Exploiting Heterogeneous Publicly Available Data Sources for Drug Safety Surveillance:

Computational Framework and Case Studies

Vassilis G. Koutkias^{1,2,*}, Agnès Lillo-Le Louët³ and Marie-Christine Jaulent²

¹Institute of Applied Biosciences, Centre for Research & Technology Hellas

²INSERM, U1142, LIMICS, F-75006, Paris, France; Sorbonne Universités, UPMC Univ Paris

06, UMR_S 1142, LIMICS, F-75006, Paris, France; Université Paris 13, Sorbonne Paris Cité,

LIMICS, (UMR_S 1142), F-93430, Villetaneuse, France

³Centre Régional de Pharmacovigilance, Hôpital Européen Georges-Pompidou, AP-HP,

F-75015, Paris, France

* *Corresponding author address*: Institute of Applied Biosciences, Centre for Research &

Technology Hellas, 6th Km. Charilaou-Thermi Road, Thermi, Thessaloniki,

P.O. Box 60631, GR 57001, GREECE, Tel.: +30 23 11 25 76 15

E-mail: vkoutkias@certh.gr

The Version of Record of this manuscript has been published and is available
in Expert Opinion on Drug Safety, Dec. 1, 2016,

<http://www.tandfonline.com/10.1080/14740338.2017.1257604>

Abstract

Objective: Driven by the need of pharmacovigilance centres and companies to routinely collect and review all available data about adverse reactions and adverse events of interest, we introduce and validate a computational framework exploiting diverse publicly available data sources.

Research design and methods: Our approach relies on appropriate query formulation for data acquisition and subsequent filtering, transformation and joint visualisation of the obtained data. We acquired data from the FDA Adverse Event Reporting System (FAERS), PubMed and Twitter. In order to assess the validity and the robustness of the approach, we elaborated on two important case studies, namely, clozapine-induced cardiomyopathy/myocarditis versus haloperidol-induced cardiomyopathy/myocarditis, and apixaban-induced cerebral haemorrhage.

Results: The analysis of the obtained data provided interesting insights (identification of potential patient and healthcare professional experiences regarding ADRs in Twitter, information/arguments against an ADR existence across all sources) while illustrating the benefits (complementing data from multiple sources to strengthen/confirm a hypothesis) and the underlying challenges (selecting search terms, data presentation) of exploiting heterogeneous information sources, thereby advocating the need for the proposed framework.

Conclusions: This work contributes in establishing a continuous learning system for drug safety surveillance by exploiting heterogeneous publicly available data sources via appropriate support tools.

Keywords: case studies, computational framework, emerging data sources for pharmacovigilance, heterogeneous public data, joint exploitation, bibliographic databases, social media, spontaneous reporting systems.

1. Introduction

Pharmacovigilance is defined as “the science and activities related with the detection, assessment, understanding and prevention of adverse effects or any other drug-related problem” [1]. An adverse drug reaction (ADR) is a noxious or unintended response to a medicinal product. Response in this context means that a causal relationship between the medicinal product and an adverse effect is at least a reasonable possibility. ADRs may arise from the use of the product within or outside the terms of the marketing authorisation, or from occupational exposure. Conditions of use outside marketing authorisation include off-label use, overdose, misuse, abuse and medication errors [2].

Main pharmacovigilance activities concern collection and analysis of drug safety data, including signal detection (i.e. information on any new or updated safety issue) [3]. Individual case reporting for ADRs in spontaneous reporting systems (SRSs) has been the cornerstone of pharmacovigilance in the last decades [1], [4]. Nevertheless, studies have pinpointed the problem of under-reporting of ADRs in SRSs [5]. Various means have been proposed to address this shortcoming, e.g. enabling patients to directly report potential ADRs to regulatory authorities [6], and the secondary use of electronic health record (EHR) data with appropriate support tools [7]. Overall, further data sources are emerging for drug safety surveillance [8], including observational healthcare databases at large [9]-[11], bibliographic databases [12]-[13], and social media platforms [14].

Considering additional sources and analysis methods to seek evidence or early indications for new or incompletely documented drug safety risks may accommodate the requirement for timely-informed pharmacovigilance [15]. However, this entails a data deluge that calls for appropriate data extraction

and analysis mechanisms to acquire and comprehensively exploit data from these sources in everyday pharmacovigilance practice [16]. Interestingly, an increasing number of such sources are made publicly accessible through well-defined data communication interfaces facilitating programmatic access. Notably, this practice has been adopted recently by FDA, exposing data from the FDA Adverse Event Reporting System (FAERS) via the openFDA platform [17].

Among the common tasks that are routinely performed in pharmacovigilance centers/departments in hospitals is the collection and review of all the available data for an adverse event case of interest. For example, a clinician may ask for a timely evaluation of the respective patient case concerning a possible ADR after a new drug administration, and the pharmacovigilance center/department shall provide a documented answer with a medical advice about the case, after assessing the eventuality of an ADR signal. In some instances, such as a possible ADR related to a new drug, or a new ADR with an “old” drug, if classical databases have no information available, exploring other sources of information is a challenge. A similar requirement for in depth data exploration concerns the task of signal verification performed by domain experts of regulatory authorities. To respond to this challenge, currently available sources of information about the drug-event pair have to be searched by pharmacovigilance experts separately and in many cases without the availability of support tools.

Driven by these pragmatic conditions, this study elaborated on the design and development of a computational framework to accommodate the requirement of exploiting diverse, publicly available data sources for pharmacovigilance. Our aim was to automate to the extent possible the data gathering process and facilitate rigorous inspection of the acquired data by depicting their evolution in time as

well as potential relations of data generated across diverse sources. Thus, we elaborated on appropriate query formulation, data acquisition and subsequent filtering, transformation and joint visualisation.

