|-Drug)/P(-AE\|-Drug)}$
Information Component (IC)	$\log_2 \frac{a/(a+b)}{(a+c)/(a+b+c+d)}$	$log_2 \frac{P(AE\|Drug)}{P(AE)}$

TABLE 2.2: Measures of disproportionality in pharmacovigilance for one drug-one AE: formulas and probabilistic interpretation.

same time, they can be more flexible in modelling complex data structures, provide more nuanced and probabilistic estimates of uncertainty, and allow for the incorporation of prior information.

In the context of disproportionality analysis using SRS data in PMS, small sample sizes, as are encountered with many drug-AE combinations with only a small number of reports, result in highly variable estimates of the reporting rate. Empirical Bayesian methods use prior distributions on these quantities that are estimated from the data and observe the effect that added information has on the posterior distributions [95]. However, the choice of prior can have a significant impact on the resulting posterior distributions and should be carefully considered since a poor choice of prior can lead to biased or inaccurate estimates.

Comparing the different available SDAs in a fair way is a challenging task. There have been previous efforts in the literature to perform comprehensive comparisons of SDAs by considering different SRS databases and multiple thresholds for the different SDAs [96]. The main conclusion was that the choice of disproportionality measure does not considerably influence the performance range of an SDA. For this reason, practical considerations should be prioritised, including how easily a method can be implemented, maintained and interpreted as well as computing resource requirements.

2.1.4.2 Method performance assessment

For classification tasks that consider the detection of signals based on the ability of a method to classify drug-AE (or drug combination-AE) cases into known associations (i.e. single-drug or DDI-related ADRs) (*positive cases*) and associations that are unlikely to be causally related (*negative cases*) using a specific discrimination threshold, the following counts are defined:

- True Positive (TP): the number of accurately classified positive cases;
- True Negative (TN): the number of accurately classified negative cases;

- False Positive (FP): the number of actual negative cases that were classified as positive;
- False Negative (FN): the number of actual positive cases that were classified as negative.

Based on the above counts, a confusion matrix (**Table 2.3**) is constructed for the specified threshold and the following evaluation metrics (which are frequently used also in the context of pharmacovigilance) can be calculated (with the formulas for the calculation of the above metrics being displayed in **Table 2.4**):

- **Sensitivity** (or **recall**): measures the proportion of actual positive cases which are correctly identified as such;
- **Specificity**: measures the proportion of actual negative cases which are correctly identified as such;
- **Precision** (or **positive predictive value (PPV)**): measures the proportion of accurately detected positive cases among all predicted positive cases;
- **Accuracy**: measures the proportion of correct predictions among all input cases;
- **F1-score**: the harmonic mean of the precision and recall.

By varying the discrimination threshold of the classification method using multiple threshold values and plotting *sensitivity* (y-axis) versus $1 - sensitivity$ (x-axis), a Receiver Operating Characteristic (ROC) curve is created and depicts the relative trade-offs between TPs and FPs. A perfect classification method (i.e. 100% sensitivity and 100% specificity) would give a point in the upper left corner, while a random guess would yield a point along a diagonal line from the bottom left (0,0) to the top right (1,1) corners.

The two-dimensional area underneath the entire ROC curve from (0,0) to (1,1), called the Area Under the ROC Curve (AUC), is commonly used as an aggregate evaluation metric of the classification method performance at multiple discrimination thresholds. AUC values range from 0 to 1, with a perfect classification method having an AUC value of 1 and a random one an AUC value of 0.5. The AUC values are often used for method comparison.

		Predicted Class	
		Positive	**Negative**
Actual Class	**Positive**	TP	FN
	Negative	FP	TN

TABLE 2.3: A confusion matrix.

Evaluation metric	**Formula**
Sensitivity	$sens = \frac{TP}{TP+FN}$
Specificity	$spec = \frac{TN}{TN+FP}$
Precision	$prec = \frac{TP}{TP+FP}$
Accuracy	$acc = \frac{TP+TN}{TP+TN+FP+FN}$
F1-score	$F_1 = 2 \times \frac{prec \times sens}{prec+sens} = \frac{2TP}{2TP+FP+FN}$

TABLE 2.4: Formulas of evaluation metrics for binary classification methods.

2.1.4.3 Considerations for practical implementation

The real-world implementation of SDAs requires a very good understanding of the methods before these can be fine-tuned and used for safety surveillance. The first step following the development of an SDA is the evaluation of its absolute, as well as relative (i.e. compared to existing SDAs), performance. This is a crucial step considering the lack of ground truth in this setting (i.e. definitive positive and negative controls), Also, it is difficult to establish suitable reference standards and to assess whether and to which extent the demonstrated performance reflects the real-world performance of the methodology. These challenges have been widely described in the literature [97, 98, 99]. Multiple frameworks (e.g. prospective versus retrospective evaluation) and types of reference sets have been explored to enable a better evaluation of SDA performance [7, 8, 10, 9].

Even once the SDA performance has been quantified using one or multiple reference standards, the trade-off between sensitivity and specificity for real-life implementation still needs to be defined before selecting an appropriate operating point. Routine SRS screening with an SDA should be set based on a careful consideration of the potential implications of both the false negative rate as well the false positive rate of an SDA. For an acceptable false negative rate, the severity and patient impact of a missed signal should be determined. The inability of an SDA to capture signals of serious complications of medicines (i.e. high-risk signals) could have deleterious consequences for patient safety. Prevalence of use for a specific medication is another valid point for consideration, as larger groups of people are exposed to a potential risk in the case of missed signals related to medications that are administered to a large population, Also, the nature and vulnerability of the population that is exposed to a drug needs to be taken into account (e.g. pregnant women, children) to define an appropriate false negative rate. Considering, on the other hand, an acceptable false positive rate, we need to factor in the available resources that can evaluate and perform a clinical review of the signals that are generated by a database screening using an SDA. Also, too many signals arising from an SDA might lead to reviewers being overwhelmed and a lack of trust

in the methodology assuming many of them are eventually refuted. In the context of signal detection in routine pharmacovigilance, a high false negative rate could be more problematic compared to a high false positive rate.

Another important element is to understand whether the SDA performance is commensurate to the one quantified using a reference set. The real-world performance of an SDA that is based on disproportionality analysis can also be affected by the background incidence (i.e. the expected rate of occurrence) of an AE. More specifically, it can be more challenging to unveil signals related to AEs with high background incidence rates using SRS data [8]. These AEs can be absent or insufficiently represented in the reference set that was used during SDA performance evaluation. The masking effect and drug competition bias can also impact the ability of an SDA to capture signals [100, 101]. To address these issues, it is possible to apply techniques such as the amendment of the background by removing known signals before calculating the disproportionality measures.

Apart from the applied methodology to an SRS database for signal detection, another clear challenge in PMS is inherent to the nature of the SRS data and the data collection process that is performed via spontaneous reporting. Under-reporting is a well-known phenomenon in pharmacovigilance [43]. Data can be: *missing completely at random*; *missing at random but dependent on specific covariates*; or *missing not at random* (i.e. informatively missing). The latter category of missing data can impact signal detection activities [102]. Also, the observed reporting rates of drug-AE associations in SRS databases depend on multiple factors, including the severity of the AE, how long the drug has been in the market [103], and any publicity of drug-related safety concerns [104]. Another challenge in the implementation of SDAs is related to how pharmacovigilance data are coded to standard terminologies (e.g. MedDRA for AEs). Variations and inconsistencies in the terms that are used to map the data have the potential to affect the signal generation process using an SDA. Finally, confounding is another factor that needs to be considered, such as confounding by indication and confounding by co-medication [105].

Finally, it is important to keep in mind that the signals identified by quantitative signal detection methods are in principle signals of disproportionate reporting and are not automatically considered safety signals [88]. There are also cases where safety signals are triggered via qualitative methods. For example, signals arising from case reports that involve rare and clinically severe complications, such as an AE from the EMA's Designated Medical Events list, are examples where urgent clinical review is needed to identify whether it can be linked to medication exposure.

2.1.5 Signal Evaluation

Once a safety signal is detected, it needs to be evaluated by clinical specialists to be either validated or rejected based on whether a causal association between the drug and the adverse event can be established. There are multiple structured approaches and

frameworks for causality assessment. The Bradford-Hill criteria were described back in 1965 for causality assessment in population studies [106] and consist of the following:

1. **Strength** (as determined by a suitable statistical analysis);
2. **Consistency** (through repeated observation of the association between the drug and the manifestation of the AE);
3. **Specificity** (through isolation of the cause to a single outcome);
4. **Temporality** (where the drug exposure precedes the AE);
5. **Biological gradient** (where the dose is positively associated with the response);
6. **Plausibility** (through linking the AE to drug exposure via a reasonable pathway from the existing biological knowledge);
7. **Coherence** (where the evidence does not critically contradicts the known natural history and biology of the disease);
8. **Experiment** (through existing experimental evidence that supports the association);
9. **Analogy** (where, in the case of pharmacoepidemiology, the effect is also associated with other drugs from the same pharmacological class).

In the context of pharmacovigilance, the main criteria for causality assessment include the following: (i) the strength of the association (i.e. the magnitude of the SDR); (ii) data consistency; (iii) exposure-response relationship; (iv) biological plausibility; (v) experimental findings (e.g. rechallenge, diagnostic markers); (vi) analogy (i.e. previous experience with related drugs or event known to frequently be drug-induced); and (vii) the nature and quality of the data [107, 108].

Other algorithmic approaches include the categorisation using the WHO-UMC causality system [109], the FDA's causality algorithm [110], the Naranjo algorithm [111], and the Liverpool ADR Causality Assessment Tool [112].

Regardless of the selected framework, signal evaluation is a challenging process and a definitive judgement on causality is difficult, if not impossible, to make. The in-depth clinical review of reports is supported by drug target information, literature search, and other types of evidence that are deemed sufficient to validate or refute a signal. Apart from causality (i.e. clinical plausibility), other considerations in the context of signal evaluation include frequency, clinical implications, and preventability [113]. Given the advances in Information Representation and the increasing availability of information via computerised systems, it is expected that additional automated filtering layers could be added between signal detection and signal evaluation steps. These filters can limit the amount of human effort needed for in-depth clinical review by trying to support some causality aspects such as biological plausibility using, for example, primary and

secondary drug target information. Additional information can also be collected, for example by performing analyses using complementary data sources or even conducting a post-authorisation safety study. A pharmacoepidemiological study that utilises routinely collected data (e.g. EHR data), is a reasonable approach. In the last decade, efforts including the OHDSI project and the FDA Sentinel Initiative aim to combine disparate data sources. The increased sample sizes and generalisability through the inclusion of heterogeneous populations are promising building blocks for performing pharmacoepidemiological studies. Cohort studies are possible, but patient recruitment and follow-up can be both challenging and costly. Disease registries can also be utilised [114].

Finally, if a safety signal is validated, this will result in regulatory action. This can be a drug label change (i.e. an update on the Summary of Product Characteristics and Patient Information Leaflet regarding precautions and contraindications), a referral procedure or, in some cases, even urgent safety restrictions such as market withdrawal.

2.1.6 Summary

Pharmacovigilance is the practice of monitoring, detecting, assessing, and preventing adverse effects or any other drug-related problems to ensure the safe and effective use of medications, enabling the detection of previously unknown drug complications. Alongside these databases, various other data sources, including electronic health records, administrative claims data, biomedical literature, social media, patient registries, and emerging novel data streams, have become increasingly important in supporting post-marketing surveillance. However, utilising and combining these databases comes with challenges that must be carefully considered. Signal detection methods, such as disproportionality analysis in spontaneous reporting system data, form the foundation of quantitative analysis in pharmacovigilance. The practical implementation of these methods should be preceded by their performance evaluation using suitable techniques and should also take into account multiple considerations, including the nature of the analysed data sources and the fact that association does not necessarily infer causation. Finally, signal evaluation is a critical step in pharmacovigilance, requiring a comprehensive examination of data and supporting evidence from multiple perspectives.

2.2 Drug-Drug Interactions

The previous section covered the basic concepts and processes in pharmacovigilance. This section presents DDIs from a pharmacological aspect, their clinical significance, and the currently available PMS methodologies for the identification of novel DDIs once drugs enter the market.

2.2.1 Definition

In real life, patients are often treated with multiple drugs at a time. Each drug molecule can follow several biological pathways within the human body before its elimination. A DDI is a clinically meaningful modification of the effect of a drug (*object drug*) caused by the presence of another drug (*precipitant drug*) [115]. Potential DDIs are assumed in the case of many drug combinations, meaning that the drugs are predicted or known to interact when they are concomitantly prescribed to a patient, regardless of whether harm ensues. For a specific drug, the volume of potential DDIs is far greater compared to the actual, clinically observable DDIs. The occurrence of a DDI with a clinically observable effect depends on multiple factors that are linked to either the drug administration conditions or the patient characteristics and profile [116].

A DDI can lead to the modification of the therapeutic effects of one or both interacting drugs. An increase in a drug's effect can potentially lead to an ADR, while a decrease in its effect can result in a lack of efficacy [117]. In some cases, DDIs are desired [118] (e.g. synergistic effects of concomitant drugs, where the overall effect of interacting drugs is greater than the sum of individual ones), but they can have severe implications for patient safety.

Apart from DDIs, drugs also interact with food, drink, nutrients (e.g. vitamins), herbal products, drug formulations (e.g. excipients) or environmental chemical agents [116].

2.2.2 Classification

DDIs can be grouped into two main categories: (i) pharmacokinetic DDIs; and (ii) pharmacodynamic DDIs [117].

2.2.2.1 Pharmacokinetic DDIs

Pharmacokinetic DDIs occur when one drug affects the absorption, distribution, metabolism, or elimination of another drug, thus causing alterations in its effective concentration. The drug that causes the interaction (*perpetrator*) alters the pharmacokinetic profile (i.e. the way the human body behaves towards a drug) of the drug that is affected by the interaction (*victim).* These interactions can lead to changes in the blood levels of the victim drug, which can affect its efficacy or toxicity [119]. Metabolism-based and transporter-based DDIs are two important and well-studied categories of pharmacokinetic DDIs with clinically observable effects.

Metabolism-based DDIs occur when one drug affects the metabolism of another drug by inhibiting or inducing metabolic enzymes in the liver (i.e. the primary site for drug metabolism) or other tissues. Drugs that act as inducers or inhibitors of these enzymes can affect the concentrations of the victim drugs metabolised by the same pathway. Metabolic inhibition causes accumulation of the victim drug, while induction results in decreased victim drug concentrations. Depending on whether the metabolite of the victim drug is active or inactive, as well as the toxicity of both the victim drug and its metabolite, the clinical outcome of a metabolism-based DDI can vary and lead to either elevated or decreased drug exposure.

One of the most common types of metabolism-based DDIs involves the cytochrome P450 (CYP) enzymes, which are responsible for the metabolism of many drugs and other xenobiotics (i.e. chemical substances not naturally produced or expected to be present within the organism). Suppose a patient is taking a drug that inhibits a particular CYP enzyme. In that case, it can slow down the metabolism of another drug that is also metabolised by that enzyme, leading to an increase in the blood levels of the victim drug. In cases where the victim drug is not converted to its active form via this metabolic pathway, its increased concentrations can augment the risk of adverse effects and toxicity. On the other hand, a drug that induces the activity of a CYP enzyme can speed up the metabolism of another drug that is also metabolised by that enzyme, leading to a decrease in the blood levels of the victim drug. This can reduce the efficacy of the victim drug and result in subtherapeutic drug levels and treatment failure if the parent form of the victim drug is the active one [120]. Five CYP family enzymes, namely CYP3A4, CYP2D6, CYP2C19, CYP2C9 and CYP1A2, are implicated in most prescribed drugs' metabolism. Genetic variations of CYP enzymes do exist among patients, causing differences in drug responses [121, 122].

Transporter-based DDIs occur when one drug affects the transport of another drug across cell membranes by interacting with drug transporters. Drug transporters are membrane proteins that can be found in most tissues of the human body, including the intestine, liver, kidney, lung, and the blood-brain barrier. Transporters, apart from serving various endogenous functions (e.g. transport of peptides, amino acids, etc), also act as important players in the absorption, distribution, and elimination of both parent drugs and their metabolites through either uptake or efflux of these molecules. Hence, tissue uptake and efflux processes in drug dispositioncan be inhibited or induced by perpetrators, changing in this way the victim's pharmacological action. Several factors, such as the relative substrate specificity and the distribution of drug transporters in the different tissues, determine the overall effect of transporter-based DDIs, making it difficult to predict [119].

P-glycoprotein (P-gp) is the most well-studied and understood drug transporter. It can be found in the intestine, liver, brain, and other epithelial tissues. Acting as a transmembrane efflux pump, it is responsible for the excretion of its substrates, which include many drugs and other xenobiotics, from inside to outside the cell. Numerous

drugs are P-gp inhibitors or inducers and account for interactions with other P-gp substrates [119]. Other transporters that can be involved in transporter-based DDIs include the breast cancer resistance protein (BCRP), organic anion-transporting polypeptides (OATPs) (e.g. OATP1B1 and OATP1B3), organic cation transporters (OCTs) (e.g. OCT2), and multidrug resistance-associated proteins (MRPs) [123].

2.2.2.2 Pharmacodynamic DDIs

Pharmacodynamic DDIs involve a direct influence on the effects of interacting drugs, by either enhancing or diminishing the effects of one or both drugs. When the effect of two drugs is greater than the sum of their individual effects, it is called a synergistic effect. On the other hand, when the effect of one drug is reduced or inhibited by another drug, it is called an antagonistic effect and can result in decreased therapeutic efficacy.

As an example of a pharmacodynamic DDI, the concomitant use of aspirin and warfarin increases the risk of bleeding. Aspirin is an antiplatelet drug with a direct hypoprothrombinaemic effect, while warfarin is an oral anticoagulant drug. When taken together, these two drugs act in a synergistic way as blood thinners through different mechanisms[117]. Aspirin blocks the production of thromboxane A2, which is a platelet activator, while warfarin inhibits the synthesis of vitamin K-dependent clotting factors in the liver. Therefore, the risk of bleeding may be significantly increased due to the additive effects of the two drugs.

In some cases, synergistic drug effects can be desired, such as the combination of ampicillin and gentamicin in the treatment of infective endocarditis [116].

2.2.3 Clinical Significance

Polypharmacy rates are rising in recent years, which leads to increasing numbers of potential interactions between administered drugs. According to a review from the United Kingdom's Department of Health and Social Care in 2021, over 1 out of 3 people in England aged 60 and older are exposed to at least 5 medicines at the same time, with more than a third of all people above 80 being on 8 or more medicines [124]. However, while a DDI can occur based on pharmacological knowledge, these potential DDIs far outnumber those leading to a clinically observable effect [125]. The clinical manifestation of a DDI can be either an ADR or a lack of efficacy for at least one of the interacting drugs. The prevention, identification, and management of DDIs in the clinic are not straightforward. In recent years, computerised systems are increasingly used to provide more detailed information on patient risk factors that can determine the occurrence of a DDI, the severity and evidence associated with a particular DDI, as well as management guidelines in cases when a DDI occurs.

There are several factors that can affect the actual behaviour of drug molecules inside the human body, making it difficult to predict the occurrence of a clinically significant DDI. They can be broadly categorised into drug-related and patient-related factors.

In terms of the drug-related factors that specify the clinical manifestation of a DDI, drug metabolism can play an important role. In some cases, a biologically inactive compound known as a prodrug is metabolised in the body to produce an active moiety, with this metabolite being involved in interactions with concomitant drugs, rather than the parent drug molecule. Therefore, the characterization of metabolic pathways through preclinical experimental models is particularly important. Also, for drugs that compete for the same binding site of a metabolic enzyme in the case of pharmacokinetic DDIs, the relative strength of their binding interaction with the enzyme (i.e. binding affinity) determines the presence and severity of a DDI [119, 126]. The therapeutic index is another important factor that determines whether a DDI will have a clinically significant effect. The therapeutic index is a ratio used to compare the blood levels at which a drug causes a therapeutic effect to the ones that cause toxicity. If a drug has a narrow therapeutic index, then there are small differences between its therapeutic and toxic dose, increasing the risk for serious DDIs [127, 126].

Patient-related factors also need to be considered to determine the clinical significance of a DDI at the individual level. Patient characteristics (e.g, demographics, diet, lifestyle), underlying diseases and polypharmacy are all important aspects that define how likely it is for a patient to experience a clinically significant DDI [116]. Genetic polymorphisms of drug-metabolising enzymes, drug transporters and drug receptors also account for the appearance of some DDIs in specific genetic subpopulations [119]. For example, the activity of a metabolic enzyme is determined by the pairing of individual alleles a person has inherited from their parents. Alleles are grouped into wild-type (WT) and mutant alleles. WT alleles are the predominant ones in the general population and are most times related to normal rates of metabolism. Mutant alleles can cause decreased activity or even complete inactivity of metabolic enzymes. There are four categories of metabolisers, based on the type of alleles they carry: *poor*, *intermediate*, *extensive* and *ultra-rapid* [120]. These genetically determined variations in drug response (e.g. differences in efficacy, ADR occurrence, dose specification, etc.) are studied in pharmacogenetics, serving as important factors for a better understanding of the patient subgroups with a higher risk for DDIs and implementing personalised medicine [128].

2.2.4 Drug-Drug Interaction Surveillance

Considering the complex mechanisms leading to a DDI and the number of factors that determine their manifestation in the clinical setting, it is easy to understand that detecting and validating a novel DDI is a challenging task. At the same time, clinicians should be able to access and utilise the latest evidence on the risks associated with the concomitant use of drugs that can lead to a clinically significant DDI. In view of the harmful consequences that DDIs can cause if they remain undetected, it is clear that pharmacovigilance activities should adequately support the identification of complications arising from DDIs to ensure increased patient safety.

The efforts to understand the safety profile of a candidate drug, including its potential to interact with other drugs, begin at the early stages of drug development. There has been a growing interest in using computational methods, such as in silico modelling and machine learning methods, to identify potential DDIs based on, for example, the chemical characteristics, target information, biological activity, and the pharmacokinetic profile of a drug [129, 130, 131]. In some cases, it is relatively straightforward to flag potential DDIs linked to a drug. An illustrative example includes pharmacokinetic DDIs that are related to metabolic enzyme activity. However, it is hard to determine the clinical significance of the predicted DDIs without having access to in vivo data. Apart from animal models and non-clinical safety experimental setups to define a drug's safety profile in the pre-clinical setting, clinical trials are also used for assessing both the safety and efficacy of a candidate drug. However, eligibility criteria for patient recruitment in clinical trials often limit the representation of patient groups more predisposed to DDIs and hinder their identification at this stage. Therefore, the detection of a large proportion of DDIs is performed following drug authorisation and PMS can play a critical role in this process.

Both the collection of adequate data and the development of methodologies are important pillars that can support PMS activities for identifying novel DDIs. In terms of available data, the resources that were previously presented for PMS in the case of single drugs are relevant and important. The sheer volume of reported drug combinations in SRS databases generates an apparent requirement for QSD methods that could filter and prioritise combinations of drugs and AEs for causality assessment. However, quantitative methods for DDI surveillance in pharmacovigilance are not as mature as those for single drugs. In terms of their chosen computational approach, they can be broadly categorised into *prediction* and *detection*, but without a clear-cut distinction between the two categories either in theory or in practice. Prediction methods focus on biological and chemical knowledge bases, drug development, and identifying beneficial interactions, whereas detection methods are mainly statistical and often based on disproportionality analysis and more conventional data sources to reveal previously unknown harmful DDIs between authorised drugs [132].

The disproportionality analysis is adapted to enable DDI surveillance using SRS data by measuring the degree of 'unexpectedness' of associations between drug combinations and AEs. The simplest scenario of two drugs and one AE considers a contingency table with four rows and two columns (**Table 2.5**). Multiple SDAs for DDI surveillance have been described in the literature [133]. One of the first efforts uses the ROR statistic with logistic regression to model two-way DDI effects along with the individual drug effects and demographic predictors (i.e. age and gender) [134, 135]. Another approach assumes a binomial distribution for the baseline model (i.e. absence of interaction), with the contributions of the individual effect of each drug to the overall AE occurrence being either additive or multiplicative. The departure from the baseline model is quantified using log-linear regressions [11]. An empirical Bayes methodological framework based

		AE		
		Mentioned	Not mentioned	**Total**
Drugs	Drug A = Yes, Drug B = Yes	n_{111}	n_{110}	$n_{11.}$
	Drug A = Yes, Drug B = No	n_{101}	n_{100}	$n_{10.}$
	Drug A = No, Drug B = Yes	n_{011}	n_{010}	$n_{01.}$
	Drug A = No, Drug B = No	n_{001}	n_{000}	$n_{00.}$
	Total	$n_{..1}$	$n_{..0}$	$n_{...}$

TABLE 2.5: A 4-by-2 contingency table in the case of two drugs and one AE.

on the MGPS method is the Interaction Signal Score (IntsSS). The detection of two-way DDI signals with IntSS is based on the EBGM values of a drug combination along with the following conditions: (i) the lower bound of the 90% confidence interval (CI) estimate (EB05 score) of the drug combination being over 1; and (ii) the magnitude of the drug combination's EB05 score being higher than the upper bound of the 90% CI estimate (EB95 score) of each individual drug [13]. Last but not least, the Omega (Ω) approach was developed by the WHO UMC and is based on a shrinkage measure, being an extension of the IC method for two-way DDI surveillance [12].

The statistical measures that were mentioned above do not allow us to directly infer whether, and which of, the individual drugs give rise to unexpected reporting on their own compared to their expected background rate. However, we would expect that modelling a pharmacokinetic DDI related to an AE, for example, would involve disproportionate reporting rates for the victim drug and the drug combination, while the respective reporting rate for the perpetrator drug would not differ from the AE background rate. On the other hand, a pharmacodynamic interaction would involve unexpected reporting rates for both individual drugs and a departure from the baseline additive model when the two drugs are concomitantly used.

Apart from purely statistical modelling approaches for DDI PMS, custom machine learning methods have also been proposed to detect signals of novel DDIs using SRS data in combination with other data resources (i.e. EHR, chemical structure data, drug target data).

Tatonetti et al. developed a method that utilised AE profiles of individual drugs to deduce the existence of unknown AEs related to DDIs [53]. Their work considered FAERS reports with up to two drugs, manually built phenotype classes for AEs by grouping MedDRA PT terms, and fitted logistic regression models to single drug-AE frequency values for model training. The trained model on single drug-AE data was then applied to drug pair-AE FAERS data to predict putative DDIs. EHR data were also used to screen for the putative DDIs originally identified in FAERS. Despite the potential applicability of this method for capturing patterns in the data indicative of a DDI signal, it cannot be considered flexible as it can only be applied to detect DDI signals related to a specific number of pre-defined AE classes. It also filters only reports from SRS data that contain up to two drugs and only drug pairs that appear in at least

5 reports, thus limiting both the available data to be used for training and the possible drug pairs for screening.

The aim of integrating biological plausibility aspects for novel DDI prediction has also led to the development of computational approaches that take into account diverse drug-related data resources. As an example, INDI is a method that aimed to predict novel DDIs using multiple data types, including chemical similarity, drug targets, and drug classes [136]. More recently, deep learning approaches have emerged to support the prediction of novel DDIs. A convolutional neural network approach, for example, models DDIs by considering protein-protein interactions, drug-protein target interactions, and ADRs resulting from polypharmacy [16]. Interestingly, the method performed particularly well with predictions of polypharmacy ADRs with a strong molecular basis.

It is noteworthy that the training of machine learning algorithms requires large datasets. Also, independent datasets should be used for prediction to avoid overfitting, where the algorithm performs well on the training data but poorly on new, unseen data. The large volume of routinely collected EHR data offers an opportunity towards the development of novel machine learning methods with real-world applicability.

2.2.5 Natural Language Processing for Extraction and Identification of Drug-Drug Interactions

The creation of knowledge databases extracted from the biomedical literature and the identification of potential novel DDIs in free text as two important tasks that have been investigated and multiple NLP and information retrieval techniques have been developed [137].

In terms of DDI extraction from biomedical literature, methods have approached DDI extraction as a classification task to distinguish between DDIs and non-DDIs. The three main systems for relationship extraction from free text include: (i) co-occurrence-based; (ii) rule-based (i.e., using linguistic rules to understand the meaning of a certain relationship); and (iii) machine learning approaches. As the availability of unstructured biomedical text grows exponentially, machine learning methods can play a pivotal role in data extraction tasks. The DDIExtraction 2013 task [138] played a crucial role in this domain, as it provided an annotated DDI corpus containing over 5000 DDIs [139]. This initiative led to the development of various DDI extraction methods, including long short-term memory (LSTM) network models [140], feature-based approaches [141], convolutional neural network (CNN)-based techniques [142, 143], recursive neural networks [144], and graph convolutional networks [145]. More recent efforts leveraging this resource have ventured into integrating heterogeneous pharmaceutical knowledge graph information, including molecular structure information, with transformer model architectures utilising pre-trained weights from BioBERT [146].

Beyond extracting DDIs from literature for knowledge databases, data mining approaches are applied to detect potential novel DDIs that pose risks to patients. This involves the extraction of DDIs from clinical notes [147], predicting novel DDIs based on

gene-drug relationships in the literature [148], and incorporation of molecular structure and target information into text data models [149]. There are also ongoing efforts related to social media and online healthcare forums for identifying novel DDIs, harnessing the experiences of patients and healthcare providers [150, 151].

Chapter 3

A similarity and consistency assessment of three online drug-drug interaction resources

The content of this chapter is published in *The British Journal of Clinical Pharmacology*.

This chapter aims to understand the existing evidence relevant to clinically-relevant drug-drug interactions (DDIs) in three well-known online clinical resources. The following types of information are explored: inclusion, severity rating, evidence rating, and clinical management recommendations.

The key contributions of this chapter are two-fold. From a clinical standpoint, the analysis is performed on complete and clinically-relevant drug information resources, thus allowing us to understand the level of variability in the information that can also lead to a clinical impact. From a methodological standpoint, it already defines the framework for web data extraction and data normalisation of the different data sources, which is utilised in **Chapter 4** to derive the DDI reference set.

Summary

Aim

To explore the level of agreement on drug-drug interaction (DDI) information listed in three major online drug information resources (DIRs) in terms of: (1) interacting drug pairs; (2) severity rating; (3) evidence rating and (4) clinical management recommendations.

Methods

We extracted information from the British National Formulary (BNF), Thesaurus, and Micromedex. Following drug name normalisation, we estimated the overlap of the DIRs in terms of DDI. We annotated clinical management recommendations either manually, where possible, or through the application of a machine learning algorithm.

Results

The DIRs contained 51,481 (BNF), 38,037 (Thesaurus), and 65,446 (Micromedex) drug pairs involved in DDIs. The number of common DDIs across the three DIRs was 6,970 (13.54% of BNF, 18.32% of Thesaurus, and 10.65% of Micromedex). Micromedex and Thesaurus overall showed higher levels of similarity in their severity ratings, while the BNF agreed more with Micromedex on the critical severity ratings and with Thesaurus on the least significant ones. Evidence-rating agreement between BNF and Micromedex was generally poor. Variation in clinical management recommendations was also identified, with some categories (i.e. Monitor and Adjust dose) showing higher levels of agreement compared to others (i.e. Use with caution, Wash-out, Modify administration).

Conclusions

There is considerable variation in the DDIs included in the examined DIRs, together with variability in categorisation of severity and clinical advice given. DDIs labelled as critical were more likely to appear in multiple DIRs. Such variability in information could have deleterious consequences for patient safety, and there is a need for harmonisation and standardisation.

What is already known about this subject

- There are a variety of online DIRs, which differ in coverage, content, and inclusion criteria, that are available to clinicians and other prescribers, mainly for prescribing decision support purposes.

- Previous studies have described major discrepancies between widely used DIRs on the inclusion of critical DDIs or interactions of specific therapeutic categories, along with discordance in their severity and evidence ratings.

What this study adds

- To the best of our knowledge, this is the first study to concurrently compare the similarity among complete datasets from DIRs in terms of inclusion of drug pairs, recommendations for clinical management, severity, and evidence of DDIs.
- Considerable variation was identified in all types of information for DDIs, which has important clinical implications for patient safety and requires efforts towards harmonisation and standardisation.

3.1 Introduction

Coadministration of multiple drugs increases the risk of drug-drug interactions (DDIs). A DDI can be defined as the modification in the therapeutic effect of one or more medications due to the presence of concomitant medications and can lead to clinically significant events, caused by either an increase in the effect of the interacting drug leading to an adverse drug reaction (ADR), or a decrease in its effect that results in lack of efficacy. Previous studies have reported that DDIs are a significant cause of hospitalisation, being responsible for 16.6% of cases where the cause was an ADR and around 1% of all hospital admissions [5]. The risk for DDIs increases during hospitalisation and after discharge, as there is a high prevalence of administration of potentially interacting drug combinations [152].
As new medicines gain approval each year, the volume of possible drug combinations is constantly growing. At the same time, the rising numbers of people with multimorbidity together with increasing life expectancy around the world are associated with the phenomenon of polypharmacy, which aggravates the impact of DDIs in clinical practice. According to a recent review, over 1 out of 3 people in England aged 60 and older are exposed to at least 5 medicines at the same time, with more than a third of all people above 80 being on 8 or more medicines [124].

The clinical manifestation of DDIs depends on several factors. Potential DDIs, based on pharmacological knowledge, far outnumber those which lead to clinically significant adverse effects [153]. Despite the theoretical potential for an ADR to occur due to a DDI, there are several factors that can affect the actual behaviour of drug molecules inside the human body, including dosage and patient characteristics (e.g. age, number and type of morbidities, etc). Also, genetic polymorphisms of drug-metabolising enzymes, drug transporters or drug receptors may be responsible for the appearance of some DDIs [119]. Therefore, it is difficult to accurately predict the occurrence of a clinically significant DDI in an individual patient. To overcome this problem, clinicians are commonly aided by drug information resources (DIRs) to assess the risk-benefit ratio of each drug added

to the treatment schedule. DIRs can be either open source or commercial, and they are often incorporated in computerised clinical decision support (CDS) tools.

The availability of DIR information related to severity, evidence availability, and clinical options for the management of DDIs (e.g. entirely avoid the combination, monitor, adjust dose, etc) are central to the development of CDS [154]. Inconsistencies between DIRs may confuse clinicians and impact clinical decisions [155]. Previous studies have assessed the level of agreement of DIRs, mainly in terms of listing of DDIs and severity ratings. However, most of them were restricted to only DDI listing for a limited number of drugs, specific therapeutic categories, or did not focus solely on clinical resources [156, 157, 158]. Moreover, the ability of a DIR to identify clinically relevant DDIs or capture critical DDIs (e.g. FDA black box warnings, ONC high priority list [159]) has also been explored [160, 161, 162]. However, it remains unclear to what extent DIRs from different geographic locations agree on their DDI listings as well as DDI-related information.

The aim of this study was to assess the concordance of leading clinical resources for DDIs from three different countries of origin in terms of: (1) inclusion of interacting drug pairs; (2) severity rating; (3) evidence rating and (4) clinical management recommendations. To the best of our knowledge, this is the first comprehensive study that attempts to compare multiple types of information pertinent to DDIs at the same time across entire DIRs. To ensure clinical utility, only clinically relevant resources were included in the present study (see **Figure A.1** for an overview of online DDI resources). Data sources of potential DDIs (e.g. DrugBank) that are mainly used for scientific research purposes were not taken into consideration.

3.2 Methods

3.2.1 Data sources

DDI data from two open-source and one commercial online DIRs were included in our evaluation: the British National Formulary [163] (hereafter called BNF), Interactions Thesaurus [164] by the French Medicines Agency (Agence Nationale de Sécurité du Médicament et des produits de santé, ANSM) (hereafter called Thesaurus) and IBM Micromedex [165] (hereafter called Micromedex). The BNF is extensively used in the United Kingdom [166, 167]. Thesaurus is maintained and updated annually by ANSM, being considered as the official source of information relevant to DDIs for French clinicians. Micromedex is a leading clinical information resource, listed as one of the statutorily named compendia in the Medicaid program and is widely used in the United States (US) [168, 169]. The BNF and Thesaurus are publicly available online, while Micromedex can only be accessed via subscription.

3.2.2 Data extraction

Automated web data collection (web scraping) was executed for BNF and Micromedex in Python 3.6 [170] with terms of use that permit data collection. Thesaurus is a Portable Document Format (PDF) file that is curated and updated annually. An R package (IMthesaurusANSM [171]) enabled the automatic data extraction from the original document (version September 2019). The types of extracted information from each DIR are summarised in **Table 3.1**.

We mapped DDIs from Thesaurus at the drug class level (e.g. beta blockers) to their constituent individual drug ingredients using a mapping table available in the ANSM website. We also excluded DDIs from Micromedex containing drug combinations (e.g. hydroxyamphetamine/tropicamide), as those simply collated DDIs from the combination's individual ingredients; hence, only single ingredient drug interactions were considered. Also, cases where drug names of an interacting pair were swapped (i.e. (D1, D2) and (D2, D1)) were considered equivalent and duplicate entries were removed from the tables that stored the extracted data (BNF original table, Thesaurus original table and Micromedex original table).

3.2.3 Drug name normalisation

Initial drug names were normalised to RxNorm Ingredients (for US-marketed medicines) [172] and RxNorm Extension Ingredients (for medicines not found in RxNorm) [173] using the OHDSI Usagi tool [174]. Some names were too general to be mapped (e.g. insulins) or were not present in either vocabulary. Thus, interacting pairs containing at least one unmapped drug were excluded from the corresponding DIR table. As the scope of this study was limited to DDIs, only interacting pairs containing drugs were included in the final DIR tables and interactions with herbs, alcohol, food, etc were excluded. The final tables (BNF final table, Thesaurus final table and Micromedex final table) contained drug interacting pairs and associated information based on normalised drug names. Any duplicate entries based on common normalised names were combined into a single entry. For example, Metoprolol Tartrate and Metoprolol Succinate were both mapped to the RxNorm entity Metoprolol, and their interactions were merged to produce a single set.

3.2.4 Comparison of resources

3.2.4.1 Listing of DDIs

The pairwise and three-way overlaps of the final DIR tables were estimated by calculating counts of common drug pairs across the DIRs as well as coverage rates (i.e. the percentage of a set A covered by B, where B is a subset of A). The directionality of interacting drug pairs was not taken into account (i.e. (D1, D2) and (D2, D1) were considered equivalent). A DIR intersection list containing common interacting drug pairs

DIR	Extracted fields	Categories
BNF (Accessed in June 2018)	- Drug Name - Interactant Name - Description - Severity (where present) - Evidence (where present)	(a) active pharmaceutical ingredients (APIs) (e.g. atropine); (b) drug classes (e.g. combined hormonal contraceptives); (c) herbs and supplements (e.g. peppermint oil); (d) foods and beverages (e.g. grapefruit juice).
Thesaurus (Accessed in July 2020; September 2019 update)	- Drug Name - Interactant Name - Mechanism (if available) - Severity Index - Additional Information (specification for drug class, etc) - Clinical Information (i.e. manifestation, management)	(a) Drug ingredient ; (b) Drug classes.
Micromedex (Accessed: August 2018) In Depth Answers Database (Detailed evidence-based information)	- Drug Name - Interactant Name - Interaction Effect - Summary - Severity - Onset - Substantiation - Clinical Management	(a) Drug ingredients; (b) Combination drugs; (c) Food; (d) Tobacco; (e) Lab tests.

TABLE 3.1: Extracted information from the drug information resources.

among all three DIRs with their corresponding text descriptions from each source was generated.

3.2.4.2 Severity and evidence ratings

All three DIRs included severity ratings (**Table 3.2a**), while only the BNF and Micromedex contained separate text fields regarding evidence ratings (**Table 3.2b**). Some DDIs from Thesaurus appeared at the drug class level in the original source, which was associated with multiple severity ratings; thus, individual drugs were assigned all applicable ratings from the drug class during the mapping process. Also, some DDIs were linked to multiple severity ratings, based on the clinical circumstances (e.g. route of administration, dose, etc). In all cases where multiple ratings were available for an individual DDI, the highest one was kept for further analysis.

To explore discrepancies among DIRs related to severity and evidence ratings for DDIs, we calculated the subset size, pairwise coverage rates and Jaccard indices for all possible pairs of DIR ratings.

3.2.4.3 Clinical management recommendations

We aimed to explore the consistency among the clinical management recommendations provided by the DIRs by analysing text descriptions from the DIR intersection list. The BNF provided a succinct description for each drug pair containing all types of available DDI information in a text field, while Thesaurus and Micromedex contained separate text fields (Conduit à tenir and Clinical Management, respectively) under each drug pair related to clinical management options.

Basic pre-processing involved text conversion to lowercase, drug name blinding (i.e. replacement of all drug names with a common string), and sentence tokenisation using the Natural Language Toolkit (NLTK) in Python 3.6 [175].

The following advice categories were initially considered:

1. Avoid;
2. Use with caution;
3. Space dosing times;
4. Wash-out;
5. Monitor;
6. Adjust dose;
7. Modify administration;
8. Use alternative;
9. Discontinue.

DIR	Level	Definition
(a) **Severity**		
BNF	1 – Severe*	The result may be a life-threatening event or have a permanent detrimental effect.
	2 – Moderate	The result could cause considerable distress or partially incapacitate a patient; they are unlikely to be life-threatening or result in long-term effects.
	3 – Mild	The result is unlikely to cause concern or incapacitate the majority of patients.
	4 – Unknown	Used for those interactions that are predicted, but there is insufficient evidence to hazard a guess at the outcome.
Thesaurus	1 – Contraindicated*	
	2 – Not recommended*	
	3 – Precautions for use	
	4 – Take into consideration	
Micromedex	1 – Contraindicated*	The drugs are contraindicated for concurrent use.
	2 – Major*	The interaction may be life-threatening and/or require medical intervention to minimise or prevent serious adverse effects.
	3 – Moderate	The interaction may result in exacerbation of the patient's condition and/or require an alteration in therapy.
	4 – Minor	The interaction would have limited clinical effects. Manifestations may include an increase in the frequency or severity of the side effects but generally would not require a major alteration in therapy.
(b) **Evidence**		
BNF	Study	For interactions where the information is based on formal study including those for other drugs with same mechanism (e.g. known inducers, inhibitors, or substrates of cytochrome P450 isoenzymes or P-glycoprotein).
	Anecdotal	Interactions based on either a single case report or a limited number of case reports.
	Theoretical	Interactions that are predicted based on sound theoretical considerations. The information may have been derived from in vitro studies or based on the way other members in the same class act.
Micromedex	Established (Excellent)	Controlled studies have clearly established the existence of the interaction.
	Theoretical (Good)	Documentation strongly suggests the interactions exists, but well-controlled studies are lacking.
	Probable (Poor)	Available documentation is poor, but pharmacologic considerations lead clinicians to suspect the interaction exists; or, documentation is good for a pharmacologically similar drug.

TABLE 3.2: Information contained in drug information resources on drug-drug interactions relating to (a) the severity ratings (in descending order as displayed in the original source); and (b) the evidence ratings.
(*): critical severity ratings

Cases that recommended clinicians to refer to literature or other resources, without mentioning any concrete clinical advice, were excluded.

The limited number of unique sentences sourced from BNF (N = 305) and Thesaurus (N = 387) following drug name blinding enabled manual sentence labelling, with each sentence being classified into one or multiple advice categories.

To annotate Clinical Management text descriptions in Micromedex (N = 4,507), we developed a bespoke text classification process in Python using a methodology that has been widely implemented in similar tasks and provided the desired functionality while keeping the level of complexity low (**Figure A.2**). First, we annotated a subset of randomly selected unique sentences (N = 200) by considering the above-mentioned categories. Then, each labelled sentence was tokenised into its constituent tokens (i.e. words) and stemming (i.e. reducing words to their word roots) was applied. We used term frequency-inverse document frequency (tf-idf) to calculate weights for each word in the annotated sentences. The goal of tf-idf is to reduce the impact of very commonly occurring words in a corpus, assuming that they are less informative. Term frequencies are calculated by counting the relative frequency of each word appearing in each of the annotated sentences. Inverse document frequencies of each word (in its root form) are estimates of the overall presence of the word across all sentences (i.e. how commonly or rarely it appears). The formula for calculation of a word's tf-idf is the following:

$$tfidf_{w,s} = tf_{w,s} \times (log\frac{N}{df_w} + 1) \tag{3.1}$$

where $tf_{w,s}$ represents the term frequency of the word, w, in the sentence, s (i.e. the number of times the word appears in the sentence divided by the total number of words in the sentence); N is the total number of sentences in the corpus and, df_w is the "document" frequency of the word, w (i.e. the number of sentences that contain the specific word). Weights were applied for sentence encoding to feed classifiers that used a supervised machine learning model called linear kernel Support Vector Machine (SVM) for binary text classification (i.e. each sentence was classified as to whether it belongs to each of the advice categories under consideration). We applied class weights to account for the imbalanced training sets (i.e. disproportion between the number of positive and negative instances) and used leave-one-out cross validation to evaluate performance of the difference classifiers through Receiver Operating Characteristic (ROC) analysis. By estimating the positive predictive value (PPV) for the different thresholds, we excluded sentence classifiers with a PPV below 80% due to poor performance; the remaining, unannotated sentences from Micromedex were automatically labelled by the classifiers using the threshold with maximum sensitivity for PPVs above 80%. A subset ($N = 100$) of the automatically annotated sentences (validation set) was also manually annotated to independently estimate the classifiers' performance in the total set of Micromedex sentences.

DIR	Initial drug names	Normalised ingredients	DDI counts
BNF	1,004	984	51,481
Thesaurus	1,049	1,001	38,037
Micromedex	2,602	1,967	65,446

TABLE 3.3: Number of initial drug names, normalised ingredients, and drug-drug interaction counts per drug information resource.

3.3 Results

3.3.1 Comparative assessment in terms of listing

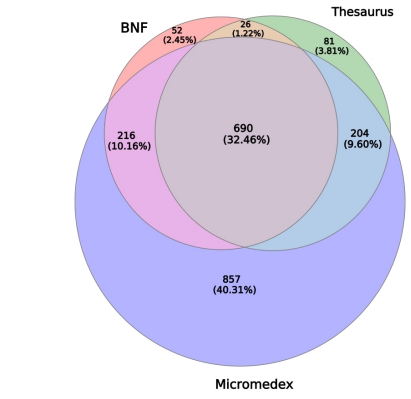
Micromedex contained the largest number of DDI drug pairs (N = 65,446), as well as normalised ingredients involved in DDIs (N = 1,967), followed by BNF (N = 51,481) and Thesaurus (N = 38,037) that covered 984 and 1,001 normalised ingredients, respectively, in their DDI section. The collation of the three final DDI tables included 121,351 DDI drug pairs. The counts of initial drug names, normalised drug ingredients, and unique DDIs in each DIR are summarised in **Table 3.3**.

There were 690 common normalised ingredients involved in DDIs across all examined DIRs, with BNF and Micromedex sharing the largest number (N = 906), followed by Thesaurus and Micromedex (N = 894) and, lastly, the BNF and Thesaurus (N = 716) (**Figure 3.1a**). Almost four out of five DDI drug pairs (78.04%, N = 94,708) in the collated list were only mentioned by a single DIR, with 57.19% of BNF, 49.58% of Thesaurus, and 70.91% of Micromedex DDI entries missing from the other two DIRs. The percentage of DDIs mentioned in exactly two out of three DIRs was lower (16.21%, N = 19,673). BNF shared 14,576 DDIs with Thesaurus (28.31% of BNF; 38.32% of Thesaurus) and 14,433 DDIs with Micromedex (28.04% of BNF; 22.05% of Micromedex), while Thesaurus and Micromedex had 11,574 common DDIs (30.43% of Thesaurus; 17.68% of Micromedex). The intersection of the three DIRs in terms of DDIs (N = 6,970) represented only 5.74% of the collated list, 13.54% of BNF, 18.32% of Thesaurus and 10.65% of Micromedex (**Figure 3.1b**).