In order to assess the validity of the obtained data and the contribution of the proposed approach, we elaborated on two case studies by acquiring publicly available data from three diverse sources, i.e. a spontaneous reporting system (SRS), a reference bibliographic database and a popular social media platform. The case studies concerned: (a) clozapine-induced cardiomyopathy or myocarditis versus haloperidol-induced cardiomyopathy or myocarditis, and (b) apixaban-induced cerebral haemorrhage. The obtained data were reviewed with respect to their relevance, type (according to the data source), evolution in time, and potential dependencies across sources. Our ultimate goal was providing insights concerning the exploitation of heterogeneous information sources as well as illustrating the benefits and the underlying challenges, advocating the need for the proposed framework.

2. Material and Methods

2.1 Data Sources

This study relied on the following data sources:

- FAERS [18], the SRS supporting the FDA post-marketing safety surveillance activities. FAERS contains information on adverse event and medication error reports submitted to FDA by healthcare professionals, consumers and manufacturers on a voluntary basis. The structure of FAERS adheres to the international safety reporting guidance, i.e. Electronic Standards for the Transfer of Regulatory Information (E2B) [19]. Adverse events are coded via the Medical

Dictionary for Regulatory Activities (MedDRA®) [20]. Publicly available data from FAERS account from 2004 and on, concerning currently more than 5.8 million reports.

- PubMed [21], the major bibliographic search engine for health and life sciences maintained by the National Library of Medicine (NLM), currently exceeding 26 million citations for biomedical literature from MEDLINE, life science journals, and online books. PubMed was launched on January 1996 and provides citations from 1809, including publication abstracts. PubMed applies quality control in its content (only publications that meet concrete scientific standards are indexed), while its citations are annotated by trained indexers using Medical Subject Headings (MeSH), subheadings, and supplementary concepts, referred as MeSH annotations. MeSH may capture clinical manifestations, drugs, or drug classes. According to [22], approximately 13,000 new ADE-related articles are indexed in PubMed each year.
- Twitter [23], the popular micro-blogging platform that is constantly attracting interest for pharmacovigilance [24]-[27]. Twitter posts contain colloquial text and non-experiential reporting, while additional challenges related to data volume must be addressed. Specifically, based on current statistics on Twitter use, real-time drug surveillance through Twitter would require screening of more than 500 million tweets each day. Additional considerations when searching Twitter for drug safety surveillance include the tweet's short character length, which restricts the amount of information that is posted, as well as custom, Twitter-specific message symbols such as hashtags ('#'), re-tweets ('RT'), replies ('@'), but also links to other Web resources. Twitter provides data from its launch (i.e. March, 2006) and thereon. A significant barrier in using social

media in general for pharmacovigilance is the lack of quality control. To this end, recommendations have been released by regulatory authorities, such as the European Medicines Agency (EMA) [28] and the FDA [29], regarding the use of such data for drug safety surveillance.

While FAERS corresponds to the dominant type of data source for drug safety surveillance, PubMed and Twitter are examples of data sources secondarily used for pharmacovigilance.

Besides their relevance, their potential and the diversity in the scope of pharmacovigilance, an important criterion for selecting the above data sources in the current study was their public availability through programmatic access, facilitating the establishment of a comprehensive computational framework for their exploitation.

2.2 Computational Framework

The overall computational workflow for exploiting the considered data sources comprises three distinct steps accounting for: 1) query formulation, 2) data acquisition, and 3) subsequent filtering, transformation and visualisation. These steps are specified in the following subsections.

2.2.1 Query Formulation

Querying varies according to the targeted data source, thus, a contextualised query mechanism is needed. For example, FAERS individual case reports encode ADRs in the MedDRA® classification and particularly at the Preferred Terms (PT) level, while drugs in each report are expressed in multiple ways, e.g. through the generic name, the medicinal product name, the brand name, the active substance and the RxNorm clinical drug concept (RxNorm CUI), to name a few. PubMed offers the option to include adverse event related MeSH headings as query terms (e.g. "adverse

effects"[Subheading] OR "chemically induced"[Subheading] OR "Chemicals and Drugs Category"[Mesh]), which may increase the specificity of the returned results [30]. While abstracts indexed in PubMed are written in scientific style, Twitter posts include layman terms, idioms and frequently typos.

In this regard, we expanded the query terms for expressing the drug(s) and the health outcome(s) of interest (DOI/HOI) by obtaining synonyms and relevant terms from reference information sources, resulting in an appropriate lexicon. In particular, synonyms for DOIs were retrieved from DrugBank [31], through the Bio2RDF Application Programming Interface (API) [32], while synonyms of HOIs were obtained by exploiting the Web services provided by BioPortal [33], a comprehensive repository of biomedical ontologies, also accounting for relevant MedDRA PTs.

While both DOI and HOI synonyms can be efficiently obtained through reference information sources, identifying appropriate layman terms for searching Twitter is quite complicated and can be hardly automated. Thus, constructing this part of the lexicon requires the involvement of domain experts.

2.2.2 Data Acquisition

We elaborated on systematic access to the data sources described in section 2.1, using public APIs which are available for each source. In particular, we used:

- The adverse event API of the openFDA platform [17], in order to retrieve individual case report data from FAERS. openFDA is built upon the principles of openness and accountability, while ensuring at the same time privacy and security of public FDA data. The output of the API is encoded in JSON (JavaScript Object Notation) format [34].

- The REST (Representational State Transfer) API to obtain data from PubMed offered by Entrez Programming Utilities [35], which provides the response format in either XML (eXtensible Markup Language) or JSON.
- The REST API of Twitter, which also provides its responses in JSON format. Real-time monitoring/processing of tweets is possible through the Twitter Streaming API [36], while search for historical tweets¹ is provided through the Twitter Search API [37].

Retrieving data in a manageable, machine-readable format enabled us to conduct the subsequent part of the computational workflow in a flexible and systematic way.

2.2.3 Data Filtering, Transformation and Visualisation

Data obtained from the APIs of the above sources contain quite detailed information about each report / article / tweet, respectively. For example, openFDA provides for each drug mentioned in an individual case report the corresponding National Drug Codes (NDC) for the product and the respective packages, the Structured Product Labelling (SPL) identifiers and the list of its manufacturers, resulting in a large recordset per case. Similarly, PubMed citations contain cross-reference data, while Twitter returns for each post the application that was used for its generation (e.g. a mobile phone app), the user profile image URL, etc. In the scope of the elaborated use case targeting pharmacovigilance, such information is not interesting for domain experts to browse. For this purpose, we developed a filtering mechanism to exclude such details using the JQ JSON processor

¹ A certain limitation in accessing historical Twitter data through the search API is the fact that a partial number of posts are typically retrieved, especially in queries that return a large number of tweets.

[38]. In addition, another type of filtering was employed to identify duplicates, especially for FAERS cases and Twitter posts. Duplicate reports in FAERS were identified by comparing simultaneously the date of death (if available), patient's characteristics (age and sex), case characteristics (adverse effect term) and the date of notification. Duplicate posts in Twitter were identified as retweets, or as similar content tweets originated from the same or another author by applying basic string matching.

After data filtering, we transformed the remaining data obtained from each individual source into a spreadsheet format, in order to facilitate their browsing and inspection by domain experts. This transformation was performed using OpenRefine [39], a tool for working with messy data. **Figure S1** illustrates an example of this transformation for an excerpt of FAERS data initially represented in JSON. Other data transformations were also applied, in order to support various visualisation options, aiming to illustrate data evolution in time and highlight potential relations of data generated across diverse sources. In particular, we elaborated on line charts as well as interactive visualisations of event series [40], built using the D3 JavaScript library [41].

2.3 Assessment Framework

The reports obtained from FAERS were analysed concerning potential duplicates, the sex and age distribution, the event seriousness, and the DOI characterisation. Data obtained from PubMed and Twitter were assessed concerning their relevance in the scope of pharmacovigilance and type (depending on the data source). In particular, PubMed citations were all reviewed and categorised, depending on the information available (from the title and abstract, to the full text), as: (a) *fully relevant* (the information available in the full text was strictly what was searched for and relevant to

the DOI-HOI relation), (b) *partially relevant* (the information available from the full text or from a part of the citation (title and/or abstract) was relevant with our query and could have an interest), and (c) *excluded* (all other situations, i.e. articles not about the topic, or articles without enough information). In addition, we further categorised the types of citations as corresponding to *animal study*, *case report*, *clinical study*, *database study*, *fundamental study*, *response/comment*, *review* or *unknown*.

Similarly, Twitter posts were categorised as: (a) *fully relevant*; (b) *partially relevant*, and (c) *excluded*.

We further categorised Twitter posts according to their content as: (a) *potential² patient experience/concern*, (b) *potential² healthcare professional experience*, (c) *reference to published study/announcement*, (d) *part of a discussion thread*, and (e) *duplicate post*. We also noted whether posts contained images and/or Web links besides text and we traced manually this material in such cases.

The data characterisation scheme employed per data source is provided in **Table 1**. Evaluations were performed manually for PubMed and Twitter data, and computationally for FAERS data, as part of a routine pharmacovigilance process.

2.4 Case Studies

The computational framework presented in section 2.2 was assessed in two different case studies, which were selected by pharmacovigilance experts based on their diversity and importance for drug safety. The first case study involved two neuroleptics, aiming to compare the available data concerning

² The term “potential” is used, since the origin of Twitter authors (and their role) cannot be certain / verified.

an old drug (haloperidol) and a newer one (clozapine) regarding the same cardiovascular effects. The second case study was chosen because the drug (apixaban) is the last direct oral anticoagulant available with great expectations regarding its safety, particularly for eliminating the risk of cerebral bleeding. Although SRSs are the dominant data source for drug safety surveillance, in the scope of the current study we did not consider one of the employed data sources as the gold standard, since we chose for our two case studies to search and exploit information regarding expected/known ADR risks, and we do not compare, but rather provide an integrative perspective of the findings from all data sources. Further details for each case study are provided in the following subsections.

2.4.1 Clozapine/Haloperidol and Cardiomyopathy/Myocarditis

The use of antipsychotics is growing: their prescription (excluding depot injections) in primary care in England has increased by 24% within a 5-year period [42]. They are used for some types of mental distress or disorder; mainly schizophrenia and manic depression (bipolar disorder). They can also be used to help severe anxiety or depression [43], including other indications e.g. psychosis in Parkinson's disease, in mania and bipolar disorders. Clozapine is the first "atypical" antipsychotic drug used to treat acute symptoms of schizophrenia, first approved in 1989 in the United States. It may be used for long-term treatment of young people (for months and even years), in order to avoid recurrence and/or for treatment maintenance.

Case reports on clozapine-induced myocarditis and cardiomyopathy have been reported in the literature [44]. Thus, cardiomyopathy and myocarditis are now considered as known ADRs of clozapine. Myocarditis is an inflammatory disease of the heart muscle, most commonly resulting from

an infectious process, but it may also be caused by drugs. Its clinical spectrum ranges from an asymptomatic state to a fulminant myocarditis with fatal issue. As myocarditis is a common cause of dilated cardiomyopathy, both diseases are linked [45].

Besides gathering data from the considered sources regarding these ADRs, in this case study we wanted to explore the considered data sources for the same adverse reaction, but with an “old” drug. Two groups of antipsychotics are currently available: the typical, older drugs, which appeared in the mid-1950s, and the atypical, such as clozapine. After investigating the data about myocarditis/cardiomyopathy and clozapine, we explored (within the same sources) data about myocarditis/cardiomyopathy and haloperidol, a typical antipsychotic, available since 1960.