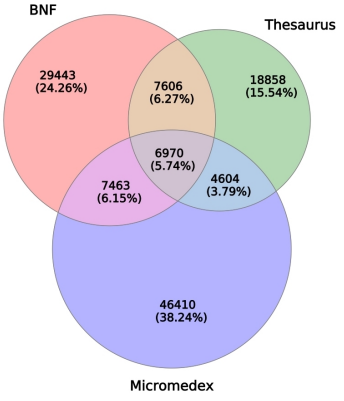
In terms of DDIs restricted to common ingredients across the three DIRs (N = 44,719), more than half (N = 24,951) were only found in a single DIR (33.74% in BNF alone, in comparison to 12.86% and 9.20% in Micromedex and Thesaurus, respectively), while 28.62% were present in two out of three DIRs. In the setting of ingredient-restricted DDIs, the BNF intersected with large proportions of both Thesaurus (71.40%) and Micromedex (59.76%), while Thesaurus overlapped with less than half of BNF (42.81%). Finally, the intersection of the three DIRs represented 15.59% of the restricted DDIs (**Figure 3.1c**).

3.3.2 Comparative assessment of severity rating

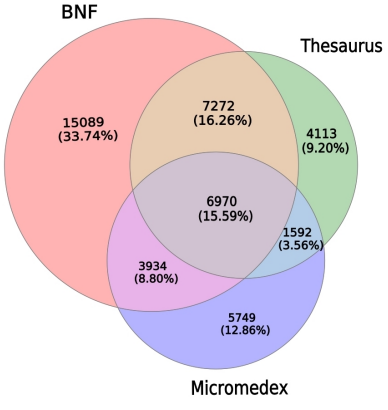
The categorisation of DDIs in each DIR in terms of severity rating is outlined in **Table 3.4**. Regarding critical severity rating categories, almost one quarter (24.56%) of unique



(a)



(b)



(c)

FIGURE 3.1: Venn diagrams illustrating the intersections in terms of: (a) drug ingredients; (b) unique drug-drug interaction pairs included in the drug information resources; and (c) drug-drug interaction pairs included in the drug information resources only for the ingredient intersection subset. Each circle represents a drug information resource and their intersections show the number of ingredients/drug-drug interactions they share with each one of the other drug information resources.

Severity rating	BNF	Thesaurus	Micromedex
1	12,644 (24.56%)	2,949 (7.75%)	5,730 (8.76%)
2	4,997 (9.46%)	12,779 (33.60%)	41,713 (63.73%)
3	273 (0.51%)	8,195 (21.54%)	15,890 (24.28%)
4	33,705 (65.47%)	14,114 (37.11%)	2,113 (3.23%)

TABLE 3.4: Number and percentage of drug-drug interactions by severity rating in each drug information resource.

DDIs in BNF were labelled as Severe, compared to 7.75% from Thesaurus characterised as Contraindicated, and 33.60% as Not recommended. In Micromedex, 8.76% of unique DDIs were mentioned as Contraindicated, while Major was the most frequent category (63.73%).

When considering the pairwise DIR overlap using coverage rates (**Figure 3.2**), the number of DDIs jointly rated as critical was:

- 2,429 between BNF and Thesaurus, representing 19.21% of the BNF and 15.44% of Thesaurus critical DDIs;
- 6,026 between the BNF and Micromedex (47.66% of BNF and 31.38% of Micromedex critical DDIs);
- 5,014 between Thesaurus and Micromedex, covering 78.39% of Thesaurus and 26.33% of Micromedex critical DDIs);
- 1,768 among all three DIRs (25.37% of the DIR intersection list).

The percentage of DDIs from the DIR intersection list that were considered critical by BNF, Thesaurus, and Micromedex was 43.39% (N = 3,024), 52.32% (N = 3,647), and 81.51% (N = 5,681), respectively.
A similarity matrix of the Jaccard index for all DIR severity rating combinations is included in **Appendix A** (**Figure A.3**).

3.3.3 Comparative assessment of evidence rating

The BNF included evidence ratings in just around one third (30.66%) of its DDIs, with the majority being flagged as Theoretical (16.99%), followed by Study (11.89%), and Anecdotal (1.78%). In Micromedex, evidence ratings were consistently present under each DDI description. Theoretical was the most common category (70.91%), while Probable and Established included 20.36% and 8.73% of the DDIs mentioned in Micromedex. Almost half (48.13%) of the DDIs from the DIR intersection list contained no evidence rating in BNF; the remainder belonged to Study (28.05%), Theoretical (19.94%), and, lastly, Anecdotal (3.87%) evidence categories. According to Micromedex, most of them (64.51%) were Theoretical, with 19.94% and 15.55% being considered as Probable and Established, respectively.

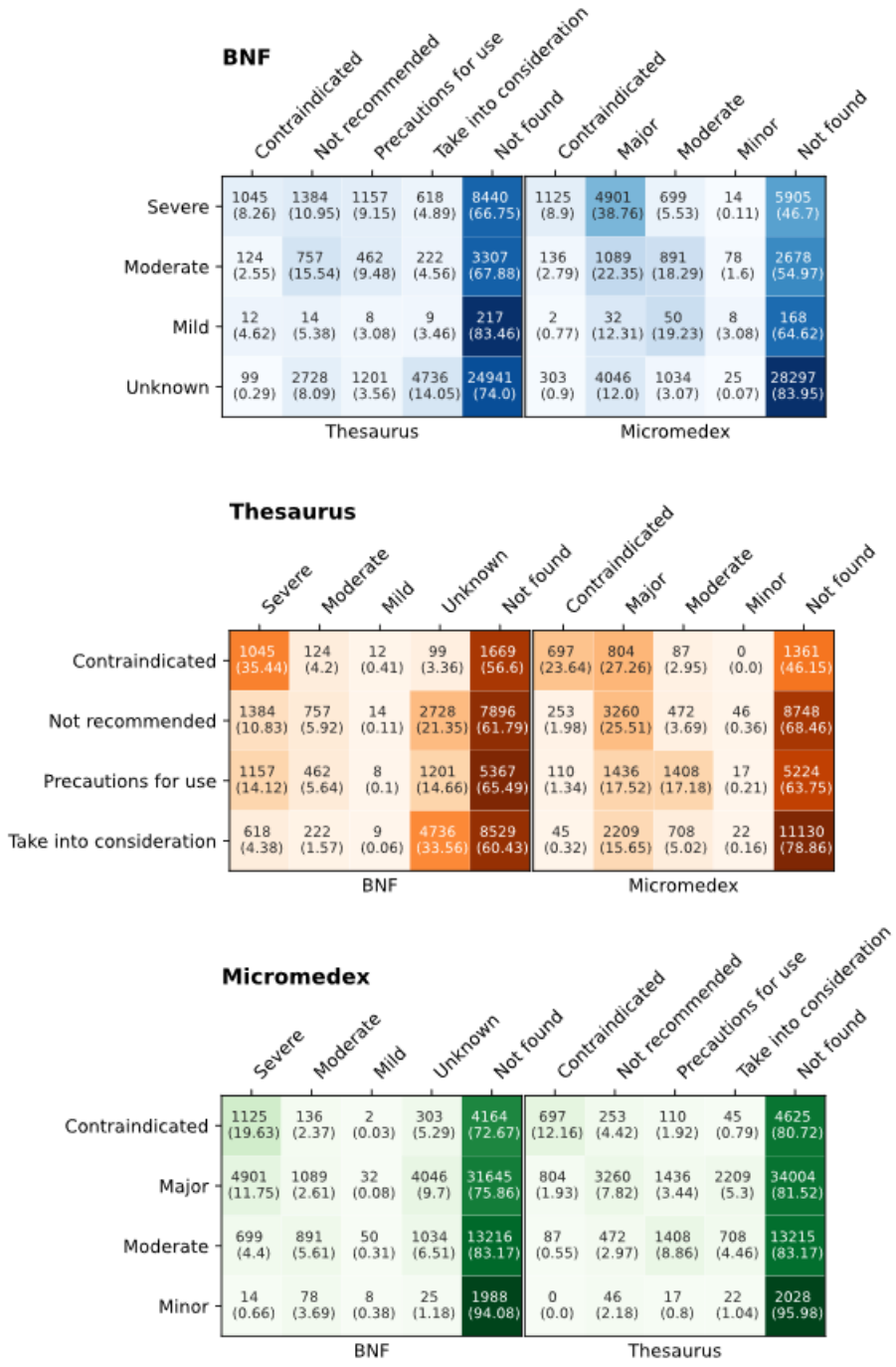


FIGURE 3.2: Pairwise comparison tables for the different drug-drug interaction severity levels. In each table, row labels contain the severity ratings of the drug information resource under consideration, while column labels represent the severity ratings of the remaining two drug information resources. A separate column has been added to include the numbers of unique drug-drug interactions missing from each of the other drug information resources. Each row contains the number of unique drug-drug interactions per severity rating of the drug information resource under consideration, subcategorised by the severity ratings of the other drug information resources. The numbers in parentheses represent the corresponding percentages of the various sets per severity rating of the drug information resource under consideration. The colour gradient shows the relative differences in the percentages mentioned among the various overlapping sets.

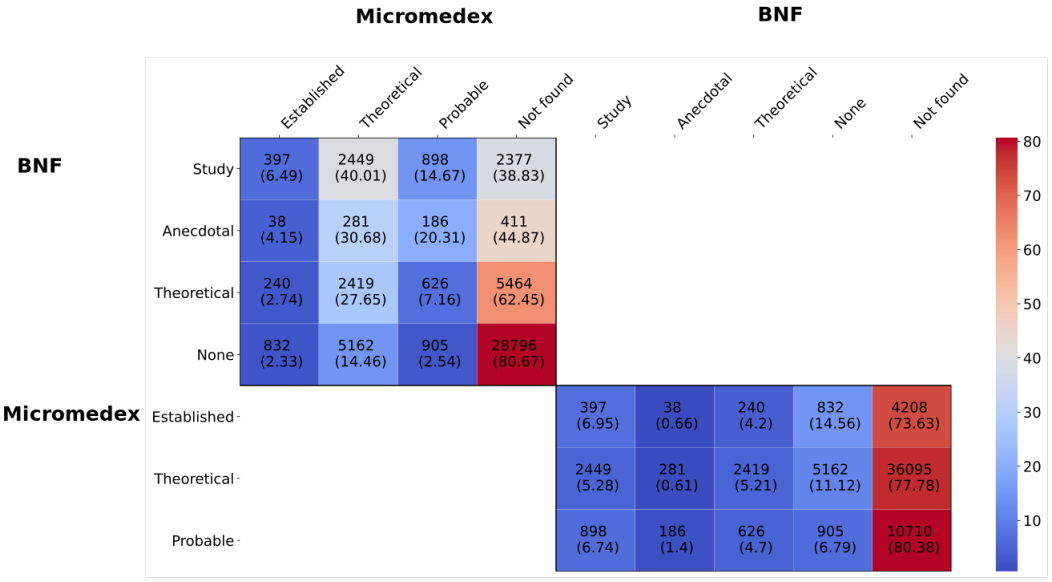


FIGURE 3.3: Heatmap for evidence rating comparison between BNF and Micromedex, including counts and coverage rates.

Figure 3.3 shows the overlap of the different evidence categories between the two DIRs as a two-by-two grid with subset counts and coverage rates. Probable DDIs from Micromedex and DDIs with no evidence rating from BNF were absent in higher percentages in the other resource. In both DIRs, the percentage of missing DDIs increased as one moved towards DDIs with a "poorer" or no evidence rating in the other resource. Using the Jaccard index, the agreement between ratings was generally low in all cases, with the BNF Study and Micromedex Theoretical categories being the most similar (0.04662), while the BNF Anecdotal and Micromedex Established had the lowest concordance (0.00573).

3.3.4 Comparative assessment of clinical management advice

In the BNF, no instances of the Discontinue advice category were identified in the DIR intersection list, while in Micromedex, counts for seven out of the nine advice categories are provided, as no sentence classifier was applied to extrapolate the remaining labels (i.e. Space dosing times and Modify administration) due to poor classifier performance (see **Supplementary Tables A.1** and **A.2** for associated metrics). The subset of Micromedex descriptions associated with BNF cases that belonged to either of those two advice categories was manually annotated as a surrogate measure of concordance.

The classification of DIR intersection list entries in each DIR in terms of clinical management advice is shown in **Table A.3**. In BNF, no advice was available in over half (56.41%) of the DDIs under consideration. The most common advice category was Avoid (32.12%) and the least frequently mentioned was Use alternative (0.01%). In Thesaurus, Monitor and Avoid jointly covered more than half of the total DDIs (34.89%

and 30.82%, respectively), while recommendations related to Space dosing times, Use with caution and Wash-out were only found in small percentages (1.98%, 1.15%, and 0.99%, respectively). In Micromedex, the labelling process that was facilitated by sentence classifiers provided the following results: 63.43% of the DDIs were characterised as containing advice related to Monitor, 47.02% related to Avoid, and 35.77% related to Adjust dose; low percentages represented Use alternative (3.79%), Discontinue (2.74%), and Wash-out (0.39%) categories. In 5.38% of the Micromedex DDIs, no advice label was assigned.

The overlap in terms of the DDI-related advice labels for the DDIs found in the DIR intersection list is illustrated using Venn diagrams (**Figure 3.4**). The BNF and Thesaurus did not share any DDIs in their Modify Administration and Use with caution categories, as opposed to the DDIs found in their Space dosing times and Adjust dose categories that showed extensive overlap. Thesaurus and Micromedex did not have any common DDIs classified into their Wash-out advice categories. Also, for Wash-out and Use with caution advice categories, there was little agreement between any two DIRs. The three DIRs overlapped to a high degree in the Monitor category. In the majority of Space dosing times BNF cases (87.96%), Micromedex also contained the respective advice. For Modify administration, Micromedex included this advice for less than half (43.18%) of the BNF cases.

3.4 Discussion

This study reports on the consistency of DDI-related information included in three major clinical DIRs from different geographic locations, namely the British National Formulary (BNF), Thesaurus and Micromedex. The DIRs differed in size and number of ingredients mentioned in the DDI sections. The number of ingredients in Micromedex was almost twice that found in the other two DIRs. This is most likely to have been due to the fact that the BNF and Thesaurus only include medicines licensed in their countries of origin (i.e. United Kingdom and France, respectively), while Micromedex includes a broader set of medicines. Although DIR ingredients overlapped to a significant extent, especially between BNF and Thesaurus, this overlap was not reflected in the DDI sets, which generally showed poor agreement. The BNF and Thesaurus shared the largest number of DDIs, in contrast with Thesaurus and Micromedex, which had the fewest DDIs in common.

Our study represents the most comprehensive assessment of the overlap in content and advice provided by different DIRs. However, our findings are consistent with previous studies. For instance, a study that analysed DDIs of fewer than 100 medicines reported less than 7% overall agreement among the examined sources [156, 157]. A more recent analysis that compared three commercial DIRs in terms of listing and severity ranking of DDIs also identified very poor overlap (5%), although DDIs flagged as minor were not considered [162].

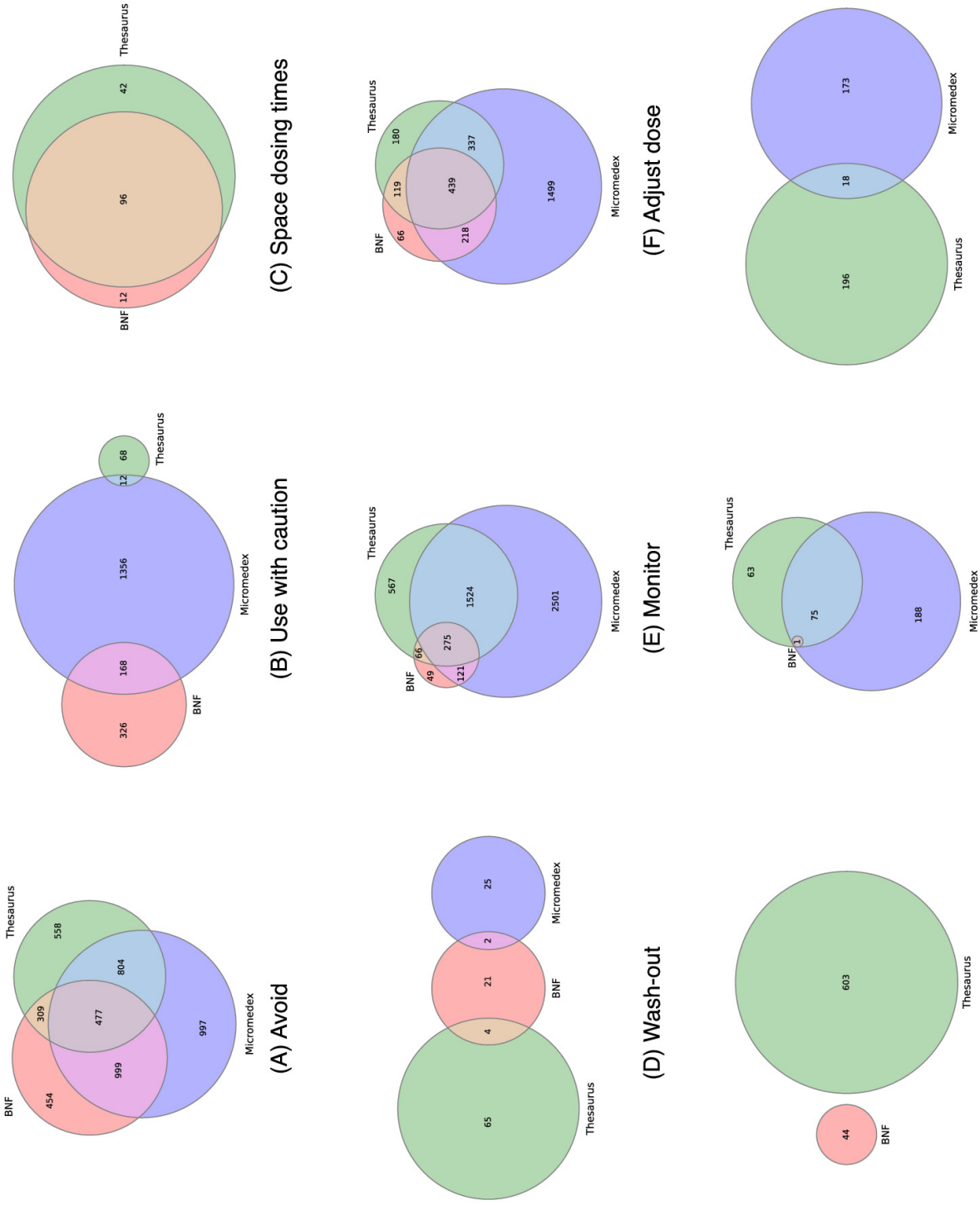


FIGURE 3.4: Venn diagrams of the overlap in clinical management advice relating to drug-drug interactions in the drug information resources.

Severity ratings were not consistently reported in the BNF, as opposed to Thesaurus and Micromedex, where ratings were available in all cases. DDIs labelled as critical comprised approximately one-fourth of BNF and more than 70% of Micromedex, in contrast to Thesaurus, where the least significant category was the most populous. Micromedex and Thesaurus showed similarity in their ratings across the different levels of severity. Between BNF and Thesaurus, their least significant categories (i.e. Unknown and Take into consideration, respectively) appeared to have a higher level of agreement. However, there was less concordance between BNF and Thesaurus on the classifications of DDIs in the high severity ratings, with the DDIs classified as Severe in BNF being spread across the different Thesaurus categories. In terms of BNF and Micromedex, there was generally better agreement between their critical ratings than with the less severe ones. Apart from the BNF-Thesaurus pair, the percentage of DDIs missing from a DIR increased as one moved to DDIs characterised as less severe by another DIR (i.e. increasing trend in the Not found column percentages as we go from top to bottom in tables from **Figure 3.2**). Micromedex categorised the largest proportion from the DIR intersection list as being critical compared to the other DIRs, while around one-fourth of the DDIs in the DIR intersection list were simultaneously labelled as critical by all three DIRs. Also, the pairwise intersections of DIRs covered larger proportions of the DDIs from the critical severity levels compared to the corresponding proportions from lower severity categories.

Although early studies concluded that significant discrepancies exist in severity ratings between DIRs, Fung et al.'s study advocated the presence of higher levels of agreement than previously reported, especially for the most severe DDIs [156, 157, 162]. Our results, suggesting better agreement between critical severity ratings between BNF and Micromedex, are partially in line with this observation.

Evidence categorisation was not available in Thesaurus, thus preventing a comprehensive assessment of the concordance of evidence rating amongst all the DIRs. In BNF, evidence ratings were available for around one-third of the DDIs, while they were consistently reported in Micromedex. In both DIRs, Theoretical was the most frequent category. However, the study revealed a lack of consistency between BNF and Micromedex, with no cases of strong agreement between any pairs of evidence ratings. An interesting observation was related to the DDIs included in one but missing from the other DIR, as the percentage of Not found DDIs in both cases increased as the evidence rating in the other DIR decreased. A study by Vitry that performed a similarity assessment of evidence ratings also highlighted inconsistencies in the grading system for evidence among the different sources [156]. In terms of clinical management recommendations, there was significant disagreement among the DIRs related to some types of advice, such as Use with caution, Wash-out, Discontinue and Modify administration. Other types (i.e. Avoid and Use alternative) showed a moderate level of agreement, while Space dosing time, Monitor and Adjust dose demonstrated higher levels of concordance.

Drug Information Resource	Ingredients	Drug-drug interactions
BNF	Aprotinin	Methylphenidate – bupropion
	Dextromethorphan	Rasagiline – metoclopramide
	Dibucaine	Bortezomib – yellow fever vaccine
	Dimercaprol	Ephedrine – midodrine
	Filgrastim	Midostaurin – lumacaftor
	Flumazenil	Pentamidine – domperidone
	Goserelin	Flecainide - propafenone
Thesaurus	Abacavir	Ondansetron – salmeterol Ondansetron – sunitinib
	Beclomethasone	
	Dimercaprol	
	Diphenoxylate	
	Filgrastim	
	Levetiracetam	
Micromedex	Adefovir	Ephedrine – moclobemide Tadalafil – voriconazole
	Dalfampridine	
	Daratumumab	
	Mifamurtide	
	Rupatadine	
	Zoledronic acid	

TABLE 3.5: Some examples of ingredients and drug-drug interactions not included in the drug-drug interaction sections of the three drug information resources.

3.4.1 Impact on clinical decision support

Examples of missed medicines and DDIs from the different DIRs can be found in **Table 3.5**. Some ingredients were surprisingly missing from the DDI section of some of the DIRs, such as levetiracetam from Thesaurus and rupatadine from Micromedex. A few ingredients were not licensed in the respective countries (e.g. enalaprilat, an intravenous formulation of enalapril, is not available in the United Kingdom or France) or had been discontinued (e.g. ketorolac in France or maprotiline in the United Kingdom) at the time of data collection. Micromedex however contained drugs that were discontinued or not approved in the US.

The difference in size might also be partly related to the nature of the different DIRs. Micromedex is a commercial knowledge base, while BNF and Thesaurus are maintained by professional and regulatory bodies, respectively. Commercial knowledge bases may be overinclusive to minimise potential legal consequences arising from their decision to omit DDIs.

There were a few important DDIs missing from one of the three DIRs (**Table 3.5**). Examples include: voriconazole (a CYP3A4 inhibitor) interacts with tadalafil (a PDE-5 inhibitor) increasing its systemic exposure [176]; bupropion and methylphenidate, indirect sympathomimetic agents that lower seizure threshold [177]; and sunitinib and ondansetron, which both prolong the QT interval which predisposes to torsade de pointes [178].

Severity ratings also varied for DDIs in the three DIRs which may impact patient safety. For example, the combination of paroxetine and tramadol was categorised at

the lowest severity level (Take into consideration) in Thesaurus while it was ranked as Severe and Major in BNF and Micromedex, respectively. This is a complex interaction, which leads to decreased plasma concentrations of the active metabolite of tramadol (M1) because of CYP2D6 inhibition by paroxetine, and also an increased risk of serotonin syndrome. Interestingly, the updated version of Thesaurus (2020) has upgraded the severity level of the drug pair to Not recommended. Other examples include (a) the interaction between cytarabine and flucytosine, which was categorised as Severe in BNF (because of decreased concentrations of flucytosine), but at the lowest level in both Micromedex and Thesaurus; (b) the interaction between niacin and statins, which increases the risk of myopathy and rhabdomyolysis, was characterised as Severe in BNF and Major in Micromedex but was completely missing in Thesaurus; and (c) the combination of non-steroid anti-inflammatory drugs with thiazide-type diuretics (e.g. chlorothiazide, chlortalidone) was ranked as Severe in BNF and Major in Micromedex (but as Precautions for use in Thesaurus) because of the risk of acute renal failure.

3.4.2 Strengths and limitations

As opposed to multiple previous efforts to assess the level of agreement among DIRs in terms of DDI information, this study examined entire resources, thus revealing the relative size of information in each of the DIRs and exploring the stratification of the included DDIs in terms of severity, evidence, and clinical management recommendations. To the best of our knowledge, this is the first comprehensive effort to compare clinical advice for managing DDIs that is provided in multiple DIRs, with a clear focus on clinically oriented sources compared to previous work [158]. While a previous study expanded the comparison of DIRs at multiple levels (i.e. clinical drug, ingredient and drug class) [162], our analysis was limited at the ingredient level. This standardisation of DIRs enabled a "fair" comparison in terms of the volume of information listed. Also, code availability for data extraction and standardisation will enable the reproducibility of the analysis.

However, there are limitations to this study. First, no updates have been taken into consideration since the date of data retrieval (i.e. offline data). Therefore, the results and conclusions of this study provide an overview of their similarity and consistency at that specific point in time, although no major updates usually occur. Second, Thesaurus contained a few DDIs originally reported at the drug class level, which were associated with multiple severity ratings. Hence, some standardised DDIs at the ingredient level were assigned more than one severity rating. In BNF, there was a limited number of DDIs having multiple severity levels depending on the described clinical outcome. In both cases, the highest severity rating was considered for further analysis. Another limitation could be relevant to the different countries of origin for the DIRs that were considered in this study, which might have contributed to a small extent to the discrepancies observed. Other limitations include the comparison of clinical management options only for the DDIs present in the intersection of the DIRs and the custom-made labelling

process applied to Micromedex. Additionally, in the BNF, there were referrals to the guidance section of the website for various drug categories that were left unmapped during the annotation process, e.g. "For Faculty of Sexually and Reproductive Healthcare (FSRH) guidance, see contraceptives, interactions.", or "See 'serotonin syndrome' and 'monoamine-oxidase inhibitor' under antidepressant drugs for more information and for specific advice on avoiding monoamine-oxidase inhibitors during and after administration of other serotonergic drugs.". In this way, the overall advice support provided by the BNF might have been underestimated, although no concrete clinical advice was provided.

For future work, it would be interesting to evaluate a larger number of DIRs and include DIRs (e.g. Medscape, Lexicomp, Stockley's Drug Interactions) which could not be accessed in this instance due to lack of a subscription or due to terms and conditions that currently prohibit the type of analysis we have conducted. It would also be interesting to explore the completeness of generic DIRs as opposed to resources tailored to specific drug categories (e.g. the Liverpool Drug Interaction Checkers for anticancer drugs, etc) that would be expected to provide more complete information. The evaluation of agreement on various types of DDI-related information among DIRs from the same country of origin would be another relevant topic for future research, although significant levels of discordance would not be surprising, similar to previous studies [162]. Additionally, more comprehensive efforts to compare clinical management recommendations among entire resources would be beneficial from a clinical perspective.

3.4.3 Implications and conclusions

It is reasonable to assume that the inclusion of clinically significant DDIs in DIRs would improve drug efficacy and reduce adverse reactions. There is however a balance to strike since the value of these tools could be diminished if too many minor or clinically insignificant DDIs are included in an effort to limit legal liability [179]. This leads to the phenomenon of alert fatigue where practitioners ignore the alerts provided by the system due to the sheer volume of generated alerts [180], with important clinical consequences for patient safety.

The Evidence workgroup from the DDI CDS Conference Series has highlighted the importance and need for higher-quality information related to DDIs and also suggested the establishment of systematic DDI search criteria in order to determine the existing evidence related to the information provided [181]. Our analysis also shows the need for consistency in the definitions of severity and evidence ratings provided by the various DIRs. The availability of DDI evidence in a standardised format with adequate literature support, where possible, can improve prescribing decisions by allowing the prescriber to refer to appropriate resources and use clinical judgement in case of doubt. The inclusion of clinical management options for DDIs in CDS tools is also quite important and especially useful in the clinical setting, as there is no single response to a potential DDI. More focus on this aspect has been suggested in multiple studies, which

advocate more detailed and actionable advice (i.e. what and when to monitor) and clear indications of the strength of the recommendation [182, 183]. We recommend that, by providing a dedicated section for clinical management recommendations that contains clear, actionable recommendations, information retrieval in the clinic can be facilitated, and potentially improve individualised risk-benefit assessment of a specific DDI. In cases where the benefits of a drug combination outweigh its risks, strategies to mitigate potential adverse outcomes (e.g. therapeutic drug monitoring, vital signs, discontinuation of one of the drugs, etc) should be provided and will improve the benefit-harm balance of the drug combination. In the future, information about specific patient risk factors for DDIs, such as genetic polymorphisms, could also be included to enhance DDI preventability.

The use of pharmacologic drug classes in Thesaurus to summarise DDIs might become a source of confusion for clinicians. In many cases, individual drugs from a drug class have different pharmacological profiles (e.g. excretion, metabolism, etc), which contribute to a markedly different DDI risk when considering the same interacting drug [183]. Consequently, drug ingredient indexing may paradoxically impede searching rather than achieve the aim of providing effective DDI summaries.

In conclusion, there is a great deal of interest in clinical decision support systems providing information on DDIs to optimise medicines use so that the use of drug combinations that affect either efficacy and/or safety can be avoided. However, there is a lack of consistency and standardisation in the information provided by different DIRs. Our study which has systematically compared three DIRs shows that there is considerable variation in the DDI information provided in these resources. Such variability in information could have deleterious consequences for patient safety, and there is a need for harmonisation and standardisation.

3.5 Data availability statement

Data from the DIRs were derived from the following web sources: the BNF website[1] (available in the public domain); ANSM website[2] (available in the public domain). Restrictions apply to the availability of Micromedex data, which were used under license for this study. Data are available at https://www.micromedexsolutions.com/ with the permission of IBM Watson Health. The code that supports the web data extraction and analysis is available at: https://github.com/elpidakon/CRESCENDDI.

[1] https://bnf.nice.org.uk/interaction/

[2] https://ansm.sante.fr/documents/reference/thesaurus-des-interactions-medicamenteuses-1

Chapter 4

Building a clinically-relevant reference set for DDIs

Building upon the methodology presented in **Chapter 3**, this chapter aims to generate a reproducible framework and a publicly available standardised reference set of DDIs that can facilitate benchmarking and support the development of novel methodologies for signal detection of DDIs. It focuses on building a methodology for automatically extracting web data from DIRs and a standardisation pipeline to align data from multiple resources to a common framework using controlled vocabularies and terminologies.

Chapter 4 contributes to the thesis by presenting a scalable approach for building a standardised resource by combining multiple clinical resources, thus supporting the clinical relevance of the reference set. The output of this approach is utilised in the next two thesis chapters (**Chapters 5 & 6**). The size of the reference set and the relatively large number of included drugs and adverse events provides a common ground, which was previously unavailable, for the comparative performance evaluation of multiple DDI surveillance methods.

Abstract

The accurate and timely detection of adverse drug-drug interactions (DDIs) during the postmarketing phase is an important yet complex task with potentially major clinical implications. The development of data mining methodologies that scan healthcare databases for drug safety signals requires appropriate reference sets for performance evaluation. Methodologies for establishing DDI reference sets are limited in the literature, while there is no publicly available resource simultaneously focusing on the clinical relevance of DDIs and individual behaviour of interacting drugs. By automatically extracting and aggregating information from multiple clinical resources, we provide a scalable approach for generating a reference set for DDIs that could support research in postmarketing safety surveillance. CRESCENDDI contains 10,286 positive and 4,544 negative controls, covering 454 drugs and 179 adverse events mapped to RxNorm and MedDRA concepts, respectively. It also includes single drug information for the included drugs (i.e. adverse drug reactions, indications, and negative drug-event associations). We demonstrate the usability of the resource by scanning a spontaneous reporting system database for signals of DDIs using traditional signal detection algorithms.

4.1 Background & Summary

Polypharmacy (i.e. the concomitant use of multiple medications in an individual) has become a common phenomenon in the Western world. In the United States, between 2015 and 2018, it has been estimated that two out of three people over 65 take at least three prescription medications during the course of a month (up from one-third in the early 1990s), with four out of ten taking five or more medications [184]. As life expectancy is increasing around the world, leading to more people living with multiple chronic diseases, together with new medicines being launched onto the market each year, giving rise to a growing volume of possible drug combinations, the implications of drug-drug interactions (DDIs) in clinical practice have become a matter of concern. A DDI can lead to the potentiation or antagonism of one drug by another, or cause another effect that is not related to the individual drug profiles. From a mechanistic perspective, DDIs are classified into two main categories: pharmacodynamic and pharmacokinetic [126]. Predicted DDIs based on pharmacological knowledge (i.e. possible drug effect alterations caused when multiple drugs are simultaneously administered) far outnumber those with clinically significant consequences, i.e. those ranging from lack of efficacy to serious and life-threatening adverse reactions [153]. The necessarily limited time and extent of pre-clinical studies and pre-marketing clinical trials may burden the identification of adverse drug reactions (ADRs) caused by single drugs or drug combinations (i.e. adverse DDIs).

Postmarketing safety surveillance (pharmacovigilance) is a vital stage in the lifecycle management of a medicine: The number of people exposed to medicines after marketing is substantially larger than the number of volunteers involved in pre-market clinical trials. Spontaneous reporting system (SRS) databases, such as the US Food and Drug

Administration (FDA) Adverse Event Reporting System (FAERS), are a particularly useful source of information, with more recent efforts focusing on the integration of multiple data sources [185, 186, 187]. Given the growing size and complexity of those databases, automated statistical tools, known as signal detection algorithms (SDAs), have become indispensable tools in an effort to distinguish real signals (i.e. information-bearing patterns) from accompanying random patterns in the background (called noise) that distract from the information [188].

The early and effective identification of drug safety signals is of paramount importance for the pharmaceutical industry and regulatory authorities. For adverse reactions caused by DDIs, there is an increasing need for improved SDAs, with the existing state-of-the-art being less mature compared to the well-established algorithms used for detecting signals of ADRs caused by a single drug [12, 11, 13]. Even after approval, detection of novel DDIs might be delayed and difficult, due to the inherent complexity of DDIs, dose-dependency (i.e. some interactions only become evident in elevated drug levels) and natural human inter-variability (as well as intra-variability, in some cases) that accounts for the onset of some DDIs (e.g. rapid and poor metabolisers, genetic subpopulations, etc.). The growing volume of real-world health data presents both a challenge and an opportunity for the pharmacovigilance community, making manual review impossible and requiring a higher level of automation in the methods that are routinely used for scanning such databases.

As the available evidence is not static, the lack of gold standards (i.e. definitive positive and negative controls) poses a challenge when it comes to defining appropriate reference sets in pharmacovigilance for performance evaluation. Also, much disagreement exists in terms of the choices and criteria under consideration (e.g. well-established versus emerging cases) [97, 98, 189].

For single drugs, a number of reference sets exist that include drug-event pairs that are either well-established (e.g. OMOP reference set [7], Harpaz [8]), belong to recent product labelling changes [10], or can be found in product labels (e.g. EU-ADR initiative [9]). More recently, efforts to automate the generation of a reference set for single-drug ADRs by combining multiple sources of evidence identified a number of limitations, including: size; consideration of a single data source for extracting positive controls; availability (i.e. not being open access); and inclusion of only a limited number of drugs and adverse events (AEs) [190]. Our understanding is that a fair algorithm comparison requires testing on a large reference set to derive performance metrics that are likely to indicate performance in the context of novel signals.

To the best of our knowledge, there is no established reference set for DDIs coupled with information on the individual behaviour of interacting drugs. Were such a reference set to exist, it could enable the classification of positive controls based on the possible underlying mechanism causing the interaction. Initial efforts for detecting signals indicative of DDIs included test cases that were limited in either size [11] or a variety of drugs and AEs [13]. A more advanced approach was implemented by Juhlin et al [191],

although their reference set relied on a single clinical resource, which might not be a good idea since discordance among DDI compendia has been identified in the literature [156, 157, 160, 161, 192].

Although a definitive reference standard including the complete set of DDIs cannot exist, the automatic extraction and aggregation of information from multiple clinical resources on DDIs and the individual behaviour of interacting drugs, along with scanning the scientific literature for negative controls, enabled us to construct, share and advocate CRESCENDDI (Clinically-relevant REference Set CENtred around Drug-Drug Interactions), a dataset that can be used to facilitate research in SDAs and allow common ground for comparing methodologies. We propose a scalable approach for generating a normalised reference set that requires less manual effort for future updates, considering the dynamic nature of data and evidence availability, and for subset selection using design criteria. **Figure 4.1** outlines the main steps for generating the reference set.

4.2 Methods

The processing pipeline for the construction of CRESCENDDI included the following steps: (1) web data extraction of information related to DDIs and single-drug ADRs from 4 different online resources for DDIs, single-drug ADRs, and drug indications; (2) mapping (normalization) of drug names appearing in the extracted data; (3) extraction of the intersection of the DDI online resources; (4) manual annotation and mapping of English language text descriptions for DDIs, single-drug ADRs, and drug indications to MedDRA concepts; (5) generation of positive controls for DDIs using the normalised intersection of the DDI online resources; (6) generation of negative controls using drugs and AEs from the positive controls set combined with PubMed search; and (7) aggregation of information on single-drug ADRs and drug indications for the drugs appearing in the DDI reference set (i.e. positive and negative control sets) and generation of negative controls for single drugs with a process similar to the one followed in the previous step for DDI negative controls.

4.2.1 Web data extraction

DDI data were derived from the following online resources: the British National Formulary (BNF) website [163], the French National Drug Safety Institute (ANSM) Portable Document Format (PDF) file (Thesaurus) [164] and the Micromedex platform [165]. For BNF and Micromedex, web data extraction tools in Python 3.6 [170] enabled the extraction of the relevant fields into a Comma Separated Values (CSV) file (June-August 2018). For Thesaurus, the R package IMThesaurusANSM [171] was used and the resulting R dataframe from the 2019 update was converted to a CSV file.

The tables contained the following fields:

- interacting drug name 1 (D1) (e.g. Metoprolol Tartrate);

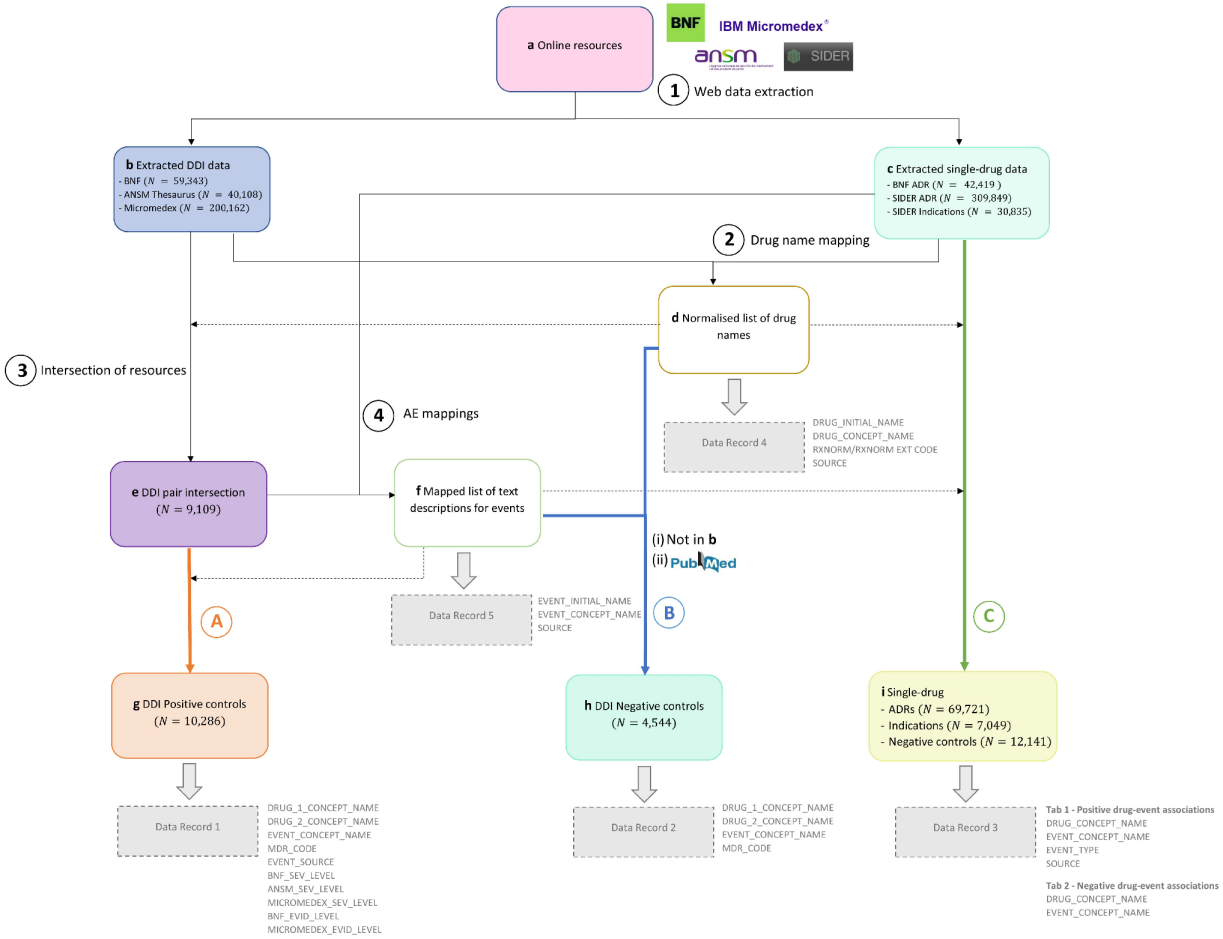


FIGURE 4.1: Data analysis workflow to generate positive and negative controls from DDI online resources with associated evidence for their component drugs. (1) DDI and single-drug online resources data (a) are extracted and stored in separate tables (b,c). (2) Drug names are normalised (d). (3) The intersection of the DDI online resources is extracted to a different table (e) and (4) English language text descriptions (for DDIs and single-drug ADRs) are annotated for AEs (after drug name masking) (f). (A) DDI pairs from (e) are assigned AEs based on the description mappings (g). Positive controls are published in Data Record 1. (B) Negative controls are generated using drugs and AEs from (g), ensuring that drug pairs cannot be found in (b) or in PubMed using a customised query (h). Negative controls are published in Data Record 2. (C) The filtered set of drugs from (d) is linked to AE and indication concepts using available evidence from (c) and negative controls for single drugs are generated following a similar process to the one described above for DDIs (i.e. drug-event is not an ADR mentioned in (c) or in PubMed using a customised query) (i). ADRs, indications and negative controls for single drugs are published in Data Record 3. Data Records 4 and 5 contain mappings of drug names and text descriptions for events, respectively. Column headers appear in grey font next to each Data Record box. Sample records from (b), (c) and (e) can be found in the bottom part of the figure.

- interacting drug name 2 (D2) (e.g. Lidocaine);
- text description for the DDI (e.g. Lidocaine is predicted to increase the risk of cardiovascular adverse effects when given with metoprolol. Manufacturer advises use with caution or avoid);
- severity label (e.g. Severe);
- evidence label (if available) (e.g. Study).

For single-drug ADRs, the following sources were considered: the BNF website [163] and SIDER dataset [193]. For drug indications, SIDER was used. BNF ADR data were extracted in a similar way as previously with DDI data (automated web data extraction) into a CSV file, while SIDER data for ADRs and drug indications were already available in CSV files.

A table containing the following fields was constructed:

- drug name (e.g. Metoprolol Tartrate);
- event text description (e.g. Bradycardia);
- event type (e.g. ADR);
- source (e.g. SIDER).

4.2.2 Drug name mapping

To facilitate usability and ensure compatibility, a standardization process was followed such that we could provide a resource with normalised concepts to standard terminologies for drugs and medical events. Specifically, the Observational Health Data Sciences and Informatics (OHDSI) Vocabulary version 5 was selected for mapping the drug names occurring in each of the DDI online resources into RxNorm and RxNorm Extension standard codes (at the Ingredient level) using OHDSI Usagi [174].

We removed combination drugs (as DDIs of their constituent drug ingredients were separately mentioned), vaccines, vitamins, herbal medicines, food, beverages, supplements, tobacco, and lab tests. Also, generic drug classes (e.g. combined hormonal contraceptives, hormonal replacement therapy) appearing in the BNF were not mapped to their individual drug ingredients, as there was no table on the BNF website specifying the drugs belonging in each drug class. We mapped the remaining unique drug names occurring in the DDI resources to OHDSI standard vocabulary concept identifiers. For example, Metoprolol Tartrate was mapped to the RxNorm Ingredient concept metoprolol. For Thesaurus, a native French speaker (pharmacist) confirmed the drug mappings of French drug names to English language OHDSI concepts.

A similar process was followed for drugs that appear in the single-drug data (ADR and indications).

4.2.3 Intersection of DDI online resources

By matching drug names to their mapped drug ingredients in the extracted DDI data tables, we obtained the set of common drug pairs across the tables and generated a new table that contains only the DDIs and associated information which could be found in each of the DDI online resources under consideration. Cases where the interacting drug mapping of D1 and D2 were swapped in the original data tables (i.e. (D1,D2) and (D2,D1)) were considered equivalent.

The final table contained the following fields:

- `drug_1 concept name` (e.g. metoprolol);
- `drug_2 concept name` (e.g. lidocaine);
- `bnf_description` (e.g. Lidocaine is predicted to increase the risk of cardiovascular adverse effects when given with metoprolol. Manufacturer advises use with caution or avoid);
- `micromedex_description` (e.g. lidocaine toxicity (anxiety, myocardial depression, cardiac arrest));
- `bnf_severity` (e.g. Severe);
- `ansm_severity` (e.g. Précautions d'emploi (Precautions for use));
- `micromedex_severity` (e.g. major);
- `bnf_evidence` (e.g. Study);
- `micromedex_evidence` (e.g. probable).

4.2.4 Adverse event and indication mappings

For DDI-related text descriptions in English from BNF and Micromedex that could be found in the DDI intersection table, a drug name blinding process was performed by replacing the interacting drug names with a common token in all cases (i.e. 'X'). In this way, the number of unique descriptions was reduced, thus facilitating the mapping process that followed. For example, the descriptions:

*Both **dexibuprofen** and **ibuprofen** can increase the risk of **nephrotoxicity**.*

and

*Both **polymyxins** and **streptomycin** can increase the risk of **nephrotoxicity**.*

were both mapped to the following blinded description:

*Both **X** and **X** can increase the risk of **nephrotoxicity**.*

The set of blinded text descriptions for BNF and Micromedex was extracted from the table and a semi-automated mapping process using OHDSI Usagi mapped them to MedDRA PT concepts. We explicitly focused on text descriptions that included clinical manifestations of DDIs, e.g. X may increase the risk of hypoglycaemia when taken with X. Text descriptions containing a potential mechanism of the interaction were left unmapped. In some cases, a single text description was linked to multiple concepts. For example:

Interaction Effect: An increased risk of cardiotoxicity (QT prolongation, torsades de pointes, cardiac arrest).

includes 3 different MedDRA PTs.

Also, serotonin syndrome was not mapped to its corresponding MedDRA PT and was not further considered as an AE for inclusion in the reference set.

Text descriptions from the BNF regarding single-drug ADRs were mapped to MedDRA PTs (where possible), but only for the drug ingredients that could be found in the DDI pair intersection table. SIDER ADR and indication data for the same list of drugs were also mapped to OHDSI concepts; however, for this resource, MedDRA PT codes were already available.

4.2.5 Positive controls

The set of positive controls was derived from the DDI intersection table, using mappings of text descriptions to AEs that were generated in the previous step. It contained 10,286 drug-drug-event (DDE) triplets, 454 unique individual drug ingredients and 179 unique AEs (as OHDSI concepts) in total.

4.2.6 Negative controls

The set of negative controls was generated by randomly pairing two drug ingredients from the 454 unique drug ingredients that can be found in the positive controls normalised drug name list (**Data Record 4**) and, in case the random drug pair did not appear in any of the DDI online resources, then it was randomly paired with an AE from the 179 unique AEs present in the positive control set event (AE type, DDI sources) from the normalised event list (**Data Record 5**). The choice of generating negative controls with common drug ingredients and AEs as the ones appearing in positive controls aimed to ensure the generation of a balanced reference set that does not contain added biases by design. For each of the created DDE triplets, a customised query was submitted to PubMed in an automated fashion and, if the search returned no results, the triplet was added to the negative control set. This process aimed to provide more confidence, to the best of our ability, about the absence of literature evidence of a potential DDI for the triplet under consideration, rather than definitive evidence to support the lack of a potential association.

The process was repeated until the number of negative controls with non-zero counts in the US Food and Drug Administration (FDA) Adverse Event Reporting System (FAERS) database (see Usage Notes) was similar in size (N=4,544) compared to the equivalent subset of positive controls. The negative control set included 161 unique AEs and 435 unique drug ingredients.

4.2.7 Single-drug ADRs, indications, and negative controls

By replacing text descriptions from BNF (N=1,538) and MedDRA PT codes to their corresponding mapped OHDSI concepts, a table with ADR and indication information related to the drug ingredients of the DDI reference set was generated. The table included: 438 unique drug ingredients, which could be found in at least one of the resources under consideration (i.e. BNF and SIDER), 3,492 AEs and 1,557 indication terms (as OHDSI concepts).

BNF and SIDER jointly contained 69,721 single-drug ADRs, with 12,318 common instances; this set could be utilised as a source for single-drug positive controls. This set covered 381 unique drugs and 835 unique AE concepts. Random pairing of those drugs and AE concepts followed by submission of a customised query to PubMed (to ensure absence of literature evidence of a potential ADR for the various drug-event associations) enabled the generation of a negative control set for single drugs (N=12,141).

4.3 Data Records

CRESCENDDI is publicly available online through Figshare[1] in 5 Excel spreadsheets. The fields contained in each of the spreadsheets are outlined below. The output of the methodology that was described in the previous section consists of 5 CSV data files and a README.txt file.

The columns contained in each of the data files are described below:

4.3.1 Data Record 1: Positive controls

`DRUG_1_CONCEPT_NAME`: Name of the first drug (active ingredient) that comprises a test case (DDE triplet).

`DRUG_2_CONCEPT_NAME`: Name of the second drug (active ingredient) that comprises a test case (DDE triplet).

`EVENT_CONCEPT_NAME`: Name of the normalised event (as a MedDRA PT) that comprises a test case (DDE triplet).

`MDR_CODE`: MedDRA code associated with the PT specified in the `EVENT_CONCEPT_NAME` field.

`EVENT_SOURCE`: Label indicating the resource where the event is mentioned (i.e. 'BNF', 'Micromedex', 'BNF+Micromedex').

[1] https://doi.org/10.6084/m9.figshare.c.5481408.v1

`BNF_SEV_LEVEL`: Label with the values of 'Severe', 'Moderate' or 'Mild' that indicates the severity level associated with the DDI as shown in BNF (if available).

`ANSM_SEV_LEVEL`: Label with the values of 'Contraindicated', 'Not recommended', 'Precautions for use' or 'Take into consideration' that indicates the severity level associated with the DDI as shown in Thesaurus.

`MICROMEDEX_SEV_LEVEL`: Label with the values of 'Contraindicated', 'Major', 'Moderate' or 'Minor' that indicates the severity level associated with the DDI as shown in Micromedex.

`BNF_EVID_LEVEL`: Label with the values of 'Study', 'Anecdotal' or 'Theoretical' that indicates the evidence level associated with the DDI as shown in BNF (if available).

`MICROMEDEX_EVID_LEVEL`: Label with the values of 'Established, 'Theoretical' or 'Probable' that indicates the evidence level associated with the DDI as shown in Micromedex.

4.3.2 Data Record 2: Negative controls

`DRUG_1_CONCEPT_NAME`: Name of the first drug (active ingredient) that comprises a test case (DDE triplet).

`DRUG_2_CONCEPT_NAME`: Name of the second drug (active ingredient) that comprises a test case (DDE triplet).

`EVENT_CONCEPT_NAME`: Name of the normalised event (as a MedDRA PT) that comprises a test case (DDE triplet).

`MDR_CODE`: MedDRA code associated with the PT specified in the `EVENT_CONCEPT_NAME` field.

4.3.3 Data Record 3: Single-drug ADRs, indications and negative controls

Tab 1 – Positive drug-event associations

`DRUG_CONCEPT_NAME`: Name of the drug (active ingredient) that comprises a single drug information case (drug-event pair).

`EVENT_CONCEPT_NAME`: Name of the normalised event (as a MedDRA PT) that comprises a single drug information case (drug-event pair).

`EVENT_TYPE`: Type of association between the drug and event of the single drug information case (i.e. 'Adverse event' or 'Indication').

`SOURCE`: Label indicating the resource of the single drug information case (i.e. 'BNF' or 'SIDER').

Tab 2 – Negative drug-event associations

`DRUG_CONCEPT_NAME`: Name of the drug (active ingredient) that comprises a single drug negative case (drug-event pair).

`EVENT_CONCEPT_NAME`: Name of the normalised event (as a MedDRA PT) that comprises a single drug negative case (drug-event pair).

4.3.4 Data Record 4: Drug mappings

`DRUG_INITIAL_NAME`: Unmapped name of the drug as extracted from the resource.

`DRUG_CONCEPT_NAME`: Normalised name of the drug (as an RxNorm Ingredient) extracted from the resource.

`RXNORM_CODE`/`RXNORM_EXTENSION_CODE` (OHDSI): RxNorm/RxNorm Extension code associated with the normalised drug name specified in the `DRUG_CONCEPT_NAME` field.

`SOURCE`: Label indicating the resource where the drug is found (i.e. '`BNF_DDI`', '`Thesaurus_DDI`', '`Micromedex_DDI`', '`BNF_Single`').

4.3.5 Data Record 5: Event mappings

`EVENT_INITIAL_TEXT`: Text description (after drug name blinding) containing an AE as extracted from the resource.

`EVENT_CONCEPT_NAME`: Mapped name of the event (as a MedDRA term) extracted from the resource.

`SOURCE`: Label indicating the resource where the text description is found (i.e. '`BNF_DDI`', '`Micromedex_DDI`', '`BNF_Single`').

4.4 Technical Validation

Given the inability to generate 'gold standards' in pharmacovigilance, the validation of the reference set included steps that supported the technical quality of the procedures followed to generate the dataset, rather than attempting to further ensure the validity of each control. The process of validating the reference set was two-way.

First, we verified the original online data as well as the unmapped extracted DDI data versus the curated reference set in order to validate the accuracy of the automated extraction and concept mapping processes. A random sample of 40 positive and 40 negative controls was manually checked in each of the DDI online resources (i.e. BNF, Thesaurus and Micromedex), to ensure the presence or absence, respectively, of the DDE triplet in the information provided. No issues were identified. Due to the time lag between data retrieval (June-August 2018) and validation (September 2021), there were issues most probably related to differing versions of the resources. More specifically, there were 6 positive controls missing from one of the resources (5 from the BNF and 1 from Micromedex), 2 of which were due to the complete removal of the drug monograph. Also, one of the negative control samples had been added to the BNF. No issues were identified during the validation of the random sample of controls from the curated reference set against the unmapped extracted DDI data.

Second, the validation of text description mapping for AEs included independent annotation of a random sample of 100 text descriptions for AEs to ensure agreement on event mapping. Again, no issues were identified.

Expert manual review of a number of cases from the reference set could be potentially performed as an additional layer of the technical validation procedure to confirm that the included DDIs are worthy of note. We have attempted to replace this need for expert manual review by combining information from multiple resources of clinical interest.

The next section outlines the application of the reference set for drug safety surveillance in an SRS database using SDAs that have been described in previous studies, in an effort to showcase its validity through practical implementation as well as prove its potential use in the real-world.

4.5 Usage Notes

This publicly available resource aims to provide a common ground for the evaluation of SDAs related to DDI signals, with a focus on assembling information from disparate sources that could provide support toward the clinical relevance of controls. The size of the reference set alongside the relatively large number of drugs and AEs considered enables a quantitative approach in algorithm performance evaluation in terms of SDAs developed for DDI signals.

Also, the supplementary single-drug positive and negative control sets could be used separately for signal detection in the context of adverse drug-event associations.

This reference set could be applied to a variety of data sources (e.g. electronic health record data, social media, literature). Here, we illustrate applicability in one such context.

4.5.1 FAERS screening for DDI signals

In pharmacovigilance, signal detection is largely based on methodologies that use disproportionality analysis, meaning that the observed counts are compared to the expected ones, assuming that the drug(s) and the events occur independently. For DDIs, similar measures of disproportionality have been described in the literature. For our analysis, the following statistical measures were considered:

1. An observed-to-expected shrunk interaction measure (Omega) [12]
2. The 'interaction coefficient' of a multiplicative additive baseline model (*delta_add*) [11]
3. A measure based on an adapted version of Multi-Gamma Poisson Shrinker (MGPS) model, called Interaction Signal Score (IntSS) [13].

A curated and standardised version of FAERS was used as the test data source. In FAERS, each spontaneous report contains information on administered drugs, AEs

				Medians	
Citation	SDA	Performance analysis in the original source	AUC (95% CI)	Score+	Score-
Noren et al., 2008 [12]	Omega	a. 5 positive examples (case studies); b. World Health Organization (WHO) database-wide screen	0.5670 (0.5534, 0.5806)	-1.6173	-2.3151
Thakrar et al., 2007 [11]	delta_add	4 positive and 4 negative controls tested on FAERS	0.4211 (0.4110, 0.4312)	0.00078	0.00417
Almenoff et al., 2003 [13]	IntSS	Beta blockers + Verapamil (as positive controls) and Angiotensin-converting enzyme (ACE) inhibitors/ Angiotensin-2 receptor blockers + Verapamil (as negative controls) tested on FAERS for impaired myocardial conduction	0.5041 (0.4921, 0.5162)	0.3669	0.3453

TABLE 4.1: Statistics related to the performance evaluation of three SDAs for DDIs using FAERS data. AUC, area under the receiver operating characteristic curve; Score+, median signal score for the set of positive controls; Score-, median signal score for the set of negative controls.

experienced, indications and demographic information (i.e. sex, age, etc). Using a slight modification of the Adverse Event Open Learning through Universal Standardization (AEOLUS) process that was described in published work [194], FAERS data files corresponding to the period 2004Q1-2018Q4 were standardised to the RxNorm Ingredient level terms and MedDRA PTs, so that compatibility of the reference set with the test data could be established. The total number of reports containing at least one drug and one AE was 9,203,239.

ROC analysis was performed using MATLAB function *perfcurve*. 95% confidence interval (CI) estimates for the Area Under the Receiver Operating Characteristic Curve (AUC) were calculated using leave-one-out cross-validation.

Table 4.1 shows statistics related to the performance of the SDAs in FAERS using our advocated reference standard. For each algorithm, the table provides the AUC scores with 95% CI estimates. As opposed to expectations, only two out of three SDAs for DDIs (i.e. Omega and IntSS) performed better than random guessing. Median signal scores for the set of positive controls were larger compared to those of negative controls in those two cases, which provides support for the validity of CRESCENDDI. Limited performance analyses in the studies describing the SDAs for DDIs as well as the absence of other efforts in the literature for quantitative performance assessment across multiple thresholds did not manage to provide an appropriate benchmark for this study.

4.5.2 FAERS screening for single-drug ADRs

For the single-drug reference set, empirical support was provided by applying three well-established SDAs to FAERS data. Positive controls (at the MedDRA PT level) were sourced from either the intersection (N= 12,318) or the union (N= 52,637) of the resources under consideration (i.e. BNF and SIDER) and ROC analysis for performance evaluation similarly to the previous section (**Table 4.2**).

Table 4.2 shows that EBGM outperformed PRR, which is in line with the results of other studies [7]. Also, improved performance was noticed in the case of resource intersection for all SDAs, which advocates the use of information that appears in multiple resources rather than relying on a single data source for deriving positive controls.

		Resource union Positive controls: 52,637 Negative controls: 12,141	**Resource intersection** Positive controls: 12,318 Negative controls: 12,141
Citation	SDA	AUC (95% CI)	AUC (95% CI)
Evans et al., 2001 [89]	Proportional Reporting Ratio (PRR)	0.5791 (0.5746, 0.5835)	0.5959 (0.5884, 0.6033)
DuMouchel, 1999 [92]	EBGM (Empirical Bayes Geometric Mean)	0.6308 (0.6259, 0.6356)	0.6593 (0.6510, 0.6675)
Bate et al., 1998 [91]	BCPNN (Bayesian Confidence Propagation Neural Network)	0. 7063 (0.7008, 0.7117)	0.7495 (0.7401, 0.7588)

TABLE 4.2: Statistics related to the performance evaluation of three SDAs for single drugs using FAERS data. AUC, area under the receiver operating characteristic curve; Score+, median signal score for the set of positive controls; Score-, median signal score for the set of negative controls.

4.5.3 Flexibility of the resource

This reference set has been mapped to standard vocabularies using the OHDSI framework, thus enabling linking to other ontologies and/or vocabularies, based on research needs, test data resources format, etc.

Considering the changing nature of available evidence for both positive and negative controls, reproduction of the resource in the future would be recommended to ensure that it is up to date. Although specific products (e.g. vaccines) were excluded given our focus on generating a reference set relevant to DDIs, slight modifications on the drug mapping process can allow the inclusion of controls pertinent to any additional products of interest.

The incorporation of additional information (e.g. shared indication, single-drug AE under the same High Level Term (HLT)/High Level Group Term (HLGT) MedDRA level, evidence levels) enables filtering and stratification of controls. The effect of stratification using those design criteria is outside the scope of this paper but is explored elsewhere.

4.6 Code Availability

The code used to generate this dataset is publicly available on a GitHub repository[2]. This code was developed and tested using: OHDSI standard vocabulary version v5.0 18-JAN-19[3], which includes: RxNorm version 20181203, RxNorm Extension version 2019-01-17, and MedDRA version 19.1. Database storage and operations were enabled using PostgreSQL 9.3. Drug and event mapping steps were performed using OHDSI Usagi version 1.2.7[4]. Web data extraction was performed using Python 3.6. Scores for the SDAs were calculated using Python 3.6 (Omega), SAS (*delta_add*) and R version 4.0.0 (IntSS, PRR, EBGM and BCPNN); AUC scores and CI estimates were calculated using MATLAB R2020b (*perfcurve* function).

[2] https://github.com/elpidakon/CRESCENDDI
[3] https://athena.ohdsi.org/
[4] https://github.com/OHDSI/usagi

Chapter 5

Exploring the impact of design criteria for reference sets on performance evaluation of signal detection algorithms: The case of drug-drug interactions

The content of this chapter is published in *Pharmacoepidemiology and Drug Safety.*

This chapter aims to identify the relative impact of different factors that could be potential sources of confounding on the performance evaluation of signal detection algorithms for DDI surveillance. It utilises the reference set of clinically-relevant DDIs that was developed in **Chapter 4**, leveraging its size and diversity to create smaller, custom-made reference sets considering multiple design criteria. These reference sets are then used to assess any differences observed in the quantitative evaluation of SDAs for two-way DDIs.

The key contribution of this chapter is the exploration of the impact of the relative composition of reference sets on evaluation metrics, which can potentially lead to modified conclusions regarding which methodologies are perceived to perform best. This is particularly relevant when using custom-made and small-in-size reference sets, which in many cases represent published work in the area of signal detection methods in pharmacovigilance, impeding a comprehensive and "fairer" evaluation. A modified version of the reference set that was established in **Chapter 4** is also developed and presented in this chapter as a supplementary resource. This is to accommodate the detection at the medical concept level, rather than just the MedDRA Preferred Term (PT) level which was directly possible using the original version of the reference set. The outcomes of the analysis presented in this chapter are taken into account for the method development

and evaluation in **Chapter 6** and act as a validation of the original assumption that large and diversified reference sets can support method benchmarking.

Abstract

Purpose

To evaluate the impact of multiple design criteria for reference sets that are used to quantitatively assess the performance of pharmacovigilance signal detection algorithms (SDAs) for drug–drug interactions (DDIs).

Methods

Starting from a large and diversified reference set for two-way DDIs, we generated custom-made reference sets of various sizes considering multiple design criteria (e.g. adverse event background prevalence). We assessed differences observed in the performance metrics of three SDAs when applied to FDA AdverseEvent Reporting System (FAERS) data.

Results

For some criteria, the impact on the performance metrics was neglectable for the different SDAs (e.g. theoretical evidence associated with positive controls), while others (e.g. restriction to designated medical events, event background prevalence) seemed to have opposing and effects of different sizes on the Area Under the Receiver Operating Characteristic Curve (AUC) and positive predictive value (PPV) estimates.

Conclusions

The relative composition of reference sets can significantly impact the evaluation metrics, potentially altering the conclusions regarding which methodologies are perceived to perform best. We therefore need to carefully consider the selection of controls to avoid misinterpretation of signals triggered by confounding factors rather than true associations as well as adding biases to our evaluation by "favouring" some algorithms while penalizing others

Key points

- Performance assessment of SDAs in pharmacovigilance has often relied on the generation of custom-made reference sets of limited size that consider ad-hoc exclusion or inclusion criteria to define eligible controls.
- SDA performance assessment might be biased based on the selected benchmarks, as each methodology can be impacted to a different extent by different confounders.
- We tested 14 design criteria for reference sets in the case of DDIs, showing that some of them considerably affected the performance and comparative evaluation of different SDAs for DDI surveillance while others did not have a significant effect.

- Overall, this analysis advocates the utilization of large, to the extent possible, reference sets that are less likely to suffer from the overrepresentation of controls that make different SDAs behave in different ways due to confounding. Any decision to restrict the evaluation set using specific criteria should be carefully justified.

Plain Language Summary

Reporting of suspected side effects experienced by patients following drug approval is a key component to identifying novel drug safety issues. Statistical methods are then used to analyze reports and reveal signals of novel associations between drugs and side effects. Performance evaluation of those methods traditionally relies on custom-made reference sets of limited size that consider ad-hoc exclusion or inclusion criteria to define eligible controls. However, each method can be impacted to a different extent by those criteria, as they can act as potential confounders. This study investigated the impact of 14 criteria on three methods that have been developed to detect signals of potential adverse drug–drug interactions, showing that some of them had opposing effects or effects of different levels of magnitude on the performance of the different methods. The relative composition of reference sets can therefore significantly affect the evaluation metrics, potentially altering the conclusions regarding which methodologies are perceived to perform best. The selection of controls should be carefully performed to avoid misinterpretation of signals triggered by confounding factors rather than true associations as well as adding biases to our evaluation by "favouring" some algorithms while penalising others.

5.1 Introduction

Monitoring drug safety issues during the post-approval phase requires reporting of suspected drug-related adverse reactions by healthcare professionals, patients, and pharmaceutical companies. The reports are collected in spontaneous reporting system (SRS) databases, such as the FDA Adverse Event Reporting System (FAERS) database in the US, the Eudravigilance database in the EU, and the Yellow Card database in the United Kingdom. These databases form an important part of the pharmacovigilance strategy since they not only contain information on adverse events (AEs) and suspected drugs, but also details regarding concomitant medications, indications, and patient demographics.

By applying statistical methods known as signal detection algorithms (SDAs), novel associations between drugs and AEs (i.e. signals) that have not been identified in clinical trials can be identified in the SRS data. Given the absence of a control group, SDAs predominantly rely on disproportionality analysis, which calculates the degree of disproportional reporting of drug-AE combinations compared to what would be expected if there were no association between them [195]. However, the presence of synthetic

associations (i.e. causative covariates that have not been taken into account or remain unobserved) can lead to confounding, either upward or downward, thus generating faulty associations between the drug and the AE and complicating the detection of safety signals [196, 197, 198]. For example, reporting quality issues arising from a poor distinction between symptoms of disease-related AEs and treatment effects of drugs (or drug combinations) is a result of a synthetic association called confounding by indication [199, 105].

The practice of using larger clusters of medical terms to perform quantitative signal detection in pharmacovigilance has been widely discussed in the literature [195, 200]. Many previous efforts investigated the impact of the Medical Dictionary for Regulatory Activities (MedDRA) granularity on signal detection tasks [201, 202]. Also, many studies have considered the use of term grouping to identify relevant reports [203, 204]. However, recommendations from the IMI-PROTECT project suggest that signal detection at the PT level should be considered the standard approach in real-life pharmacovigilance [202, 205].

The development of novel SDAs in pharmacovigilance requires the existence of appropriate reference sets that can be utilised both for absolute performance evaluation as well as for comparison with existing methodologies. Given that each SDA, depending on the applied modelling, might be impacted to a different extent by a confounder, the performance evaluation might be biased based on the selected benchmarks. The challenge of building appropriate reference sets in pharmacovigilance has been previously acknowledged in the literature [206, 97, 98, 207]. Most studies have attempted to comparatively evaluate SDAs by testing their performance against custom-made reference sets, often limited in size [13, 8, 191] or not publicly available [208, 209] which commonly consider ad-hoc inclusion or exclusion criteria to generate positive and negative controls. Examples of such criteria include those related to AE background prevalence (given that, in disproportionality analysis, the denominator signifies the expected rate of occurrence) [7], disease-related AEs [210], AE seriousness [210, 211] or evidence associated with positive controls [7, 210, 211, 212, 9]. The criteria are typically used to attempt to address the limitations of disproportionality analysis and to tackle issues with potential confounders.

In the case of adverse drug–drug interactions (DDIs), signal detection is considered more complicated, with the existing methodology being less mature compared to the one in the case of signals for single drugs. A previous study has suggested that the detection of DDI-related signals might suffer from multiple confounders [14]. For example, concomitant medications appear to be a significant source of con-founding (i.e. the signal associated with a drug combination triggered by drugs that are usually given concomitantly but not signifying true adverse drug–drug-event associations). In addition, only limited efforts exist in the literature to generate reference sets related to two-way DDIs [13, 191, 14, 147].

In this study, we aim to explore the relative impact of different factors that could be potential sources of confounding on the performance evaluation of existing methods for

signal detection of DDIs. By utilising a large and diversified reference set, we were able to create custom-made reference sets considering multiple ia to assess any differences observed in the quantitative evaluation of SDAs tailored for two-way DDIs.

5.2 Methods

5.2.1 Data sources

5.2.1.1 FAERS data - spontaneous reports

We used a curated and standardised version of the publicly available FAERS database. The data pre-processing pipeline was based on the Adverse Event Open Learning through Universal Standardisation (AEOLUS) process and included removal of duplicate reports, drug name normalization at the RxNorm ingredient level, and AE mapping to MedDRA Preferred Terms (PTs) [194]. The curated data set included 9,203,239 reports containing at least one drug and one AE between 2004 (Q1) and 2018 (Q4), with 3,973,749 (43.18%) reports mentioning more than one drug. Each drug was considered equivalent in the analysis irrespective of its reported role (i.e. primary suspect; secondary suspect; concomitant; and interacting).

5.2.1.2 Reference sets for DDIs

CRESCENDDI, a reference set for two-way DDIs, was the primary source of controls [21]. This reference sets covers 454 drugs and 179 adverse events mapped to RxNorm Ingredient and MedDRA PT concepts, respectively, from the Observational Medical Outcomes Partnership (OMOP) Common Data Model (version 5). We used 4455 positive and 4544 negative controls from CRESCENDDI that were also present in the curated FAERS dataset (hereafter called PT Reference Set).

To accommodate and test the impact of MedDRA granularity to detect signals at the medical concept (MC) level, we extended CRESCENDDI by building PT groups (event groups), where possible, that are relevant to the adverse events described in the original reference set. These groups were formed by examining Standardised MedDRA Queries (SMQs) and event definitions from a time-indexed reference standard by Harpaz et al. [10] and were manually reviewed for clinical relevance. In total, 20 adverse events from CRESCENDDI were deemed suitable for extension to the MC level (**Table 5.1**). A full list of the event groups is available online[1] in the Supporting Information of the publication (**Appendix S1**). The new reference set (hereafter called MC Reference Set) contained 1,097 positive and 614 negative controls and is also available online[2] in the Supporting Information of the publication (**Appendix S2**).

[1] https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Fpds.5609&file=pds5609-sup-0002-AppendixS1.xlsx

[2] https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Fpds.5609&file=pds5609-sup-0003-AppendixS2.xlsx

Name		
Acute kidney injury	Drug-induced liver injury	Myopathy
Acute psychosis	Hypoglycaemia	Priapism
Angioedema	Hypertension	Rhabdomyolysis
Arrhythmia	Hypoglycaemia	Tachycardia
Bradycardia	Hyponatraemia	Thrombocytopenia
Cardiac failure	Hypothyroidism	Torsade de pointes
Drug withdrawal syndrome	Lactic acidosis	

TABLE 5.1: Medical Concepts in the MC Reference Set.

5.2.2 Data mining

We performed the case/non-case analysis at two different levels, based on the reference sets that we utilised. The first one was restricted to the reports that included the PT that was related to each control from the PT Reference Set. The second one considered as cases all the reports that contained any of the PTs that were part of the MC linked to the control in the MC Reference Set.

For example, the case/non-case analysis for a control related to torsade de pointes resulted in two contingency tables: the first one only considered the PT "Torsade de pointes" to retrieve case reports, while the second one included the following terms (as PTs): "Electro-cardiogram QT interval abnormal", "Electrocardiogram QT prolonged", "Long QT syndrome", "Torsade de pointes", "Ventricular tachycardia". Non-cases included the reports without the aforementioned PTs, while reports containing more than one of the relevant PTs linked to the MC were not double-counted.

5.2.3 Design criteria

Table 5.2 shows the design criteria that were considered as potential confounding factors, which fall into the following categories: (i) evidence level; (ii) event seriousness; (iii) event frequency; (iv) potential confounding by indication; and (v) potential confounding by concomitant medication. PT Reference Set controls were stratified based on each of the design criteria, forming suitable restricted sub-sets of different sizes in each case, depending on the criterion under consideration. MC Reference Set could not be stratified using categories (ii) and (iii).

5.2.4 PT prevalence

The impact of reference set restriction by PT prevalence on the Area Under the Curve (AUC) estimates was also examined. The PT prevalence was calculated in the curated FAERS data set as the frequency of PTs from reports containing at least one drug. We grouped the 179 PTs from the PT Reference Set using quartile binning of their prevalence. The controls were then stratified into four groups (Groups Q1–Q4) based on their PTs by considering the respective PT prevalence quartile.

Category	Design criterion (DC)	Description
Evidence level	BNF - Study	Interactions where the information is based on formal study including those for other drugs with the same mechanism, for example, known inducers, inhibitors, or substrates of cytochrome P450 isoenzymes or P-glycoprotein.
	BNF - Theoretical	Interactions that are predicted based on sound theoretical considerations. The information may have been derived from in vitro studies or based on the way other members of the same class act.
	BNF - Anecdotal	Interactions based on either a single case report or a limited number of case reports.
	Micromedex - Established	Controlled studies have clearly established the existence of the interaction.
	Micromedex - Theoretical	The available documentation is poor, but pharmacologic considerations lead clinicians to suspect the interaction exists; or documentation is good for a pharmacologically similar drug.
	Micromedex - Probable	Documentation strongly suggests that the interactions exist, but well-controlled studies are lacking.
Event seriousness*	EMA Important Medical Event (IME) Terms	Any untoward medical occurrence that at any dose: - results in death, - is life-threatening, - requires inpatient hospitalization or prolongation of existing hospitalization, - results in persistent or significant disability/incapacity, or - is a congenital anomaly/birth defect.
	EMA Designated Medical Event (DME) Terms	Medical conditions that are inherently serious and often medicine-related (e.g. Stevens-Johnson syndrome). This list does not address product-specific issues or medical conditions with high prevalence in the general population.
Event frequency*	Common PTs	PT prevalence $\geq$ 90th percentile of the prevalence of PTs reported in FAERS.
	Rare PTs	PT prevalence $\leq$ 10th percentile of the prevalence of PTs reported in FAERS.
Potential confounding by indication	AE is an indication - True	The AE is also an indication for at least one of the two drugs from the drug–drug-event triplet under consideration.
	AE is an indication - False	The AE is not an indication for either of the drugs from the drug–drug-event triplet under consideration.
Potential confounding by concomitant medication	Shared indications - False	Drug pairs that share at least one indication are excluded.
	Shared indications - True	Only drug pairs that share at least one indication are considered.

TABLE 5.2: Categories and descriptions of design criteria for reference sets that could affect performance evaluation of SDAs for DDI surveillance. *Note: The categories marked with an asterisk (*) contain design criteria that were not applicable to the MC Reference Set.*

5.2.5 SDAs

Three SDAs that have been previously described in the literature were considered:

1. An observed-to-expected shrunk interaction measure (Omega) [12];
2. The "interaction coefficient" in a linear regression model with additive baseline (delta_add) [11];
3. measure based on an adapted version of Multi-Gamma Pois-son Shrinker (MGPS) model, called Interaction Signal Score (IntSS) [13].

5.2.6 Impact of MedDRA granularity on SDAperformance evaluation

To assess the impact of MedDRA granularity on the SDAs that were considered in this study, we performed a Receiver Operating Characteristic (ROC) analysis to examine the difference in AUC when considering matched controls from the two reference sets.

5.2.7 Estimation of design criteria impact on SDA performance evaluation

For each reference set and design criterion, we simulated the generation of a constrained reference set by randomly drawing an equal number (1:1) of positive and negative controls from the restricted control subset that used the specified design criterion for control stratification. An unconstrained reference set of equal size was generated in each case by following a similar process but using the original reference set. This sampling generation process took into account the correlation between the two sets, as the probability of drawing one control for the constrained reference set did not affect the probability of drawing any control for the unconstrained reference set. The size of the simulated reference sets varied from 100 to $2 \times N_{max}$, where N_{max} was determined by either the number of positive or negative controls (depending on which one was smaller) in each of the restricted subsets. For each SDA, we calculated: (i) AUC scores; and (ii) positive predictive value (PPV) for fixed sensitivity values (i.e. 0.60, 0.75, and 0.90) for both reference set types (i.e. constrained and unconstrained) by performing 1000 simulations. The statistics of the samples were summarised by fitting a Normal distribution, for which we report the mean and variance. The difference of the means of AUC (AUC_{diff}), and PPV (PPV_{diff}) (with 95% confidence intervals) were the target measures. The probability of AUC_{diff} being non-zero, $P(|AUC_{diff}| > 0)$, was also estimated under the normality assumption:

$$|AUC_{diff}| \sim \mathcal{N}(|\mu_{AUC_{Restricted\_ROC}} - \mu_{AUC_{Unrestricted\_ROC}}|, \sqrt{\sigma^2_{AUC_{Restricted\_ROC}} + \sigma^2_{AUC_{Unrestricted\_ROC}}}) \tag{5.1}$$

where μ is the mean, σ is the standard deviation, and $F_{AUC_{diff}}$ is the normal cumulative distribution function (CDF) of AUC_{diff}. **Figure 5.1** illustrates the simulation

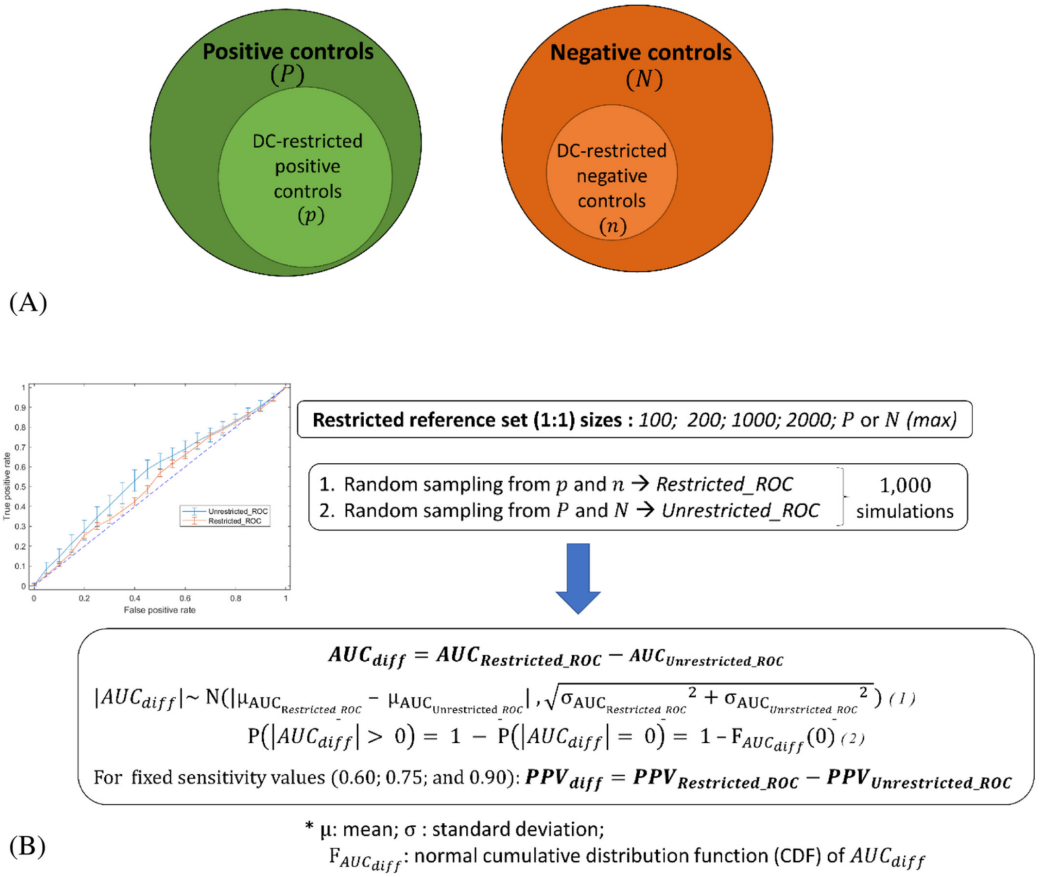


FIGURE 5.1: (A) Initial positive and negative control sets (P and N) and their respective restricted subsets (DC-restricted, p and n) when applying a design criterion; (B) Simulation workflow for the differences in AUC (AUC_{diff}) and PPV (PPV_{diff}) when considering the specified design criterion.

workflow for the calculation of differences in AUC scores and PPV when considering the various design criteria.

5.3 Results

The total number of positive and negative controls when applying each of the design criteria to the PT Reference Set is presented in **Figure 5.2**. In cases where restricted subsets contained both positive and negative controls (**Figure 5.2A**), the maximum number of controls considered from each type (i.e. positive or negative) to form simulated reference sets (N_{max}) is denoted with white colour in the respective bar. For the design criteria under the Evidence level category, where the restriction was only applied to positive controls (**Figure 5.2B**), N_{max} was defined as the total number of positive controls in the respective restricted subsets. Apart from two cases (i.e. *Shared indications - False* and *AE is an indication - False*), positive controls out-numbered negative controls in the restricted subsets. The simulated reference sets varied in size, with N_{max} ranging from 131 to 3,568. Hence, more than 250 positive and negative controls were

SDA	PT Reference Set AUC (95% CI)	MC Reference Set AUC (95% CI)
Omega	0.6011 (0.5704, 0.6317)	0.5406 (0.5150, 0.5662)
delta_add	0.4645 (0.4408, 0.4882)	0.4956 (0.4721, 0.5191)
IntSS	0.5374 (0.5100, 0.5648)	0.4885 (0.4654, 0.5117)

TABLE 5.3: Statistics related to the performance evaluation of three SDAs for DDIs using matched controls from the PT Reference Set and MC Reference Set.

considered for every design criterion. For the MC Reference Set, the restricted subsets were smaller in size (**Table A.4**). Three design criteria (*BNF - Anecdotal*, *BNF - Theoretical*, and *AE is an indication - True*) were not tested with the MC Reference Set, as their N_{max} was less than or equal to 100. **Figure 5.3** provides the frequency distribution of PT prevalence in: (a) the set of unique PTs in the PT Reference Set; (b) PT Reference Set positive controls; and (c) PT Reference Set negative controls. The right-tailed distribution of unique PTs in CRESCENDDI shows that the data set was populated with less common PTs, with only a small number of them having a prevalence over 0.01 in FAERS. Similar trends were present in the curves of the positive and negative controls, with the latter consisting of more cases with a higher PT prevalence in FAERS. The 1st, 2nd, and 3rd quartiles for the PT prevalence were 0.000343, 0.00135, and 0.00410, respectively. The total number of positive and negative controls for each group formed using PT prevalence quartile binning is shown in **Figure 5.4**. Group Q3 contained the largest volume in the case of positive controls, with Group Q1 and Group Q2 being considerably smaller, while negative controls showed an increasing trend while moving to groups of higher PT prevalence.

The MedDRA granularity affected the SDA performance metrics in different ways (**Table 5.3**). Omega and IntSS performed worse at the MC level as opposed to the PT level, with their mean AUC score dropping by 0.0605 and 0.0489, respectively. For Omega, there was a statistically significant decrease in the AUC between the PT and MC level evaluations. In the case of delta_add, the mean AUC slightly increased (0.0311) when considering the MC level, however without outperforming Omega.

By plotting AUC_{diff} for a fixed constrained reference set size of 100 and ordering design criteria by increasing range of AUC_{diff} values among the three SDAs (**Figures 5.5** and **A.4**), points that lie above the x-axis signify positive estimates for AUC_{diff}, meaning that the design criterion had a positive effect on the calculated AUC. Conversely, points below the x-axis were associated with a negative effect on the AUC when the specific design criterion was applied to constrain the reference set. Also, for the different sizes of restricted reference sets using the PT Reference Set and the MC Reference Set, AUC_{diff} value estimates and associated probabilities of a non-zero AUC_{diff} estimate were plotted (**Figures A.5** and **A.6**). With the PT Reference Set, the largest AUC_{diff} values were associated with the *EMA Designated Medical Event Terms* criterion (between 0.071 and 0.095), while *Common PTs* resulted in negative values in the range of -0.041 to -0.021 for the AUC_{diff} measure for all SDAs. In the case of the MC

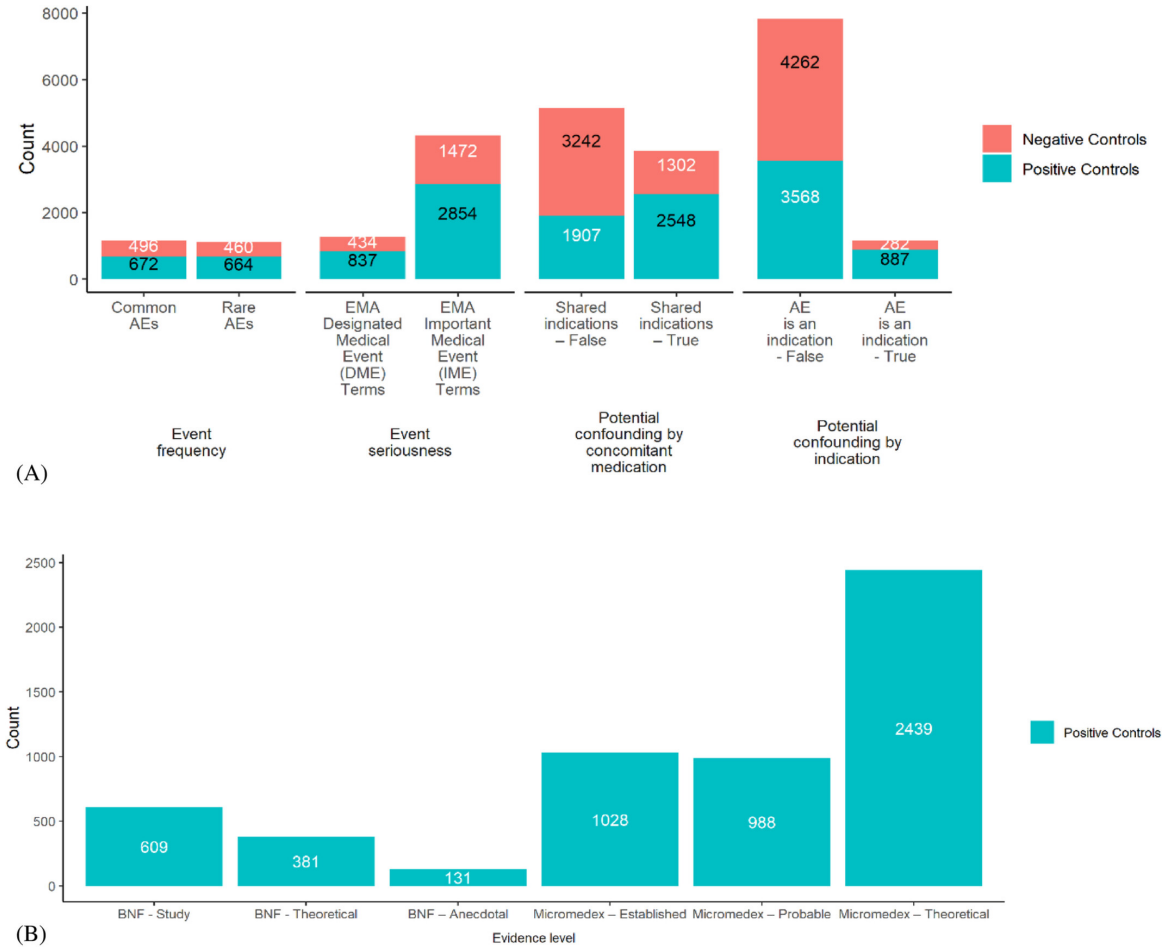


FIGURE 5.2: (A) Number of positive and negative controls from the PT Reference Set for each of the different design criteria when the restricted subsets contained both control types. The maximum number of controls considered from each type to form simulated reference sets (N_{max}) is denoted with white colour in the respective bar; (B) Number of PT Reference Set positive controls for the Evidence level design criteria, where the restriction could not be applied to negative controls.

Reference Set, *BNF - Study* had the largest positive impact on all AUC_{diff} values (between 0.098 and 0.051), while negative AUC_{diff} values derived from *Shared indications - True* and *AE is an indication - False* (up to -0.043). Some design criteria affected the performance evaluation of all three SDAs in a similar way and level of magnitude (e.g. *BNF - Anecdotal*, *BNF - Study*), while others (e.g. *Shared indication - False*) seemed to have opposing and different in size effects on AUC estimates.

Tables A.5 and **A.6** report the PPV_{diff} estimates (with 95% CIs) for the different design criteria, and a fixed reference set size of 100, for the PT Reference Set and MC Reference Set, respectively. For both reference sets and sensitivity equal to 0.60, some design criteria affected PPV in opposing ways among the different SDAs. For example, *Shared indications - False* resulted in negative PPV_{diff} estimates for Omega and IntSS (in the range between -0.029 and -0.021) as opposed to positive ones for delta_add (around 0.051). For other design criteria (i.e. *BNF - Study* and *EMA - Designated Medical Events*), PPV_{diff} estimates were positive across the different sensitivity values

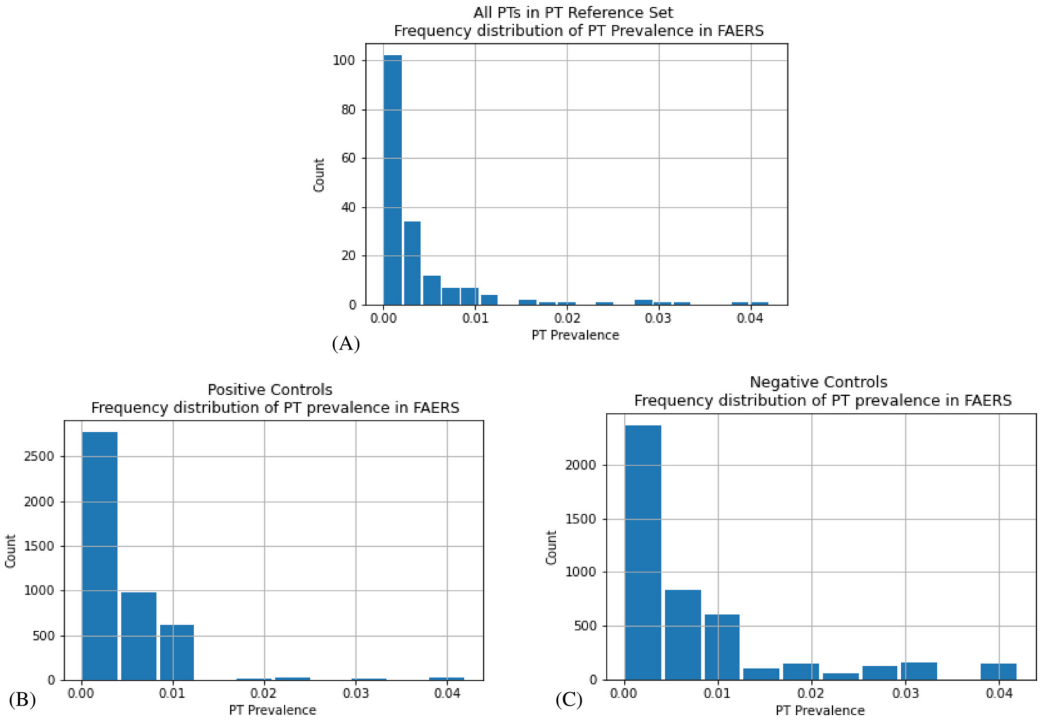


FIGURE 5.3: Frequency distribution of PT prevalence in FAERS for: (A) the set of unique PTs in the PT Reference Set; (B) PTs contained in the PT Reference Set positive controls; and (C) PTs contained in the PT Reference Set negative controls.

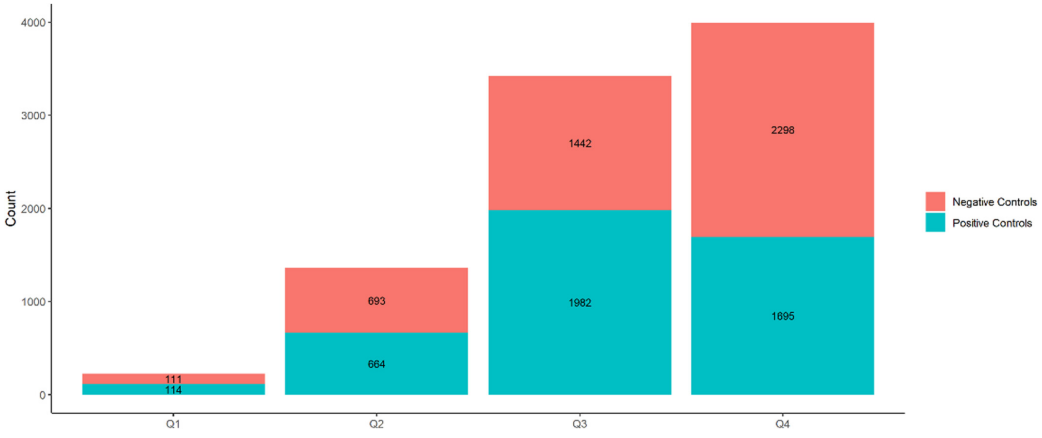


FIGURE 5.4: Number of positive and negative controls for groups Q1–Q4 that were formed using PT prevalence quartile binning, with Q1 containing the controls with the lowest prevalence and Q4 the highest one.

FIGURE 5.5: AUC_{diff} for a fixed restricted reference set size of 100 for: (A) the PT Reference Set; (B) the MC Reference Set. Design criteria are ordered by increasing range of AUC_{diff} values among the three SDAs. The dot size represents the probability of the estimated score, AUC_{diff}, being non-zero.

for all three SDAs. For a sensitivity value of 0.90, PPV_{diff} for the different design criteria were close to zero in all cases (values between 0.029 and -0.009).

With the PT Reference Set, we identified three main categories:

(i) Positive AUC_{diff} values

  (a) BNF - Anecdotal
  (b) EMA IME Terms
  (c) BNF - Study
  (d) Micromedex - Probable
  (e) EMA DME Terms
  (f) Rare PTs

(ii) Negative AUC_{diff} values

  (a) Common PTs
  (b) Micromedex - Theoretical

(iii) Mixed effect on AUC_{diff} values

  (a) AE is an indication - False
  (b) AE is an indication - True
  (c) Micromedex - Established
  (d) BNF - Theoretical
  (e) Only drug pairs that share at least one indication are included
  (f) Drug pairs that share at least one indication are excluded.