2.4.2 Apixaban and Cerebral Haemorrhage

Apixaban is an oral, direct and highly selective factor Xa (FXa) inhibitor, used for the prevention and treatment of thromboembolic diseases. Its marketing authorisation was first granted in US in December 2012. It is considered as part of the “new” class of direct oral anticoagulants, with clear advantages (e.g. rapid onset of action, wide therapeutic range, predictable therapeutic effect with fixed dosing, no food interactions, no biological monitoring) over the old oral anticoagulant, namely, the vitamin-K-antagonist (VKA) warfarin. One of the major clinical benefits of apixaban is that it is associated with fewer intracranial hemorrhages, the most devastating complication of VKA use [46]. We therefore explored the information available about apixaban and cerebral haemorrhage. The fact that the considered data sources preexisted apixaban’s market release was another reason for selecting this case study, anticipating that all possible data would be available across all the considered data

sources, contrary to the first case study.

3. Results

In this section we describe the findings concerning the case studies presented in section 2.4, which were obtained by applying the computational and assessment framework presented in sections 2.2 and 2.3, respectively. The lexicon used for query formulation in the scope of the elaborated case studies is provided as supplementary material (**Listing L1**). This lexicon has been reviewed by domain experts working in the Centre Régional de Pharmacovigilance, Hôpital Européen Georges-Pompidou (Paris, France) concerning its appropriateness and coverage. The queries used per data source and case study are also provided as supplementary material (**Queries Q1**). For both case studies, data collection was performed in a 6-month period, in which we tested the respective query and data acquisition mechanisms. Data referred in this study correspond to those available from the three data sources employed in this work on April 30, 2015. The assessment was performed via manual review by a pharmacovigilance expert with a systematic review by the first author.

Table 2 provides the number of reports/publications/posts obtained from FAERS/PubMed/Twitter, respectively, by querying the considered sources regarding the DOI-HOI pairs of the case studies, as well as the assessment outcome. An analysis per DOI-HOI pair follows.

3.1 Clozapine/Haloperidol and Cardiomyopathy/Myocarditis

3.1.1 Clozapine and Myocarditis

441 cases of myocarditis were obtained from FAERS. Discarding the identified duplicate cases,

resulted in 407 cases, 76% (310/407) of which concerned men, while the median age was 37 [15-77]. The outcome was fatal for 67 patients (16.5%). Clozapine was the only suspected drug for the 89% of the cases (364/407).

44% (45/103) of the citations retrieved from PubMed were fully accessible; 91% of these were fully relevant (41/45) and 9% (4/45) partially relevant (they referred to another cardiac event or drug, containing also comments to other studies). The first, fully relevant citation was a database study, which was made available in PubMed on January, 1992. From the set of fully relevant citations, 39% (16/41) were case reports, 29% (12/41) were review papers, 15% (6/41) were clinical studies, 12% (5/41) were database studies, while we also obtained 1 animal study and 1 fundamental study.

Out of the 56 Twitter posts, 33 were categorised as fully relevant (referring at published studies, clozapine labeling warning changes, news about the ADR, as well as personal experiences), 10 as partially relevant (parts of a relevant discussion thread), and 13 were excluded (from which 9 messages as part of a discussion thread about the genetic mechanism of clozapine). The first, fully relevant Twitter post was made available on March, 2010. Interestingly, among the fully relevant Twitter posts, we discriminated 3 posts (2 generated from the same user) corresponding to potential patient experiences of the ADR, and 2 potential healthcare professional experiences regarding this ADR. We also identified 2 duplicate messages.

3.1.2 Clozapine and Cardiomyopathy

296 cases of cardiomyopathy were obtained from FAERS. Discarding the identified duplicate cases, resulted in 272 cases, from which 72% (197/272) concerned men, while the median age was 43

[15-72]. The outcome was fatal in 28% (77/272) of the reports. Clozapine was the only suspected drug for 88% of the cases (242/272).

51% (26/53) of the citations retrieved from PubMed were fully accessible; 92% of these were fully relevant (24/26), supporting the existence of the targeted ADR, and 8% (2/26) were classified as partially relevant, referring to another cardiac event. The first, fully relevant citation was a database study, which was made available in PubMed on December, 1999, following 1 prior case report published in June, 1996 (categorised as partially relevant, since its full text was not accessible). From the set of fully relevant citations, 38% (9/24) were case reports, 29% (7/24) were review papers, 16.5% (4/24) were database studies, 2 were clinical studies, 1 was a fundamental study, and 1 was an animal study.

Out of the 9 Twitter posts obtained, 2 were categorised as fully relevant (referring at published studies), 5 as partially relevant (including 1 with relevant text though linking to a Web page of a medical portal requiring an account for accessing the full details, 1 written in Indonesian and 3 duplicate posts referring to a case report article about a similar cardiac adverse effect), while 2 were excluded due to unreachable Web links. The first, fully relevant Twitter post was made available on March, 2011.

3.1.3 Haloperidol and Myocarditis

50 cases of myocarditis were obtained from FAERS. Discarding the identified duplicate cases, resulted in 34 distinct cases of myocarditis, from which 23 concerned males, 9 females, and 3 did not include gender information. The median age was 42.5 [23-66] and the outcome was fatal in 41% (14/34) of the cases. Haloperidol was always associated with other suspected drugs, and clozapine was also

suspected in 68% of the cases (23/34). The first case was reported on March 2004.

From the 3 articles retrieved from PubMed, 2 were considered as fully relevant (1 animal study and 1 database study), while the later was a case report classified as partially relevant, as it concerned probably a myocarditis after an overdose (we only had access to the title and the abstract). The first, fully relevant citation (available in May 2001) was the database study referring to antipsychotic drugs and heart muscle disorder.

No Twitter posts were retrieved for haloperidol-induced myocarditis.