With the MC Reference Set study, *Omega* and $IntSS$ were affected in a similar way by the different design criteria. *BNF - Study* and *Micromedex - Established* had a positive impact on the target measure for all SDAs, while excluding AEs related to drugs' indications (*AE is an indication - False*) or only considering drug pairs with shared indications as controls (*Shared indications – True*) negatively affected the SDA performance in all cases. In terms of PT prevalence (**Figure 5.6**), there was a similar trend for Groups Q1–Q3, with AUC_{diff} metric increasing for all algorithms as we moved to more common PTs. However, this relationship appears to be reversed in Group Q4, which contains the most frequent PTs in FAERS from the original data set, for Omega and *delta_add*, showing a negative impact on their AUC.

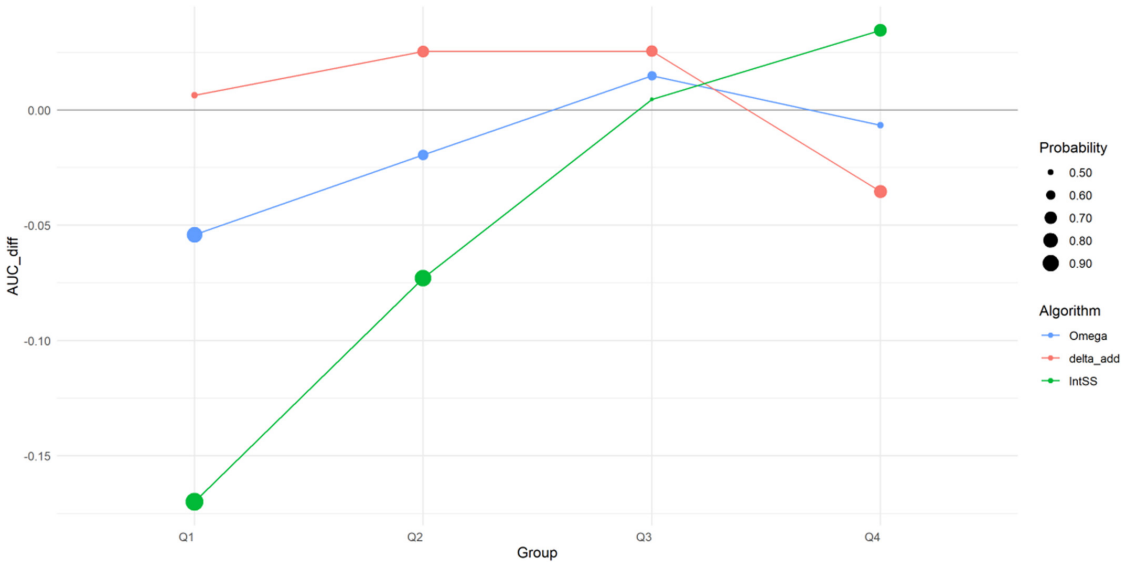


FIGURE 5.6: AUC_{diff} values for Groups Q1–Q4 relevant to PT prevalence. The dot size represents the probability of the estimated score, AUC_{diff}, being non-zero.

5.4 Discussion

This study provides a systematic evaluation of the impact of multiple design criteria for reference sets on the comparative assessment of signal detection methodologies of adverse DDIs in SRS data. Performance assessment of SDAs in pharmacovigilance has often relied on the generation of custom-made reference sets that consider exclusion or inclusion criteria to define eligible controls. Thus, the motivation behind this research was to examine how different criteria could affect the evaluation, potentially altering the conclusions regarding which algorithms perform best. Our study highlighted that the relative composition of reference sets might significantly impact the evaluation metrics. Some criteria affect the comparison of different methodologies, such as the restriction of controls to only include PTs from the EMA's designated medical event list. Other criteria that were thought to have a potential effect on the evaluation process (e.g. anecdotal evidence supporting a positive control) were not found to significantly change the observed difference in metrics among the methodologies, as all of them were influenced in a similar way (**Figure 5.5**). Moreover, we found that the size of the reference set did not have a considerable effect on the AUC_{diff}, although the associated probability of that metric being non-zero increased when considering larger sizes (**Figures A.4** and **A.5**). Apart from the AUC, commonly applied sensitivity values were considered to identify the impact of design criteria on PPV. For most of the design criteria (e.g. EMA Designated Medical Events, Micromedex evidence categories), PPV_{diff} values were affected consistently with the AUC_{diff} estimates across the three different SDAs. For the highest sensitivity that was considered (0.90), the difference in PPV was in most cases neglectable.

Given the inability of SDAs to account for all potential confounding factors that are present in SRS data, each methodology might be impacted to a different extent by a

confounder. At the same time, there might be cases where signals are triggered by those confounding factors. As an illustrative example, the majority of DDI signals identified using IntSS in the original research paper were composed of drug pairs that are usually given concomitantly (e.g. antibiotics) [14].

We therefore need to consider the selection of appropriate controls to avoid misinterpretation of signals triggered by confounding factors rather than true associations as well as adding biases to our evaluation by "favouring" some algorithms while penalising others. On the other hand, by attempting to completely remove all potential sources of confounding in our evaluation sets, we are more likely to fail to demonstrate their utility in real-life application, which should be determined by their ability to perform at a commensurate level when it is applied prospectively to identify novel signals in SRS databases [97, 98]. Overall, this analysis advocates the utilization of large, to the extent possible, reference sets when it comes to comparative performance assessment, that are less likely to suffer from the overrepresentation of controls that make different SDAs behave in different ways due to confounding. Also, regarding novel reference sets, the decision to restrict the evaluation set using specific design criteria should be adequately supported.

A major concern about reference sets used for prospective signal detection in pharmacovigilance revolves around the validity of established (i.e. well-known) positive controls to test the performance of algorithms. This aspect has been widely discussed in the literature [97, 98, 99]. It has been acknowledged that the combination of established and emerging positive controls might be a better choice when we try to evaluate the prospective performance and compare different methodologies because merely emerging positive controls (i.e. recently detected ADRs) cannot establish a reliable reference standard [8]. Especially for DDIs, the establishment of reference sets by only using emerging positive controls turns out to be particularly challenging, as we would end up having a very limited number of controls to be able to quantitatively assess differences in the performance of the SDAs under comparison. A solution to this issue would be to perform a backdated analysis to detect the time point that a signal of a true positive association (positive control) was first highlighted, as proposed in previous studies [213]. However, this back-dated analysis was not possible in this study due to the lack of a time-indexed reference set for DDIs. A previous study compared the performance of SDA algorithms for DDI surveillance between established and emerging positive controls, with *Omega* and *delta_add* showing increased specificity but diminished sensitivity in the latter case [191]. In our analysis, the results related to the evidence level are consistent with what we would expect to see. In terms of theoretical DDIs, it is common for drug interaction compendia to extend the included DDIs to the drug class level, therefore covering drugs under the same drug class that sometimes, but not necessarily, have a similar interaction profile. Our results showed declining AUC values when considering theoretical DDIs (i.e. *Micromedex - Theoretical*) as opposed to improvements with established ones (i.e. *BNF - Study* and *Micromedex - Established*). On the other hand, all three examined

methodologies demonstrated enhanced performance against anecdotal DDIs from BNF and probable DDIs from Micromedex. However, the former category represented only a small fraction of the overall positive cases contained in the PT Reference Set (2.94%).

In terms of event background prevalence, the simulation results suggest that, if we restricted the evaluation set to specific ranges of PT prevalence, the conclusions would change, that is, the sole choice of common PTs would have an inverse impact on the comparative evaluation as to rare AEs. We know that SRS data are predominantly used in the post-marketing setting to spot rare adverse reactions that ave not been revealed during clinical trials. However, the use of SRS data for the detection of DDIs can be considered a different scenario, given that clinical trial data are not sufficient to detect adverse reactions of drug combinations due to inherent limitations (e.g. patient recruitment processes that exclude people taking multiple medications). Hence, the detection of novel DDI-related adverse reactions, even with a common background rate, in SRS data should be of special interest.

Disease-related AEs are a challenging issue in the effort to generate signals using SRS data, as confounding by indication can occur. A previous study reported that around 5% of the total reports for any drug in FAERS mention a drug's indication as an adverse event [214]. This might be related to poor reporting quality or intended to report a disease's exacerbations due to a drug. Our results support that the choice of excluding disease-related AEs (i.e. *AE is an indication - False*) did not have a significant effect on the AUC across the SDAs with the PT Reference Set, while it decreased the performance of all SDAs with the MC Reference Set. On the other hand, Omega demonstrated deteriorated performance in the scenario of detecting controls with AEs that were drugs' indications at the same time (i.e. *AE is an indication - True*), while the other two SDAs did not seem to be substantially affected by this design criterion.

Event seriousness has been used to build reference sets and assess SDA performance, as it could be utilised to filter signals in real-life pharmacovigilance settings [210, 211]. Our study suggests that, by only considering "significant" events, bias is introduced to evaluating SDAs that could be potentially used in routine pharmacovigilance to detect a broader set of events. Also, given that DMEs are rare events (i.e. have low prevalence) with a high drug-attributable risk, it is important to note that this category might have been confounded to an extent by other design criteria categories that were considered in our study, such as the event frequency.

Quantitative signal detection is only one aspect of the more complex framework before a safety signal is validated. In the case of adverse DDI surveillance, previous studies have considered triage filters alongside disproportionality analysis to direct preliminary signal assessment [215, 216]. These filters might be less suitable depending on the type of DDI. For example, there are more filters relevant to pharmacokinetic DDIs (e.g. cytochrome P450 activity) as opposed to pharmacodynamic interactions. Although the clinical significance of the differences between SDAs that are reported in this study might be questioned, it is important to note that quantitative methods

for adverse DDI surveillance remain way less mature compared to those for single-drug safety surveillance, also considering the additional complexity that is inherent to DDIs. In this way, the potential impact on real-world pharmacovigilance could not be refuted, as even small changes in the performance of an SDA might have a considerable impact on the number of generated signals that are captured for further evaluation, leading to either missed signals or large amounts of potential signals that need to be evaluated, thus increasing the manual effort needed. It is also important to note that the three SDAs that were included in our study are not implemented to the same extent in the real world. *Omega* and *IntSS* are two of the major methods that we understand to be used for routine pharmacovigilance screening for DDIs. *delta_add* is a less mature method that is described in the literature, for which, as far as we are aware, is not as widely used in practice. Although this study provides a novel framework for studying how SDA performance may change by considering different criteria for eligibility of controls, there are some limitations worth mentioning. First, only a single test data set (i.e. FAERS) was utilised for the purposes of this study. Also, CRESCENDDI was the only reference set utilised to generate estimates of the impact on AUC, in the absence of another comprehensive data set that could be used as a comparative source. We acknowledge that, by modifying the CRESCENDDI data set to consider adverse events at the MC level, we ended up with a smaller reference set that only included controls that could be represented by event groups (e.g. angioedema). This can have an impact on the extrapolation of the results and conclusions drawn from our analysis when considering single PTs as opposed to event groups. Additionally, for the determination of hit versus miss, it is important to consider how the results calculated at the PT level can depict the signal generation at the MC level. For example, if one SDA signals polymorphous ventricular tachycardia and another one signals torsade des points at the PT level, they have both made the same classification in real-world pharmacovigilance, as both would have triggered the same case review by a diligent pharmacovigilance organization. The performance of SDAs was only assessed using the default values provided in the original research papers describing those methods (e.g. tuning parameter for shrinkage, α, equal to 0.5 in the case of *Omega*). Finally, the aspect of unbalanced reference sets was not explored in this study (i.e. positive to negative control ratio different from 1:1), since previous studies in pharmacovigilance have evaluated SDAs using asymmetrical reference sets [8, 211, 10].

5.5 Conclusions

This study revealed a varying impact of design criteria for reference sets on the performance metrics of three SDAs that are used for DDI post-marketing surveillance. This analysis showcases that the design of reference sets should be performed carefully, as the comparison of SDA performance might be affected by the choices made when building a

reference set and the decision to restrict the evaluation to specific controls. Also, it highlights the need to establish frameworks that can make use of large and disparate data sources to support the generation of open-source, flexible benchmarks in pharmacovigilance. These benchmarks can not only ensure transparency and enable a fair evaluation of SDA performance, but also provide a strong foundation that promotes productive research in pharmacovigilance signal detection methodologies.

5.6 Data availability statement

The CRESCENDDI data set that supports the findings of this study is openly available in Figshare at https://doi.org/10.6084/m9.figshare.c.5481408.v1.

Chapter 6

Identifying drug-drug interactions in spontaneous reports utilising signal detection and biological plausibility aspects

The content of this chapter is a manuscript under preparation for submission to *Clinical Pharmacology and Therapeutics*

Building upon the reference set that was developed in **Chapter 4** and the exploration of existing approaches for quantitative signal detection of DDIs from **Chapter 5**, this chapter aims to develop and evaluate a novel framework for identification of drug-drug interaction (DDI) signals in post-marketing surveillance data. This framework combines disproportionate reporting with a signal refinement step to assess the biological plausibility of the generated signals. Signal refinement is based on the establishment of a network that aims to leverage both pharmacokinetic (enzyme and transporter) and pharmacodynamic (drug target) information to inform biological plausibility, which had not previously been explored in the scope of DDIs.

The key contributions of this chapter are: (i) a novel signal detection method for DDIs based on Bayesian hypothesis testing aiming to produce a pharmacology-driven output; (ii) the development of a systems pharmacology framework of established associations between biological nodes (i.e. drug targets, drug ingredients, and AEs) that is used to refine the signals of potential DDIs and assess their biological plausibility; and (iii) the exploration of two case studies demonstrating the applicability of the novel approach for real-life detection and refinement of potential signals of DDIs linked to QT interval prolongation and rhabdomyolysis.

Abstract

The identification of safety signals in the post-marketing setting largely relies on signal detection algorithms (SDAs) to reveal disproportionally reported adverse events (AEs) following drug exposure. In the case of drug-drug interactions (DDIs), the detection of novel signals is complex, with SDAs being less mature. At the same time, drug-target associations have not been adequately utilised to enhance signal detection and evaluation processes in this setting. This research aims to detect signals of disproportionate reporting that are attributed to drug combinations and discern those from signals arising from the constituent drugs by applying a Bayesian hypothesis testing framework. For a subset of AEs of interest, potential DDI signals generated by a FAERS database-wide screening were evaluated in terms of biological plausibility, with the aid of drug-target, target-AE, as well as metabolism (i.e. enzyme and transporter) information.

6.1 Introduction

Polypharmacy rates have been increasing in the Western world over the last few decades, with important implications for patient safety and a higher risk of drug-drug interactions (DDIs). A DDI occurs when a drug's effect on the body changes in the presence of another drug and can be categorised as either pharmacodynamic (i.e. at the level of drug receptors or other drug targets) or pharmacokinetic (i.e. due to changes in the drug absorption, metabolism or excretion) [118]. DDIs can be beneficial or adverse. For the remainder of this paper, we will focus on adverse DDIs.

When a DDI is clinically observable, the possible outcome is either a lack of drug efficacy or an adverse drug reaction (ADR) that cannot be attributed to the individual drugs separately. Pre-marketing clinical trials and drug interaction studies during drug development are inadequate to fully capture adverse effects linked to DDIs, which seem to contribute to a significant proportion of drug-related adverse effects in clinical practice [5, 217]. The prediction of clinically-relevant DDIs is also a difficult task due to human inter-individual variability (e.g. poor versus rapid metabolisers, genetic susceptibility) and differences between how the human body acts compared to animal or in silico models. Thus, post-marketing surveillance is essential to identify complications due to interacting drug combinations.

Post-marketing safety surveillance databases, called spontaneous reporting system (SRS) databases, collate reports that contain information on suspected drug complications. Examples include VigiBase (maintained by the Uppsala Monitoring Centre), the US Food and Drug Administration Adverse Event Reporting System (FAERS) database, and Eudravigilance from the European Medicines Agency. Disproportionality analysis is a popular approach for signal detection in SRS data that aims to identify 'unexpectedness' in reported data by comparing the observed rate of adverse event (AE) occurrence to the expected one using the background of the rest of the database [195]. While these

databases and methods remain at the forefront to identify drug safety issues in the post-marketing setting, there are multiple inherent challenges. Those include the inability to get an accurate estimate of the incidence or reporting rates of AEs, the potential presence of reporting biases due to the voluntary nature of reporting (i.e. either under- or over-reporting), stimulated reporting, duplicate reporting arising from multiple sources reporting the same adverse incident, data quality issues, data loss when moving from an unstructured to a structured format of information in SRS data, data incompleteness, etc [218].

For DDI surveillance, existing SDAs test the degree of 'unexpectedness' of associations between two drugs and an AE, assuming that the baseline model (i.e. the contributions of the individual effect of each drug in the absence of an interaction) is either multiplicative [11] or additive [11, 12]. Other approaches have compared the disproportionality measure for a drug combination (e.g. EB05 and EB95 scores) relative to the measures for the individual drugs for a specific AE to identify signals of potential DDIs [13]. At the same time, the calculated measure does not allow us to directly infer whether, and which of, the individual drugs give rise to unexpected reporting on their own compared to their expected background rate. However, we would expect that, for example, a pharmacokinetic DDI would involve disproportionate reporting rates for the victim drug and the combination, while the reporting rate for the perpetrator drug would not differ from the background rate. On the other hand, a pharmacodynamic interaction would involve both unexpected reporting rates for the two drugs individually, plus a departure from the baseline additive model when the two drugs are used concomitantly. The drug safety signal lifecycle involves several steps following the detection of a signal, such as signal prioritization, signal evaluation and, if appropriate, risk communication and management interventions [113]. The processes of signal evaluation rely heavily on in-depth manual clinical review to assess the biological plausibility of the identified signal. The well-known Bradford-Hill criteria are traditionally used at this stage [106, 219]. Some previous efforts attempted to integrate some biological aspects into the signal generation process for DDIs by considering metabolic enzymes and transporters to assess the plausibility of a signal from a mechanistic perspective [215]. However, as DDIs, apart from pharmacokinetic, can also be pharmacodynamic, leveraging supporting information relevant to the biological mechanisms that might be involved in the generation of this signal in the first place could enable mechanism-based filtering of signals [220].

Recent efforts aimed to integrate systems pharmacology aspects into pharmacovigilance signal detection, such as connections between drug targets and AEs as putative AE mechanistic pathways to identify signals of single-drug side effects [15, 18]. Other studies have attempted to predict drug safety profiles in the post-marketing setting based on target similarity with comparator drugs that have known safety profiles [221, 222, 196]. In the context of DDIs, the number of possible drug combinations in the clinical setting and the amount of data that are constantly accumulating indicate the potential

of coupling data mining and biological plausibility aspects to incorporate a mechanistic understanding of the associations in our data mining methodologies. However, target-AE associations have not been utilised in related studies to assess the performance of SDAs for DDI surveillance. This approach could potentially limit the number of spurious associations that are identified as signals and help uncover novel DDIs that cannot be solely detected in SRS data.

The aims of this study were: (a) to advocate an SDA for adverse DDIs that could produce a pharmacologically-motivated output by detecting increasing reporting rates in SRS data while being able to distinguish signals that might arise from constituent drugs; and (b) to utilise a systems pharmacology framework of established associations between biological nodes (i.e. drug targets, drug ingredients, and AEs) to refine the signals of potential DDIs and automate an assessment of their biological plausibility.

6.2 Methods

6.2.1 Data sources

6.2.1.1 CRESCENDDI reference set

We used CRESCENDDI, an open-access reference set for adverse DDIs, as the source of positive and negative controls for DDIs [21]. All controls were additionally stratified using information from CRESCENDDI regarding their individual drug safety profile (i.e. single-drug ADRs). For example, the combination of paroxetine and ibuprofen can lead to an increased risk of bleeding. At the same time, both drugs are individually associated with haemorrhagic events as adverse effects. Thus, it was possible to classify the control based on both the combined behaviour (i.e. whether there is evidence that the drug combination interacts leading to a specific AE) and the individual ones (i.e. whether each of the drugs is separately associated with the AE).

For a DDI control to be classified for the behaviour of either of its constituent drugs (**Table 6.1**), we checked whether the ADR list of the drug contained any closely related MedDRA Preferred Terms (PTs) to the DDI control's PT. These 'closely related' PTs were found under the same SMQ groups as the PT under consideration. For example, if a DDI control was associated with Hypertension (PT), the ADR lists of the constituent drugs were checked for the presence of any PTs from the SMQ 'Hypertension' (e.g. hypertensive crisis). This classification enabled a stratified analysis to assess the ability to detect disproportionate reporting for the different categories of individual drug behaviour: (i) Both (i.e. both drugs are independently related to the AE), (ii) One (i.e. only one of the constituent drugs is linked to the AE), and (iii) None (i.e. neither of the drugs is related to the AE).

Category	Description
00,0	A DDI negative control where the adverse event (AE) is not a side effect for either of the drugs
00,1	A DDI positive control where the AE is not a side effect for either of the drugs.
10,0	A DDI negative control where the AE is a side effect only for the first but not the second drug.
10,1	A DDI positive control where the AE is a side effect only for the first but not the second drug.
01,0	A DDI negative control where the AE is a side effect only for the second but not the first drug.
01,1	A DDI positive control where the AE is a side effect only for the second but not the first drug.
11,0	A DDI negative control where the AE is a side effect for both drugs.
11,1	A DDI positive control where the AE is a side effect for both drugs.

TABLE 6.1: Categories of individual drug behaviour for control stratification.

6.2.1.2 FAERS database

We curated and standardised the publicly available version of the FAERS database[1] using the Adverse Event Open Learning through Universal Standardization (AEOLUS) process [194] and considered spontaneous reports covering the period from the 1st quarter of 2004 through to the 4th quarter of 2018. Drug concepts were standardised to the RxNorm Ingredient level terms and AE concepts to MedDRA PTs, to ensure compatibility with the reference set. The curated dataset contained 9,203,239 reports that included at least one drug and one AE, with 3,973,749 (43.18%) reports mentioning more than one drug. Each drug was considered equivalent in the analysis irrespective of its reported role (i.e. primary suspect; secondary suspect; concomitant; and interacting).

6.2.1.3 Open Targets

Open Targets [17] is a freely available resource that combines multiple public data sources regarding potential therapeutic drug targets and associated information, including evidence regarding target safety (i.e. associations between drug targets and potential unintended adverse consequences). After downloading the core annotations for drug molecules and targets as Parquet files[2], we extracted the following information:

i Drug-target associations

ii Target-AE associations representing target safety liabilities.

Drugs were mapped to RxNorm Ingredients, while target mappings were available as both Ensembl stable IDs (e.g. ENSG00000157764) (as the primary identifiers) and UniProtKB accession numbers (e.g. P15056). In Open Targets, AE terms were reported

[1] https://www.fda.gov/drugs/questions-and-answers-fdas-adverse-event-reporting-system-faers/fda-adverse-event-reporting-system-faers-latest-quarterly-data-files

[2] https://platform.opentargets.org/downloads

inconsistently in various ontologies[3], thus we retrieved relevant MedDRA synonym terms where possible via querying the OLS WEBAPI[4] or using a manually curated mapping table[5] for Human Phenotype Ontology (HPO) [223] entities.

6.2.1.4 DrugBank

DrugBank [224] is an open-access knowledge base that contains information related to drugs and drug targets. We downloaded DrugBank version 5.1.9 as an XML file[6]and extracted drug enzyme, transporter, and target data. Apart from therapeutic (i.e. primary) targets, DrugBank also included secondary ones that were considered to complement primary target information for drugs from Open Targets.

6.2.1.5 Search Tool for the Retrieval of Interacting Genes/Proteins (STRING)

STRING 11.5 [225] is an open-access database that contains protein-protein interactions (PPIs) in various organisms mined from multiple evidence channels. Each PPI is associated with a confidence score in the dataset, which is computed by combining the probabilities from the different channels and correcting for the probability of randomly observing an interaction. We only selected PPIs of higher confidence in humans (i.e. scores over 700 out of a maximum score of 1,000) for further consideration.

6.2.2 Development of a pharmacology-driven signal detection algorithm for adverse DDIs

This approach can be also adapted to detect signals of two-way DDIs. Considering a drug-drug-event (DDE) triplet, $D_1 - D_2 - AE$, FAERS reports can be categorised based on the presence or absence of: (i) the first drug (D_1); (ii) the second drug (D_2); and (iii) the event (AE), thus enabling the storage of the various category counts in a 4-by-2 contingency table (**Figure 6.1a**). We developed a novel SDA for detecting signals of DDIs in SRS data (**Figure 6.1b**). The focus of this novel approach is on detecting signals related to two-way DDIs in SRS data and distinguishing them from signals that arise due to individual complications of the constituent drugs. Given the presence or absence of either of the drugs, D_1 and D_2, there are four (potentially identical) rates of occurrence of the event (AE). There are eight distinct hypotheses for which of the rates differ (and which, if any, are identical): each hypothesis relates to whether each of the individual drugs gives rise to complications and whether a two-way DDI exists. By assuming a Beta-Binomial model, we assigned Beta prior distributions to the different rates in the context of each hypothesis. We defined the hyperparameters of these priors by maximising the likelihood of the AE counts in FAERS for all AEs that are mentioned in

[3]Experimental Factor Ontology (EFO), Human Phenotype Ontology (HPO), Gene Ontology (GO), Mondo Disease Ontology, Orphanet Rare Disease Ontology, National Cancer Institute thesaurus (NCIt)

[4]https://www.ebi.ac.uk/ols/docs/api

[5]https://github.com/elpidakon/BANet/blob/main/appendix/Appendix_S1.xlsx

[6]https://go.drugbank.com/releases/latest#full

the database (n = 19,931). For each DDE triplet, we calculated the posterior probability (and so log-likelihood) for a DDI being present by summing the probabilities related to all hypotheses that relate to the rate of the *AE* being reported when both drugs are present being different to the rate of the *AE* being reported in all cases other than when both drugs are present. The log-likelihood ratio and log posterior odds ratio were the chosen metrics to observe a changing or an increased probability of *AE* occurrence indicative of a DDI. A detailed mathematical description of the algorithm and the prior hyperparameter estimation method are available in **Appendix B.1**.

6.2.3 Performance assessment - Model validation

To assess the performance of the novel SDA, we implemented the following evaluations using Receiver Operating Characteristic (ROC) analysis: (1) a sensitivity analysis to check the impact of the hyperparameters of the prior distributions on the SDA performance (see **Appendix B.2**); (2) a comparison of the novel SDA with three other existing methodologies for DDI surveillance, namely *Omega* [12], the *Interaction Signal Score (IntSS)* [13], and *delta_add* [11]; (3) PT-restricted analysis for specific AEs of interest with a sufficient number of controls (see **Table A.7**): *bradycardia*; *gastrointestinal (GI) haemorrhage*; *haemorrhage*; *hypertension*; *hypoglycaemia*; *myopathy*; *QT prolongation*; *rhabdomyolysis*; and *torsade de pointes*; and (4) stratified analysis based on the individual drug safety profile by utilising the classification of controls from CRESCENDDI.

6.2.4 A systems pharmacology framework for signal refinement

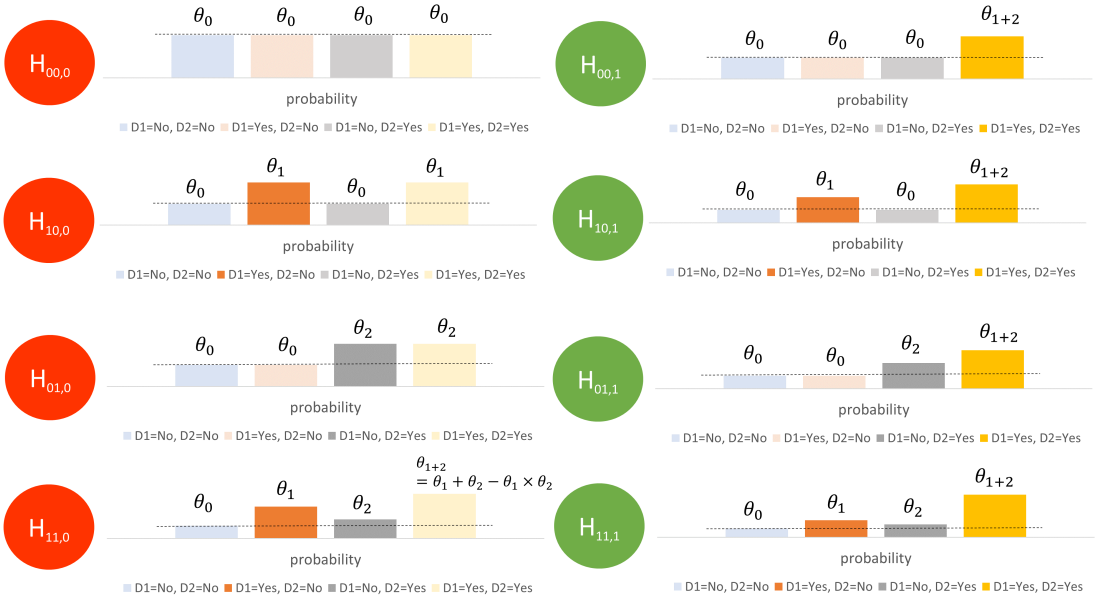
We constructed the ***Biological Attribute Network*** (**Figure 6.2**) by integrating multiple types of information pertinent to systems pharmacology. The network included the following types of nodes: (a) drug ingredients (RxNorm/RxNorm Extension concepts); (b) AEs (MedDRA PT concepts); (c) targets (ensembl/UniProt/HGNC IDs); (d) enzymes; and (e) transporters (UniProt/HGNC IDs). The links between the nodes represented: (a) drug-target information from Open Targets and DrugBank; (b) target-AE associations from Open Targets; and (c) target-target associations (i.e. PPI) data from the STRING database.

We considered three different measures using the ***Biological Attribute Network*** for signal refinement of signals for DDIs. For a DDE triplet, $D_1 - D2 - AE$:

- if the network contained all three nodes $(D_1, D2, AE)$ and links connecting each node with any number of targets, we calculated the total number of nodes that were included in the union of the individual shortest paths between each drug and the AE under consideration (hereafter called *shortest path measure*);
- if the network contained the drug nodes $(D_1, D2)$ and links connecting each drug node with any number of enzymes or transporters, we calculated the number of

	AE = Yes	AE = No
D_1 = No, D_2 = No	n_{001}	n_{000}
D_1 = Yes, D_2 = No	n_{101}	n_{100}
D_1 = No, D_2 = Yes	n_{011}	n_{010}
D_1 = Yes, D_2 = Yes	n_{111}	n_{110}

(a)



(b)

FIGURE 6.1: (a) Contingency table for signal detection of two-way DDIs (two drugs and an adverse event); (b) Illustrative presentation of the Bayesian hypothesis testing framework for two drugs. Hypotheses marked with red indicate that no signal related to a DDI is detected, while hypotheses marked with green indicate the presence of a signal indicative of an interaction between D_1 and D_2. If we consider two drugs, D_1 and D_2, then we can assume that one of the following scenarios will hold: (1) The probability of AE reporting is equal irrespective of the presence of D_1 and D_2; (2) The rate of occurrence of AE only changes when both D_1 and D_2 are present; (3) The rate of occurrence of AE changes when D_1 is present, irrespective of the presence of D_2; (4) The rate of occurrence of AE does not change when only D_2 is present, but it changes when only D_1 is present and it is different to the one where D_1 and D_2 are both present; (5) The rate of occurrence of AE changes when D_2 is present, irrespective of the presence of D_1; (6) The rate of occurrence of AE does not change when only D_1 is present, but it changes when only D_2 is present and it is different to the one where D_1 and D_2 are both present; (7) The rate of occurrence of AE changes to a different degree when either D_1 or D_2 is present, with the rate being equal to an independent risk assumption of no interaction when both D_1 and D_2 are present; (8) The rate of occurrence of AE changes to a different degree when either D_1 or D_2 is present and different to the independent risk assumption of no interaction when both D_1 and D_2 are present.

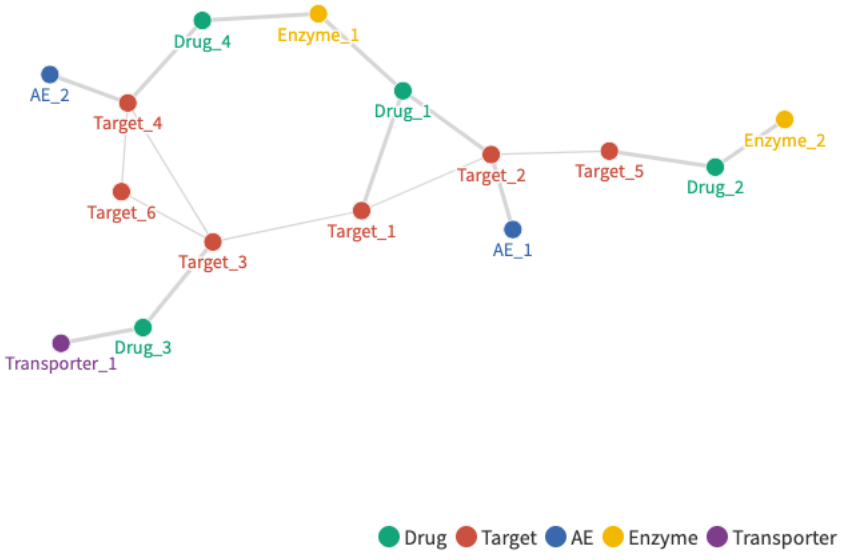


FIGURE 6.2: Illustrative presentation of the Biological Attribute Network.

common enzymes and transporters between them (hereafter called *enzyme measure* and *transporter measure*, respectively, and collectively called *PK measures*).

We then estimated via ROC curve analysis the combined performance of each of the three types of measures with the log-likelihood ratio metric using logistic regression.

6.2.5 Novel signal evaluation – Case studies

We selected two AEs of interest (*QT interval prolongation*, *rhabdomyolysis*) to run a FAERS-wide screening of signals. For each AE, we extracted all drug pairs found in at least 5 FAERS reports and generated suitable contingency tables. We calculated the log-likelihood ratios, shortest path measures and both PK measures to rank the DDE triplets. We removed DDE triplets that were present in any of three established DDI online resources (i.e. the British National Formulary [163], IBM Micromedex [165], and the French Medicines Agency Thesaurus for DDIs [164]). We also did not consider drug pairs that were under the same Anatomic Therapeutic Class (ATC) 4th level category, as those belonging to the same chemical/pharmacological class and are not taken concomitantly. We then extracted the top 20 associations: (a) when only applying the SDA (i.e. log-likelihood ratio); and (b) when applying a binary logistic regression model that considered the SDA and either of the signal refinement framework measures (i.e. *shortest path measure* for QT interval prolongation and *PK measures* for rhabdomyolysis) to estimate predicted probabilities by utilising the logistic regression model coefficients

derived from the reference set (see previous section). We compared the rankings of the top associations from either approach to assess the impact on the relative change of ranking of the drug pairs.

For rhabdomyolysis, we calculated the log-likelihood ratio scores related to drug pairs containing a statin (ATC code: C10AA), which is a drug class known to cause an increased risk of rhabdomyolysis due to interaction with other drugs (i.e. fibrates, macrolides, and fusidic acid) and compared them with the scores of ezetimibe (i.e. another lipid-lowering agent) with the same drugs.

6.3 Results

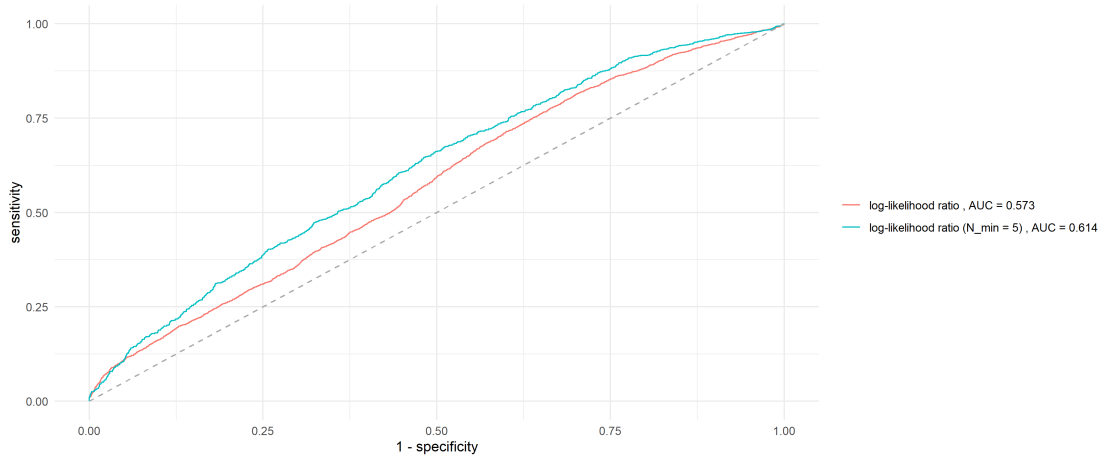
6.3.1 Evaluation of the novel signal detection algorithm for adverse DDIs

We assessed the novel SDA performance in FAERS using 4,455 positive and 4,544 negative DDI controls from CRESCENDDI that involved 442 drug ingredients and 168 AEs as MedDRA PTs in total. The log-likelihood ratio, detecting a changing probability indicative of a DDI, performed slightly better compared to the log posterior odds ratio, which monitored increases in that probability (AUC: 0.574 and 0.548, respectively). Thus, the log-likelihood ratio was then used for comparison with other SDAs. By taking the subset of controls from the evaluation set that were found in at least 5 FAERS reports (3,507 in total; 2,213 positive and 1,294 negative), the AUC score of the log-likelihood ratio increased to 0.614 (**Figure 6.3**).

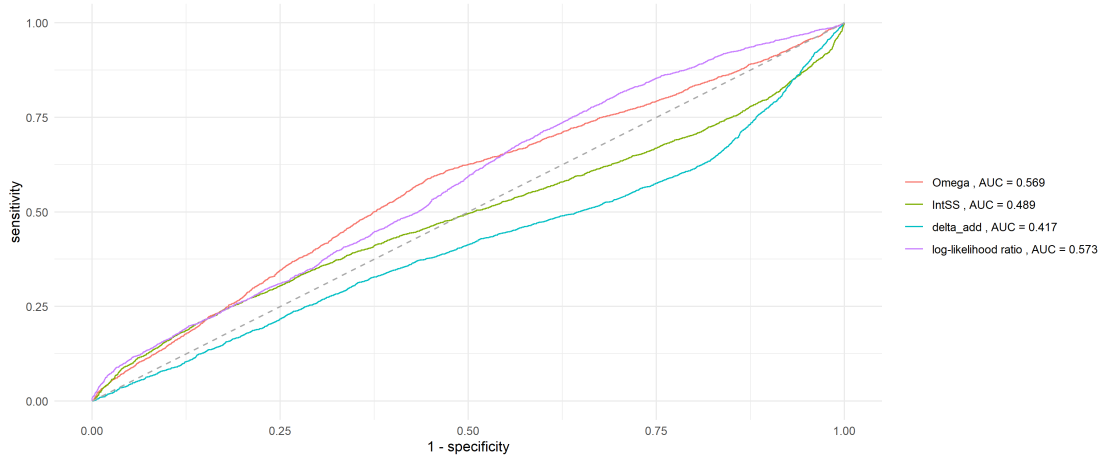
The selection of beta prior's hyperparameters impacts the results produced by the novel SDA. We investigated the effect of the chosen prior hyperparameter set (α_0 and β_0) on the SDA performance assessment using ROC curve analysis. The AUC scores varied between 0.502 and 0.599, with the lowest performance corresponding to a Beta prior for the null hypothesis with hyperparameters $\alpha_{0|00,0} = \beta_{0|00,0} = 4$ and the highest one to a Beta prior with hyperparameters estimated from the FAERS AE rate distribution (see **Appendix B.2**).

In terms of comparison with other SDAs for DDI surveillance, the log-likelihood ratio outperformed Omega, which showed the highest performance (AUC: 0.569) among the existing SDAs, followed by IntSS (AUC: 0.489) and $delta_a dd$ (AUC: 0.417). Also, the log-likelihood ratio achieved better specificity for high sensitivity values compared to Omega (**Figure 6.3**).

From the selected AEs that were individually considered for ROC curve analysis, the novel approach produced the following AUC scores: hypoglycaemia (AUC: 0.742), rhabdomyolysis (AUC: 0.674), bradycardia (AUC: 0.652), myopathy (AUC: 0.628) (**Figure 6.4**). The SDA performance was lower for haemorrhage (AUC: 0.528), GI haemorrhage (AUC: 0.542), hypertension (AUC: 0.546), QT prolongation (AUC: 0.568), and Torsade de pointes (AUC: 0.515).



(a)



(b)

FIGURE 6.3: ROC curve analysis for: (a) log-likelihood ratio, using the whole evaluation set (red line) or the subset of controls that were found in at least 5 FAERS reports (blue line); (b) comparative assessment of SDAs for DDI surveillance: Omega (red line); Interaction Signal Score (green line); *delta_add* (blue line); and log-likelihood ratio (purple line).

The stratified version of the reference set in terms of individual drug behaviour contained 4,212 controls in the *Both* category, 2,312 in the *One* category and 1,365 in the *None* category. With this reference set, the overall log-likelihood ratio performance was higher in the *None* category (AUC: 0.645) compared to the *One* (AUC: 0.578) and *Both* (AUC: 0.533) categories (**Figure 6.5a**). The relative performance among the three different categories was similar in the case of log posterior odds ratio (**Figure 6.5b**).

6.3.2 Improving signal detection of DDIs using systems pharmacology aspects

The number of nodes and links from the individual data resources that were added to the Biological Attribute Network are provided in **Table 6.2**. In total, the network contained 1,311 drugs, 351 AEs, 16,814 targets, 325 enzymes and 204 transporters. There were

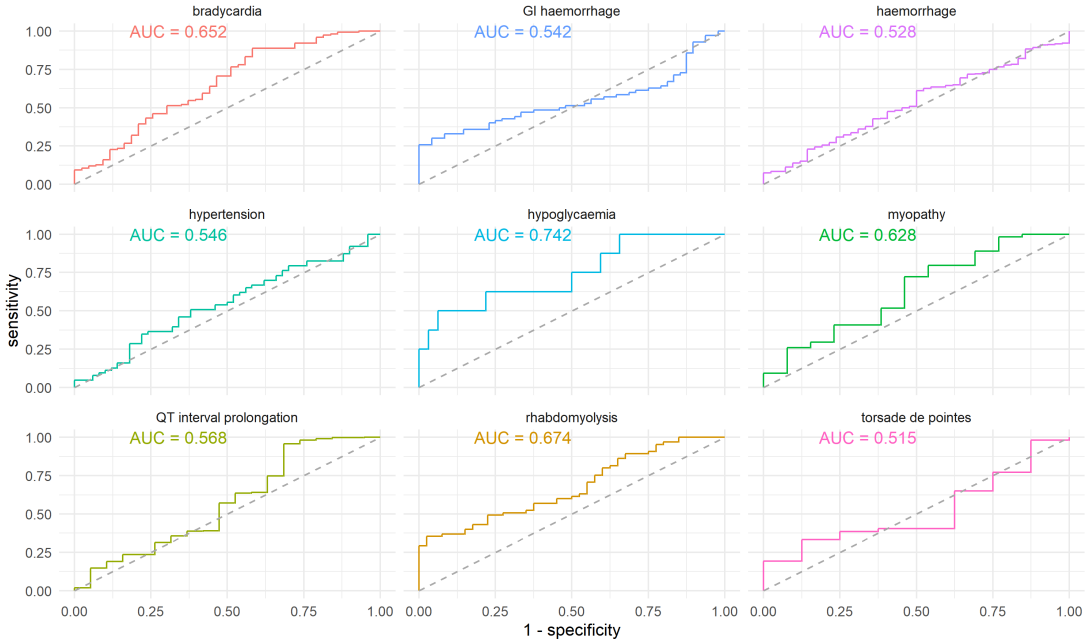
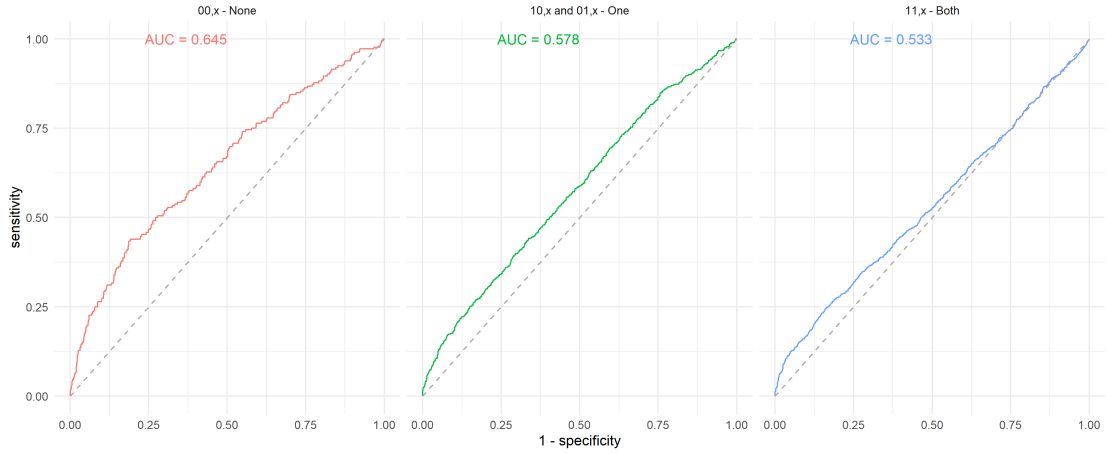


FIGURE 6.4: Receiver Operating Characteristic (ROC) curve analysis for specific MedDRA Preferred Terms (PTs): (a) rhabdomyolysis; (b) QT interval prolongation; (c) myopathy; (d) hypertension; (e) hypoglycaemia; (f) haemorrhage; (g) gastrointestinal haemorrhage; and (h) bradycardia.

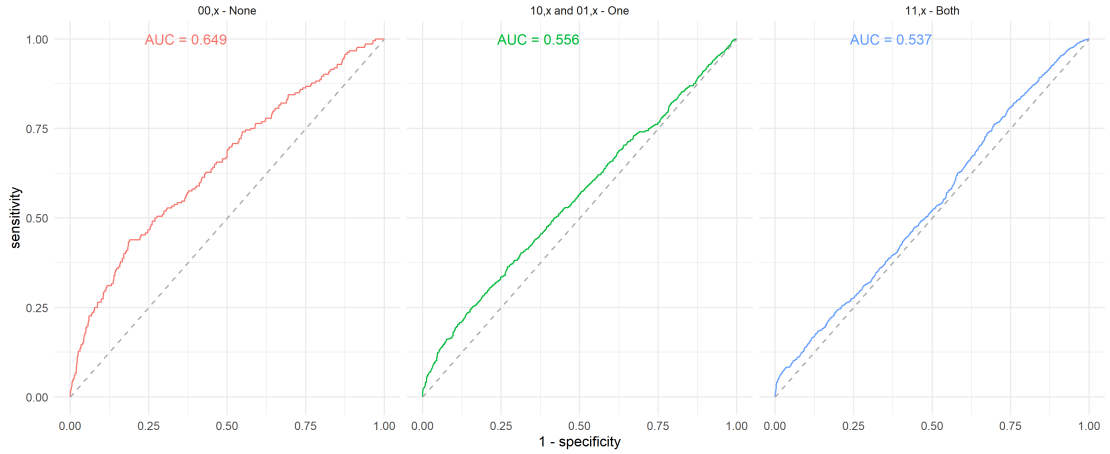
Type	Source	Count
Nodes		
Drugs	Open Targets	1,311
Targets	Open Targets	819
Targets	DrugBank	2,180
AEs	Open Targets	351
Targets	STRING	16,814
Enzymes	DrugBank	325
Transporters	DrugBank	204
Links		
Drug-Target	Open Targets	3,663
Drug-Target	DrugBank	10,035
Target-AE	Open Targets	1,060
Target-Target	STRING	505,968
Drug-Enzyme	DrugBank	4,966
Drug-Transporter	DrugBank	2,108

TABLE 6.2: Number of nodes and links contained in the Biological Attribute Network.

3,663 drug-target and 1,060 target-AE links from Open Targets, while the number of drug-enzyme and drug-transporter links was 4,966 and 2,108, respectively. The number of high-confidence STRING target-target associations was 505,968.



(a)



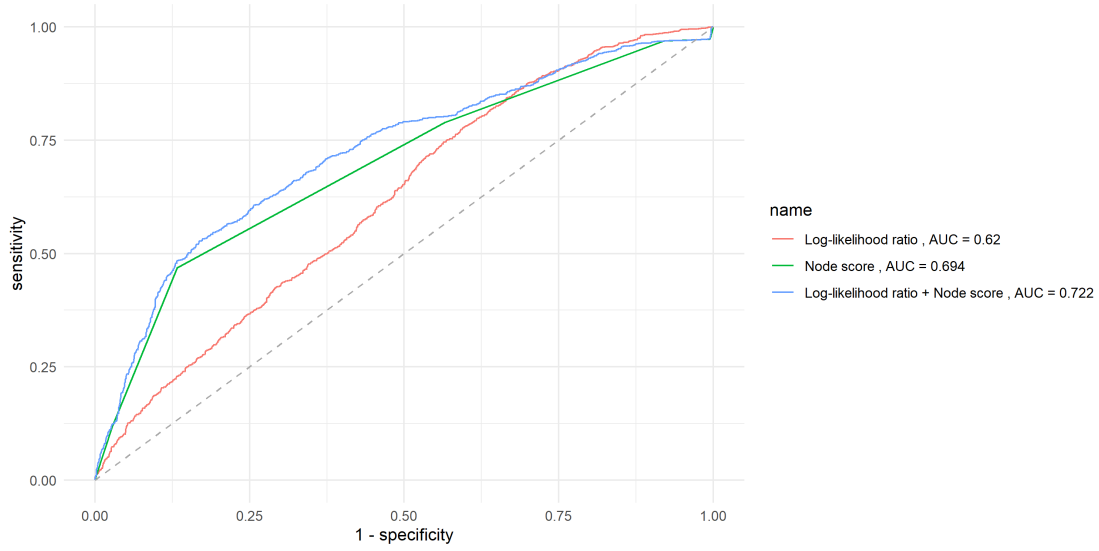
(b)

FIGURE 6.5: Individual drug behaviour Receiver Operating Characteristic (ROC) curves (a) using log-likelihood ratios; (b) using log posterior odds ratios.

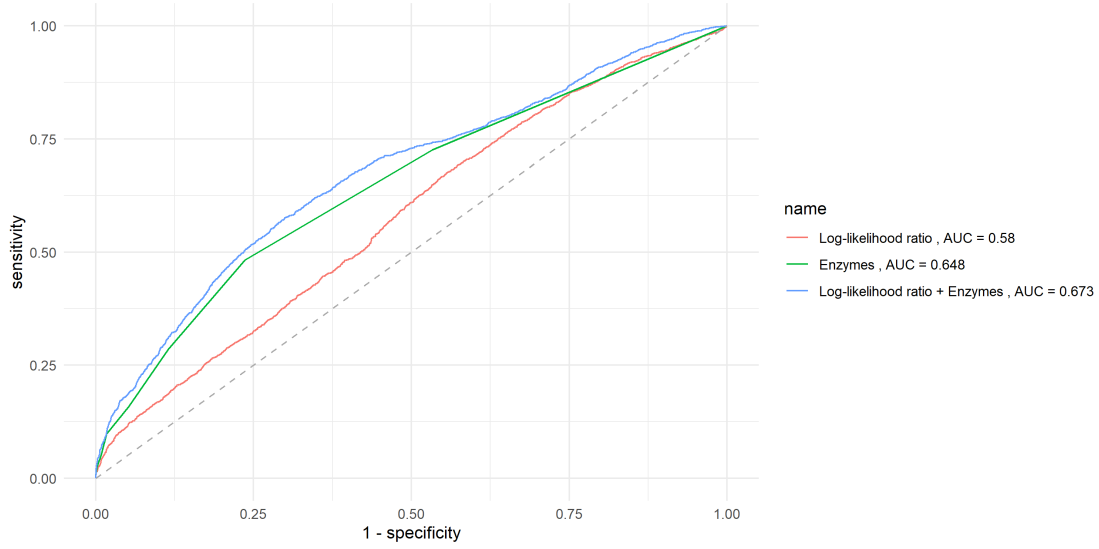
The application of the systems pharmacology framework for signal refinement produced ROC curves with higher AUC values when the log-likelihood ratio was combined with any of the three measures (shortest path, enzyme, and transporter measures) (**Figure 6.6**). The shortest path measure combined with the log-likelihood ratio produced an increase to the AUC score from 0.620 to 0.722 using the relevant controls with nodes that were present in the ***Biological Attribute Network***. Similarly, for both PK measures, the AUC scores of the combined models (i.e. log-likelihood ratio + enzymes; log-likelihood ratio + transporters) were higher to the ones considering only the log-likelihood ratio (from 0.580 to 0.673 for consideration of enzymes; from 0.568 to 0.653 for consideration of transporters).

6.3.3 Exploring signal evaluation for selected AEs

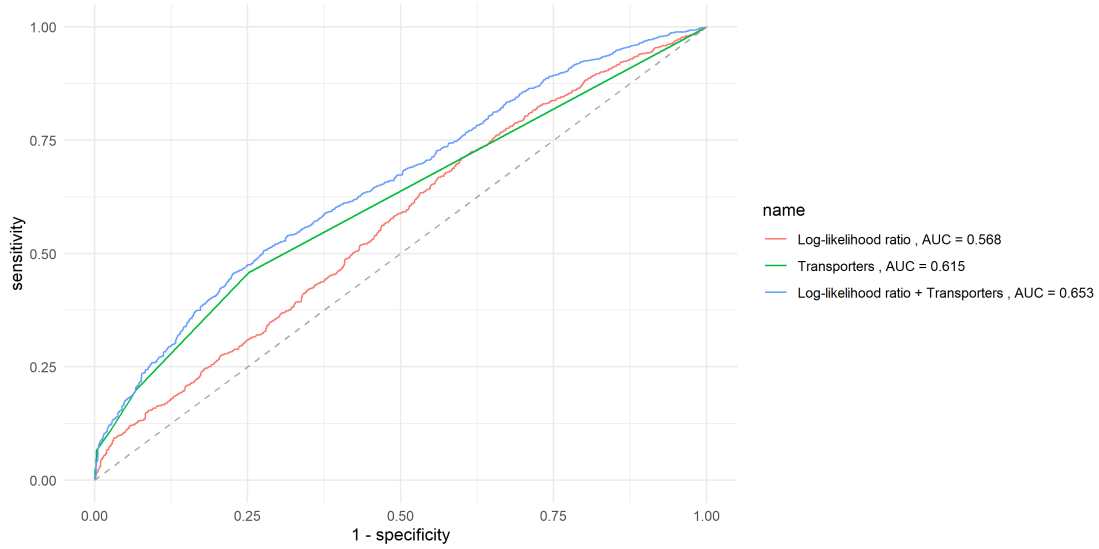
For QT interval prolongation, the top 20 associations using only the log-likelihood ratio were consistently ranked lower using the combined approach with the shortest path



(a)



(b)



(c)

FIGURE 6.6: Receiver Operating Characteristic (ROC) analysis that considers the use of the log-likelihood ratio combined with: (a) the *shortest path measure*; (b) the *enzyme measure*; and (c) the *transporter measure*.

measure (**Figure 6.7**). Only one positive control from CRESCENDDI was present in the top 20 associations using the log-likelihood ratio for ranking, while 6 drug pairs were found in at least one of the clinical resources that were considered in this study (i.e. BNF, Micromedex, or ANSM Thesaurus). On the other hand, the top 40 associations using the combined approach contained 5 positive controls from CRESCENDDI, while only 6 drug pairs out of 40 did not belong to any of the clinical resources (**Figure 6.8**). For the top 31 drug pairs when applying the combined approach, the differences in rankings ranged from 981 to 7,809. For the remainder of the associations (32-40), small ranking differences were observed. The combination of amlodipine with dofetilide has not been reported in any clinical resource but was ranked 14th when considering the shortest path measure (and moved 4,279 positions up). Similarly, the combination of clonazepam with acamprosate also moved 70 positions (36th) with the application of the combined approach.

For rhabdomyolysis, we compared the rankings using the log-likelihood ratio scores of statins, which is a drug class known to cause an increased risk for rhabdomyolysis due to interaction with other drugs (i.e. fibrates, macrolides, and fusidic acid) as opposed to rankings of ezetimibe, which is another lipid-lowering agent, with the same drugs. In the ranking of the 16,799 filtered drug pairs (i.e. with at least 5 FAERS reports) that were screened for rhabdomyolysis, drug pairs that contain statins consistently populated the top places in the ranking table (e.g. simvastatin with gemfibrozil or clarithromycin, atorvastatin with fusidic acid), while drug pairs that contain the comparator drug (ezetimibe) were lower in rankings (below 1,000) (**Table 6.3**). In all four cases of interacting drugs, simvastatin and atorvastatin were placed above ezetimibe, with a very high log-likelihood ratio (values between 3.69 and 241.2).

6.4 Discussion

Signal detection in pharmacovigilance encounters multiple challenges due to the nature of the data and reporting, leading to methodologies either producing falsely generated signals or being incapable of spotting relevant ones. This issue is also augmented by the fact that SDAs mainly rely on disproportionality analysis and do not involve any pharmacological considerations. As this is particularly relevant in the case of DDIs, the main goal of this study was to assess a novel SDA for identifying novel DDIs using a Bayesian hypothesis testing framework and adding a signal refinement step utilising systems pharmacology data. First, we performed a quantitative comparison of existing SDAs along with the novel one using a large and diversified publicly available reference set. The novel method outperformed all three existing ones in terms of AUC scores. We also noticed adequate or above-average algorithm performance for specific AEs of interest, especially for DDI surveillance, such as QT interval prolongation, rhabdomyolysis, bradycardia, and hypoglycaemia. The novel SDA showed enhanced performance when combined with any of the three measures derived from the ***Biological Attribute***

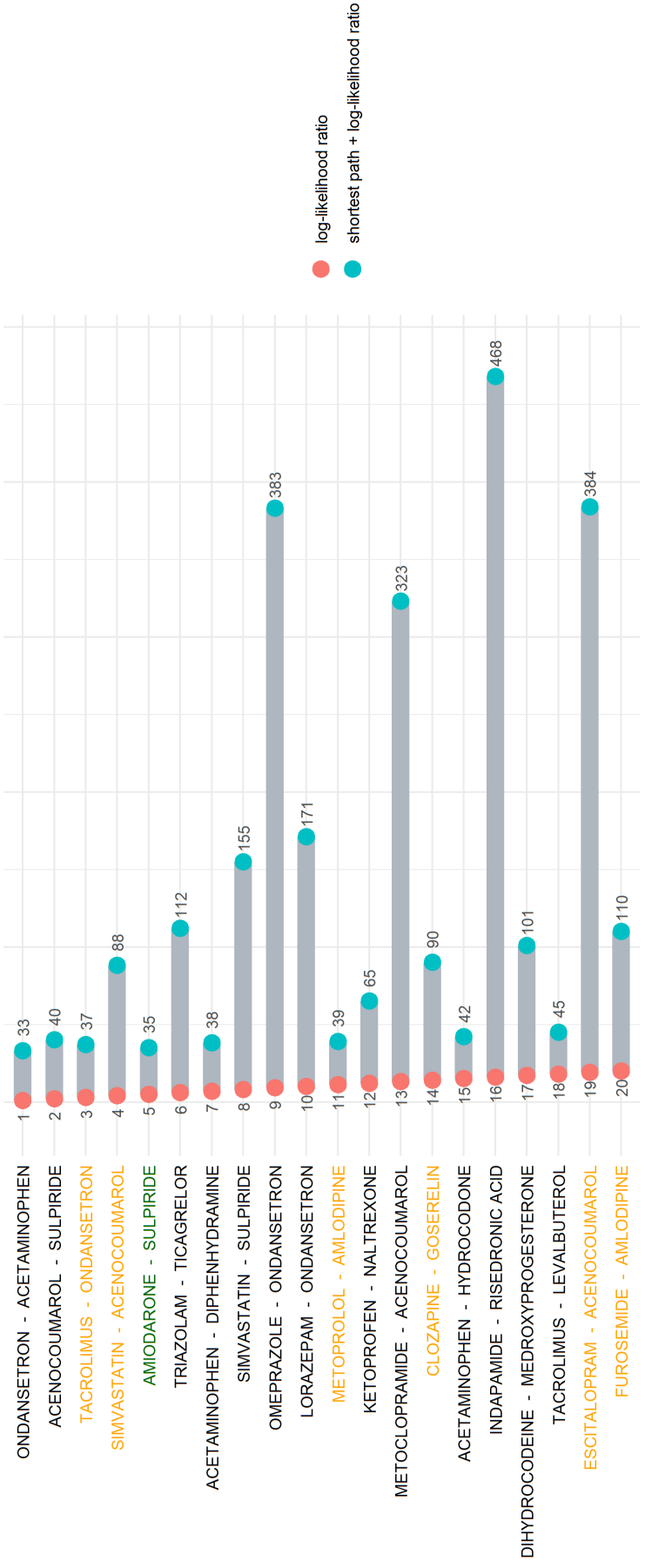


FIGURE 6.7: Top 20 associations using the log-likelihood ratio approach and their respective rankings with the shortest path + log-likelihood ratio approach. Drug pairs in green represent the ones that were positive controls for QT interval prolongation in the CRESCENDDI reference set. Drug pairs in orange represent the ones that were mentioned as known to interact in any of the clinical resources that were considered in this study (i.e. BNF, Micromedex, or ANSM Thesaurus), although those drug pairs could be associated with other medical events apart from QT interval prolongation.

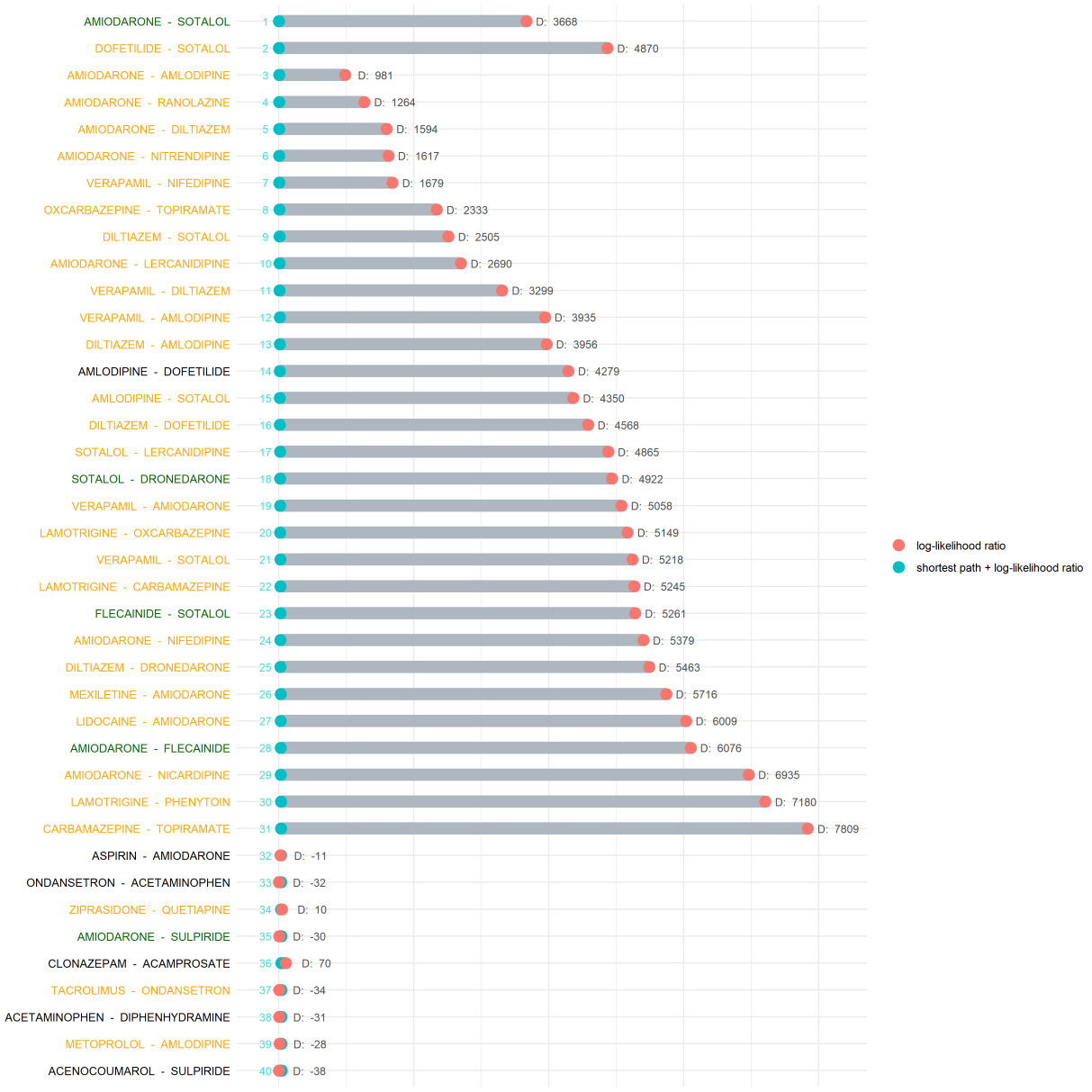


FIGURE 6.8: Top 40 associations using the *shortest path + log-likelihood ratio* approach and their respective rankings with the log-likelihood ratio approach. The difference between the two rankings for each drug pair is denoted in the label on the right side of the plot. Drug pairs in green represent the ones that were positive controls for QT interval prolongation in the CRESCENDDI reference set. Drug pairs in orange represent the ones that were mentioned as known to interact in any of the clinical resources that were considered in this study (i.e. BNF, Micromedex, or ANSM Thesaurus), although those drug pairs could be associated with other medical events apart from QT interval prolongation.

Drug 1	Drug 2	Log-likelihood ratio	Drug pair ranking
GEMFIBROZIL	SIMVASTATIN	241.2192869	2
	CERIVASTATIN	15.45426136	1315
	PRAVASTATIN	9.746103259	2301
	ATORVASTATIN	3.687716196	4656
	ROSUVASTATIN	-1.424158741	13149
	LOVASTATIN	-1.444116569	13200
	EZETIMIBE	**-1.902040641**	14046
CLARITHROMYCIN	SIMVASTATIN	148.8437063	12
	ATORVASTATIN	11.26737274	1967
	EZETIMIBE	**2.523279591**	5574
	PRAVASTATIN	-0.449481203	9767
	ROSUVASTATIN	-0.767005855	10653
FUSIDIC ACID	ATORVASTATIN	105.3754776	22
	SIMVASTATIN	88.46783535	34
	PRAVASTATIN	4.715557493	4082
	EZETIMIBE	**1.187202838**	6968
	ROSUVASTATIN	-0.301944238	9388
FENOFIBRATE	SIMVASTATIN	37.277021	263
	ATORVASTATIN	26.35263041	518
	EZETIMIBE	**17.91501155**	1054
	ROSUVASTATIN	-0.620035622	10189
	PRAVASTATIN	-1.915656202	14065

TABLE 6.3: Drug pair log-likelihood ratio scores and rankings from screening rhabdomyolysis FAERS cases that contain: and (i) a fibrate, macrolide or fusidic acid (i.e. potentially interacting drug) (**Drug 1**) and (ii) a statin or ezetimibe (as a comparator drug) (**Drug 2**). Rankings were calculated out of the 16,799 eligible drug pairs (i.e. at least 5 FAERS reports) that were screened for rhabdomyolysis.

Network (i.e. *shortest path*, *enzyme*, and *transporter*). Also, two case studies demonstrated the applicability of the novel approach for real-life signal detection purposes: the first one was related to signal prioritization for QT interval prolongation; the second one showed the relative magnitude of rhabdomyolysis signals of the novel SDA associated with statins and other lipid-lowering agents.

Systems pharmacology can support signal detection in pharmacovigilance to identify more signals with biological plausibility. While this is important for single drugs, it is also particularly relevant for the detection of novel DDIs, considering the even larger number of potential drug combinations that can arise, many of which can be flagged as potential signals. The increasing availability of data related to the mode of action of drugs, and their metabolic and elimination pathways, and the human protein interactome can allow the incorporation of this knowledge to inform statistical models that can flag potential novel drug(-drug) complications. However, appropriate methods are needed to be able to transform this knowledge into quantitative evidence that can be fed into a statistical framework.

The strength of this study includes the use of a comprehensive and clinically relevant reference set. By having access to a large set of controls that also considers multiple AEs, a quantitative comparison of existing SDAs with the novel approach was possible.

The novel SDA provides outputs that could be utilised in combination or separately to monitor the different probabilities that could provide a pharmacology-driven framework. Also, the signal detection framework could be extended to consider higher-order drug interactions. The use of open data (SRS database, reference set, systems pharmacology data) is another strength of this study. This study introduced the concept of incorporating biological plausibility aspects as a signal refinement step, which has been explored in other studies but not in the scope of DDIs.

The SDA yielded reasonable results in terms of ranking drug pairs for signals of rhabdomyolysis based on existing pharmacological knowledge. These findings suggest that the novel SDA could be useful in screening SRS data in real-world applications. In terms of the signals of ezetimibe (that was used as a comparator drug), those were always below the respective signals from atorvastatin and simvastatin. Simvastatin and atorvastatin are predominantly metabolised by CYP3A4 and their levels increase significantly when co-administered with strong CYP3A4 inhibitors, such as clarithromycin [226]. Moreover, both statins are substrates of the organic anion-transporting polypeptide (OATP)1B1, which is responsible for their hepatic uptake and is also inhibited by clarithromycin. Therefore, the presence of the macrolide substantially affects the concentration of both statins [227]. Another example is the signal of cerivastatin with gemfibrozil, which ranked high in the analysis. Cerivastatin was withdrawn from the market worldwide in 2001 due to its association with rhabdomyolysis, with a higher risk observed when taken concurrently with gemfibrozil [228]. Furthermore, although most statins generated relatively strong signals with gemfibrozil, the same was not observed with fenofibrate. In fact, according to clinical guidelines and literature, of the two fibrates, gemfibrozil has a higher risk of interacting with statins and leading to rhabdomyolysis [229, 230]. A surprising finding was the rhabdomyolysis signal of fenofibrate with ezetimibe, which was in a higher ranking (1054th) in comparison to signals of other statins. Previous studies have examined the safety of this combination and have reported no clinically important elevations in creatine phosphokinase (CPK) (which are indicative of rhabdomyolysis) or additional risk of myopathy due to the combination therapy [231, 232].

This study also identified two potential signals of novel DDIs linked to QT interval prolongation (amlodipine – dofetilide and clonazepam – acamprosate), which are currently unknown but are supported by both statistical screening and biological information. More precisely, both clonazepam and acamprosate are positive modulators of the anion channel of the GABA-A receptor GABRG3. In the ***Biological Attribute Network***, GABRG3 is linked to the potassium voltage-gated channel KCNH2 that is associated with QT interval prolongation (target-AE association) via two nodes (GABARAP and AMK2)[7]. In terms of the combination of amlodipine and dofetilide, the associated nodes in the network are interconnected. Amlodipine blocks the voltage-gated L-type calcium channel (CACNA1C), while dofetilide blocks the voltage-activated potassium

[7] https://string-db.org/cgi/network?taskId=byHEDTBYX46p&sessionId=bZwf7vZ81xl7&allnodes=1

channel (KCNH2). These two targets are directly linked and KCNH2 is also associated with QT interval prolongation[8].

A systematic evaluation of different SDAs for DDI surveillance was missing from the literature. A previous study used Stockley's as a source of positive controls, but the resulting reference set was not made available, hindering the reproducibility and extension of the study and the possibility to further dive into the nature of the controls [191]. Other efforts have used benchmarking only for a very limited number of AEs of interest to measure and compare SDA performance [11, 12, 53]. In our study, the consideration of a large and diversified reference set enabled us to compare the performance of the novel SDA across multiple AEs. We noticed substantial differences in method performance depending on the AE. As an illustrative example, for common AEs, such as haemorrhage, the masking effect might have been responsible for lower performance.

Systems pharmacology has been incorporated in drug development. Multiple machine learning and artificial intelligence (AI) methodologies have also examined DDI prediction by integrating various data types and information sources as features, such as drug target profiles (i.e. drug-protein interactions), metabolising enzymes, and transporters [196, 16, 233, 234]. However, systems pharmacology coupled with pharmacovigilance has only been recently considered in studies and has not been explored in the case of adverse DDIs. For single drugs, we have seen some recent efforts to develop similar frameworks that, apart from main pharmacological targets, also consider off-targets to aid the detection of drug-related side effects [18, 235]. The use of off-target data might be particularly relevant in the context of drug safety, as many drug complications leading to adverse drug reactions are related to secondary pathways and off-target activity of the drug molecule.

6.4.1 Limitations of the study

The focus of this study was on two-way DDIs, although high-order DDIs (i.e. involving more than two drugs) would be an area of focus for future studies. Some previous work has already attempted to explore this area [236, 237].

The masking effect that can be present in SRS data was not considered in this study in the data mining step. Revising the concept of masking in the case of contingency tables for two-way DDIs, which has been extensively described and explored in previous studies for single drugs [101, 238, 239, 240], would also be interesting. The modifications of the data mining step in an effort to minimise the bias resulting from masking in SRS data could potentially increase the performance measures of the different SDAs for DDI surveillance. Additionally, it could have an impact on the signal prioritization step, altering the drug pair rankings.

The existing evidence that currently appears in Open Targets regarding target safety liabilities is limited to a small number of targets and only validated associations; thus

[8] https://string-db.org/cgi/network?taskId=bQeC3KyUtXJ6&sessionId=bocWSP6Up7f7

it might correspond to well-known safety complications arising from drug combinations that appeared in the reference set that was utilised for performance evaluation.

This study considered the combination of the different systems pharmacology measures with the SDA scores using binary logistic regression. However, non-linear approaches could also be relevant.

6.5 Conclusion

This study provides a novel framework for detecting DDI signals using disproportionate reporting in FAERS combined with a biological information network. With an increasing volume of systems pharmacology information now available, we show that this information has the potential to enhance signal detection in pharmacovigilance, with DDIs being an important and promising area of application. This study also identified two potential DDI signals related to QT interval prolongation. Further studies, including the consideration of additional SRS databases, real-world evidence, in vitro or in vivo experiments, are needed to validate the potential signals.

Chapter 7

Conclusions

The first section of this chapter presents an overview of the research findings and contributions discussed in the previous chapters of the thesis as per the objectives the thesis objectives outlined in **Section 1.2**. The second section briefly discusses open problems and potential future research directions based on the findings of this thesis.

7.1 Summary of Thesis Findings and Contributions

Objective I: Understand the existing evidence related to clinically relevant and observable drug-drug interactions (DDIs) by exploring the level of agreement on information listed in different drug information resources (DIRs).

Chapter 3 addressed this objective by performing a similarity and consistency assessment of three leading clinical resources for DDIs from different geographic locations, namely the British National Formulary (BNF), Thesaurus and Micromedex. One of the key findings was that, despite the considerable overlap in the listed ingredients of the three resources, there was significant variation in the DDIs included. Additionally, there was considerable variability in the other three examined types of information pertinent to DDIs (severity rating, evidence rating, and clinical management recommendations). This variability, which can have critical implications for patient safety, also emphasises the necessity of using multiple data sources to establish a reliable dataset for scientific research that represents the existing evidence on clinically relevant and observable DDIs. This reference set could serve as a benchmark for evaluating the performance of signal detection algorithms (SDAs) in pharmacovigilance.

This study stands out as the first comprehensive assessment that has considered both complete and clinically relevant DIRs, rather than data sources of potential DDIs that may not be observable. Moreover, the study presented a methodology for automatically extracting web data from DIRs and a standardisation pipeline to align data from multiple resources to a common framework using controlled vocabularies and terminologies.

Objective II: Build a clinically relevant reference set for DDIs.

Chapter 4 tackled the development of a Clinically-relevant Reference Set CENtred around Drug-Drug Interactions (CRESCENDDI). CRESCENDDI is a publicly available resource that adheres to the FAIR data principles (Findable, Accessible, Interoperable, and Reusable). It was built by assembling information from disparate clinical resources that could provide support for the clinical relevance of controls. CRESCENDDI considered a large number of controls related to multiple drugs and adverse events (AEs) that were mapped to controlled vocabularies and terminologies.

Chapter 4 presented several key contributions. First, it described a scalable approach for building a standardised resource by combining multiple harmonised resources. The pipeline can be re-run to get an updated reference set based on the latest modifications of the DIRs. Hence, this approach would require minimal manual effort for future updates. Also, CRESCENDDI was mapped to standardised terms from the Observational Health Data Sciences and Informatics (OHDSI) Common Data Model, allowing it to be linked to other vocabularies and terminologies as needed based on the test data resources, including, for example, electronic health records (EHRs) and patient registries. Additionally, the size of CRESCENDDI and the relatively large number of

included drugs and AEs provides a common ground for the performance evaluation of SDAs for DDIs in pharmacovigilance. Finally, the inclusion of additional information such as drug indications, single-drug adverse drug reactions (ADRs), and evidence levels enables filtering and stratification of controls.

Objective III: Assess the impact of different choices on the nature of the controls included in a reference standard on the performance assessment of existing SDAs for DDI surveillance.

Chapter 5 utilised the CRESCENDDI reference set that was developed in **Chapter 4** to address this objective. It systematically evaluated 14 design criteria for reference sets as potential sources of confounding for signal detection of DDIs in pharmacovigilance. The study revealed that certain criteria, such as restricting controls to only include MedDRA Preferred Terms (PTs) from the European Medicines Agency (EMA) Designated Medical Event list or rare AEs, had a significant effect on the performance and comparative evaluation of different SDAs for DDI surveillance in spontaneous reporting system (SRS) data. Conversely, other criteria, such as limiting the reference set to positive controls associated with theoretical evidence, did not have a significant impact on the absolute and relative performance metrics of the three SDAs.

The analysis highlighted the importance of considering the choices made when constructing a reference set and restricting the evaluation to specific controls when comparing SDA performance. Additionally, it emphasised the need for frameworks that can leverage large and diverse data sources to generate open and flexible benchmarks in pharmacovigilance.

Objective IV: Evaluate the ability of a novel Bayesian hypothesis testing framework to identify signals of disproportionate reporting indicative of DDIs in post-marketing surveillance data and assess the potential for signal refinement using biological plausibility aspects.

This objective was addressed in **Chapter 6** by presenting and evaluating a new approach to identifying DDI signals in SRS data that combines disproportionate reporting with a network of biological information. A Bayesian hypothesis testing framework was compared against existing SDAs for DDI surveillance in SRS data using the CRESCENDDI reference set that was built in **Chapter 4**. The novel method outperformed all three existing SDAs in terms of the Area Under the Receiver Operating Characteristic Curve (AUC) scores and achieved better specificity for high sensitivity values. Furthermore, for specific adverse events of interest, such as QT interval prolongation, rhabdomyolysis, bradycardia, and hypoglycaemia, the novel SDA showed adequate or above-average performance. Additionally, by incorporating biological plausibility aspects as a signal refinement step through the Biological Attribute Network, the novel SDA demonstrated

enhanced performance when combined with any of the three network measures (shortest path, enzyme, and transporter). Notably, two case studies demonstrated the applicability of the novel approach for real-life signal detection purposes, including signal prioritization for QT interval prolongation and evaluating rhabdomyolysis signals associated with statins and other lipid-lowering agents.

Chapter 6 introduced several novel contributions. First, it provided a systematic evaluation of different SDAs for DDI surveillance that was missing from the literature. It also introduced the Biological Attribute Network, which contains systems pharmacology data from two different publicly available resources (i.e. DrugBank and Open Targets). The establishment of this network enabled the utilisation of both pharmacokinetic (enzyme and transporter) and pharmacodynamic (drug target) information to inform biological plausibility aspects as a signal refinement step, which had not previously been explored in the scope of DDIs. Finally, the study highlighted potential signals that require further investigation, such as utilizing other SRS databases or conducting a literature search for validation.

7.2 Future Perspectives

The findings presented in this thesis can provide a valuable foundation for exploring at least three possible directions for future research. Firstly, enhancing the existing evidence and standards for clinical decision-making. Secondly, improving the quantitative methods used for the detection of novel DDIs, by addressing computational challenges and utilising disparate information sources to support quantitative analysis. Lastly, integrating biological knowledge and systems pharmacology evidence to gain a mechanistic understanding of the signals identified.

Further exploration of the evidence related to clinically-observable DDIs can lead to two important outcomes. It can first inform harmonisation efforts in information reporting across resources that are used in the clinical setting with the goal of improving patient safety. Also, it can support pharmacovigilance activities by generating reference sets that aid in developing algorithms and statistical methods for detecting and predicting novel DDIs.

From a clinical perspective, expanding the field of interoperable standards for DDI information in clinical resources is an important direction for patient safety that requires a combination of collaborative partnerships as well as data standardisation efforts. There are multiple sensible activities that could draw upon the research and methodology described in this thesis to further explore and understand the current situation and practices with the aim of supporting the establishment of standards that can be adopted for real-world implementation. For example, it is crucial to broaden the scope of analysis to include clinical resources that are not readily accessible due to subscription or usage restrictions. Investigating the completeness of generic resources for DDIs relative to those tailored to specific drug categories, such as the Liverpool Drug Interaction Checkers

for anticancer drugs, can also provide valuable insights. Conducting comprehensive comparisons of clinical management recommendations across entire resources, rather than subsets, can also help identify discrepancies in reported information.

The perspective of developing collaborative networks and partnerships between DIR vendors, developers, healthcare providers, clinicians but also researchers would enable the creation of information reporting standards that meet the requirements of various stakeholders. These standards should comprehensively define the required data fields, types, and ways of information reporting in clinical DIRs. Some important aspects would be related to data provenance (e.g. information from drug labels, literature, etc), information update procedures, as well as clear definitions and categorisation of evidence and severity categories for clinically relevant DDIs that can be uniformly adopted. The overall goal of these activities is to provide guidelines and frameworks for building more comprehensive DDI information resources for clinical practice. In the context of potential DDI data sources for scientific research, there is a recently developed minimal information model [241] that could be utilised as an exemplar for clinically-relevant DDI standards development.

In terms of data standardisation, while bioinformatics and cheminformatics resources such as DrugBank and the Open Targets Platform use controlled vocabularies and ontologies to standardise their data where possible, clinical DIRs are considerably less standardised to public terminologies. However, standardising the mapping of terms associated with the clinical manifestation of DDIs in DIRs could enable interoperability with other resources, contribute to standards development in the field, but also facilitate the integration of resources with routinely collected data. For example, the Systematised Nomenclature of Medicine Clinical Terms (SNOMED-CT) could be used for mapping medical events associated with the clinical manifestation of DDIs and RxNorm or WHO-Drug for drug names.

From a methodological perspective, building reference sets that incorporate the most current evidence by leveraging scientific literature through techniques such as identifying causality patterns in free text using natural language processing (NLP) could help in identifying eligible controls. Additionally, current reference sets only consider drug interactions on a pair-wise drug combination basis and do not consider other patient-specific contextual information (e.g. genetic polymorphisms) that can impact the risk of experiencing a clinically observable adverse DDI [242].

A second direction for future research concerns quantitative methods for signal detection of DDIs. A current gap in the literature concerns the development of data-driven techniques to overcome the masking effect in disproportionality analysis caused by the one-versus-rest approach for the AE background estimation (see **Chapter 6**). This would be particularly important in the context of processing data from novel data streams, which may include true signals that have not been seen before. In principle, such a data-driven approach could be developed by extending the hypothesis testing

techniques considered in **Chapter 6** but with a focus on hypotheses for the set of surprising rates (across all AEs and drug combinations). Searching that space would be likely to require sophisticated numerical Bayesian techniques (e.g. Markov Chain Monte Carlo [243]).

Another opportunity lies in the use of more data streams. The inherent limitations in SRS data place constraints on the scope of research that can be conducted using this data source for PMS. One significant limitation lies in the SRS data's incapability to capture the temporal utilization of drugs, including the sequence of medication introduction into a patient's treatment regime and the time to onset of adverse events (AEs) following drug administration. This deficiency in tracking temporal patterns hinders the assessment of the causal relationship between drug interactions and the onset of AEs. Also, the incomplete information available in SRS data regarding drug doses, durations, and routes of administration, which relies on the details provided by the reporter, limits the inclusion of these variables in data mining efforts. Furthermore, under-reporting and selection bias are two other substantial challenges associated with SRS data [244]. Under-reporting skews the estimation of the prevalence of AEs within the population. Additionally, SRS data may be susceptible to selection bias due to selective reporting patterns. The awareness and publicity surrounding specific alerts can distort disproportionality measures in spontaneous reporting databases. This notoriety bias may also trigger a 'ripple' effect, altering the reporting patterns of other drugs associated with the same effect [245].

One avenue to address these limitations of SRS data is leveraging alternative data streams, including EHR and administrative claims data that can enable longitudinal analysis. EHR databases such as the Clinical Practice Research Datalink (CPRD) can also assist in the causality assessment of potential signals by providing additional clinical details. The combination of multiple data sources also requires more sophisticated and flexible methods than the disproportionality measures for DDI surveillance that have been previously described in the literature. The hypothesis testing framework developed in **Chapter 6** would be a good starting point to enable consideration of data coming from disparate data sources. Previous research (e.g. [53]) has used SRS data for training and EHR data for prediction, but the combined use of multiple databases for DDI surveillance has not been fully explored. The utilisation of contextual information from EHR data can also provide further support and explanation of signal generation processes that result in predicted novel DDIs and are triggered by statistical and other computational methods. For instance, patient-specific factors, including renal function, genetic polymorphisms, etc, can be available in EHR data and may influence the manifestation of DDIs. In addition, non-conventional data sources, such as social media, could also be investigated in combination with traditional data streams. However, text mining and NLP techniques would be necessary for encoding free text from data sources that are not coded to medical terms, including clinical notes in EHR data and social media posts.

Moreover, the signal detection framework that was developed and applied in **Chapter 6** can be extended to consider higher-order drug interactions. However, this activity also requires appropriate reference sets of multi-way (e.g. three-way) drug interactions for performance evaluation that have not been developed to date (as mentioned above).

Current research is also limited by the fact that DDIs are typically considered independently of each other. In other words, knowing about the existence of one DDI provides no information about the signal detection of any DDI that can involve one of the interacting drugs, drugs of the same class or drugs with similar mechanisms of action. Grouping drugs using established terminologies and classification systems such as RxNorm and ATC could help in identifying potential DDIs by utilising drug similarity in terms of mechanism of action or therapeutic class. Previous research has considered similarity analysis in the context of cheminformatic data sets such as PubChem to predict novel DDIs based on structural drug similarities with drug molecules that are known to interact [246].

Finally, the integration of systems pharmacology aspects into signal detection for DDIs can inform the statistical methods with the biological plausibility of the examined associations. This thesis considered the combination of the different systems pharmacology measures with the SDA scores using binary logistic regression. However, non-linear approaches could also be relevant. Utilising additional data resources for constructing the network of biological information, such as a recently developed knowledge graph [247], could also be considered. Regarding graph mining approaches, apart from the shortest path measure (Dijkstra's algorithm) that was examined in this thesis, there are other more sophisticated methods for graph mining tasks such as node classification and link prediction, including random walks and neural networks.

Some of the above-described research directions can be potential areas to be addressed in the DynAIRx project [248], which is led by the University of Liverpool. DynAIRx aims to support medicines optimisation by leveraging artificial intelligence (AI) and statistical methods for causal inference and prediction of clinical outcomes using structured clinical data from general practice, hospitals, and social care.

In summary, the identification of novel complications of drugs related to DDIs in the post-marketing setting is a complex task that requires a combination and appropriate fusion of various sources of data using methods that can utilise existing biomedical knowledge and integrate it with routinely collected healthcare data and voluntarily reported safety data. With the advancement of computational methods capable of handling complex problems, there is an opportunity to identify novel DDIs that could be a result of more than a two-way interaction in the future.

Appendix A

Supplementary Figures and Tables

Chapter 3

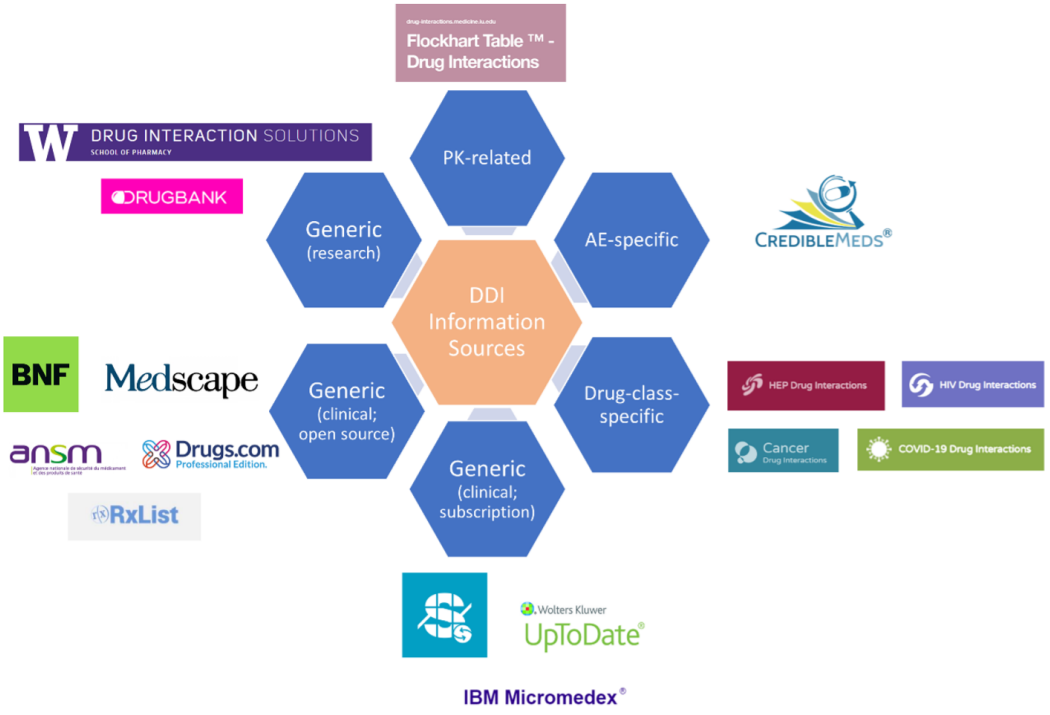


FIGURE A.1: An overview of drug-drug interaction online resources.

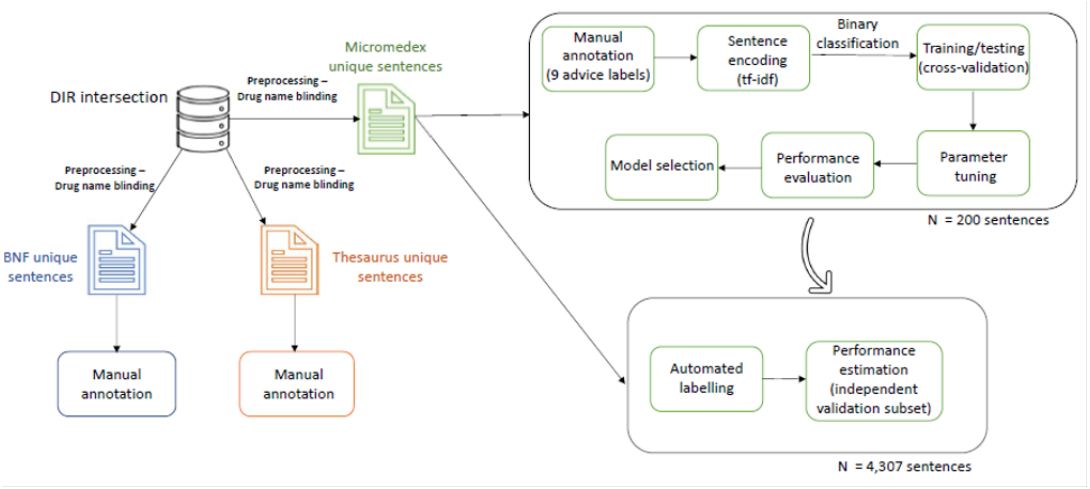


FIGURE A.2: Pipeline for the clinical recommendation labelling.

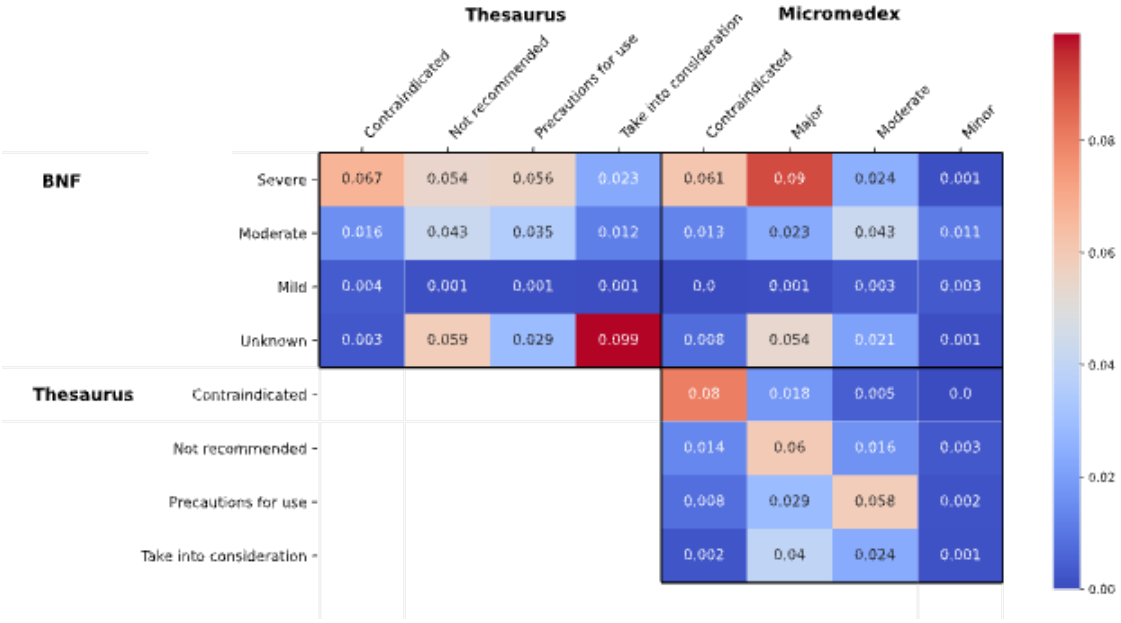


FIGURE A.3: Similarity matrix of the Jaccard index for all drug information resource severity ratings.

TABLE A.1: Performance metrics and applied thresholds of the selected sentence classifiers for Micromedex descriptions. *PPV: Positive Predictive Value*

Advice label	PPV	Sensitivity	Threshold
ADJUST DOSE	0.860000	1.000000	0.352397
AVOID	0.804878	0.750000	0.459679
DISCONTINUE	1.000000	0.500000	0.653243
MONITOR	0.892308	1.000000	0.356476
USE ALTERNATIVE	0.833333	0.714286	0.367563
USE WITH CAUTION	0.800000	0.923077	0.210383
WASH-OUT	1.000000	0.142857	0.858192

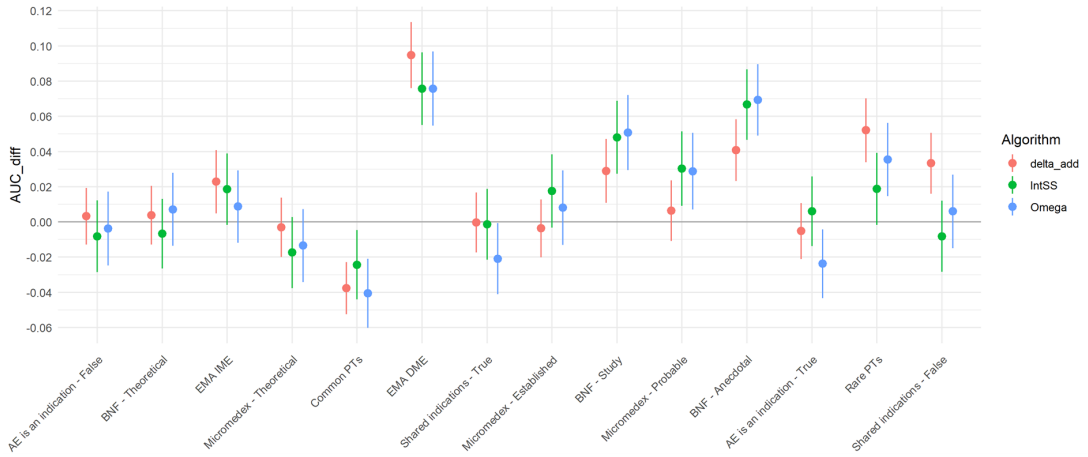
TABLE A.2: Evaluation of selected classifiers using an independent validation subset in terms of positive predictive value (PPV), sensitivity, and F1-score metrics (%).