3.1.4 Haloperidol and Cardiomyopathy

66 cases of cardiomyopathy were retrieved from FAERS, from which 1 duplicate was identified. 70% of the cases concerned men (45/65), and the median age was 52 [20-82]. The outcome was fatal in 44% (29/65) of the cases. Haloperidol was always associated with other suspected drugs, and clozapine was also suspected in 45% (29/65).

Out of 6 articles retrieved from PubMed, we had access to the full text in 4 of them. Upon these, 3 were considered as fully relevant supporting the existence of the ADR and 1 as partially relevant (referred to another cardiac effect). The fully relevant citations corresponded to 1 animal study, 1 database study and 1 fundamental study, the first one chronologically being the database study (the one identified also for haloperidol and myocarditis).

The single Twitter post retrieved was written in Indonesian (as identified by an online translation service) containing though the DOI-HOI pair in English, potentially corresponding to a healthcare professional experience of the ADR.

3.2 Apixaban and Cerebral Haemorrhage

As depicted in **Table 2**, from the 76 cases retrieved from FAERS, we identified 2 pairs of duplicates. 60% of patients were men (44/74), median age was 75 [6-92], including 2 pediatric cases. Apixaban was the only suspected drug in 81% of the cases (60/74). The outcome was fatal in 39% of the cases (29/74), and the first case was reported on August 2011.

From the 102 citations retrieved from PubMed, 48% were fully available (49/102). From these, only 22.5% (11/49) were considered as fully relevant, 26.5% (13/49) as partially relevant, while 51% (25/49) were excluded. We found only 1 case report (fully relevant) and 1 partially relevant from all the articles. Apixaban has been authorised recently, and the first articles have been published in the beginning of 2011. Most of the retrieved articles were therefore reviews, analyses of clinical data about direct anticoagulants (including apixaban), while the bleeding risk, including cerebral haemorrhage, is discussed as a benefit of this therapeutic class compared to old anticoagulants. This is the main explanation for the low number of fully relevant citations obtained.

From the 8 tweets retrieved, 3 referred explicitly to a published study indicating lower risk for intracranial bleeding as the most important advantage of apixaban over VKAs, 4 were similar comments not citing a published study, while 1 tweet was a case report - personal experience of a healthcare professional concerning intracranial bleeding induced by a VKA, in which less bleeding risk was welcomed as anticipated from apixaban use.

3.3 An Integrative Perspective on the Findings

Evidently, very limited information was obtained for potential haloperidol and

myocarditis/cardiomyopathy, in comparison to clozapine. For instance, haloperidol and cardiomyopathy cases obtained from FAERS represented 12.5% of clozapine cases, while we found only 3 articles and one post compared to 103 articles and 56 posts for clozapine. These results illustrate that, despite the fact that haloperidol has been used for several years, no new information was acquired using the proposed approach.

Besides separately exploring the data obtained from each source, we elaborated on their joint visualisation to illustrate them across a common timeline. **Figure 1** depicts this visualisation for the clozapine-induced myocarditis, for which the larger amount of data was acquired. Although references in the literature for clozapine and cardiomyopathy/myocarditis appeared quite early, FAERS provides public data from 2004 and thereon. Similarly, Twitter was launched on March 2006, thus, relevant tweets appeared from 2010. Especially for clozapine and myocarditis, a number of tweets were identified, clearly indicating potential personal experiences of their authors (**Figure 1**). This indication supports the arguments for synthesis of all possible information sources, in order to build the safety profile of each drug. In order to facilitate data inspection in large datasets, we also elaborated on the interactive visualisation of data (**Figure 2**; based on [40]), enabling pharmacovigilance experts to navigate in time by scrolling vertically (zoom-in/out) and at a fixed time window (moving the graph content horizontally).

Similarly, **Figure 3** illustrates the data obtained across sources for apixaban and cerebral haemorrhage. Given that all the data sources pre-existed apixaban's commercial availability, data are available from all sources. The first FAERS report was obtained quite early (2011) from China. A number of reports

appeared from 2013 and on, all indicating serious events. Notably, many publications cited in PubMed referred to apixaban prior to its commercial release, especially due to the anticipated reduction of bleeding risk. Nevertheless, the obtained Twitter posts were very limited, with most of them referring to a relevant study stressing the benefits of apixaban, rather than its risks. As a result, this case study constitutes a characteristic example of obtaining information against the existence of the considered ADR, which was supported by the majority of PubMed citations and Twitter posts. From another viewpoint, these data indicate the awareness among healthcare professionals and patients regarding the risk of cerebral haemorrhage induced by oral anticoagulants.

4. Discussion

4.1 Contribution of the Study

Accurate and timely identification of drug safety risks is a major issue in pharmacovigilance. Driven by this challenge, significant advancements have been made lately towards active drug safety surveillance, for example, in the scope of the EU-ADR [9], the Mini-Sentinel [10], and the OMOP projects [11]. The above efforts illustrated the potential of exploiting electronic healthcare databases for drug safety surveillance, exceeding the search space for pharmacovigilance experts beyond data from SRSs, but also highlighted the benefits of potential complementary use [47]. Along the line of exploiting further data sources for drug safety, other efforts concentrated on exploiting bibliographic databases [12]-[13], and quite lately user-generated content either shared among networked communities in social media [14], [24]-[27], or implicitly captured through Web search logs [48].

In the current work, we followed a pragmatic approach in exploiting diverse publicly available data

sources for drug safety surveillance, particularly driven by the task that is routinely performed in pharmacovigilance centers/departments in pharmaceutical companies and government organisations concerning collection and review of any possible evidence given an inquiry adverse drug reaction case. The same model may be used by regulatory authorities or by the industry, seeking for information about drug safety and drug use. To this end, we elaborated on the design and development of a computational framework, enabling the systematic exploitation of data obtained from FAERS, PubMed and Twitter. The emphasis was given on automating the data gathering process and facilitating rigorous inspection of data in time, while also depicting potential associations across diverse sources. For this purpose, we built a workflow comprising of appropriate query formulation, data acquisition through available APIs and subsequent filtering, transformation and joint visualisation mechanisms. Overall, we demonstrated an approach that can be applicable to additional public data sources for pharmacovigilance.