Classifier	PPV	Sensitivity	F1-score
AVOID	78.26	85.71	81.82
USE WITH CAUTION	100.00	85.71	92.31
MONITOR	82.76	100.00	90.57
ADJUST DOSE	90.32	100.00	94.92
USE ALTERNATIVE	71.43	71.43	71.43
WASH-OUT	N/A	0	N/A
DISCONTINUE	N/A	0	N/A

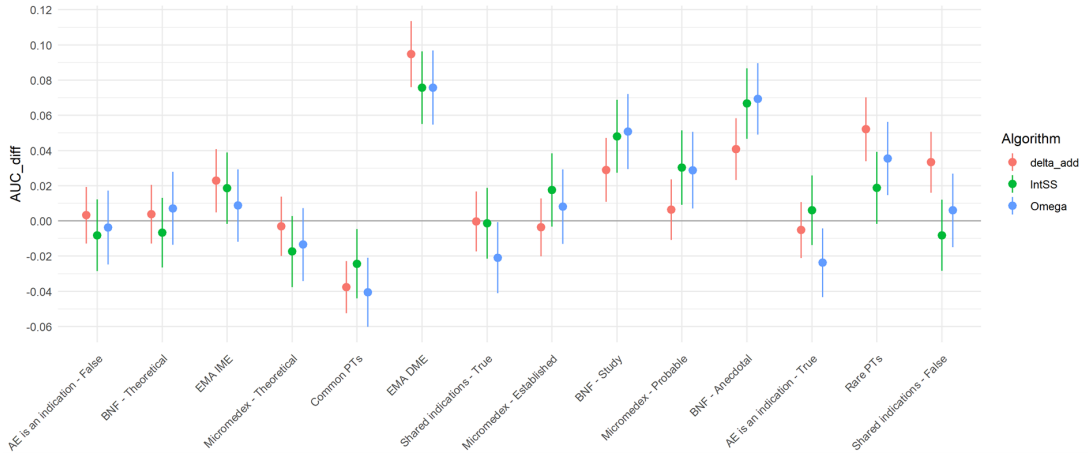
TABLE A.3: Number and percentage of drug-drug interactions included in the *DIR intersection list* by advice label for each drug information resource. (*) BNF cases for *Space dosing times*; (**) BNF cases for *Modify administration*

Advice label	BNF	Thesaurus	Micromedex
AVOID	2,239 (32.12%)	2,148 (30.82%)	3,277 (47.02%)
USE WITH CAUTION	494 (7.09%)	80 (1.15%)	1,536 (22.04%)
MONITOR	511 (7.33%)	2,432 (34.89%)	4,421 (63.43%)
ADJUST DOSE	842 (12.08%)	1,075 (15.42%)	2,493 (35.77%)
WASH-OUT	27 (0.39%)	69 (0.99%)	27 (0.39%)
SPACE DOSING TIMES	108 (1.55%)	138 (1.98%)	95/108* (87.96%)
MODIFY ADMINISTRATION	44 (0.63%)	603 (8.65%)	19/44** (43.18%)
DISCONTINUE	N/A	214 (3.07%)	191 (2.74%)
USE ALTERNATIVE	1 (0.01%)	139 (1.99%)	264 (3.79%)
NOT MENTIONED	3,932 (56.41%)	2,793 (40.07%)	375 (5.38%)

Chapter 5



(A)



(B)

FIGURE A.4: AUC_{diff} for a fixed restricted reference set size of 100 with 95% confidence intervals for: (A) the PT Reference Set; (b) the MC Reference Set. Design criteria are ordered by increasing range of AUC_{diff} values among the three signal detection algorithms.

FIGURE A.5: AUC_{diff} values for the different design criteria, signal detection algorithms, and sizes of restricted reference set for the PT Reference Set. In cases where the number of available controls in the restricted subset using a design criterion was smaller than 2,000, there are missing points in the respective graph. Points that lie above the x-axis signify positive estimates for AUC_{diff} (i.e. the design criterion had a positive effect on the calculated area under the curve), while those below the x-axis were associated with a negative effect of the design criterion on the area under the curve score. The dot size represents the probability of the estimated score, AUC_{diff}, being non-zero.

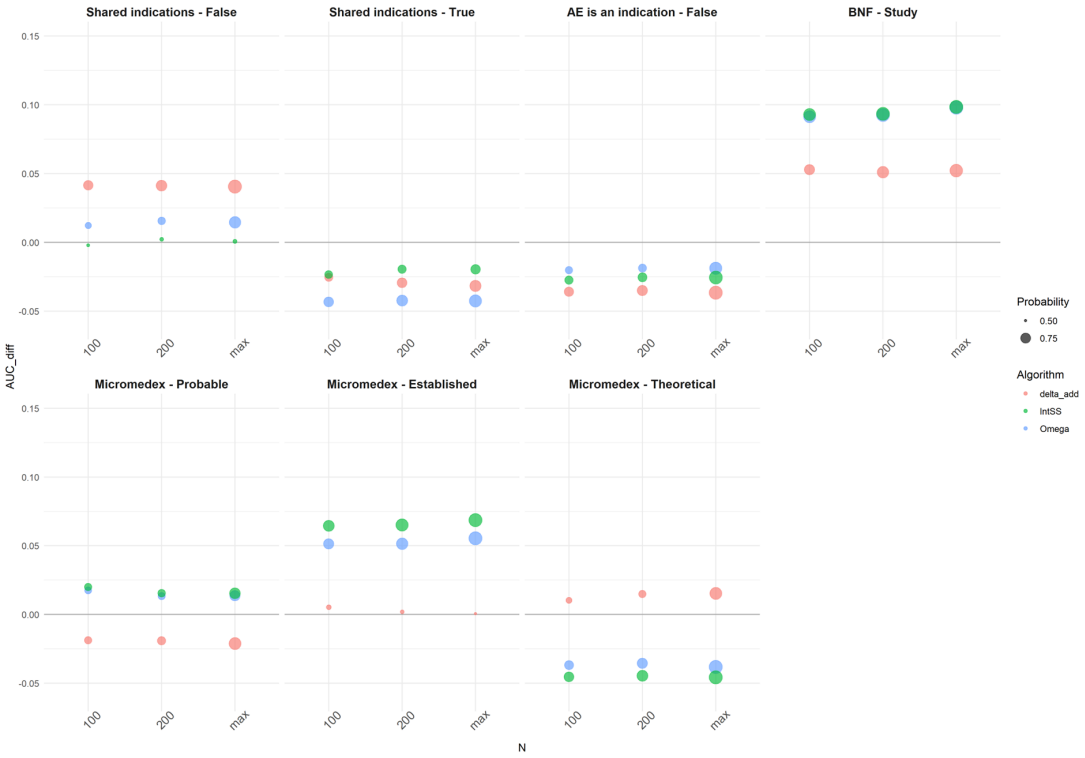


FIGURE A.6: AUC_{diff} estimated values and associated probabilities of a non-zero AUC_{diff} estimate for the different design criteria, signal detection algorithms, and sizes of restricted reference set for the MC Reference Set. In cases where the number of available controls in the restricted subset using a design criterion was smaller than 200, there are missing points in the respective graph. Points that lie above the x-axis signify positive estimates for AUC_{diff} (i.e. the design criterion had a positive effect on the calculated area under the curve), while those below the x-axis were associated with a negative effect of the design criterion on the area under the curve score. The dot size represents the probability of the estimated score, AUC_{diff}, being non-zero.

TABLE A.4: Number of positive and negative controls from the MC Reference Set for each of the different design criteria. The maximum number of controls considered from each type to form simulated reference sets (N_max) is denoted with bold. The design criteria marked with an asterisk (*) were not tested due to the small number of their restricted sets.

Design Criterion	N_pos	N_neg
BNF - Study	**184**	N/A
BNF - Theoretical*	100	N/A
BNF - Anecdotal*	66	N/A
Micromedex - Established	**235**	N/A
Micromedex - Theoretical	**498**	N/A
Micromedex - Probable	**364**	N/A
AE is an indication - True*	248	97
AE is an indication - False	849	**517**
Shared indications – True	569	**169**
Shared indications – False	528	**445**

TABLE A.5: PPV_{diff} values with 95% confidence intervals (CIs) for the different design criteria, and a fixed restricted reference set size of 100 using the PT Reference Set. Green colour represents estimates with a CI range containing only positive values. Red colour represents estimates with a CI range containing only negative values.

Design Criterion (DC)	**SDA**	PPV_{diff} **(95% CI)**		
		Sensitivity		
		0.60	0.75	0.90
Shared indications - True	Omega	-0.0089 (-0.0294, 0.0115)	0.0066 (-0.0096, 0.0228)	0.0064 (-0.0015, 0.0144)
	delta_add	-0.0122 (-0.0243, -0.0001)	-0.0075 (-0.0135, -0.0014)	-0.0039 (-0.0073, -0.0006)
	IntSS	0.0177 (-0.0004, 0.0357)	0.0169 (0.0053, 0.0285)	0.0059 (0.0013, 0.0104)
Shared indications - False	Omega	-0.0225 (-0.0443, -0.0007)	-0.0135 (-0.0288, 0.0019)	-0.003 (-0.0105, 0.0044)
	delta_add	0.029 (0.0141, 0.0439)	0.0099 (0.0026, 0.0172)	0.0039 (0.0002, 0.0077)
	IntSS	-0.027 (-0.0436, -0.0104)	-0.0148 (-0.0242, -0.0054)	-0.004 (-0.0076, -0.0005)
AE is an indication - True	Omega	-0.0056 (-0.0261, 0.0148)	0.0078 (-0.0083, 0.024)	0.0007 (-0.0069, 0.0083)
	delta_add	0.0193 (0.0072, 0.0314)	0.0216 (0.0144, 0.0288)	0.009 (0.0043, 0.0137)
	IntSS	0.0358 (0.0174, 0.0542)	0.0251 (0.013, 0.0372)	0.003 (-0.0014, 0.0074)
AE is an indication - False	Omega	-0.0092 (-0.0311, 0.0128)	-0.0055 (-0.0213, 0.0103)	-0.0022 (-0.0095, 0.0052)
	delta_add	0.0021 (-0.01, 0.0143)	0.0011 (-0.0052, 0.0074)	0.0004 (-0.0033, 0.004)
	IntSS	-0.0126 (-0.0301, 0.0048)	-0.005 (-0.0147, 0.0048)	-0.0007 (-0.0043, 0.003)
EMA IME	Omega	0.0179 (-0.005, 0.0409)	0.0131 (-0.004, 0.0303)	0.0073 (-0.001, 0.0157)
	delta_add	0.0066 (-0.0074, 0.0206)	-0.0009 (-0.0071, 0.0054)	-0.0017 (-0.0052, 0.0017)

	IntSS	0.0263 (0.007, 0.0456)	0.017 (0.0056, 0.0285)	0.0032 (-0.0009, 0.0073)
EMA DME	Omega	0.0545 (0.0316, 0.0775)	0.0517 (0.0344, 0.0689)	0.0292 (0.0195, 0.039)
	delta_add	0.0729 (0.0544, 0.0915)	0.0255 (0.0157, 0.0352)	0.0091 (0.005, 0.0133)
	IntSS	0.0573 (0.0361, 0.0785)	0.0277 (0.0153, 0.04)	0.0047 (0, 0.0094)
BNF - Anecdotal	Omega	0.0252 (0.003, 0.0475)	0.0097 (-0.0067, 0.0261)	0.0041 (-0.0038, 0.012)
	delta_add	0.0084 (-0.0047, 0.0215)	0.0012 (-0.0053, 0.0078)	-0.0004 (-0.0039, 0.003)
	IntSS	0.0312 (0.012, 0.0505)	-0.0014 (-0.0129, 0.0101)	-0.007 (-0.01, -0.0041)
BNF - Study	**Omega**	0.037 (0.0148, 0.0593)	0.0358 (0.0187, 0.0529)	0.0116 (0.0026, 0.0205)
	delta_add	0.025 (0.0101, 0.0399)	0.0109 (0.0036, 0.0182)	0.005 (0.001, 0.009)
	IntSS	0.0449 (0.0255, 0.0642)	0.0312 (0.0196, 0.0428)	0.0112 (0.0061, 0.0163)
BNF - Theoretical	Omega	0.0013 (-0.0205, 0.023)	-0.0052 (-0.0213, 0.011)	-0.0015 (-0.0085, 0.0055)
	delta_add	-0.0057 (-0.0177, 0.0064)	-0.0079 (-0.014, -0.0017)	-0.0029 (-0.0063, 0.0004)
	IntSS	-0.0213 (-0.0377, -0.005)	-0.0096 (-0.0185, -0.0007)	-0.0029 (-0.0064, 0.0005)
Micromedex - Established	Omega	0.0202 (-0.001, 0.0415)	0.0117 (-0.0057, 0.029)	0.0049 (-0.0032, 0.0129)
	delta_add	0.0077 (-0.0057, 0.0212)	0.0025 (-0.0044, 0.0094)	0.0022 (-0.0016, 0.006)
	IntSS	0.028 (0.0095, 0.0465)	0.0254 (0.0138, 0.0371)	0.0117 (0.0069, 0.0165)
Micromedex - Theoretical	Omega	-0.0159 (-0.0377, 0.0058)	-0.0076 (-0.0233, 0.008)	-0.0023 (-0.0096, 0.005)
	delta_add	-0.0016 (-0.0136, 0.0104)	0.001 (-0.0054, 0.0075)	0.0007 (-0.003, 0.0043)
	IntSS	-0.0172 (-0.0343, -0.0001)	-0.0105 (-0.0198, -0.0012)	-0.0026 (-0.0059, 0.0007)
Micromedex - Probable	Omega	0.0274 (0.0047, 0.05)	0.0148 (-0.0033, 0.0329)	-0.0013 (-0.0092, 0.0065)

	delta_add	0.0022 (-0.0109, 0.0152)	-0.0048 (-0.0111, 0.0015)	-0.0028 (-0.0062, 0.0006)
	IntSS	0.0188 (-0.0012, 0.0388)	0.005 (-0.0063, 0.0164)	-0.0007 (-0.0045, 0.003)
Common PTs	Omega	-0.031 (-0.0507, -0.0114)	-0.0078 (-0.0227, 0.0072)	-0.0037 (-0.0111, 0.0036)
	delta_add	0.007 (-0.0046, 0.0186)	0.0106 (0.0036, 0.0175)	0.0048 (0.0009, 0.0088)
	IntSS	-0.0076 (-0.0244, 0.0092)	0.0049 (-0.0051, 0.0149)	0.0021 (-0.0016, 0.0059)
Rare PTs	Omega	-0.0104 (-0.036, 0.0151)	-0.0182 (-0.0336, -0.0028)	-0.006 (-0.0128, 0.0009)
	delta_add	0.008 (-0.0066, 0.0225)	-0.0129 (-0.0194, -0.0063)	-0.0089 (-0.0117, -0.0062)
	IntSS	-0.0266 (-0.0433, -0.0099)	-0.0137 (-0.0229, -0.0044)	-0.0037 (-0.0068, -0.0006)

TABLE A.6: $PPVdiff$ values with 95% CIs for the different design criteria, and a fixed restricted reference set size of 100 using the MC Reference Set. Green colour represents estimates with a CI range containing only positive values. The red colour represents estimates with a CI range containing only negative values.

Design Criterion (DC)	**SDA**	**PPV_diff (95% CI)**		
		Sensitivity		
		0.60	0.75	0.90
Shared indications - True	Omega	-0.0073 (-0.0376, 0.0087)	0.0067 (-0.0064, 0.0198)	0.0011 (-0.0051, 0.0073)
	delta_add	-0.0382 (-0.0685, -0.0127)	-0.0106 (-0.0149, 0.0005)	-0.0012 (-0.0056, 0.0032)
	IntSS	0.0066 (-0.008, 0.0212)	0.0166 (0.0063, 0.0269)	0.0085 (0.0041, 0.0128)
Shared indications - False	Omega	-0.0203 (-0.0375, -0.003)	-0.0158 (-0.0283, -0.0032)	-0.0017 (-0.0081, 0.0047)
	delta_add	0.0506 (0.029, 0.0722)	0.0164 (0.0055, 0.0272)	0.0021 (-0.0026, 0.0068)
	IntSS	-0.0288 (-0.0436, -0.014)	-0.0209 (-0.03, -0.0118)	-0.0054 (-0.0086, -0.0021)
AE is an indication - False	Omega	-0.01 (-0.0274, 0.0073)	-0.0093 (-0.0232, 0.0046)	-0.0044 (-0.0107, 0.002)
	delta_add	-0.0296 (-0.0459, -0.0133)	-0.0149 (-0.0231, -0.0067)	-0.0069 (-0.0108, -0.003)
	IntSS	-0.0142 (-0.029, 0.0006)	-0.0067 (-0.0168, 0.0034)	0.0004 (-0.0033, 0.0042)
BNF - Study	Omega	0.0789 (0.0602, 0.0976)	0.0621 (0.0448, 0.0793)	0.0143 (0.0065, 0.0221)
	delta_add	0.0917 (0.0708, 0.1126)	0.0417 (0.0276, 0.0558)	0.0096 (0.0045, 0.0147)
	IntSS	0.0834 (0.066, 0.1008)	0.0477 (0.0315, 0.0639)	0.0082 (0.0039, 0.0125)
Micromedex - Established	Omega	0.0404 (0.0216, 0.0592)	0.0329 (0.0179, 0.048)	0.0059 (-0.0012, 0.013)
	delta_add	0.0023 (-0.0156, 0.0202)	0.0112 (0.002, 0.0203)	0.0053 (0.0002, 0.0105)
	IntSS	0.0531 (0.0354, 0.0708)	0.0424 (0.0311, 0.0537)	0.0203 (0.0144, 0.0261)

Micromedex - Theoretical	Omega	-0.0269 (-0.0438, -0.0099)	-0.0212 (-0.0338, -0.0086)	-0.005 (-0.0114, 0.0014)
	delta_add	0.0101 (-0.0091, 0.0293)	0.0043 (-0.0051, 0.0137)	0.0024 (-0.0023, 0.0071)
	IntSS	-0.0275 (-0.042, -0.0129)	-0.0182 (-0.0277, -0.0087)	-0.0039 (-0.0071, -0.0006)
Micromedex - Probable	Omega	0.0105 (-0.0069, 0.0279)	0.0148 (0.0005, 0.0292)	0.0044 (-0.0029, 0.0116)
	delta_add	-0.0183 (-0.0361, -0.0004)	-0.0111 (-0.0191, -0.003)	-0.004 (-0.0083, 0.0002)
	IntSS	0.014 (-0.0034, 0.0314)	0.0037 (-0.0069, 0.0144)	0.0015 (-0.0023, 0.0053)

Chapter 6

TABLE A.7: Number of positive and negative controls for the adverse events (as MedDRA PTs) that were used in the PT-restricted analysis.

Adverse Event (MedDRA PT)	**Positive controls (N)**	**Negative controls (N)**	**Total (N)**
Rhabdomyolysis	65	40	105
QT interval prolongation	431	19	450
Myopathy	54	13	67
Hypertension	63	50	113
Hypoglycaemia	8	32	40
Haemorrhage	476	42	518
Gastrointestinal haemorrhage	70	48	118
Bradycardia	150	43	193
Torsade de pointes (TdP)	57	8	65

Appendix B

Mathematical description of the novel signal detection algorithm for drug-drug interactions - Model calibration

B.1 Mathematical Description of the Signal Detection Algorithm

To help define the notation and introduce the mathematics in a way that can be easily extended to the context of DDIs, we begin by considering signal detection in the context of a single drug.

B.1.1 Single Drugs

B.1.1.1 Counts

We assume we observe several counts. The counts reflect the fact that each individual person may or may not experience an adverse event (AE) and may or may not take a drug. We define the count, $N_{i|j}$, as the counts where:

- $i = 1$ for the count of people who took the drug and $i = 0$ for the count of people who did not take the drug;
- $j = 1$ for the count of people who experienced the AE and $j = 0$ for the count of people who did not experience the AE.

B.1.1.2 Detecting Changes

We assume two hypotheses:

- H_0: the probability that an individual report relates to the AE is θ_0 irrespective of whether the drug is taken;

- H_1: the probability that an individual report relates to the AE is θ_1 if the drug is taken and θ_0 otherwise[1].

Were each θ to be known then, assuming that the events contributing to the counts are independent random events, the corresponding likelihood would be $Bi\left(x;\theta_m,N\right)$ where this denotes a Binomial distribution of x positive examples from N events with a probability of θ_m of a positive example occurring, where θ_m could represent any of the parameters related to the hypothesis under consideration.

Since we don't know θ's, we assume conjugate Beta priors on θ_0 and θ_1 in the context of H_1 as follows:

$$p\left(\theta_0|H_1\right) = Be\left(\theta_0;\alpha_{0|1},\beta_{0|1}\right) \tag{B.1}$$

$$p\left(\theta_1|H_1\right) = Be\left(\theta_1;\alpha_{1|1},\beta_{1|1}\right) \tag{B.2}$$

We have assumed a parameterisation such that $Be\left(\theta_m;\alpha_{m|n},\beta_{m|n}\right)$ is a Beta distribution for a probability of θ_m with hyperparameters $\alpha_{m|n}$ and $\beta_{m|n}$, where m and n are indicative of the related probability (θ_m) and hypothesis (H_n), correspondingly:

$$Be\left(\theta_m;\alpha_{m|n},\beta_{m|n}\right) = \frac{1}{B(\alpha;\beta)}\theta_m^{\alpha_{m|n}-1}(1-\theta_m)^{\beta_{m|n}-1} \tag{B.3}$$

with $B(\alpha;\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ [2].

We can then assume the prior for H_0 is such that the total of the hyperparameters in the prior is equal in the two hypotheses:

$$p\left(\theta_0|H_0\right) = Be\left(\theta_0;\alpha_{0|0},\beta_{0|0}\right) \tag{B.4}$$

where

$$\alpha_{0|0} = \alpha_{0|1} + \alpha_{1|1} \tag{B.5}$$

$$\beta_{0|0} = \beta_{0|1} + \beta_{1|1}. \tag{B.6}$$

We then calculate the likelihood of observing the counts $N_{0|0}, N_{0|1}, N_{1|0}$ and $N_{1|1}$ (denoted as $N_{0:1|0:1}$ for short) for each hypothesis given the total number of reports $(N_{.|.})$ and the number of reports that include $(N_{1|.})$ and don't include $(N_{0|.})$ the drug as follows:

$$p\left(N_{0:1|0:1}|H_0,N_{0:1|.}\right) = BeBi\left(N_{0|1};\alpha_{0|0},\beta_{0|0},N_{0|.}\right) BeBi\left(N_{1|1};\alpha_{0|0}+N_{0|1},\beta_{0|0}+N_{0|0},N_{1|.}\right) \tag{B.7}$$

$$p\left(N_{0:1|0:1}|H_1,N_{0:1|.}\right) = BeBi\left(N_{0|1};\alpha_{0|1},\beta_{0|1},N_{0|.}\right) BeBi\left(N_{1|1};\alpha_{1|1},\beta_{1|1},N_{1|.}\right) \tag{B.8}$$

[1] It is implicit that $\theta_1 \neq \theta_0$.

[2] $\Gamma(N) = (N-1)!$, if N is a positive integer.

where $N_{i|.} = \sum_j N_{i|j}$ and where $BeBi\left(x;\alpha,\beta,N\right)$ is the Beta-binomial (prior predictive density) associated with observing a count of x positive examples and $N-x$ negative examples given previous experience of seeing α positive examples and β negative examples (i.e. the likelihood of observing the counts given the experience of observing other counts in the past):

$$\begin{aligned} p\left(x\right) &= \int p\left(x|\theta_m\right) p\left(\theta_m\right) d\theta_m \\ &= \int Bi\left(x;\theta_m,N\right) Be\left(\theta_m;\alpha,\beta\right) d\theta_m \\ &= BeBi\left(x;\alpha,\beta,N\right) \\ &= \binom{N}{x} \frac{B\left(x+\alpha, N-x+\beta\right)}{B\left(\alpha,\beta\right)}. \end{aligned} \tag{B.9}$$

For each combination of drug and AE, we can calculate the log-likelihood ratio (LLR):

$$LLR = \ln\left(\frac{p\left(N_{0:1|0:1}|H_1, N_{0:1|.}\right)}{p\left(N_{0:1|0:1}|H_0, N_{0:1|.}\right)}\right) \tag{B.10}$$

so that pairs can be ranked on the basis of their log-likelihood ratio.

Equivalently, we can utilise the log posterior odds ratio ($LogPostOddsRatio$):

$$LogPostOddsRatio = ln\left(\frac{p\left(H_1|N_{0:1|0:1}\right)}{p\left(H_0|N_{0:1|0:1}\right)}\right) \tag{B.11}$$

where the posterior probabilities are calculated as follows:

$$p\left(H_1|N_{0:1|0:1}\right) = \frac{p\left(N_{0:1|0:1}|H_1\right) p\left(H_1\right)}{p\left(N_{0:1|0:1}|H_1\right) p\left(H_1\right) + p\left(N_{0:1|0:1}|H_0\right) p\left(H_0\right)} \tag{B.12}$$

$$p\left(H_0|N_{0:1|0:1}\right) = \frac{p\left(N_{0:1|0:1}|H_0\right) p\left(H_0\right)}{p\left(N_{0:1|0:1}|H_1\right) p\left(H_1\right) + p\left(N_{0:1|0:1}|H_0\right) p\left(H_0\right)} \tag{B.13}$$

and $p(H_0)$, $p(H_1)$ are the priors on the respective hypotheses.

B.1.1.3 Detecting Increases

We assume three hypotheses:

- H_0: the probability that an individual report relates to the AE is θ_0 irrespective of whether the drug was taken;
- H_+: the probability that an individual report relates to the AE is θ_1 if the drug was taken and θ_0 otherwise, where $\theta_1 > \theta_0$;

- H_-: the probability that an individual report relates to the AE is θ_1 if the drug was taken and θ_0 otherwise, where $\theta_1 < \theta_0$.

We define the following terms:

$$\begin{aligned}
p\left(H_+|H_1, N_{0:1|0:1}\right) &= 1 - p\left(H_-|H_1, N_{0:1|0:1}\right) \\
&= p\left(\theta_1 > \theta_0|H_1, N_{0:1|0:1}\right) \\
&= \int_0^1 \int_0^1 p\left(\theta_0|H_1, N_{0:1|0:1}\right) p\left(\theta_1|H_1, N_{0:1|0:1}\right) \mathbb{I}_{\theta_1>\theta_0} d\theta_1 d\theta_0 \\
&= \int_0^1 p\left(\theta_0|H_1, N_{0:1|0:1}\right) \int_{\theta_0}^1 p\left(\theta_1|H_1, N_{0:1|0:1}\right) d\theta_1 d\theta_0 \\
&\approx \frac{1}{N_s} \sum_{i=1}^{N_s} (1 - I_{\theta_0^{(i)}}(N_{1|1} + \alpha_{1|1}, N_{1|0} + \beta_{1|1})) \qquad \text{(B.14)} \\
&= \int_0^1 p\left(\theta_1|H_1, N_{0:1|0:1}\right) \int_0^{\theta_1} p\left(\theta_0|H_1, N_{0:1|0:1}\right) d\theta_0 d\theta_1 \\
&\approx \frac{1}{N_s} \sum_{i=1}^{N_s} I_{\theta_1^{(i)}}(N_{0|1} + \alpha_{0|1}, N_{0|0} + \beta_{0|1}) \qquad \text{(B.15)}
\end{aligned}$$

where $\mathbb{I}_X = 1$, if X is true, N_s is the number of samples to be used in the Monte-Carlo approximation to the integrals, $I_x(\alpha, \beta)$ is the incomplete Beta function, and the integrals are approximated by sampling as follows:

$$\theta_0^{(i)} \sim Be\left(\theta_0; N_{0|1} + \alpha_{0|1}, N_{0|0} + \beta_{0|1}\right) \qquad \text{(B.16)}$$

$$\theta_1^{(i)} \sim Be\left(\theta_1; N_{1|1} + \alpha_{1|1}, N_{1|0} + \beta_{1|1}\right). \qquad \text{(B.17)}$$

We sample from the distribution with the lower variance so we use (B.14) if $N_{0|1} + \alpha_{0|1} + N_{0|0} + \beta_{0|1} > N_{1|1} + \alpha_{1|1} + N_{1|0} + \beta_{1|1}$ and (B.15) otherwise.

We can then write:

$$p\left(H_+|N_{0:1|0:1}\right) = p\left(H_+, H_1|N_{0:1|0:1}\right) = p\left(H_+|H_1, N_{0:1|0:1}\right) p\left(H_1|N_{0:1|0:1}\right) \qquad \text{(B.18)}$$

such that we can replace $p(H_1|N_{0:1|0:1})$ with $p(H_+|N_{0:1|0:1})$ for the analysis that involves the calculation of $LogPostOddsRatio$ in (B.11).

B.1.2 Two drugs

B.1.2.1 Counts

We now define eight counts, $N_{i,j|k}$, as the counts where:

- $i = 1$ for the count of people who took the first drug and $i = 0$ for the count of people who did not take the first drug;

- $j = 1$ for the count of people who took the second drug and $j = 0$ for the count of people who did not take the second drug;
- $k = 1$ for the count of people who experienced the AE and $k = 0$ for the count of people who did not experience the AE.

B.1.2.2 Hypotheses

We assume eight hypotheses, $H_{ij,k}$, with $i, j, k \in \{0, 1\}$, where $i = 1$ implies the probability that an individual report relates to the AE when the first drug is taken is surprising; $j = 1$ implies the probability that an individual report relates to the AE when the second drug is taken is surprising; $k = 1$ implies the probability that an individual report relates to the AE when the two drugs are both taken is surprising;

such that:

- $H_{00,0}$: the probability that an individual report relates to the AE is θ_0 irrespective of whether either drug is taken;
- $H_{00,1}$: the probability that an individual report relates to the AE is θ_{1+2} if both drugs are taken and θ_0 otherwise;
- $H_{01,0}$: the probability that an individual report relates to the AE is θ_2 if the second drug is taken (whether or not the first drug is also taken) and θ_0 otherwise;
- $H_{01,1}$: the probability that an individual report relates to the AE is θ_2 if the second drug is taken (if the first drug is not also taken), θ_{1+2} if both drugs are taken and θ_0 otherwise;
- $H_{10,0}$: the probability that an individual report relates to the AE is θ_1 if the first drug is taken (whether or not the second drug is also taken) and θ_0 otherwise;
- $H_{10,1}$: the probability that an individual report relates to the AE is θ_1 if the first drug is taken (if the second drug is not also taken), θ_{1+2} if both drugs are taken and θ_0 otherwise;
- $H_{11,0}$: the probability that an individual report relates to the AE is θ_1 if the first drug is taken (if the second drug is not also taken), θ_2 if the second drug is taken (if the first drug is not also taken) and θ_0 otherwise unless both drugs are taken, in which case there is no interaction such that the probability is $\theta_{1+2} = \theta_1 + (1 - \theta_1) \times \theta_2$ if both drugs are taken;
- $H_{11,1}$: the probability that an individual report relates to the AE is θ_1 if the first drug is taken (if the second drug is not also taken), θ_2 if the second drug is taken (if the first drug is not also taken), θ_{1+2} if both drugs are taken and θ_0 otherwise.

B.1.2.3 Priors

We assume conjugate Beta priors for one of the hypotheses, $H_{11,1}$, as follows:

$$p(\theta_0|H_{11,1}) = Be(\theta_0; \alpha_{0|11,1}, \beta_{0|11,1}) \tag{B.19}$$
$$p(\theta_1|H_{11,1}) = Be(\theta_1; \alpha_{1|11,1}, \beta_{1|11,1}) \tag{B.20}$$
$$p(\theta_2|H_{11,1}) = Be(\theta_2; \alpha_{2|11,1}, \beta_{2|11,1}) \tag{B.21}$$
$$p(\theta_{1+2}|H_{11,1}) = Be(\theta_{1+2}; \alpha_{1+2,1|11,1} + \alpha_{1+2,2|11,1}, \beta_{1+2,1|11,1} + \beta_{1+2,2|11,1}). \tag{B.22}$$

We choose the total of the hyperparameters in the prior to be equal in all hypotheses, starting with $H_{00,0}$:

$$p(\theta_0|H_{00,0}) = Be(\theta_0; \alpha_{0|00,0}, \beta_{0|00,0}) \tag{B.23}$$

where

$$\alpha_{0|00,0} = \alpha_{0|11,1} + \alpha_{1|11,1} + \alpha_{2|11,1} + \alpha_{1+2,1|11,1} + \alpha_{1+2,2|11,1} \tag{B.24}$$
$$\beta_{0|00,0} = \beta_{0|11,1} + \beta_{1|11,1} + \beta_{2|11,1} + \beta_{1+2,1|11,1} + \beta_{1+2,2|11,1}. \tag{B.25}$$

We then proceed to $H_{00,1}$:

$$p(\theta_0|H_{00,1}) = Be(\theta_0; \alpha_{0|00,1}, \beta_{0|00,1}) \tag{B.26}$$
$$p(\theta_{1+2}|H_{00,1}) = Be(\theta_{1+2}; \alpha_{1+2|00,1}, \beta_{1+2|00,1}) \tag{B.27}$$

where

$$\alpha_{0|00,1} = \alpha_{0|11,1} + \alpha_{1|11,1} + \alpha_{2|11,1} \tag{B.28}$$
$$\beta_{0|00,1} = \beta_{0|11,1} + \beta_{1|11,1} + \beta_{2|11,1} \tag{B.29}$$
$$\alpha_{1+2|00,1} = \alpha_{1+2,1|11,1} + \alpha_{1+2,2|11,1} \tag{B.30}$$
$$\beta_{1+2|00,1} = \beta_{1+2,1|11,1} + \beta_{1+2,2|11,1}. \tag{B.31}$$

Then we consider $H_{01,0}$:

$$p(\theta_0|H_{01,0}) = Be(\theta_0; \alpha_{0|01,0}, \beta_{0|01,0}) \tag{B.32}$$
$$p(\theta_2|H_{01,0}) = Be(\theta_2; \alpha_{2|01,0}, \beta_{2|01,0}) \tag{B.33}$$

where

$$\alpha_{0|01,0} = \alpha_{0|11,1} + \alpha_{1|11,1} + \alpha_{1+2,1|11,1} \tag{B.34}$$
$$\beta_{0|01,0} = \beta_{0|11,1} + \beta_{1|11,1} + \beta_{1+2,1|11,1} \tag{B.35}$$
$$\alpha_{2|01,0} = \alpha_{2|11,1} + \alpha_{1+2,2|11,1} \tag{B.36}$$
$$\beta_{2|01,0} = \beta_{2|11,1} + \beta_{1+2,2|11,1}. \tag{B.37}$$

And then we consider $H_{01,1}$:

$$p\left(\theta_0|H_{01,1}\right) = Be\left(\theta_0; \alpha_{0|01,1}, \beta_{0|01,1}\right) \tag{B.38}$$
$$p\left(\theta_2|H_{01,1}\right) = Be\left(\theta_2; \alpha_{2|01,1}, \beta_{2|01,1}\right) \tag{B.39}$$
$$p\left(\theta_{1+2}|H_{01,1}\right) = Be\left(\theta_{1+2}; \alpha_{1+2|01,1}, \beta_{1+2|01,1}\right) \tag{B.40}$$

where

$$\alpha_{0|01,1} = \alpha_{0|11,1} + \alpha_{1|11,1} \tag{B.41}$$
$$\beta_{0|01,1} = \beta_{0|11,1} + \beta_{1|11,1} \tag{B.42}$$
$$\alpha_{2|01,1} = \alpha_{2|11,1} \tag{B.43}$$
$$\beta_{2|01,1} = \beta_{2|11,1} \tag{B.44}$$
$$\alpha_{1+2|01,1} = \alpha_{1+2,1|11,1} + \alpha_{1+2,2|11,1} \tag{B.45}$$
$$\beta_{1+2|01,1} = \beta_{1+2,1|11,1} + \beta_{1+2,2|11,1}. \tag{B.46}$$

Then we consider $H_{10,0}$:

$$p\left(\theta_0|H_{10,0}\right) = Be\left(\theta_0; \alpha_{0|10,0}, \beta_{0|10,0}\right) \tag{B.47}$$
$$p\left(\theta_1|H_{10,0}\right) = Be\left(\theta_1; \alpha_{1|10,0}, \beta_{1|10,0}\right) \tag{B.48}$$

where

$$\alpha_{0|10,0} = \alpha_{0|11,1} + \alpha_{2|11,1} + \alpha_{1+2,2|11,1} \tag{B.49}$$
$$\beta_{0|10,0} = \beta_{0|11,1} + \beta_{2|11,1} + \beta_{1+2,2|11,1} \tag{B.50}$$
$$\alpha_{1|10,0} = \alpha_{1|11,1} + \alpha_{1+2,1|11,1} \tag{B.51}$$
$$\beta_{1|10,0} = \beta_{1|11,1} + \beta_{1+2,1|11,1}. \tag{B.52}$$

Penultimately, we consider $H_{10,1}$:

$$p\left(\theta_0|H_{10,1}\right) = Be\left(\theta_0; \alpha_{0|10,1}, \beta_{0|10,1}\right) \tag{B.53}$$
$$p\left(\theta_1|H_{10,1}\right) = Be\left(\theta_1; \alpha_{1|10,1}, \beta_{1|10,1}\right) \tag{B.54}$$
$$p\left(\theta_{1+2}|H_{10,1}\right) = Be\left(\theta_{1+2}; \alpha_{1+2|10,1}, \beta_{1+2|10,1}\right) \tag{B.55}$$

where

$$\alpha_{0|10,1} = \alpha_{0|11,1} + \alpha_{2|11,1} \tag{B.56}$$
$$\beta_{0|10,1} = \beta_{0|11,1} + \beta_{2|11,1} \tag{B.57}$$
$$\alpha_{1|10,1} = \alpha_{1|11,1} \tag{B.58}$$
$$\beta_{1|10,1} = \beta_{1|11,1} \tag{B.59}$$
$$\alpha_{1+2|10,1} = \alpha_{1+2,1|11,1} + \alpha_{1+2,2|11,1} \tag{B.60}$$
$$\beta_{1+2|10,1} = \beta_{1+2,1|11,1} + \beta_{1+2,2|11,1}. \tag{B.61}$$

Finally, we consider $H_{11,0}$:

$$p\left(\theta_0|H_{11,0}\right) = Be\left(\theta_0; \alpha_{0|11,0}, \beta_{0|11,0}\right) \tag{B.62}$$
$$p\left(\theta_1|H_{11,0}\right) = Be\left(\theta_1; \alpha_{1|11,0}, \beta_{1|11,0}\right) \tag{B.63}$$
$$p\left(\theta_2|H_{11,0}\right) = Be\left(\theta_2; \alpha_{2|11,0}, \beta_{2|11,0}\right) \tag{B.64}$$

where

$$\alpha_{0|11,0} = \alpha_{0|11,1} \tag{B.65}$$
$$\beta_{0|11,0} = \beta_{0|11,1} \tag{B.66}$$
$$\alpha_{1|11,0} = \alpha_{1|11,1} + \alpha_{1+2,1|11,1} \tag{B.67}$$
$$\beta_{1|11,0} = \beta_{1|11,1} + \beta_{1+2,1|11,1} \tag{B.68}$$
$$\alpha_{2|11,0} = \alpha_{2|11,1} + \alpha_{1+2,2|11,1} \tag{B.69}$$
$$\beta_{2|11,0} = \beta_{2|11,1} + \beta_{1+2,2|11,1}. \tag{B.70}$$

B.1.2.4 Likelihoods

We then calculate the likelihoods of observing the counts $N_{0:1,0:1|0:1}$ for each hypothesis given the total number of reports ($N_{.,.|.}$) and the number of reports that include ($N_{1,.|.}$, $N_{.,1|.}$) and don't include ($N_{0,.|.}$, $N_{.,0|.}$) each drug as follows:

$$p\left(N_{0:1,0:1|0:1}|H_{00,0},N_{0:1,0:1|.}\right)=BeBi\left(N_{0,0|1};\alpha_{0|00,0},\beta_{0|00,0},N_{0,0|.}\right)$$
$$\times BeBi\left(N_{1,0|1};\alpha_{0|00,0}+N_{0,0|1},\beta_{0|00,0}+N_{0,0|0},N_{1,0|.}\right)$$
$$\times BeBi\left(N_{0,1|1};\alpha_{0|00,0}+N_{0,0|1}+N_{1,0|1},\beta_{0|00,0}+N_{0,0|0}+N_{1,0|0},N_{0,1|.}\right)$$
$$\times BeBi\left(N_{1,1|1};\alpha_{0|00,0}+N_{0,0|1}+N_{1,0|1}+N_{0,1|1},\beta_{0|00,0}+N_{0,0|0}+N_{1,0|0}+N_{0,1|0},N_{1,1|.}\right) \tag{B.71}$$

$$p\left(N_{0:1,0:1|0:1}|H_{00,1},N_{0:1,0:1|.}\right)=BeBi\left(N_{0,0|1};\alpha_{0|00,1},\beta_{0|00,1},N_{0,0|.}\right)$$
$$\times BeBi\left(N_{1,0|1};\alpha_{0|00,1}+N_{0,0|1},\beta_{0|00,1}+N_{0,0|0},N_{1,0|.}\right)$$
$$\times BeBi\left(N_{0,1|1};\alpha_{0|00,1}+N_{0,0|1}+N_{1,0|1},\beta_{0|00,1}+N_{0,0|0}+N_{1,0|0},N_{0,1|.}\right)$$
$$\times BeBi\left(N_{1,1|1};\alpha_{1+2|00,1},\beta_{1+2|00,1},N_{1,1|.}\right) \tag{B.72}$$

$$p\left(N_{0:1,0:1|0:1}|H_{01,0},N_{0:1,0:1|.}\right)=BeBi\left(N_{0,0|1};\alpha_{0|01,0},\beta_{0|01,0},N_{0,0|.}\right)$$
$$\times BeBi\left(N_{1,0|1};\alpha_{0|01,0}+N_{0,0|1},\beta_{0|01,0}+N_{0,0|0},N_{1,0|.}\right)$$
$$\times BeBi\left(N_{0,1|1};\alpha_{2|01,0},\beta_{2|01,0},N_{0,1|.}\right)$$
$$\times BeBi\left(N_{1,1|1};\alpha_{2|01,0}+N_{0,1|1},\beta_{2|01,0}+N_{0,1|0},N_{1,1|.}\right) \tag{B.73}$$

$$p\left(N_{0:1,0:1|0:1}|H_{01,1},N_{0:1,0:1|.}\right)=BeBi\left(N_{0,0|1};\alpha_{0|01,1},\beta_{0|01,1},N_{0,0|.}\right)$$
$$\times BeBi\left(N_{1,0|1};\alpha_{0|01,1}+N_{0,0|1},\beta_{0|01,1}+N_{0,0|0},N_{1,0|.}\right)$$
$$\times BeBi\left(N_{0,1|1};\alpha_{2|01,1},\beta_{2|01,1},N_{0,1|.}\right)$$
$$\times BeBi\left(N_{1,1|1};\alpha_{1+2|01,1},\beta_{1+2|01,1},N_{1,1|.}\right) \tag{B.74}$$

$$p\left(N_{0:1,0:1|0:1}|H_{10,0},N_{0:1,0:1|.}\right)=BeBi\left(N_{0,0|1};\alpha_{0|10,0},\beta_{0|10,0},N_{0,0|.}\right)$$
$$\times BeBi\left(N_{1,0|1};\alpha_{1|10,0},\beta_{1|10,0},N_{1,0|.}\right)$$
$$\times BeBi\left(N_{0,1|1};\alpha_{0|10,0}+N_{0,0|1},\beta_{0|10,0}+N_{0,0|0},N_{0,1|.}\right)$$
$$\times BeBi\left(N_{1,1|1};\alpha_{1|10,0}+N_{1,0|1},\beta_{1|10,0}+N_{1,0|0},N_{1,1|.}\right) \tag{B.75}$$

$$p\left(N_{0:1,0:1|0:1}|H_{10,1},N_{0:1,0:1|.}\right)=BeBi\left(N_{0,0|1};\alpha_{0|10,1},\beta_{0|10,1},N_{0,0|.}\right)$$
$$\times BeBi\left(N_{1,0|1};\alpha_{1|10,1},\beta_{1|10,1},N_{1,0|.}\right)$$
$$\times BeBi\left(N_{0,1|1};\alpha_{0|10,1}+N_{0,0|1},\beta_{0|10,1}+N_{0,0|0},N_{0,1|.}\right)$$
$$\times BeBi\left(N_{1,1|1};\alpha_{1+2|10,1},\beta_{1+2|10,1},N_{1,1|.}\right) \tag{B.76}$$

$$p\left(N_{0:1,0:1|0:1}|H_{11,0},N_{0:1,0:1|.}\right)=BeBi\left(N_{0,0|1};\alpha_{0|11,0},\beta_{0|11,0},N_{0,0|.}\right)$$
$$\times BeBi\left(N_{1,0|1};\alpha_{1|11,0},\beta_{1|11,0},N_{1,0|.}\right)$$
$$\times BeBi\left(N_{0,1|1};\alpha_{2|11,0},\beta_{2|11,0},N_{0,1|.}\right)$$
$$\times\int Bi\left(N_{1,1|1};\theta_1+(1-\theta_1)\times\theta_2,N_{1,1|.}\right)Be\left(\theta_1;\alpha_{1|11,0}+N_{1,0|1},\beta_{1|11,0}+N_{1,0|0}\right)$$
$$Be\left(\theta_2;\alpha_{2|11,0}+N_{0,1|1},\beta_{2|11,0}+N_{0,1|0}\right)d\theta_1 d\theta_2 \tag{B.77}$$

$$p\left(N_{0:1,0:1|0:1}|H_{11,1},N_{0:1,0:1|.}\right)=BeBi\left(N_{0,0|1};\alpha_{0|11,1},\beta_{0|11,1},N_{0,0|.}\right)$$
$$\times BeBi\left(N_{1,0|1};\alpha_{1|11,1},\beta_{1|11,1},N_{1,0|.}\right)$$
$$\times BeBi\left(N_{0,1|1};\alpha_{2|11,1},\beta_{2|11,1},N_{0,1|.}\right)$$
$$\times BeBi\left(N_{1,1|1};\alpha_{1+2|11,1},\beta_{1+2|11,1},N_{1,1|.}\right) \tag{B.78}$$

where the integral in **Equation B.77** is approximated via Monte-Carlo integration using 1000 simulations for the Beta distributions.