The case studies that we conducted concerned two important drug safety aspects for which we were able to identify relevant information in the elaborated data sources, using the proposed approach. Useful information was obtained from each source and in accordance, reinforcing the validity of the results. From the Twitter posts obtained, a number were considered as fully / partially relevant with the DOI-HOI. This was probably related with the strict search strategy that we employed and the lexicon that we defined. The absence of Twitter posts in one of our case studies (i.e. posts for haloperidol) is also useful, indicating the lack of data/evidence from this data source and the limited time-coverage of Twitter for this particular case (Twitter was launched on 2006, while haloperidol is available in the

market from 1960). Even more important, among the tweets on the clozapine-myocarditis case, some concerned potential patient experiences/concerns as well as potential healthcare professional experiences regarding the ADR. Some posts referred at publications, which were identifiable in PubMed. While this could be simply seen as a reference to the literature, it can also express/imply additional aspects, e.g. that the problem is known and important at the time of the posts (a “current” problem), people speak about the literature probably because a case (confirming the literature) or counter-case (contradictory to what is described in the literature) appeared somewhere (even if not described), etc. The apixaban-cerebral haemorrhage case study, constituted a representative example of collecting information against the existence of the ADR supported by the majority of PubMed citations (resulting in excluding 51% of fully available papers) and Twitter posts. Nevertheless, even this outcome can be considered as useful, indicating the awareness/interest among healthcare professionals and patients regarding the risk of cerebral haemorrhage associated with the use of oral anticoagulants.

4.2 Challenges and Limitations

An important factor influencing the amount of the retrieved data is the time coverage of each source. This was particularly noticeable for the haloperidol-cardiomyopathy/myocarditis case, since haloperidol was marketed many years before the release of Twitter and the public availability of FAERS data. From another viewpoint, this shortcoming supports the argument for the synthesis of diverse data sources for drug safety surveillance. The inherent shortcomings of data sources like social media (e.g. lack of quality control, uncertainty and trustworthiness, noise and missing data,

complexity in the analysis, data interpretation framework, automatic tracking of Web resources contained in the posts, etc. that have been widely discussed [14], [49]-[51]), were verified in our study confirming the necessity to support pharmacovigilance professionals with comprehensive tools to exploit such data.

The starting point for performing the introduced computational workflow concerns the definition of a comprehensive query, which shall be designed by taking into account the characteristics of each data source. For example, using search terms from controlled vocabularies like MeSH headings and subheadings when querying PubMed is a reasonable approach that can increase the specificity of the obtained data. However, a recent study indicated that such a strategy entails various challenges besides benefits, thus, requiring careful consideration [30]. Building a lexicon for searching social media content is a complex issue, as it has to account for issues like layman terms and typos, idioms, multilinguality, and so forth. Common sense corpora like SenticNet [52], linked databases like Wikidata [53], as well as resources such as Consumer Health Vocabularies [54], could be useful in identifying such terms, without avoiding though manual expert review.

4.3 Related Works - Future Work

The insights obtained from Twitter in the elaborated case studies (with respect to the relatively low number of posts concerning potential personal adverse event experiences) are in accordance with the findings of Coloma et al. [55], which analysed data from three different social media platforms (including Twitter). Given that Twitter is a general purpose social media platform, its use by patients or healthcare professionals for sharing experiences on potential adverse drug events is a rather special

use case. Digital literacy can be a barrier for many patients and healthcare professionals, potentially hampering the use of social media for reporting. In addition, the particular case studies and the strict search strategy employed (cf. section 4.1) could have an impact in the overall low number of the retrieved posts.

Another relevant study was recently conducted by Topaz et al. [56], which aimed to answer whether patients and clinicians report the same concerns by comparing EHR data with social media data concerning two drugs, namely, aspirin and lipitor. Among the key findings was that the most frequently reported ADRs in the EHR matched the most frequent patients' concerns on social media, while several less frequently reported ADRs were more prevalent on social media, illustrating this way the value of considering additional data sources for pharmacovigilance.

Yeleswarapu et al. presented a semi-automated pipeline to extract drug-adverse event pairs from multiple data sources, including adverse event databases, health-related websites and MEDLINE abstracts [57]. The study relied on data obtained for 12 drugs, and focused on the application of Natural Language Processing (NLP) techniques to identify any kind of adverse event information. Thus, a different search strategy was employed compared to our study, in which we explicitly searched for specific DOI-HOI pairs driven by the common task that is routinely performed in pharmacovigilance centers/departments concerning collection and review of all the evidence given an adverse event case of interest. Our work also emphasised on systematising this effort, under a uniform computational framework.

Our study is inline with the above works, conveying insights as regards the benefits and the underlying

challenges originated from exploiting emerging data sources for pharmacovigilance. In our case, the complexity increases as we considered data sources in a combinatorial setting, with a clear focus on publicly accessible data sources under a pragmatic setting. In this respect, it is evident that presenting the obtained data in a way that one not only can make sense of but also take appropriate actions from is an issue that requires further research. Besides conducting additional case studies which will enable us to potentially generalize and enrich our findings, we elaborate on developing machine learning techniques aiming to automate data classification and systematize the interpretation of the results building upon the presented assessment framework [51], [58], and on the incorporation of the presented computational workflow within a user-friendly, integrated tool to support pharmacovigilance experts in signal detection and evaluation.