The log-likelihood ratio is calculated as follows:

$$LLR = \ln\left(\frac{\sum_i \sum_j p\left(N_{0:1,0:1|0:1}|H_{ij,1}, N_{0:1,0:1|.}\right)}{\sum_i \sum_j p\left(N_{0:1,0:1|0:1}|H_{ij,0}, N_{0:1,0:1|.}\right)}\right) \tag{B.79}$$

and the log posterior odds ratio is equal to:

$$LogPostOddsRatio = \ln\left(\frac{\sum_i \sum_j p\left(H_{ij,1}|N_{0:1,0:1|0:1}\right)}{\sum_i \sum_j p\left(H_{ij,0}, N_{0:1,0:1|0:1}\right)}\right) \tag{B.80}$$

where the posterior probability in the case of $H_{00,0}$, for example, is calculated as follows:

$$\begin{aligned} p\left(H_{00,0}|N_{0:1,0:1|0:1}\right) &= \frac{p\left(N_{0:1,0:1|0:1}|H_{00,0}\right) p\left(H_{00,0}\right)}{p\left(N_{0:1,0:1|0:1}\right)} \\ &= \frac{p\left(N_{0:1,0:1|0:1}|H_{00,0}\right) p\left(H_{00,0}\right)}{\sum_i \sum_j \sum_k p\left(N_{0:1,0:1|0:1}|H_{ij,k}\right) p\left(H_{ij,k}\right)}. \end{aligned} \tag{B.81}$$

B.1.2.5 Detecting Increases

We now consider the following twelve hypotheses, $H_{ij,k}$, with $i, j, k \in \{0, +, -\}$, where:

- i implies the probability that an individual report relates to the AE when the first drug is taken is not surprising (0), higher (+) or lower (-) than expected;
- j implies the probability that an individual report relates to the AE when the second drug is taken is not surprising (0), higher (+) or lower (-) than expected;
- k implies the probability that an individual report relates to the AE when the two drugs are both taken is not surprising (0), higher (+) or lower (-) than expected;

as follows:

- $H_{00,0}$: the probability that an individual report relates to the adverse event is θ_0 irrespective of whether either drug is taken;
- $H_{00,+}$: the probability that an individual report relates to the adverse event is θ_{1+2} if both drugs are taken and θ_0 otherwise, where $\theta_{1+2} > \theta_0$;
- $H_{00,-}$: the probability that an individual report relates to the adverse event is θ_{1+2} if both drugs are taken and θ_0 otherwise, where $\theta_{1+2} < \theta_0$;
- $H_{10,0}$: the probability that an individual report relates to the adverse event is θ_1 if the first drug is taken (whether or not the second drug is also taken) and θ_0 otherwise;

- $H_{10,+}$: the probability that an individual report relates to the adverse event is θ_1 if the first drug is taken (if the second drug is not also taken), θ_{1+2} if both drugs are taken and θ_0 otherwise, where $\theta_{1+2} > \theta_0$ and $\theta_{1+2} > \theta_1$ (it might hold that $\theta_{1+2} > \theta_1 > \theta_0$ or $\theta_{1+2} > \theta_0 > \theta_1$);

- $H_{10,-}$: the probability that an individual report relates to the adverse event is θ_1 if the first drug is taken (if the second drug is not also taken), θ_{1+2} if both drugs are taken and θ_0 otherwise, where $\theta_{1+2} < \theta_0$ or $\theta_{1+2} < \theta_1$;

- $H_{01,0}$: the probability that an individual report relates to the adverse event is θ_2 if the second drug is taken (whether or not the first drug is also taken) and θ_0 otherwise;

- $H_{01,+}$: the probability that an individual report relates to the adverse event is θ_2 if the second drug is taken (if the first drug is not also taken), θ_{1+2} if both drugs are taken and θ_0 otherwise, where $\theta_{1+2} > \theta_0$ and $\theta_{1+2} > \theta_2$ (it might hold that $\theta_{1+2} > \theta_2 > \theta_0$ or $\theta_{1+2} > \theta_0 > \theta_2$);

- $H_{01,-}$: the probability that an individual report relates to the adverse event is θ_2 if the second drug is taken (if the first drug is not also taken), θ_{1+2} if both drugs are taken and θ_0 otherwise, where $\theta_{1+2} < \theta_0$ or $\theta_{1+2} < \theta_2$;

- $H_{11,0}$: the probability that an individual report relates to the adverse event is θ_1 if the first drug is taken (if the second drug is not also taken), θ_2 if the second drug is taken (if the first drug is not also taken) and θ_0 otherwise unless both drugs are taken, in which case there is no interaction such that the probability is $\theta_{1+2} = \theta_1 + (1 - \theta_1) \times \theta_2$ if both drugs are taken;

- $H_{11,+}$: the probability that an individual report relates to the adverse event is θ_1 if the first drug is taken (if the second drug is not also taken), θ_2 if the second drug is taken (if the first drug is not also taken), θ_{1+2} if both drugs are taken and θ_0 otherwise, where $\theta_{1+2} > \theta_0$ and $\theta_{1+2} > \theta_1$ and $\theta_{1+2} > \theta_2$ (it might hold that $\theta_{1+2} > \theta_0 > \theta_1 > \theta_2$ or $\theta_{1+2} > \theta_0 > \theta_2 > \theta_1$ or $\theta_{1+2} > \theta_1 > \theta_2 > \theta_0$ or $\theta_{1+2} > \theta_1 > \theta_0 > \theta_2$ or $\theta_{1+2} > \theta_2 > \theta_1 > \theta_0$ or $\theta_{1+2} > \theta_2 > \theta_0 > \theta_1$);

- $H_{11,-}$: the probability that an individual report relates to the adverse event is θ_1 if the first drug is taken (if the second drug is not also taken), θ_2 if the second drug is taken (if the first drug is not also taken), θ_{1+2} if both drugs are taken and θ_0 otherwise, where $\theta_{1+2} < \theta_0$ or $\theta_{1+2} < \theta_1$ or $\theta_{1+2} < \theta_2$.

To focus our attention on the probability of increased rates of occurrence rather than simply changes to those rates, we now need to define the following terms:

$$\begin{aligned}
p\left(H_{00,+}|H_{00,1}, N_{0:1,0:1|.}\right) &= 1 - p\left(H_{00,-}|H_{00,1}, N_{0:1,0:1|.}\right) \\
&= p\left(\theta_{1+2} > \theta_0|H_{00,1}, N_{0:1,0:1|.}\right) \\
&= \int_0^1 \int_0^1 p\left(\theta_0|H_{00,1}, N_{0:1,0:1|.}\right) p\left(\theta_{1+2}|H_{00,1}, N_{0:1,0:1|.}\right) \mathbb{I}_{\theta_{1+2}>\theta_0} d\theta_{1+2} d\theta_0 \\
&= \int_0^1 p\left(\theta_0|H_{00,1}, N_{0:1,0:1|.}\right) \int_{\theta_0}^1 p\left(\theta_{1+2}|H_{00,1}, N_{0:1,0:1|.}\right) d\theta_{1+2} d\theta_0 \\
&\approx \frac{1}{N_s} \times \sum_{i=1}^{N_s} (1 - I_{\theta_0^{(i)}}(N_{1,1|1} + \alpha_{1+2|00,1}, N_{1,1|0} + \beta_{1+2|00,1})) \qquad \text{(B.82)} \\
&= \int_0^1 p\left(\theta_{1+2}|H_{00,1}, N_{0:1,0:1|.}\right) \int_0^{\theta_{1+2}} p\left(\theta_0|H_{00,1}, N_{0:1,0:1|.}\right) d\theta_0 d\theta_{1+2} \\
&\approx \frac{1}{N_s} \times \\
&\sum_{i=1}^{N_s} I_{\theta_{1+2}^{(i)}}(N_{1,0|1} + N_{0,1|1} + N_{0,0|1} + \alpha_{0|00,1}, N_{1,0|0} + N_{0,1|0} + N_{0,0|0} + \beta_{0|00,1}) \qquad \text{(B.83)}
\end{aligned}$$

where $\mathbb{I}_X = 1$, if X is true.

Again, we can approximate the integrals by sampling as follows:

$$\theta_0^{(i)} \sim Be\left(\theta_0; N_{1,0|1} + N_{0,1|1} + N_{0,0|1} + \alpha_{0|00,1}, N_{1,0|0} + N_{0,1|0} + N_{0,0|0} + \beta_{0|00,1}\right) \quad \text{(B.84)}$$

$$\theta_{1+2}^{(i)} \sim Be\left(\theta_{1+2}; N_{1,1|1} + \alpha_{1+2|00,1}, N_{1,1|0} + \beta_{1+2|00,1}\right). \quad \text{(B.85)}$$

In a similar way, we have that:

$$\begin{aligned}
p\left(H_{10,+}|H_{10,1}, N_{0:1,0:1|.}\right) &= 1 - p\left(H_{10,-}|H_{10,1}, N_{0:1,0:1|.}\right) \\
&= p\left(\theta_{1+2} > \theta_0|H_{10,1}, N_{0:1,0:1|.}\right) \times p\left(\theta_{1+2} > \theta_1|H_{10,1}, N_{0:1,0:1|.}\right)
\end{aligned} \qquad \text{(B.86)}$$

where

$$
\begin{aligned}
p\left(\theta_{1+2} > \theta_0 | H_{10,1}, N_{0:1,0:1|.}\right) &= \int_0^1 \int_0^1 p\left(\theta_0, \theta_{1+2} | H_{10,1}, N_{0:1,0:1|.}\right) \mathbb{I}_{\theta_{1+2}>\theta_0} d\theta_{1+2} d\theta_0 \\
&= \int_0^1 p\left(\theta_{1+2} | H_{10,1}, N_{0:1,0:1|.}\right) \int_0^{\theta_{1+2}} p\left(\theta_0 | H_{10,1}, N_{0:1,0:1|.}\right) d\theta_0 d\theta_{1+2} \\
&\approx \frac{1}{N_s} \sum_{i=1}^{N_s} I_{\theta_{1+2}^{(i)}}(N_{0,1|1} + N_{0,0|1} + \alpha_{0|10,1}, N_{0,1|0} + N_{0,0|0} + \beta_{0|10,1})
\end{aligned} \tag{B.87}
$$

$$
\begin{aligned}
p\left(\theta_{1+2} > \theta_1 | H_{10,1}, N_{0:1,0:1|.}\right) &= \int_0^1 \int_0^1 p\left(\theta_1, \theta_{1+2} | H_{10,1}, N_{0:1,0:1|.}\right) \mathbb{I}_{\theta_{1+2}>\theta_1} d\theta_{1+2} d\theta_1 \\
&= \int_0^1 p\left(\theta_{1+2} | H_{10,1}, N_{0:1,0:1|.}\right) \int_0^{\theta_{1+2}} p\left(\theta_1 | H_{10,1}, N_{0:1,0:1|.}\right) d\theta_1 d\theta_{1+2} \\
&\approx \frac{1}{N_s} \sum_{i=1}^{N_s} I_{\theta_{1+2}^{(i)}}(N_{1,0|1} + \alpha_{1|10,1}, N_{1,0|0} + \beta_{1|10,1})
\end{aligned} \tag{B.88}
$$

and

$$
\theta_{1+2}^{(i)} \sim Be\left(\theta_{1+2}; N_{1,1|1} + \alpha_{1+2|10,1}, N_{1,1|0} + \beta_{1+2|10,1}\right). \tag{B.89}
$$

It also follows that:

$$
\begin{aligned}
p\left(H_{01,+} | H_{01,1}, N_{0:1,0:1|.}\right) &= 1 - p\left(H_{01,-} | H_{01,1}, N_{0:1,0:1|.}\right) \\
&= p\left(\theta_{1+2} > \theta_0 | H_{01,1}, N_{0:1,0:1|.}\right) \times p\left(\theta_{1+2} > \theta_2 | H_{01,1}, N_{0:1,0:1|.}\right)
\end{aligned} \tag{B.90}
$$

where

$$
\begin{aligned}
p\left(\theta_{1+2} > \theta_0 | H_{01,1}, N_{0:1,0:1|.}\right) &= \int_0^1 \int_0^1 p\left(\theta_0, \theta_{1+2} | H_{10,1}, N_{0:1,0:1|.}\right) \mathbb{I}_{\theta_{1+2}>\theta_0} d\theta_{1+2} d\theta_0 \\
&= \int_0^1 p\left(\theta_{1+2} | H_{10,1}, N_{0:1,0:1|.}\right) \int_0^{\theta_{1+2}} p\left(\theta_0 | H_{10,1}, N_{0:1,0:1|.}\right) d\theta_0 d\theta_{1+2} \\
&\approx \frac{1}{N_s} \sum_{i=1}^{N_s} I_{\theta_{1+2}^{(i)}}(N_{1,0|1} + N_{0,0|1} + \alpha_{0|01,1}, N_{1,0|0} + N_{0,0|0} + \beta_{0|01,1})
\end{aligned} \tag{B.91}
$$

$$
\begin{aligned}
p\left(\theta_{1+2} > \theta_2 | H_{01,1}, N_{0:1,0:1|.}\right) &= \int_0^1 \int_0^1 p\left(\theta_2, \theta_{1+2} | H_{01,1}, N_{0:1,0:1|.}\right) \mathbb{I}_{\theta_{1+2}>\theta_2} d\theta_{1+2} d\theta_2 \\
&= \int_0^1 p\left(\theta_{1+2} | H_{01,1}, N_{0:1,0:1|.}\right) \int_0^{\theta_{1+2}} p\left(\theta_2 | H_{01,1}, N_{0:1,0:1|.}\right) d\theta_2 d\theta_{1+2} \\
&\approx \frac{1}{N_s} \sum_{i=1}^{N_s} I_{\theta_{1+2}^{(i)}}(N_{0,1|1} + \alpha_{2|01,1}, N_{0,1|0} + \beta_{2|01,1}) \qquad \text{(B.92)}
\end{aligned}
$$

and

$$
\theta_{1+2}^{(i)} \sim Be\left(\theta_{1+2}; N_{1,1|1} + \alpha_{1+2|01,1}, N_{1,1|0} + \beta_{1+2|01,1}\right). \qquad \text{(B.93)}
$$

Finally, we have that:

$$
\begin{aligned}
p\left(H_{11,+} | H_{11,1}, N_{0:1,0:1|.}\right) &= 1 - p\left(H_{11,-} | H_{11,1}, N_{0:1,0:1|.}\right) \\
&= p\left(\theta_{1+2} > \theta_0 | H_{11,1}, N_{0:1,0:1|.}\right) \times p\left(\theta_{1+2} > \theta_1 | H_{11,1}, N_{0:1,0:1|.}\right) \\
&\times p\left(\theta_{1+2} > \theta_2 | H_{11,1}, N_{0:1,0:1|.}\right) \qquad \text{(B.94)}
\end{aligned}
$$

where

$$
\begin{aligned}
p\left(\theta_{1+2} > \theta_0 | H_{11,1}, N_{0:1,0:1|.}\right) &= \int_0^1 \int_0^1 p\left(\theta_0, \theta_{1+2} | H_{11,1}, N_{0:1,0:1|.}\right) \mathbb{I}_{\theta_{1+2}>\theta_0} d\theta_{1+2} d\theta_0 \\
&= \int_0^1 p\left(\theta_{1+2} | H_{11,1}, N_{0:1,0:1|.}\right) \int_0^{\theta_{1+2}} p\left(\theta_0 | H_{11,1}, N_{0:1,0:1|.}\right) d\theta_0 d\theta_{1+2} \\
&\approx \frac{1}{N_s} \sum_{i=1}^{N_s} I_{\theta_{1+2}^{(i)}}(N_{0,0|1} + \alpha_{0|11,1}, N_{0,0|0} + \beta_{0|11,1}) \qquad \text{(B.95)}
\end{aligned}
$$

$$
\begin{aligned}
p\left(\theta_{1+2} > \theta_1 | H_{11,1}, N_{0:1,0:1|.}\right) &= \int_0^1 \int_0^1 p\left(\theta_1, \theta_{1+2} | H_{11,1}, N_{0:1,0:1|.}\right) \mathbb{I}_{\theta_{1+2}>\theta_1} d\theta_{1+2} d\theta_1 \\
&= \int_0^1 p\left(\theta_{1+2} | H_{11,1}, N_{0:1,0:1|.}\right) \int_0^{\theta_{1+2}} p\left(\theta_1 | H_{11,1}, N_{0:1,0:1|.}\right) d\theta_1 d\theta_{1+2} \\
&\approx \frac{1}{N_s} \sum_{i=1}^{N_s} I_{\theta_{1+2}^{(i)}}(N_{1,0|1} + \alpha_{1|11,1}, N_{1,0|0} + \beta_{1|11,1}) \qquad \text{(B.96)}
\end{aligned}
$$

$$
\begin{aligned}
p\left(\theta_{1+2} > \theta_2 | H_{11,1}, N_{0:1,0:1|.}\right) &= \int_0^1 \int_0^1 p\left(\theta_2, \theta_{1+2} | H_{11,1}, N_{0:1,0:1|.}\right) \mathbb{I}_{\theta_{1+2}>\theta_2} d\theta_{1+2} d\theta_2 \\
&= \int_0^1 p\left(\theta_{1+2} | H_{11,1}, N_{0:1,0:1|.}\right) \int_0^{\theta_{1+2}} p\left(\theta_2 | H_{11,1}, N_{0:1,0:1|.}\right) d\theta_2 d\theta_{1+2} \\
&\approx \frac{1}{N_s} \sum_{i=1}^{N_s} I_{\theta_{1+2}^{(i)}}(N_{0,1|1} + \alpha_{2|11,1}, N_{0,1|0} + \beta_{2|11,1})
\end{aligned} \tag{B.97}
$$

and

$$
\theta_{1+2}^{(i)} \sim Be\left(\theta_{1+2}; N_{1,1|1} + \alpha_{1+2|11,1}, N_{1,1|0} + \beta_{1+2|11,1}\right). \tag{B.98}
$$

We can then write:

$$
p\left(H_{00,+} | N_{0:1|0:1}\right) = p\left(H_{00,+}, H_{00,1} | N_{0:1|0:1}\right) = p\left(H_{00,+} | H_{00,1}, N_{0:1|0:1}\right) p\left(H_{00,1} | N_{0:1|0:1}\right) \tag{B.99}
$$

$$
p\left(H_{10,+} | N_{0:1|0:1}\right) = p\left(H_{10,+}, H_{10,1} | N_{0:1|0:1}\right) = p\left(H_{10,+} | H_{10,1}, N_{0:1|0:1}\right) p\left(H_{10,1} | N_{0:1|0:1}\right) \tag{B.100}
$$

$$
p\left(H_{01,+} | N_{0:1|0:1}\right) = p\left(H_{01,+}, H_{01,1} | N_{0:1|0:1}\right) = p\left(H_{01,+} | H_{01,1}, N_{0:1|0:1}\right) p\left(H_{01,1} | N_{0:1|0:1}\right) \tag{B.101}
$$

$$
p\left(H_{11,+} | N_{0:1|0:1}\right) = p\left(H_{11,+}, H_{11,1} | N_{0:1|0:1}\right) = p\left(H_{11,+} | H_{11,1}, N_{0:1|0:1}\right) p\left(H_{11,1} | N_{0:1|0:1}\right) \tag{B.102}
$$

such that we can replace the posterior probabilities for changing rates for the analysis that involves the calculation of *LogPostOddsRatio* in **Equation B.80**.

B.2 Prior parameter estimation

B.2.1 Observed adverse event rate calcuation

The observed rate of an AE that was reported in FAERS, *ObservedRate*, can be estimated as follows:

$$
ObservedRate = \frac{\sum_{i=0}^{1} \sum_{j=0}^{1} N_{i,j|1}}{\sum_{i=0}^{1} \sum_{j=0}^{1} \sum_{k=0}^{1} N_{i,j|k}} \tag{B.103}
$$

A histogram that shows the observed rates in FAERS for the AEs that are mentioned in more than 10 reports is shown in **Figure B.1**.

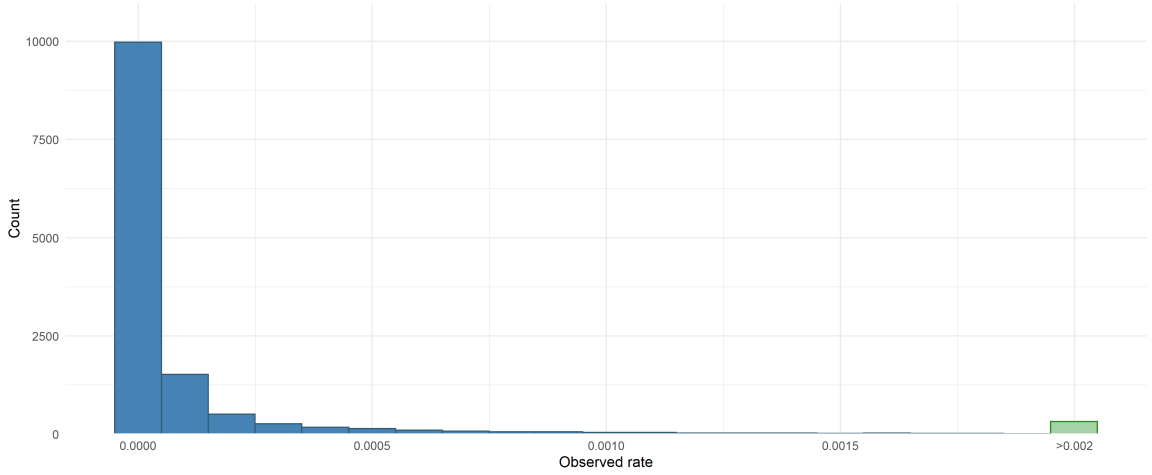


FIGURE B.1: Histogram of observed AE rates in FAERS.

For $H_{11,1}$, we can estimate the hyperparameters $\alpha_{0|11,1}$ and $\beta_{0|11,1}$ from the mean and standard deviation of the observed AE rates in FAERS using the method of moments[3]:

$$E(\theta_0) = \frac{\alpha}{\alpha + \beta} \tag{B.104}$$

$$Var(\theta_0) = \frac{\alpha \times \beta}{(\alpha + \beta)^2 \times (\alpha + \beta + 1)} \tag{B.105}$$

$$\alpha + \beta = \frac{E(\theta_0) \times (1 - E(\theta_0))}{Var(\theta_0)} - 1 \tag{B.106}$$

The estimated Beta prior hyperparameters are equal to:

$$\alpha^\star = 0.0171 \text{ and } \beta^\star = 112$$

and can be used to run the hypothesis testing framework in the case of two drugs, as it can be seen as reflecting the observed AE rates across all reports.

The following hyperparameter set was selected to run the analysis:

[3]Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., & Rubin, D.B. (2013). Bayesian Data Analysis (3rd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/b16018

$$\alpha_{0|11,1} = \alpha^\star$$
$$\alpha_{1|11,1} = 1$$
$$\alpha_{2|11,1} = 1$$
$$\alpha_{12,1|11,1} = 0.5$$
$$\alpha_{12,2|11,1} = 0.5$$
$$\beta_{0|11,1} = \beta^\star$$
$$\beta_{1|11,1} = 1$$
$$\beta_{2|11,1} = 1$$
$$\beta_{12,1|11,1} = 0.5$$
$$\beta_{12,2|11,1} = 0.5.$$

B.2.2 Sensitivity analysis

As part of model checking, it is important to check how much the predictions change when modifying the hyperparameters of the prior distribution. Therefore, we selected six different sets of hyperparameters (**Table B.1**) to test the above-described hypothesis testing framework for two drugs and performed receiver operating characteristic (ROC) analysis.

	Hyperparameter set					
Hyperparameter	**1**	**2**	**3**	**4**	**5**	**6**
$\alpha_{0\|11,1}$	1	1	3	$\alpha^\star$	$\alpha^\star$	$0.95 \times \alpha^\star$
$\alpha_{1\|11,1}$	1	1	1	1	1	1
$\alpha_{2\|11,1}$	1	1	1	1	1	1
$\alpha_{12,1\|11,1}$	0.5	0.5	0.5	0.5	0.5	0.5
$\alpha_{12,2\|11,1}$	0.5	0.5	0.5	0.5	0.5	0.5
$\beta_{0\|11,1}$	1	0.5	0.5	$\beta^\star$	$\beta^\star$	$0.95 \times \beta^\star$
$\beta_{1\|11,1}$	1	1	1	1	0.01	1
$\beta_{2\|11,1}$	1	1	1	1	0.01	1
$\beta_{12,1\|11,1}$	0.5	0.5	0.5	0.5	0.005	0.5
$\beta_{12,2\|11,1}$	0.5	0.5	0.5	0.5	0.005	0.5

TABLE B.1: Tested hyperparameter sets for the Beta prior distributions.

In **Figure B.2**, we can observe that the hyperparameter set (4) that is used in the analysis produces a better area under the curve (AUC) score than the sets 1-3.

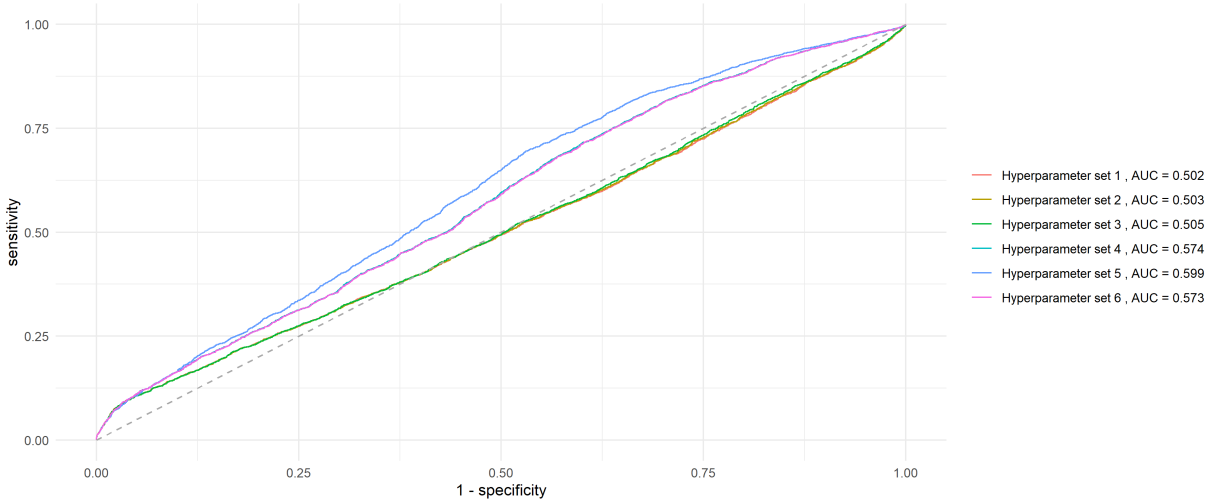


FIGURE B.2: ROC curves for the different hyperparameter sets (from Table B.1).

B.3 Recommendations for future research

The above-described framework could be extended in the context of multiple (i.e. three or more) drugs to detect signals of disproportionate reporting indicative of an interaction caused by any pairwise or multi-wise combination of the individual drugs under consideration. However, extending the framework to accommodate N drugs would require several considerations:

- *Defining the number of hypotheses:* Given the consideration of N drugs, a combinatorial enumeration approach is necessary to carefully define the number of hypotheses to be tested. This involves systematically considering all possible combinations of the individual drugs under investigation.
- *Automating the prior and likelihood construction:* Instead of hardcoding prior distributions and likelihoods for each drug combination, an automated approach seems necessary for more than two drugs, thus allowing for efficient and scalable analysis.
- *Defining or estimating prior hyperparameters:* With N drugs, there is a need to define or estimate a large number of prior hyperparameters associated with the multiple enumerated hypotheses. A strategy should be devised to determine these hyperparameters, potentially using data-driven techniques.
- *Obtaining contingency tables and database counts:* In the context of N drugs, a method needs to be devised to obtain contingency tables and associated FAERS (or other) database counts specifically tailored to the multiple drug combinations being investigated.

Bibliography

[1] Michael S. Kinch, Austin Haynesworth, Sarah L. Kinch, and Denton Hoyer. An overview of FDA-approved new molecular entities: 1827-2013. *Drug Discovery Today*, 19(8):1033–1039, 2014.

[2] Philippe Grandjean. Paracelsus Revisited: The Dose Concept in a Complex World. *Basic and Clinical Pharmacology and Toxicology*, 119(2):126–132, 8 2016.

[3] World Health Organization. International drug monitoring: the role of national centres, report of a who meeting [held in geneva from 20 to 25 september 1971].

[4] Duxin Sun, Wei Gao, Hongxiang Hu, and Simon Zhou. Why 90% of clinical drug development fails and how to improve it? *Acta Pharmaceutica Sinica B*, 12(7):3049–3062, 7 2022.

[5] Munir Pirmohamed, Sally James, Shaun Meakin, and Chris Green. Adverse drug reactions as cause of admission to hospital. *BMJ : British Medical Journal*, 329:460, 2004.

[6] Rostam Osanlou, Lauren Walker, Dyfrig A Hughes, Girvan Burnside, and Munir Pirmohamed. Adverse drug reactions, multimorbidity and polypharmacy: a prospective analysis of 1 month of medical admissions. *BMJ Open*, 12(7):55551, 2022.

[7] Patrick B Ryan, Martijn J. Schuemie, Emily Welebob, Jon Duke, Sarah Valentine, and Abraham G Hartzema. Defining a reference set to support methodological research in drug safety. *Drug Safety*, 36, 2013.

[8] Rave Harpaz, William DuMouchel, Paea LePendu, Anna Bauer-Mehren, Patrick Ryan, and Nigam H. Shah. Performance of pharmacovigilance signal detection algorithms for the fda adverse event reporting system. *Clinical Pharmacology Therapeutics*, 93:539–46, 2013.

[9] Preciosa M. Coloma, Paul Avillach, Francesco Salvo, Martijn J. Schuemie, Carmen Ferrajolo, Antoine Pariente, Annie Fourrier-Réglat, Mariam Molokhia, Vaishali Patadia, Johan Van Der Lei, Miriam Sturkenboom, and Gianluca Trifirò. A reference standard for evaluation of methods for drug safety signal detection using electronic healthcare record databases. *Drug Safety*, 36:13–23, 2013.

[10] Rave Harpaz, David Odgers, Greg Gaskin, William DuMouchel, Rainer Winnenburg, Olivier Bodenreider, Anna Ripple, Ana Szarfman, Alfred Sorbello, Eric Horvitz, Ryen W White, and Nigam H. Shah. A time-indexed reference standard of adverse drug reactions. *Scientific Data*, 1:140043, 2014.

[11] Bharat T Thakrar, Sabine Borel Grundschober, and Lucette Doessegger. Detecting signals of drug-drug interactions in a spontaneous reports database. *British journal of clinical pharmacology*, 64:489–95, 10 2007.

[12] G. Niklas Norén, Rolf Sundberg, Andrew Bate, and I. Ralph Edwards. A statistical methodology for drug–drug interaction surveillance. *Statistics in Medicine*, 27:3057–3070, 2008.

[13] June S. Almenoff, William DuMouchel, L. Allen Kindman, Xionghu Yang, and David Fram. Disproportionality analysis using empirical bayes data mining: a tool for the evaluation of drug interactions in the post-marketing setting. *Pharmacoepidemiology and Drug Safety*, 12:517–521, 9 2003.

[14] Rave Harpaz, Herbert S Chase, and Carol Friedman. Mining multi-item drug adverse effect associations in spontaneous reporting systems. *BMC Bioinformatics*, 11:S7, 10 2010.

[15] T. Lorberbaum, M. Nasir, M. J. Keiser, S. Vilar, G. Hripcsak, and N. P. Tatonetti. Systems pharmacology augments drug safety surveillance. *Clinical Pharmacology and Therapeutics*, 97:151–158, 2 2015.

[16] Marinka Zitnik, Monica Agrawal, and Jure Leskovec. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*, 34:i457–i466, 2018.

[17] David Ochoa, Andrew Hercules, Miguel Carmona, Daniel Suveges, Asier Gonzalez-Uriarte, Cinzia Malangone, Alfredo Miranda, Luca Fumis, Denise Carvalho-Silva, Michaela Spitzer, Jarrod Baker, Javier Ferrer, Arwa Raies, Olesya Razuvayevskaya, Adam Faulconbridge, Eirini Petsalaki, Prudence Mutowo, Sandra MacHlitt-Northen, Gareth Peat, Elaine McAuley, Chuang Kee Ong, Edward Mountjoy, Maya Ghoussaini, Andrea Pierleoni, Eliseo Papa, Miguel Pignatelli, Gautier Koscielny, Mohd Karim, Jeremy Schwartzentruber, David G Hulcoop, Ian Dunham, and Ellen M. McDonagh. Open targets platform: Supporting systematic drug-target identification and prioritisation. *Nucleic Acids Research*, 49:D1302–D1310, 2021.

[18] Alexander Sebastian Hauser, Albert Jelke Kooistra, Eva Sverrisdóttir, and Maurizio Sessa. Utilizing drug-target-event relationships to unveil safety patterns in pharmacovigilance. *Expert Opinion on Drug Safety*, 19:961–968, 2020.

[19] Peter Schotland, Rebecca Racz, David B. Jackson, Theodoros G. Soldatos, Robert Levin, David G. Strauss, and Keith Burkhart. Target adverse event profiles for predictive safety in the postmarket setting. *Clinical Pharmacology and Therapeutics*, 109:1232–1243, 2021.

[20] Elpida Kontsioti, Simon Maskell, Amina Bensalem, Bhaskar Dutta, and Munir Pirmohamed. Similarity and consistency assessment of three major online drug–drug interaction resources. *British Journal of Clinical Pharmacology*, 88:4067–4079, 9 2022.

[21] E. Kontsioti, S. Maskell, B. Dutta, and M. Pirmohamed. A reference set of clinically relevant adverse drug-drug interactions. *Scientific Data*, 9, 2022.

[22] William McBride. Thalidomide and congenital abnormalities. *The Lancet*, 278(7216):1358, 1961.

[23] Patricia M. Reed, Stuart J. Mair, and Stephen Freestone. *Non-Clinical Safety Evaluation and Adverse Events in Phase I Trials*, pages 75–84. 2nd edition, 2006.

[24] J. M. Luteijn, B. C. White, H. Gunnlaugsdóttir, F. Holm, N. Kalogeras, O. Leino, S. H. Magnússon, G. Odekerken, M. V. Pohjola, M. J. Tijhuis, J. T. Tuomisto, Ueland, P. A. McCarron, and H. Verhagen. State of the art in benefit-risk analysis: Medicines. *Food and Chemical Toxicology*, 50(1):26–32, 1 2012.

[25] C S Aaron, Pr R Harbach, S S Mattano, J K Mayo, Y Wang, R L Yu, and D M Zimmer. Risk and Benefit Evaluation in Development of Pharmaceutical Products. *Environmental Health Perspectives Supplements*, 101:291–295, 1993.

[26] I. Ralph Edwards, Bengt Erik Wiholm, and Carlos Martinez. Concepts in risk-benefit assessment. A simple merit analysis of a medicine? *Drug Safety*, 15(1):1–7, 10 1996.

[27] Council for International Organizations of Medical Sciences. *Benefit-risk balance for marketed drugs : evaluating safety signals : report of CIOMS Working Group IV*. CIOMS, 1998.

[28] Juhaeri Juhaeri. Benefit–risk evaluation: The past, present and future. *Therapeutic Advances in Drug Safety*, 10, 2019.

[29] James Leong, Sam Salek, and Stuart Walker. *Benefit-risk assessment of medicines: The development and application of a universal framework for decision-making and effective communication*. 2015.

[30] MHRA. MHRA Corporate Plan 2013-2018. Technical report, 2013.

[31] Ola Caster. Benefit-risk assessment in pharmacovigilance. In *Methods in Pharmacology and Toxicology*, pages 233–257. Humana Press Inc., 2018.

[32] I. Ralph Edwards. Spontaneous reporting—of what? Clinical concerns about drugs. *British Journal of Clinical Pharmacology*, 48(2):138, 1999.

[33] A. J. Avery, C. Anderson, C. M. Bond, H. Fortnum, A. Gifford, P. C. Hannaford, L. Hazell, J. Krska, A. J. Lee, D. J. McLernon, E. Murphy, S. Shakir, and M. C. Watson. Evaluation of patient reporting of adverse drug reactions to the UK 'Yellow card scheme': Literature review, descriptive and qualitative analyses, and questionnaire surveys. *Health Technology Assessment*, 15(20):1–234, 2011.

[34] Gale G. White and Lori Love. The MedWatch program. *Clinical Toxicology*, 36(7):769–770, 1998.

[35] Diogo Rafael Amorim Mendes. *Data Mining for Studying Drug Interactions and Adverse Effects Prediction*. PhD thesis, University of Porto, 2023.

[36] Marie Lindquist. VigiBase, the WHO Global ICSR Database System: Basic facts. *Drug Information Journal*, 42(5):409–419, 12 2008.

[37] Florence van Hunsel, Linda Härmark, and Leàn Rolfes. Fifteen years of patient reporting –what have we learned and where are we heading to? *https://doi.org/10.1080/14740338.2019.1613373*, 18(6):477–484, 6 2019.

[38] Kruger Paola and Gasperini Claudio. The value of direct patient reporting in pharmacovigilance, 2020.

[39] G Niklas Norén, Roland Orre, Andrew Bate, and I Ralph Edwards. Duplicate detection in adverse drug reaction surveillance. *Data Mining and Knowledge Discovery*, 14(3):305–328, 2007.

[40] Manfred Hauben, Lester Reich, James DeMicco, and Katherine Kim. 'Extreme duplication' in the US FDA adverse events reporting system database. *Drug Safety*, 30(6):551–554, 2007.

[41] Kory Kreimeyer, David Menschik, Scott Winiecki, Wendy Paul, Faith Barash, Emily Jane Woo, Meghna Alimchandani, Deepa Arya, Craig Zinderman, Richard Forshee, and Taxiarchis Botsis. Using Probabilistic Record Linkage of Structured and Unstructured Data to Identify Duplicate Cases in Spontaneous Adverse Event Reporting Systems. *Drug Safety*, 40(7):571–582, 2017.

[42] Keith B. Hoffman, Andrea R. Demakas, Mo Dimbil, Nicholas P. Tatonetti, and Colin B. Erdman. Stimulated Reporting: The Impact of US Food and Drug Administration-Issued Alerts on the Adverse Event Reporting System (FAERS). *Drug Safety*, 37(11):971–980, 10 2014.

[43] Lorna Hazell and Saad A W Shakir. Under-Reporting of Adverse Drug Reactions A Systematic Review. *Drug Safety*, 29(5):385–396, 2006.

[44] Martin R. Cowie, Juuso I. Blomster, Lesley H. Curtis, Sylvie Duclaux, Ian Ford, Fleur Fritz, Samantha Goldman, Salim Janmohamed, Jörg Kreuzer, Mark Leenay, Alexander Michel, Seleen Ong, Jill P. Pell, Mary Ross Southworth, Wendy Gattis Stough, Martin Thoenes, Faiez Zannad, and Andrew Zalewski. Electronic health records to facilitate clinical research. *Clinical Research in Cardiology*, 106(1):1–9, 2017.

[45] Keith Marsolo and S. Andrew Spooner. Clinical genomics in the world of the electronic health record. *Genetics in Medicine*, 15(10):786–791, 10 2013.

[46] Fanny Leroy, Jean Yves Dauxois, Hélène Théophile, Françoise Haramburu, and Pascale Tubert-Bitter. Estimating time-to-onset of adverse drug reactions from spontaneous reporting databases. *BMC Medical Research Methodology*, 14(1):1–11, 2 2014.

[47] Kevin Haynes. Electronic Health Record, Transactional Insurance Claims, and Distributed Databases. In *Evidence-Based Pharmacovigilance*, pages 135–148. Humana Press Inc., New York, NY, 2018.

[48] Rave Harpaz, Alison Callahan, Suzanne Tamang, Yen Low, David Odgers, Sam Finlayson, Kenneth Jung, Paea LePendu, and Nigam H. Shah. Text mining for adverse drug events: the promise, challenges, and state of the art. *Drug Safety*, 37:777–790, 10 2014.

[49] R. Ball, M. Robb, S. A. Anderson, and G. Dal Pan. The FDAs sentinel initiative A comprehensive approach to medical product surveillance, 3 2016.

[50] Ryan M. Carnahan, Carlos J. Bell, and Richard Platt. Active Surveillance: The United States Food and Drug Administration's Sentinel Initiative. *Mann's Pharmacovigilance: Third Edition*, pages 429–437, 2014.

[51] Rebecca E. Ghosh, Elizabeth Crellin, Sue Beatty, Katherine Donegan, Puja Myles, and Rachael Williams. How Clinical Practice Research Datalink data are used to support pharmacovigilance. *Therapeutic Advances in Drug Safety*, 10:204209861985401, 1 2019.

[52] Jenna Marie Reps, Jonathan M. Garibaldi, Uwe Aickelin, Daniele Soria, Jack Gibson, and Richard Hubbard. Comparison of algorithms that detect drug side effects using electronic healthcare databases. *Soft Computing*, 17(12):2381–2397, 12 2013.

[53] Nicholas P. Tatonetti, Guy Haskin Fernald, and Russ B. Altman. A novel signal detection algorithm for identifying hidden drug-drug interactions in adverse event reports. *Journal of the American Medical Informatics Association*, 19:79–85, 2012.

[54] Alan R Aronson. Effective Mapping of Biomedical Text to the UMLS Metathesaurus: The MetaMap Program. In *AMIA Annual Symposium Proceedings*, pages 17–21, 2001.

[55] Alan R Aronson and François Michel Lang. An overview of MetaMap: Historical perspective and recent advances. *Journal of the American Medical Informatics Association*, 17(3):229–236, 2010.

[56] Thomas C. Rindflesch and Marcelo Fiszman. The interaction of domain knowledge and linguistic structure in natural language processing: Interpreting hypernymic propositions in biomedical text. *Journal of Biomedical Informatics*, 36(6):462–477, 2003.

[57] Halil Kilicoglu, Graciela Rosemblat, Marcelo Fiszman, and Dongwook Shin. Broad-coverage biomedical relation extraction with SemRep. *BMC Bioinformatics*, 21(1):1–28, 2020.

[58] Paul Avillach, Jean Charles Dufour, Gayo Diallo, Francesco Salvo, Michel Joubert, Frantz Thiessard, Fleur Mougin, Gianluca Trifirò, Annie Fourrier-Réglat, Antoine Pariente, and Marius Fieschi. Design and validation of an automated method to detect known adverse drug reactions in MEDLINE: A contribution from the EU-ADR project. *Journal of the American Medical Informatics Association*, 20(3):446–452, 2013.

[59] Rong Xu and Quanqiu Wang. Automatic construction of a large-scale and accurate drug-side-effect association knowledge base from biomedical literature HHS Public Access Keywords text mining; drug side effect; drug discovery; drug repositioning; drug toxicity prediction. *J Biomed Inform*, 51:191–199, 2014.

[60] Ahmad P Tafti, Jonathan Badger, Eric LaRose, Ehsan Shirzadi, Andrea Mahnke, John Mayer, Zhan Ye, David Page, and Peggy Peissig. Adverse drug event discovery using biomedical literature: A big data neural network adventure. *JMIR Medical Informatics*, 5(4), 2017.

[61] Trung Huynh, Yulan He, Alistair Willis, and Stefan Uger. Adverse drug reaction classification with deep neural networks. *Coling*, pages 877–887, 2016.

[62] Harsha Gurulingappa, Abdul Mateen Rajput, Angus Roberts, Juliane Fluck, Martin Hofmann-Apitius, and Luca Toldo. Development of a benchmark corpus to support the automatic extraction of drug-related adverse effects from medical case reports. *Journal of Biomedical Informatics*, 45(5):885–892, 10 2012.

[63] Yoon Kim. Convolutional neural networks for sentence classification. In *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, pages 1746–1751, 2014.

[64] Diego Saldana Miranda. Automated detection of adverse drug reactions in the biomedical literature using convolutional neural networks and biomedical word Embeddings. In *CEUR Workshop Proceedings*, volume 2226, pages 33–41, 2018.

[65] Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. Biobert: A pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36:1234–1240, 2020.

[66] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 2020-December, 5 2020.

[67] Renqian Luo, Liai Sun, Yingce Xia, Tao Qin, Sheng Zhang, Hoifung Poon, and Tie Yan Liu. BioGPT: generative pre-trained transformer for biomedical text generation and mining. *Briefings in bioinformatics*, 23(6):1–12, 2022.

[68] Rainer Winnenburg, Alfred Sorbello, Anna Ripple, Rave Harpaz, Joseph Tonning, Ana Szarfman, Henry Francis, and Olivier Bodenreider. Leveraging MEDLINE indexing for pharmacovigilance - Inherent limitations and mitigation strategies. *Journal of Biomedical Informatics*, 57:425–435, 10 2015.

[69] Danushka Bollegala, Simon Maskell, Richard Sloane, Joanna Hajne, and Munir Pirmohamed. Causality patterns for detecting adverse drug reactions from social media: Text mining approach. *JMIR Public Health and Surveillance*, 4(5), 5 2018.

[70] Richard Sloane, Orod Osanlou, David Lewis, Danushka Bollegala, Simon Maskell, and Munir Pirmohamed. Social media and pharmacovigilance: A review of the opportunities and challenges. *British Journal of Clinical Pharmacology*, 80(4):910–920, 10 2015.

[71] Robert Leaman, Laura Wojtulewicz, Ryan Sullivan, Annie Skariah, Jian Yang, and Graciela Gonzalez. Towards Internet-Age Pharmacovigilance: Extracting Adverse Drug Reactions from User Posts to Health-Related Social Networks. Technical report, 2010.

[72] Rajesh Ghosh and David Lewis. Aims and approaches of Web-RADR: A consortium ensuring reliable ADR reporting via mobile devices and new insights from social media. *Expert Opinion on Drug Safety*, 14(12):1845–1853, 2015.

[73] John van Stekelenborg, Johan Ellenius, Simon Maskell, Tomas Bergvall, Ola Caster, Nabarun Dasgupta, Juergen Dietrich, Sara Gama, David Lewis, Victoria Newbould, Sabine Brosch, Carrie E. Pierce, Gregory Powell, Alicia Ptaszyńska-Neophytou, Antoni F.Z. Wiśniewski, Phil Tregunno, G. Niklas Norén, and Munir Pirmohamed. Recommendations for the Use of Social Media in Pharmacovigilance: Lessons from IMI WEB-RADR. *Drug Safety*, 42(12):1393–1407, 12 2019.

[74] Adam Lavertu, Bianca Vora, Kathleen M. Giacomini, Russ Altman, and Stefano Rensi. A New Era in Pharmacovigilance: Toward Real-World Data and Digital Monitoring, 5 2021.

[75] Juergen Dietrich, Lucie M. Gattepaille, Britta Anne Grum, Letitia Jiri, Magnus Lerch, Daniele Sartori, and Antoni Wisniewski. Adverse Events in Twitter-Development of a Benchmark Reference Dataset: Results from IMI WEB-RADR. *Drug Safety*, 43(5):467–478, 5 2020.

[76] Alasdair M. Breckenridge, Ross A. Breckenridge, and Carl C. Peck. Report on the current status of the use of real-world data (rwd) and real-world evidence (rwe) in drug development and regulation. *British Journal of Clinical Pharmacology*, 85:1874–1877, 2019.

[77] January Weiner, David J.M. Lewis, Jeroen Maertzdorf, Hans Joachim Mollenkopf, Caroline Bodinham, Kat Pizzoferro, Catherine Linley, Aldona Greenwood, Alberto Mantovani, Barbara Bottazzi, Philippe Denoel, Geert Leroux-Roels, Kent E. Kester, Ingileif Jonsdottir, Robert van den Berg, Stefan H.E. Kaufmann, and Giuseppe Del Giudice. Characterization of potential biomarkers of reactogenicity of licensed antiviral vaccines: randomized controlled clinical trials conducted by the biovacsafe consortium. *Scientific Reports*, 9, 12 2019.

[78] Luis Garcia-Gancedo and Andrew Bate. Digital biomarkers for post-licensure safety monitoring. *Drug Discovery Today*, 27(11):103354, 2022.

[79] Rave Harpaz, William DuMouchel, Martijn Schuemie, Olivier Bodenreider, Carol Friedman, Eric Horvitz, Anna Ripple, Alfred Sorbello, Ryen W. White, Rainer Winnenburg, and Nigam H. Shah. Toward multimodal signal detection of adverse drug reactions. *Journal of Biomedical Informatics*, 76:41–49, 2017.

[80] Rolf Gedeborg, Wilmar Igl, Bodil Svennblad, Peter Wilén, Bénédicte Delcoigne, Karl Michaëlsson, Rickard Ljung, and Nils Feltelius. Federated analyses of multiple data sources in drug safety studies. *Pharmacoepidemiology and Drug Safety*, 3 2022.

[81] Amir Karami, Michael Zhu, Bailey Goldschmidt, Hannah R Boyajieff, and Mahdi M Najafabadi. Covid-19 vaccine and social media in the u.S.: Exploring emotions and discussions on twitter. *Vaccines*, 9(10), 2021.

[82] Steven Lloyd Wilson and Charles Wiysonge. Social media and vaccine hesitancy. *BMJ Global Health*, 5(10):4206, 2020.

[83] Rave Harpaz, William DuMouchel, Robbert Van Manen, Alexander Nip, Steve Bright, Ana Szarfman, Joseph Tonning, and Magnus Lerch. Signaling COVID-19 Vaccine Adverse Events. *Drug Safety*, 45(7):765–780, 7 2022.

[84] Marion Mueller, David J. Lewis, and Amalia Alexe. The evolution of pharmacovigilance ecosystems: Does Moore's law invite the use of Occam's razor? *British Journal of Clinical Pharmacology*, 89(2):470–482, 2 2023.

[85] WHO. Safety of Medicines. In *World Health Organization Geneva*, volume 2002.2, page 20. 2002.