5. Conclusions

In this paper, we illustrated a pragmatic approach in exploiting diverse publicly available data sources for drug safety surveillance. Our study was driven by the common task that is routinely performed in pharmacovigilance centers/departments concerning collection and review of any possible evidence given an inquiry case, similar to the tasks that regulatory authorities or the industry perform for routine or targeted drug safety surveillance. To this end, we elaborated on the design and development of a computational framework, enabling the systematic exploitation of data, with a prototype development relying on FAERS, PubMed and Twitter data. The emphasis was given on automating the data gathering process and facilitating rigorous inspection of data in time, while also depicting potential associations across diverse sources. Through two important case studies, we illustrated interesting

insights (e.g. identification of potential patient and healthcare professional experiences regarding the ADRs in Twitter, information against the existence of the risk regarding apixaban-induced cerebral haemorrhage across all sources), benefits (e.g. complementing data from multiple sources to strengthen/confirm a hypothesis) as well as underlying challenges (selecting search terms, joint presentation of the obtained data, etc.) of exploiting heterogeneous information sources, advocating the need for the proposed framework. This work contributes to the establishment of a comprehensive, continuous learning system for drug safety surveillance [59].

References

- [1] World Health Organisation. The safety of medicines in public health programmes: Pharmacovigilance an essential tool. Geneva: WHO; 2006.
- [2] European Medicines Agency. Guideline on good pharmacovigilance practices (GVP). Module III – Pharmacovigilance inspections. Available at:

http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2014/09/WC500172400.pdf [Last accessed 20 April 2016].
- [3] Council for International Organisations of Medical Sciences. Practical aspects of signal detection in pharmacovigilance: report of CIOMS Working Group VIII. Geneva: Council for International Organisations of Medical Sciences; 2010.
- [4] World Health Organisation. The importance of pharmacovigilance - Safety monitoring of medicinal products. UK: World Health Organisation; 2002.
- [5] Hazell L, Shakir SA. Under-reporting of adverse drug reactions: a systematic review. *Drug Saf.*

2006;29:385–96.

- [6] Anderson C, Krska J, Murphy E, Avery A, Yellow Card Study Collaboration. The importance of direct patient reporting of suspected adverse drug reactions: a patient perspective. *Br J Clin Pharmacol.* 2011;72:806–22.
- [7] Sinaci AA, et al. Postmarketing safety study tool: a web based, dynamic, and interoperable system for postmarketing drug surveillance studies. *Biomed Res Int.* 2015;2015:976272.
- [8] Harpaz R, DuMouchel W, Shah NH, Madigan D, Ryan P, Friedman C. Novel data-mining methodologies for adverse drug event discovery and analysis. *Clin Pharmacol Ther.* 2012;91:1010–21.
- [9] Patadia VK, et al. Using real-world healthcare data for pharmacovigilance signal detection - the experience of the EU-ADR project. *Expert Rev Clin Pharmacol.* 2015;8:95–102.
- [10] Platt R, Carnahan R, editors. The US Food and Drug Administration's Mini-Sentinel Program. *Pharmacoepidemiol Drug Saf.* 2012;21:S1–303.
- [11] Evans SJW, editor. Studying the science of observational research: empirical findings from the Observational Medical Outcomes Partnership. *Drug Saf.* 2013;36:S1–204.
- [12] Shang N, Xu H, Rindflesch TC, Cohen T. Identifying plausible adverse drug reactions using knowledge extracted from the literature. *J Biomed Inform.* 2014;52:293–310.
- [13] Shetty KD, Dalal SR. Using information mining of the medical literature to improve drug safety. *J Am Med Inform Assoc.* 2011;18:668–74.
- [14] Ghosh R, Lewis D. Aims and approaches of Web-RADR: a consortium ensuring reliable ADR

reporting via mobile devices and new insights from social media. *Expert Opin Drug Saf.* 2015;14(12):1845–53.

[15] Koutkias VG, Jaulent MC. Computational approaches for pharmacovigilance signal detection: toward integrated and semantically-enriched frameworks. *Drug Saf.* 2015;38:219–32.

[16] Koutkias V, Jaulent MC. A multiagent system for integrated detection of pharmacovigilance signals. *J Med Syst.* 2016;40(2):37.

[17] Reference of the openFDA adverse drug event API. Available at: <https://api.fda.gov/drug/event/>
[Last accessed 20 April 2016]

[18] FDA Adverse Event Reporting System (FAERS). Available at:
[http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrug
Effects/](http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/) [Last accessed 20 April 2016].

[19] Electronic Standards for the Transfer of Regulatory Information. Available at: <http://estri.ich.org/>
[Last accessed 20 April 2016]

[20] Medical Dictionary for Regulatory Activities. Available at: <http://www.meddra.org/> [Last
accessed 20 April 2016]

[21] PubMed. Available at: <http://www.ncbi.nlm.nih.gov/pubmed> [Last accessed 20 April 2016]

[22] Harpaz R, et al. Text mining for adverse drug events: the promise, challenges, and state of the art. *Drug Saf.* 2014;37:777–90.

[23] Twitter. Available at: <https://twitter.com/> [Last accessed 20 April 2016]

[24] Bian J, Topaloglu U, Yu F. Towards large-scale Twitter mining for drug-related adverse events. In

Proc. of the 2012 Int. Workshop on Smart Health and Wellbeing (SHB '12). ACM, New York, NY, USA, pp. 25–32.

[25] O'Connor K, Pimpalkhute P, Nikfarjam A, Ginn R, Smith KL, Gonzalez G. Pharmacovigilance on Twitter? Mining tweets for adverse drug reactions. AMIA Annu Symp Proc. 2014, pp. 924–33.

[26] Jiang K, Zheng Y. Mining Twitter data for potential drug effects. in: Advanced Data Mining and Applications, Lecture Notes in Computer Science, vol. 8346, 2013, pp. 434–43.