[86] Council for International Organizations of Medical Sciences. *Practical Aspects of Signal Detection in Pharmacovigilance: Report of CIOMS Working Group VIII*. CIOMS, 2010.

[87] A. C.G. Egberts, R. H.B. Meyboom, and Eugène P Van Puijenbroek. Use of measures of disproportionality in pharmacovigilance: Three Dutch examples. *Drug Safety*, 25(6):453–458, 2002.

[88] Andrew Bate, Antoine Pariente, Manfred Hauben, and Bernard Bégaud. *Quantitative Signal Detection and Analysis in Pharmacovigilance*, pages 331–354. 2014.

[89] S. J. W. Evans, P. C. Waller, and S. Davis. Use of proportional reporting ratios (prrs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiology and Drug Safety*, 10:483–486, 10 2001.

[90] Kenneth J. Rothman, Stephan Lanes, and Susan T. Sacks. The reporting odds ratio and its advantages over the proportional reporting ratio. *Pharmacoepidemiology and Drug Safety*, 13(8):519–523, 8 2004.

[91] A Bate, M Lindquist, I R Edwards, S Olsson, R Orre, A Lansner, and R M De Freitas. A bayesian neural network method for adverse drug reaction signal generation. *European journal of clinical pharmacology*, 54:315–21, 6 1998.

[92] William DuMouchel. Bayesian data mining in large frequency tables, with an application to the fda spontaneous reporting system. *The American Statistician*, 53:177–190, 8 1999.

[93] Ana Szarfman, Stella G Machado, and Robert T O'neill. 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. Technical Report 6, 2002.

[94] J. Martin Bland and Douglas G. Altman. Bayesians and frequentists. *BMJ*, 317(7166):1151, 10 1998.

[95] Christopher Gravel. Statistical methods for signal detection in pharmacovigilance, 2009.

[96] Gianmario Candore, Kristina Juhlin, Katrin Manlik, Bharat Thakrar, Naashika Quarcoo, Suzie Seabroke, Antoni Wisniewski, and Jim Slattery. Comparison of statistical signal detection methods within and across spontaneous reporting databases. *Drug Safety*, 38:577–587, 2015.

[97] G. Niklas Norén, Ola Caster, Kristina Juhlin, and Marie Lindquist. Zoo or savannah? choice of training ground for evidence-based pharmacovigilance. *Drug Safety*, 37:655–659, 2014.

[98] Rave Harpaz, William DuMouchel, and Nigam H. Shah. Comment on: "zoo or savannah? choice of training ground for evidence-based pharmacovigilance". *Drug Safety*, 38:113–114, 2015.

[99] G. Niklas Norén, Ola Caster, Kristina Juhlin, and Marie Lindquist. Authors' reply to harpaz et al. comment on: "zoo or savannah? choice of training ground for evidence-based pharmacovigilance". *Drug Safety*, 38:115–116, 2015.

[100] A. Lawrence Gould. Practical pharmacovigilance analysis strategies. *Pharmacoepidemiology and Drug Safety*, 12:559–574, 2003.

[101] Francois Maignen, Manfred Hauben, Eric Hung, Lionel Van Holle, and Jean Michel Dogne. A conceptual approach to the masking effect of measures of disproportionality. *Pharmacoepidemiology and Drug Safety*, 23:208–217, 2014.

[102] Andrew Bate. The use of a bayesian confidence propagation neural network in pharmacovigilance andrew bate umeå 2003. 2003.

[103] Keith B Hoffman, Mo Dimbil, Colin B Erdman, Nicholas P Tatonetti, and Brian M Overstreet. The weber effect and the united states food and drug administration's adverse event reporting system (faers): Analysis of sixty-two drugs. 2014.

[104] Keith B Hoffman, Andrea R Demakas, Mo Dimbil, Nicholas P Tatonetti, and Colin B Erdman. Stimulated reporting: The impact of us food and drug administration-issued alerts on the adverse event reporting system (faers). *Drug Safety*, 37:971–980, 2014.

[105] Maribel Salas, Albert Hofman, and Bruno H.Ch Stricker. Confounding by indication: An example of variation in the use of epidemiologic terminology. *American Journal of Epidemiology*, 149:981–983, 1999.

[106] Austin Bradford Hill. The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine*, 58:295–300, 1965.

[107] Ronald HB Meyboom, Antoine C.G. Egberts, I. Ralph Edwards, Yechiel A Hekster, Fred H.P. De Koning, and Frank W.J. Gribnau. Principles of signal detection in pharmacovigilance. *Drug Safety*, 16(6):355–365, 1997.

[108] Judith K Jones and Elyse Kingery. How We Assess Causality. In *Mann's Pharmacovigilance: Third Edition*, pages 319–330. 2014.

[109] Uppsala Monitoring Centre. The use of the WHO-UMC system for standardised case causality assessment. Technical report.

[110] Wayne M. Turner. The food and drug administration algorithm: Special workshop—Regulatory. *Drug Information Journal*, 18(3-4):259–266, 1984.

[111] C. A. Naranjo, U. Busto, E. M. Sellers, P. Sandor, I. Ruiz, E. A. Roberts, E. Janacek, C. Domecq, and D. J. Greenblatt. A method for estimating the probability of adverse drug reactions. *Clinical Pharmacology and Therapeutics*, 30(2):239–245, 1981.

[112] Ruairi M Gallagher, Jamie J Kirkham, Jennifer R Mason, Kim A Bird, Paula R Williamson, Anthony J. Nunn, Mark A. Turner, Rosalind L. Smyth, and Munir Pirmohamed. Development and inter-rater reliability of the Liverpool adverse drug reaction causality assessment tool. *PLoS ONE*, 6(12):28096, 2011.

[113] Council for International Organizations of Medical Sciences. *Practical Aspects of Signal Detection in Pharmacovigilance: Report of CIOMS Working Group VIII*. 2010.

[114] Lesley Wise. Risks and benefits of (pharmaco)epidemiology, 2011.

[115] Lisa E. Hines, Daniel C. Malone, and John E. Murphy. Recommendations for generating, evaluating, and implementing drug-drug interaction evidence. In *Pharmacotherapy*, volume 32, pages 304–313, 4 2012.

[116] Keith Gallicano and George Drusano. *Introduction to Drug Interactions*, pages 1–11. 2005.

[117] Munir Pirmohamed. Drug-drug interactions and adverse drug reactions: separating the wheat from the chaff. *Wiener klinische Wochenschrift*, 122:62–64, 2010.

[118] Ingolf Cascorbi. Drug interactions—principles, examples and clinical consequences. *Deutsches Arzteblatt International*, 109:546–556, 2012.

[119] Thorir D. Bjornsson, John T. Callaghan, Heidi J. Einolf, Volker Fischer, Lawrence Gan, Scott Grimm, John Kao, S. Peter King, Gerald Miwa, Lan Ni, Gondi Kumar, James McLeod, R. Scott Obach, Stanley Roberts, Amy Roe, Anita Shah, Fred Snikeris, John T. Sullivan, Donald Tweedie, Jose M. Vega, John Walsh, and Steven A. Wrighton. The conduct of in vitro and in vivo drug-drug interaction

studies: A pharmaceutical research and manufacturers of america (phrma) perspective. *Drug Metabolism and Disposition*, 31:815–832, 2003.

[120] Anne M McDonnell and Cathyyen H Dang. Basic review of the cytochrome p450 system. *Journal of the advanced practitioner in oncology*, 4:263–8, 7 2013.

[121] Munir Pirmohamed. Pharmacogenetics: Past, present and future, 10 2011.

[122] M. Wadelius and M. Pirmohamed. Pharmacogenetics of warfarin: Current status and future challenges. *Pharmacogenomics Journal*, 7:99–111, 4 2007.

[123] Fabian Müller and Martin F. Fromm. Transporter-mediated drug-drug interactions, 7 2011.

[124] Department of Health and Social Care. Good for you, good for us, good for everybody: A plan to reduce overprescribing to make patient care better and safer, support the nhs, and reduce carbon emissions, 2021.

[125] Nina Fokter, Martin Možina, and Miran Brvar. Potential drug-drug interactions and admissions due to drug-drug interactions in patients treated in medical departments. *Wiener Klinische Wochenschrift*, 122:81–88, 2010.

[126] Cara Tannenbaum and Nancy L Sheehan. Understanding and preventing drug-drug and drug-gene interactions. *Expert Review of Clinical Pharmacology*, 7:533–544, 2014.

[127] Hege S Blix, Kirsten K Viktil, Tron A Moger, and Aasmund Reikvam. Drugs with narrow therapeutic index as indicators in the risk management of hospitalised patients. *Pharmacy practice*, 8:50–5, 1 2010.

[128] Munir Pirmohamed. Pharmacogenomics: current status and future perspectives. *Nature Reviews Genetics*, 63 Suppl 1:50–56, 2023.

[129] Lei Zhang, Yuanchao Zhang, Ping Zhao, and Shiew Mei Huang. Predicting drug-drug interactions: An fda perspective. *AAPS Journal*, 11:300–306, 2009.

[130] Hongbin Yang, Lixia Sun, Weihua Li, Guixia Liu, and Yun Tang. In silico prediction of chemical toxicity for drug design using machine learning methods and structural alerts. *Frontiers in Chemistry*, 6, 2 2018.

[131] Siqi Chen, Tiancheng Li, Luna Yang, Fei Zhai, Xiwei Jiang, Rongwu Xiang, and Guixia Ling. Artificial intelligence-driven prediction of multiple drug interactions. *Briefings in bioinformatics*, 23(6), 11 2022.

[132] Manfred Hauben. Artificial intelligence and data mining for the pharmacovigilance of drug–drug interactions. *Clinical Therapeutics*, 45:117–133, 2 2023.

[133] Yoshihiro Noguchi, Tomoya Tachi, and Hitomi Teramachi. Review of statistical methodologies for detecting drug–drug interactions using spontaneous reporting systems. *Frontiers in Pharmacology*, 10, 11 2019.

[134] Eugene F. Van Puijenbroek, Antoine C.G. Egberts, Eibert R. Heerdink, and Huhert G.M. Leufkens. Detecting drug-drug interactions using a database for spontaneous adverse drug reactions: An example with diuretics and non-steroidal anti-inflammatory drugs. *European Journal of Clinical Pharmacology*, 56:733–738, 2000.

[135] Eugène P. Van Puijenbroek, Antoine C. G. Egberts, Ronald H. B. Meyboom, and Hubert G. M. Leufkens. Signalling possible drug-drug interactions in a spontaneous reporting system: delay of withdrawal bleeding during concomitant use of oral contraceptives and itraconazole. *British Journal of Clinical Pharmacology*, 47:689–693, 12 1999.

[136] Assaf Gottlieb, Gideon Y Stein, Yoram Oron, Eytan Ruppin, and Roded Sharan. Indi: A computational framework for inferring drug interactions and their associated recommendations. *Molecular Systems Biology*, 8:592, 1 2012.

[137] Santiago Vilar, Carol Friedman, and George Hripcsak. Detection of drug-drug interactions through data mining studies using clinical sources, scientific literature and social media, 2018.

[138] Isabel Segura-Bedmar, Paloma Martínez Fernández, and María Herrero Zazo. Semeval-2013 task 9 : Extraction of drug-drug interactions from biomedical texts. 2:341–350, 2013.

[139] María Herrero-Zazo, Isabel Segura-Bedmar, Paloma Martínez, and Thierry Declerck. The ddi corpus: An annotated corpus with pharmacological substances and drug-drug interactions. *Journal of Biomedical Informatics*, 46:914–920, 2013.

[140] Sunil Kumar Sahu and Ashish Anand. Drug-drug interaction extraction from biomedical texts using long short-term memory network. *Journal of Biomedical Informatics*, 86:15–24, 10 2018.

[141] Quoc Chinh Bui, Peter M.A. Sloot, Erik M. Van Mulligen, and Jan A. Kors. A novel feature-based approach to extract drug–drug interactions from biomedical text. *Bioinformatics*, 30:3365–3371, 12 2014.

[142] Zhehuan Zhao, Zhihao Yang, Ling Luo, Hongfei Lin, and Jian Wang. Drug drug interaction extraction from biomedical literature using syntax convolutional neural network. *Bioinformatics*, 32:3444–3453, 2016.

[143] Shengyu Liu, Buzhou Tang, Qingcai Chen, and Xiaolong Wang. Drug-drug interaction extraction via convolutional neural networks. *Computational and Mathematical Methods in Medicine*, 2016, 2016.

[144] Sangrak Lim, Kyubum Lee, and Jaewoo Kang. Drug drug interaction extraction from the literature using a recursive neural network. *PLoS ONE*, 13, 2018.

[145] Chanhee Park, Jinuk Park, and Sanghyun Park. Agcn: Attention-based graph convolutional networks for drug-drug interaction extraction. *Expert Systems with Applications*, 159:113538, 11 2020.

[146] Masaki Asada, Makoto Miwa, and Yutaka Sasaki. Integrating heterogeneous knowledge graphs into drug–drug interaction extraction from the literature. *Bioinformatics*, 39, 1 2023.

[147] Srinivasan V. Iyer, Rave Harpaz, Paea LePendu, Anna Bauer-Mehren, and Nigam H. Shah. Mining clinical text for signals of adverse drug-drug interactions. *Journal of the American Medical Informatics Association*, 21:353–362, 2014.

[148] Bethany Percha, Yael Garten, and Russ B. Altman. Discovery and explanation of drug-drug interactions via text mining. *Pacific Symposium on Biocomputing*, pages 410–421, 2012.

[149] Su Yan, Xiaoqian Jiang, and Ying Chen. Text mining driven drug-drug interaction detection. *Proceedings - 2013 IEEE International Conference on Bioinformatics and Biomedicine, IEEE BIBM 2013*, pages 349–354, 2013.

[150] Haodong Yang and Christopher C. Yang. Mining a weighted heterogeneous network extracted from healthcare-specific social media for identifying interactions between drugs. *Proceedings - 15th IEEE International Conference on Data Mining Workshop, ICDMW 2015*, pages 196–203, 1 2016.

[151] Ahmed Abdeen Hamed, Xindong Wu, Robert Erickson, and Tamer Fandy. Twitter k-h networks in action: Advancing biomedical literature for drug search. *Journal of Biomedical Informatics*, 56:157–168, 8 2015.

[152] Sabin S Egger, Jürgen Drewe, and Raymond G Schlienger. Potential drug–drug interactions in the medication of medical patients at hospital discharge. *European Journal of Clinical Pharmacology*, 58:773–778, 3 2003.

[153] Lara Magro, Ugo Moretti, and Roberto Leone. Epidemiology and characteristics of adverse drug reactions caused by drug-drug interactions. *Expert Opinion on Drug Safety*, 11:83–94, 2012.

[154] Alberto Romagnoni, Simon Jégou, Kristel Van Steen, Gilles Wainrib, Jean Pierre Hugot, Laurent Peyrin-Biroulet, Mathias Chamaillard, Jean Frederick Colombel, Mario Cottone, Mauro D'Amato, Renata D'Incà, Jonas Halfvarson, Paul Henderson, Amir Karban, Nicholas A. Kennedy, Mohammed Azam Khan, Marc Lémann, Arie Levine, Dunecan Massey, Monica Milla, Sok Meng Evelyn Ng, Ioannis Oikonomou, Harald Peeters, Deborah D. Proctor, Jean Francois Rahier, Paul

Rutgeerts, Frank Seibold, Laura Stronati, Kirstin M. Taylor, Leif Törkvist, Kullak Ublick, Johan Van Limbergen, Andre Van Gossum, Morten H. Vatn, Hu Zhang, Wei Zhang, Jane M. Andrews, Peter A. Bampton, Murray Barclay, Timothy H. Florin, Richard Gearry, Krupa Krishnaprasad, Ian C. Lawrance, Gillian Mahy, Grant W. Montgomery, Graham Radford-Smith, Rebecca L. Roberts, Lisa A. Simms, Katherine Hanigan, Anthony Croft, Leila Amininijad, Isabelle Cleynen, Olivier Dewit, Denis Franchimont, Michel Georges, Debby Laukens, Harald Peeters, Jean Francois Rahier, Paul Rutgeerts, Emilie Theatre, André Van Gossum, Severine Vermeire, Guy Aumais, Leonard Baidoo, Arthur M. Barrie, Karen Beck, Edmond Jean Bernard, David G. Binion, Alain Bitton, Steve R. Brant, Judy H. Cho, Albert Cohen, Kenneth Croitoru, Mark J. Daly, Lisa W. Datta, Colette Deslandres, Richard H. Duerr, Debra Dutridge, John Ferguson, Joann Fultz, Philippe Goyette, Gordon R. Greenberg, Talin Haritunians, Gilles Jobin, Seymour Katz, Raymond G. Lahaie, Dermot P. McGovern, Linda Nelson, Sok Meng Ng, Kaida Ning, Ioannis Oikonomou, Pierre Paré, Deborah D. Proctor, Miguel D. Regueiro, John D. Rioux, Elizabeth Ruggiero, L. Philip Schumm, Marc Schwartz, Regan Scott, Yashoda Sharma, Mark S. Silverberg, Denise Spears, A. Hillary Steinhart, Joanne M. Stempak, Jason M. Swoger, Constantina Tsagarelis, Wei Zhang, Clarence Zhang, Hongyu Zhao, Jan Aerts, Tariq Ahmad, Hazel Arbury, Anthony Attwood, Adam Auton, Stephen G. Ball, Anthony J. Balmforth, Chris Barnes, Jeffrey C. Barrett, Inês Barroso, Anne Barton, Amanda J. Bennett, Sanjeev Bhaskar, Katarzyna Blaszczyk, John Bowes, Oliver J. Brand, Peter S. Braund, Francesca Bredin, Gerome Breen, Morris J. Brown, Ian N. Bruce, Jaswinder Bull, Oliver S. Burren, John Burton, Jake Byrnes, Sian Caesar, Niall Cardin, Chris M. Clee, Alison J. Coffey, John MC Connell, Donald F. Conrad, Jason D. Cooper, Anna F. Dominiczak, Kate Downes, Hazel E. Drummond, Darshna Dudakia, Andrew Dunham, Bernadette Ebbs, Diana Eccles, Sarah Edkins, Cathryn Edwards, Anna Elliot, Paul Emery, David M. Evans, Gareth Evans, Steve Eyre, Anne Farmer, I. Nicol Ferrier, Edward Flynn, Alistair Forbes, Liz Forty, Jayne A. Franklyn, Timothy M. Frayling, Rachel M. Freathy, Eleni Giannoulatou, Polly Gibbs, Paul Gilbert, Katherine Gordon-Smith, Emma Gray, Elaine Green, Chris J. Groves, Detelina Grozeva, Rhian Gwilliam, Anita Hall, Naomi Hammond, Matt Hardy, Pile Harrison, Neelam Hassanali, Husam Hebaishi, Sarah Hines, Anne Hinks, Graham A. Hitman, Lynne Hocking, Chris Holmes, Eleanor Howard, Philip Howard, Joanna M.M. Howson, Debbie Hughes, Sarah Hunt, John D. Isaacs, Mahim Jain, Derek P. Jewell, Toby Johnson, Jennifer D. Jolley, Ian R. Jones, Lisa A. Jones, George Kirov, Cordelia F. Langford, Hana Lango-Allen, G. Mark Lathrop, James Lee, Kate L. Lee, Charlie Lees, Kevin Lewis, Cecilia M. Lindgren, Meeta Maisuria-Armer, Julian Maller, John Mansfield, Jonathan L. Marchini, Paul Martin, Dunecan Co Massey, Wendy L. McArdle, Peter McGuffin,

Kirsten E. McLay, Gil McVean, Alex Mentzer, Michael L. Mimmack, Ann E. Morgan, Andrew P. Morris, Craig Mowat, Patricia B. Munroe, Simon Myers, William Newman, Elaine R. Nimmo, Michael C. O'Donovan, Abiodun Onipinla, Nigel R. Ovington, Michael J. Owen, Kimmo Palin, Aarno Palotie, Kirstie Parnell, Richard Pearson, David Pernet, John Rb Perry, Anne Phillips, Vincent Plagnol, Natalie J. Prescott, Inga Prokopenko, Michael A. Quail, Suzanne Rafelt, Nigel W. Rayner, David M. Reid, Anthony Renwick, Susan M. Ring, Neil Robertson, Samuel Robson, Ellie Russell, David St Clair, Jennifer G. Sambrook, Jeremy D. Sanderson, Stephen J. Sawcer, Helen Schuilenburg, Carol E. Scott, Richard Scott, Sheila Seal, Sue Shaw-Hawkins, Beverley M. Shields, Matthew J. Simmonds, Debbie J. Smyth, Elilan Somaskantharajah, Katarina Spanova, Sophia Steer, Jonathan Stephens, Helen E. Stevens, Kathy Stirrups, Millicent A. Stone, David P. Strachan, Zhan Su, Deborah P.M. Symmons, John R. Thompson, Wendy Thomson, Martin D. Tobin, Mary E. Travers, Clare Turnbull, Damjan Vukcevic, Louise V. Wain, Mark Walker, Neil M. Walker, Chris Wallace, Margaret Warren-Perry, Nicholas A. Watkins, John Webster, Michael N. Weedon, Anthony G. Wilson, Matthew Woodburn, B. Paul Wordsworth, Chris Yau, Allan H. Young, Eleftheria Zeggini, Matthew A. Brown, Paul R. Burton, Mark J. Caulfield, Alastair Compston, Martin Farrall, Stephen C.L. Gough, Alistair S. Hall, Andrew T. Hattersley, Adrian V.S. Hill, Christopher G. Mathew, Marcus Pembrey, Jack Satsangi, Michael R. Stratton, Jane Worthington, Matthew E. Hurles, Audrey Duncanson, Willem H. Ouwehand, Miles Parkes, Nazneen Rahman, John A. Todd, Nilesh J. Samani, Dominic P. Kwiatkowski, Mark I. McCarthy, Nick Craddock, Panos Deloukas, Peter Donnelly, Jenefer M. Blackwell, Elvira Bramon, Juan P. Casas, Aiden Corvin, Janusz Jankowski, Hugh S. Markus, Colin Na Palmer, Robert Plomin, Anna Rautanen, Richard C. Trembath, Ananth C. Viswanathan, Nicholas W. Wood, Chris C.A. Spencer, Gavin Band, Céline Bellenguez, Colin Freeman, Garrett Hellenthal, Eleni Giannoulatou, Matti Pirinen, Richard Pearson, Amy Strange, Hannah Blackburn, Suzannah J. Bumpstead, Serge Dronov, Matthew Gillman, Alagurevathi Jayakumar, Owen T. McCann, Jennifer Liddle, Simon C. Potter, Radhi Ravindrarajah, Michelle Ricketts, Matthew Waller, Paul Weston, Sara Widaa, and Pamela Whittaker. Comparative performances of machine learning methods for classifying crohn disease patients using genome-wide genotyping data. *Scientific Reports*, 9, 2019.

[155] P.D. Hansten. Drug interaction management. *Pharmacy World Science*, 25:94–97, 2003.

[156] Agnes I. Vitry. Comparative assessment of four drug interaction compendia. *British Journal of Clinical Pharmacology*, 63:709–714, 2007.

[157] Thomas R. Fulda, Robert J. Valuck, Jeanne Vander Zanden, Sandra Parker, and Patricia J. Byrns. Disagreement among drug compendia on inclusion and ratings of

drug-drug interactions. *Current Therapeutic Research - Clinical and Experimental*, 61:540–548, 2000.

[158] Serkan Ayvaz, John Horn, Oktie Hassanzadeh, Qian Zhu, Johann Stan, Nicholas P. Tatonetti, Santiago Vilar, Mathias Brochhausen, Matthias Samwald, Majid Rastegar-Mojarad, Michel Dumontier, and Richard D. Boyce. Toward a complete dataset of drug-drug interaction information from publicly available sources. *Journal of Biomedical Informatics*, 55:206–217, 2015.

[159] Shobha Phansalkar, Amrita A. Desai, Douglas Bell, Eileen Yoshida, John Doole, Melissa Czochanski, Blackford Middleton, and David W. Bates. High-priority drug-drug interactions for use in electronic health records. *Journal of the American Medical Informatics Association*, 19:735–743, 9 2012.

[160] Priska Vonbach, André Dubied, Stephan Krähenbühl, and Jürg H. Beer. Evaluation of frequently used drug interaction screening programs. *Pharmacy World and Science*, 30:367–374, 2008.

[161] Lorraine M. Wang, Maple Wong, James M. Lightwood, and Christine M. Cheng. Black box warning contraindicated comedications: Concordance among three major drug interaction screening programs. *Annals of Pharmacotherapy*, 44:28–34, 2010.

[162] Kin Wah Fung, Joan Kapusnik-Uner, Jean Cunningham, Stefanie Higby-Baker, and Olivier Bodenreider. Comparison of three commercial knowledge bases for detection of drug-drug interactions in clinical decision support. *Journal of the American Medical Informatics Association*, 24:806–812, 2017.

[163] National Institute for Health and Care Excellence. Bnf: British national formulary, 2018.

[164] Agence nationale de sécurité du médicament et des produits de santé. Thésaurus des interactions médicamenteuses, 2019.

[165] IBM Watson Health. Micromedex® (electronic version), 2018.

[166] Tony Avery. A brief insight into the bnf. *Prescriber*, 19:7–8, 2008.

[167] Joy Ogden. The british national formulary : past , present and future health professionals. *Prescriber*, 28:20–24, 2017.

[168] Christopher A. Aakre, Laurie J. Pencille, Kristi J. Sorensen, Jane L. Shellum, Guilherme Del Fiol, Lauren A. Maggio, Larry J. Prokop, and David A. Cook. Electronic knowledge resources and point-of- care learning: A scoping review, 2018.

[169] Tina Roblek, Tomaz Vaupotic, Ales Mrhar, and Mitja Lainscak. Drug-drug interaction software in clinical practice: A systematic review. *European Journal of Clinical Pharmacology*, 71:131–142, 2015.

[170] Python Software Foundation. Python language reference, 2009.

[171] Sebastien Cossin. Imthesaurusansm: Thesaurus des interactions medicamenteuses de l' ansm, 2016. R package version 0.1.

[172] Stuart J Nelson, Kelly Zeng, John Kilbourne, Tammy Powell, and Robin Moore. Normalized names for clinical drugs: Rxnorm at 6 years. *Journal of the American Medical Informatics Association*, 18:441–448, 2011.

[173] Rxnorm extension: an ohdsi resource to represent international drugs.

[174] OHDSI Team. Usagi: an application to help create mappings between coding systems and the vocabulary standard concepts., 2020.

[175] Steven Bird, Ewan Klein, and Edward Loper. *Natural language processing with Python: analyzing text with the natural language toolkit.* " O'Reilly Media, Inc.", 2009.

[176] G. Corona, E. Razzoli, G. Forti, and M. Maggi. The use of phosphodiesterase 5 inhibitors with concomitant medications, 2008.

[177] Abel Ickowicz. Methylphenidate and the cytochrome p450 system. *Canadian Journal of Psychiatry*, 48:426, 2003.

[178] Christian Kollmannsberger, Georg Bjarnason, Patrick Burnett, Patricia Creel, Mellar Davis, Nancy Dawson, Darren Feldman, Suzanne George, Jerome Hershman, Thomas Lechner, Amy Potter, Eric Raymond, Nathaniel Treister, Laura Wood, Shenhong Wu, and Ronald Bukowski. Sunitinib in metastatic renal cell carcinoma: Recommendations for management of noncardiovascular toxicities. *The Oncologist*, 16:543–553, 2011.

[179] Michael Greenberg and M Susan Ridgely. Clinical decision support and malpractice risk. *Jama*, 306:90–91, 2011.

[180] Shobha Phansalkar, Heleen van der Sijs, Alisha D. Tucker, Amrita A. Desai, Douglas S. Bell, Jonathan M. Teich, Blackford Middleton, and David W. Bates. Drug-drug interactions that should be noninterruptive in order to reduce alert fatigue in electronic health records. *Journal of the American Medical Informatics Association*, 20:489–493, 2013.

[181] Richard T. Scheife, Lisa E. Hines, Richard D. Boyce, Sophie P. Chung, Jeremiah D. Momper, Christine D. Sommer, Darrell R. Abernethy, John R. Horn, Stephen J.

Sklar, Samantha K. Wong, Gretchen Jones, Mary L. Brown, Amy J. Grizzle, Susan Comes, Tricia Lee Wilkins, Clarissa Borst, Michael A. Wittie, and Daniel C. Malone. Consensus recommendations for systematic evaluation of drug–drug interaction evidence for clinical decision support. *Drug Safety*, 38:197–206, 2015.

[182] John R. Horn, Karl F. Gumpper, J. Chad Hardy, Patrick J. McDonnell, Shobha Phansalkar, and Cynthia Reilly. Clinical decision support for drug-drug interactions: improvement needed. *American journal of health-system pharmacy : AJHP : official journal of the American Society of Health-System Pharmacists*, 70:905–909, 2013.

[183] Hugh Tilson, Lisa E. Hines, Gerald McEvoy, David M. Weinstein, Philip D. Hansten, Karl Matuszewski, Marianne Le Comte, Stefanie Higby-Baker, Joseph T. Hanlon, Lynn Pezzullo, Kathleen Vieson, Amy L. Helwig, Shiew Mei Huang, Anthony Perre, David W. Bates, John Poikonen, Michael A. Wittie, Amy J. Grizzle, Mary Brown, and Daniel C. Malone. Recommendations for selecting drug-drug interactions for clinical decision support. *American Journal of Health-System Pharmacy*, 73:576–585, 2016.

[184] National Center for Health Statistics. Health, united states, 2019, 2021.

[185] Ying Li, Ryan B Patrick, Ying Wei, and Carol Friedman. A method to combine signals from spontaneous reporting systems and observational healthcare data to detect adverse drug reactions. *Drug Safety*, 38:895–908, 2015.

[186] Ying Li, Antonio Jimeno Yepes, and Cao Xiao. Combining social media and fda adverse event reporting system to detect adverse drug reactions. *Drug Safety*, 43:893–903, 2020.

[187] Rave Harpaz, William DuMouchel, Martijn Schuemie, Olivier Bodenreider, Carol Friedman, Eric Horvitz, Anna Ripple, Alfred Sorbello, Ryen W. White, Rainer Winnenburg, and Nigam H. Shah. Toward multimodal signal detection of adverse drug reactions. *Journal of Biomedical Informatics*, 76:41–49, 12 2017.

[188] Andrew Bate, Antoine Pariente, Manfred Hauben, and Bernard Bégaud. Quantitative Signal Detection and Analysis in Pharmacovigilance. In *Mann's Pharmacovigilance: Third Edition*, pages 331–354. 2014.

[189] G Niklas Norén, Ola Caster, Kristina Juhlin, and Marie Lindquist. Authors' reply to harpaz et al. comment on: "zoo or savannah? choice of training ground for evidence-based pharmacovigilance". 2014.

[190] E. A. Voss, R. D. Boyce, P. B. Ryan, J. van der Lei, P. R. Rijnbeek, and M. J. Schuemie. Accuracy of an automated knowledge base for identifying drug adverse reactions. *Journal of Biomedical Informatics*, 66:72–81, 2017.

[191] Kristina Juhlin, Daniel Soeria-Atmadja, Bharat Thakrar, and G Niklas Norén. Evaluation of statistical measures for adverse drug interaction surveillance. *Pharmacoepidemiology and Drug Safety Drug Saf*, 23:294–5, 2014.

[192] Raziyeh Kheshti, Mohammadsadegh Aalipour, and Soha Namazi. A comparison of five common drug–drug interaction software programs regarding accuracy and comprehensiveness. *Journal of Research in Pharmacy Practice*, 5:257, 2016.

[193] Michael Kuhn, Ivica Letunic, Lars Juhl Jensen, and Peer Bork. The sider database of drugs and side effects. *Nucleic Acids Research*, 44:D1075–D1079, 2016.

[194] Juan M Banda, Lee Evans, Rami S Vanguri, Nicholas P Tatonetti, Patrick B Ryan, and Nigam H Shah. Data descriptor: A curated and standardized adverse drug event resource to accelerate drug safety research. *Scientific Data*, 3, 2016.

[195] Andrew Bate and S. J.W. Evans. Quantitative signal detection using spontaneous adr reporting. *Pharmacoepidemiology and Drug Safety*, 18:427–436, 2009.

[196] Nicholas P. Tatonetti, Patrick P. Ye, Roxana Daneshjou, and Russ B. Altman. Data-driven prediction of drug effects and interactions. *Science Translational Medicine*, 4, 2012.

[197] Louis Dijkstra, Marco Garling, Ronja Foraita, and Iris Pigeot. Adverse drug reaction or innocent bystander? a systematic comparison of statistical discovery methods for spontaneous reporting systems. *Pharmacoepidemiology and Drug Safety*, 29:396–403, 4 2020.

[198] Johan Hopstadius, G. Niklas Norén, Andrew Bate, and I. Ralph Edwards. Impact of stratification on adverse drug reaction surveillance. *Drug Safety*, 31:1035–1048, 2008.

[199] Catalogue of bias collaboration, Jeffrey K. Aronson, C Bankhead, K R Mahtani, and D Nunan. Confounding by indication, 2018.

[200] Cédric Bousquet, Corneliu Henegar, Agnès Lillo Le Louët, Patrice Degoulet, and Marie Christine Jaulent. Implementation of automated signal generation in pharmacovigilance using a knowledge-based approach. *International Journal of Medical Informatics*, 74:563–571, 2005.

[201] Ronald K. Pearson, Manfred Hauben, David I. Goldsmith, A. Lawrence Gould, David Madigan, Donald J. O'Hara, Stephanie J. Reisinger, and Alan M. Hochberg. Influence of the meddra® hierarchy on pharmacovigilance data mining results. *International Journal of Medical Informatics*, 78:97–103, 2009.

[202] Richard Hill, Johan Hopstadius, Magnus Lerch, and G. Niklas Noren. An attempt to expedite signal detection by grouping related adverse reaction terms. *Drug Safety*, 35:1194–1195, 2012.

[203] Cédric Bousquet, Georges Lagier, Agnès Lillo-Le Louët, Christine Le Beller, Alain Venot, and Marie-Christine Jaulent. Appraisal of the meddra conceptual structure for describing and grouping adverse drug reactions. *Drug Safety*, 28:19–34, 2005.

[204] Hélène Géniaux, Denise Assaf, Ghada Miremont-Salamé, Bénédicte Raspaud, Amandine Gouverneur, Philip Robinson, Antoine Pariente, and Francesco Salvo. Performance of the standardised meddra® queries for case retrieval in the french spontaneous reporting database. *Drug Safety*, 37:537–542, 2014.

[205] Antoni F.Z. Wisniewski, Andrew Bate, Cedric Bousquet, Andreas Brueckner, Gianmario Candore, Kristina Juhlin, Miguel A Macia-Martinez, Katrin Manlik, Naashika Quarcoo, Suzie Seabroke, Jim Slattery, Harry Southworth, Bharat Thakrar, Phil Tregunno, Lionel Van Holle, Michael Kayser, and G. Niklas Norén. Good signal detection practices: Evidence from imi protect. *Drug Safety*, 39:469–490, 2016.

[206] Richard D. Boyce, Patrick B. Ryan, G. Niklas Norén, Martijn J. Schuemie, Christian Reich, Jon Duke, Nicholas P. Tatonetti, Gianluca Trifirò, Rave Harpaz, J. Marc Overhage, Abraham G. Hartzema, Mark Khayter, Erica A. Voss, Christophe G. Lambert, Vojtech Huser, and Michel Dumontier. Bridging islands of information to establish an integrated knowledge base of drugs and health outcomes of interest. *Drug Safety*, 37:557–567, 2014.

[207] Manfred Hauben, Jeffrey K. Aronson, and Robin E. Ferner. Evidence of misclassification of drug–event associations classified as gold standard 'negative controls' by the observational medical outcomes partnership (omop). *Drug Safety*, 39:421–432, 5 2016.

[208] Johanna Strandell, Ola Caster, Andrew Bate, Niklas Norén, and I Ralph Edwards. Reporting patterns indicative of adverse drug interactions a systematic evaluation in vigibase, 2011.

[209] Alan M. Hochberg, Manfred Hauben, Ronald K. Pearson, Donald J. OHara, Stephanie J. Reisinger, David I. Goldsmith, A. Lawrence Gould, and David Madigan. An evaluation of three signal-detection algorithms using a highly inclusive reference event database. *Drug Safety*, 32:509–525, 2009.

[210] Keith B. Hoffman, Mo Dimbil, Nicholas P. Tatonetti, and Robert F. Kyle. A pharmacovigilance signaling system based on fda regulatory action and post-marketing adverse event reports. *Drug Safety*, 39:561–575, 2016.

[211] Mickael Arnaud, Bernard Bégaud, Frantz Thiessard, Quentin Jarrion, Julien Bezin, Antoine Pariente, and Francesco Salvo. An automated system combining safety signal detection and prioritization from healthcare databases: A pilot study. *Drug Safety*, 41:377–387, 2018.

[212] Suzie Seabroke, Gianmario Candore, Kristina Juhlin, Naashika Quarcoo, Antoni Wisniewski, Ramin Arani, Jeffery Painter, Philip Tregunno, G. Niklas Norén, and Jim Slattery. Performance of stratified and subgrouped disproportionality analyses in spontaneous databases. *Drug Safety*, 39:355–364, 2016.

[213] Yolanda Alvarez, Ana Hidalgo, Francois Maignen, and Jim Slattery. Validation of statistical signal detection procedures in eudravigilance post-authorization data a retrospective evaluation of the potential for earlier signalling.

[214] Mateusz Maciejewski, Eugen Lounkine, Steven Whitebread, Pierre Farmer, William DuMouchel, Brian K Shoichet, and Laszlo Urban. Reverse translation of adverse event reports paves the way for de-risking preclinical off-targets. *eLife*, 6:1–25, 2017.

[215] Johanna Strandell, Ola Caster, Johan Hopstadius, I. Ralph Edwards, and G. Niklas Norén. The development and evaluation of triage algorithms for early discovery of adverse drug interactions. *Drug Safety*, 36:371–388, 2013.

[216] Sara Hult, Daniele Sartori, Tomas Bergvall, Sara Hedfors Vidlin, Birgitta Grundmark, Johan Ellenius, and G. Niklas Norén. A feasibility study of drug–drug interaction signal detection in regular pharmacovigilance. *Drug Safety*, 2020.

[217] S. M. Huang, R. Temple, D. C. Throckmorton, and L. J. Lesko. Drug interaction studies: Study design, data analysis, and implications for dosing and labeling, 2 2007.

[218] Sara K. Quinney. Opportunities and challenges of using big data to detect drug-drug interaction risk. *Clinical Pharmacology Therapeutics*, pages 1–3, 6 2019.

[219] Kristen M. Fedak, Autumn Bernal, Zachary A. Capshaw, and Sherilyn Gross. Applying the bradford hill criteria in the 21st century: How data integration has changed causal inference in molecular epidemiology. *Emerging Themes in Epidemiology*, 12, 9 2015.

[220] Theodoros G. Soldatos, Sarah Kim, Stephan Schmidt, Lawrence J. Lesko, and David B. Jackson. Advancing drug safety science by integrating molecular knowledge with post-marketing adverse event reports. *CPT: Pharmacometrics and Systems Pharmacology*, 2022.

[221] Peter Schotland, Rebecca Racz, David Jackson, Robert Levin, David G. Strauss, and Keith Burkhart. Target-adverse event profiles to augment pharmacovigilance: A pilot study with six new molecular entities. *CPT: Pharmacometrics and Systems Pharmacology*, 7:809–817, 2018.

[222] Peter Schotland, Rebecca Racz, David B. Jackson, Theodoros G. Soldatos, Robert Levin, David G. Strauss, and Keith Burkhart. Target adverse event profiles for predictive safety in the postmarket setting. *Clinical Pharmacology and Therapeutics*, 109:1232–1243, 5 2021.

[223] Sebastian Köhler, Michael Gargano, Nicolas Matentzoglu, Leigh C Carmody, David Lewis-Smith, Nicole A Vasilevsky, Daniel Danis, Ganna Balagura, Gareth Baynam, Amy M Brower, Tiffany J. Callahan, Christopher G. Chute, Johanna L. Est, Peter D. Galer, Shiva Ganesan, Matthias Griese, Matthias Haimel, Julia Pazmandi, Marc Hanauer, Nomi L. Harris, Michael J. Hartnett, Maximilian Hastreiter, Fabian Hauck, Yongqun He, Tim Jeske, Hugh Kearney, Gerhard Kindle, Christoph Klein, Katrin Knoflach, Roland Krause, David Lagorce, Julie A McMurry, Jillian A Miller, Monica C Munoz-Torres, Rebecca L Peters, Christina K Rapp, Ana M Rath, Shahmir A Rind, Avi Z Rosenberg, Michael M Segal, Markus G Seidel, Damian Smedley, Tomer Talmy, Yarlalu Thomas, Samuel A Wiafe, Julie Xian, Zafer Yüksel, Ingo Helbig, Christopher J Mungall, Melissa A Haendel, and Peter N Robinson. The human phenotype ontology in 2021. *Nucleic Acids Research*, 49(D1):D1207–D1217, 2021.

[224] D. S. Wishart. Drugbank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Research*, 34:D668–D672, 1 2006.

[225] Damian Szklarczyk, Annika L Gable, Katerina C Nastou, David Lyon, Rebecca Kirsch, Sampo Pyysalo, Nadezhda T Doncheva, Marc Legeay, Tao Fang, Peer Bork, Lars J Jensen, and Christian von Mering. The string database in 2021: Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Research*, 49:D605–D612, 2021.

[226] Rachel Howard. Risk of myopathy with statin and fibrate treatment. *Prescriber*, 18(7):75–77, 2007.

[227] Mette Marie Hougaard Christensen, Maija Bruun Haastrup, Thomas Øhlenschlæger, Peter Esbech, Sidsel Arnspang Pedersen, Ann Cathrine Bach Dunvald, Tore Bjerregaard Stage, Daniel Pilsgaard Henriksen, and Andreas James Thestrup Pedersen. Interaction potential between clarithromycin and individual statins—A systematic review. *Basic & Clinical Pharmacology & Toxicology*, 126(4):307–317, 4 2020.

[228] Curt D. Furberg and Bertram Pitt. Withdrawal of cerivastatin from the world market. *Current Controlled Trials in Cardiovascular Medicine*, 2(5):205–207, 2001.

[229] Terry A. Jacobson. Myopathy with statin-fibrate combination therapy: Clinical considerations. *Nature Reviews Endocrinology*, 5(9):507–518, 2009.

[230] Terry A. Jacobson and Franklin H. Zimmerman. Fibrates in combination with statins in the management of dyslipidemia, 2006.

[231] James M. McKenney, Michel Farnier, Kwok Wing Lo, Harold E. Bays, Inna Perevozkaya, Gary Carlson, Michael J. Davies, Yale B. Mitchel, and Barry Gumbiner. Safety and Efficacy of Long-Term Co-Administration of Fenofibrate and Ezetimibe in Patients With Mixed Hyperlipidemia. *Journal of the American College of Cardiology*, 47(8):1584–1587, 4 2006.

[232] Shinichi Oikawa, Shizuya Yamashita, Noriaki Nakaya, Jun Sasaki, and Suminori Kono. Efficacy and safety of long-term coadministration of fenofibrate and ezetimibe in patients with combined hyperlipidemia: Results of the EFECTL study. *Journal of Atherosclerosis and Thrombosis*, 24(1):77–94, 2017.

[233] Suyu Mei and Kun Zhang. A machine learning framework for predicting drug–drug interactions. *Scientific Reports*, 11:1–12, 9 2021.

[234] Reza Ferdousi, Reza Safdari, and Yadollah Omidi. Computational prediction of drug-drug interactions based on drugs functional similarities. *Journal of Biomedical Informatics*, 70:54–64, 6 2017.

[235] Anna Bauer-Mehren, Erik M. van Mullingen, Paul Avillach, María del Carmen Carrascosa, Ricard Garcia-Serna, Janet Piñero, Bharat Singh, Pedro Lopes, José L. Oliveira, Gayo Diallo, Ernst Ahlberg Helgee, Scott Boyer, Jordi Mestres, Ferran Sanz, Jan A. Kors, and Laura I. Furlong. Automatic filtering and substantiation of drug safety signals. *PLoS Computational Biology*, 8(4):1002457, 2012.

[236] Xiao Qin, Tabassum Kakar, Elke A Rundensteiner, Susmitha Wunnava, and Lei Cao. Maras: Signaling multi-drug adverse reactions. 17, 2017.

[237] Danai Chasioti, Xiaohui Yao, Pengyue Zhang, Samuel Lerner, Sara K. Quinney, Xia Ning, Lang Li, and Li Shen. Mining directional drug interaction effects on myopathy using the faers database. *IEEE Journal of Biomedical and Health Informatics*, 23:2156–2163, 9 2019.

[238] Francois Maignen, Manfred Hauben, Eric Hung, Lionel Van Holle, and Jean Michel Dogne. Assessing the extent and impact of the masking effect of disproportionality analyses on two spontaneous reporting systems databases. *Pharmacoepidemiology and Drug Safety*, 23:195–207, 2 2014.

[239] Antoine Pariente, Marie Didailler, Paul Avillach, Ghada Miremont-Salame, Annie Fourrier-Réglat, Francoise Haramburu, and Nicholas Moore. A potential competition bias in the detection of safety signals from spontaneous reporting databases. *Pharmacoepidemiology and Drug Safety*, 19:1166–1171, 2010.

[240] Francesco Salvo, Florent Leborgne, Frantz Thiessard, Nicholas Moore, Bernard Bégaud, and Antoine Pariente. A potential event-competition bias in safety signal detection: Results from a spontaneous reporting research database in france. *Drug Safety*, 36:565–572, 2013.

[241] Harry Hochheiser, Xia Jing, Elizabeth A. Garcia, Serkan Ayvaz, Ratnesh Sahay, Michel Dumontier, Juan M. Banda, Oya Beyan, Mathias Brochhausen, Evan Draper, Sam Habiel, Oktie Hassanzadeh, Maria Herrero-Zazo, Brian Hocum, John Horn, Brian LeBaron, Daniel C. Malone, Øystein Nytrø, Thomas Reese, Katrina Romagnoli, Jodi Schneider, Louisa Zhang, and Richard D. Boyce. A Minimal Information Model for Potential Drug-Drug Interactions. *Frontiers in Pharmacology*, 11, 3 2021.

[242] Eric Chou, Richard D. Boyce, Baran Balkan, Vignesh Subbian, Andrew Romero, Philip D. Hansten, John R. Horn, Sheila Gephart, and Daniel C. Malone. Designing and evaluating contextualized drug-drug interaction algorithms. *JAMIA Open*, 4(1), 1 2021.

[243] Don van Ravenzwaaij, Pete Cassey, and Scott D. Brown. A simple introduction to Markov Chain Monte–Carlo sampling. *Psychonomic Bulletin and Review*, 25(1):143–154, 2 2018.

[244] Meera M. Dhodapkar , Joseph S. Ross , and Reshma Ramachandran. Spontaneous reporting of post-market safety signals: what evidence should support regulatory action? *BMJ*, 10 2022.

[245] Antoine Pariente, Fleur Gregoire, Annie Fourrier-Reglat, Franҫoise Haramburu, and Nicholas Moore. Impact of safety alerts on measures of disproportionality in spontaneous reporting databases: The notoriety bias. *Drug Safety*, 30:891–898, 1 2007.

[246] Selma Dere and Serkan Ayvaz. Prediction of drug–drug interactions by using profile fingerprint vectors and protein similarities. *Healthcare Informatics Research*, 26(1):42–49, 1 2020.

[247] Payal Chandak, Kexin Huang, and Marinka Zitnik. Building a knowledge graph to enable precision medicine. *Scientific Data*, 10(1), 12 2023.

[248] Lauren E Walker, Aseel S Abuzour, Danushka Bollegala, Andrew Clegg, Mark Gabbay, Alan Griffiths, Cecil Kullu, Gary Leeming, Frances S Mair, Simon Maskell, Samuel Relton, Roy A Ruddle, Eduard Shantsila, Matthew Sperrin, Tjeerd Van Staa, Alan Woodall, and Iain Buchan. The DynAIRx Project Protocol: Artificial Intelligence for dynamic prescribing optimisation and care integration in multimorbidity. *Journal of Multimorbidity and Comorbidity*, 12:263355652211454, 1 2022.