[27] Freifeld CC, et al. Digital drug safety surveillance: monitoring pharmaceutical products in Twitter. Drug Saf. 2014;37:343–50.

[28] European Medicines Agency. Guideline on good pharmacovigilance practices (GVP). Module VI—Management and reporting of adverse reactions to medicinal products. Available at: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2014/09/WC500172402.pdf [Last accessed 20 April 2016]

[29] US FDA. For industry: using social media. Available at: <http://www.fda.gov/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CDER/ucm397791.htm> [Last accessed 20 April 2016]

[30] Winnenburg R, et al. Leveraging MEDLINE indexing for pharmacovigilance – Inherent limitations and mitigation strategies. J Biomed Inform. 2015;57:425–35.

[31] DrugBank. Available at: <http://www.drugbank.ca/> [Last accessed 20 April 2016]

[32] Bio2RDF. Available at: <http://bio2rdf.org/> [Last accessed 20 April 2016]

[33] BioPortal Web services API documentation. Available at:

<http://data.bioontology.org/documentation> [Last accessed 20 April 2016]

[34] JavaScript Object Notation. Available at: <http://www.json.org/> [Last accessed 20 April 2016]

[35] Entrez Programming Utilities. Available at: <http://eutils.ncbi.nlm.nih.gov/entrez/>. [Last accessed 20 April 2016]

[36] Twitter Streaming API. Available at: <https://dev.twitter.com/streaming/overview> [Last accessed 20 April 2016]

[37] Twitter Search API. Available at: <https://dev.twitter.com/rest/public/search> [Last accessed 20 April 2016]

[38] The JQ JSON Processor. Available at: <https://stedolan.github.io/jq/> [Last accessed 20 April 2016]

[39] OpenRefine. Available at: <http://openrefine.org/> [Last accessed 20 April 2016]

[40] EventDrops. Available at: <https://github.com/marmelab/EventDrops> [Last accessed 20 April 2016]

[41] D3. Available at: <http://d3js.org/> [Last accessed 20 April 2016]

[42] Mackin P, Thomas SH. Atypical antipsychotic drugs. *BMJ*. 2011 Mar 4;342:d1126.

[43] Royal College of Psychiatrists. Antipsychotics. Available at: <http://www.rcpsych.ac.uk/healthadvice/treatmentswellbeing/antipsychoticmedication.aspx> [Last accessed 20 April 2016]

[44] Kilian JG, Kerr K, Lawrence C, Celermajer DS. Myocarditis and cardiomyopathy associated with clozapine. *Lancet*. 1999 Nov 27;354(9193):1841–5.

[45] Sagar S, Liu PP, Cooper LT Jr. Myocarditis. *Lancet* 2012;379(9817):738–47.

- [46] Connolly SJ, et al. Apixaban in patients with atrial fibrillation. *N Engl J Med*. 2011 Mar 3;364(9):806–17.
- [47] Pacurariu AC, et al. Useful interplay between spontaneous ADR reports and electronic healthcare records in signal detection. *Drug Saf*. 2015 Dec;38(12):1201–10.
- [48] White RW, Harpaz R, Shah NH, DuMouchel W, Horvitz E. Toward enhanced pharmacovigilance using patient-generated data on the Internet. *Clin Pharmacol Ther*. 2014;96:239–46.
- [49] Edwards IR, Lindquist M. Social media and networks in pharmacovigilance: boon or bane? *Drug Saf*. 2011;34:267–71.
- [50] Sloane R, Osanlou O, Lewis D, Bollegala D, Maskell S, Pirmohamed M. Social media and pharmacovigilance: a review of the opportunities and challenges. *Br J Clin Pharmacol*. 2015;80:910–20.
- [51] Sarker A, et al. Social media mining for toxicovigilance: automatic monitoring of prescription medication abuse from Twitter. *Drug Saf*. 2016;39:231–40.
- [52] SenticNet. <http://sentic.net/downloads/>. Last access 20 Feb 2016.
- [53] Vrandečić D, Krötzsch M. Wikidata: a free collaborative knowledgebase. *Communications of the ACM* 2014;7:78–85.
- [54] Zeng QT, Tse T. Exploring and developing consumer health vocabularies. *J Am Med Inform Assoc*. 2006;13:24–9.
- [55] Coloma PM, Becker B, Sturkenboom MC, van Mulligen EM, Kors JA. Evaluating social media networks in medicines safety surveillance: two case studies. *Drug Saf*. 2015;38:921–30.

- [56] Topaz M, et al. Clinicians' reports in electronic health records versus patients' concerns in social media: a pilot study of adverse drug reactions of aspirin and atorvastatin. *Drug Saf.* 2016;39:241–50.
- [57] Yeleswarapu S, Rao A, Joseph T, Saipradeep VG, Srinivasan R. A pipeline to extract drug-adverse event pairs from multiple data sources. *BMC Med Inform Decis Mak.* 2014;14:13.
- [58] Akay A, Dragomir A, Erlandsson BE. A novel data-mining approach leveraging social media to monitor consumer opinion of sitagliptin. *IEEE J Biomed Health Inform.* 2015;19:389–96.
- [59] World Health Organisation. Reporting and learning systems for medication errors: the role of pharmacovigilance centres. WHO; 2014.

Table 1. Characterisation scheme for (a) FAERS individual case reports, (b) PubMed citations, and (c) Twitter posts.

Attribute	Value	Description
Duplicate	Yes / No	Indicates whether the report is a duplicate in the dataset retrieved from FAERS
Sex	F / M / Unknown	Female, male, or unknown
Age	Number / Unknown	The age of the patient for whom the report was made
Fatal outcome	Yes / No / Unknown	Indicates whether the report refers to “death” in the seriousness level field
Drug characterisation	Suspected / Concomitant / Interacting / Unknown	Characterisation of the drug role with respect to the reported ADR

(a)

Attribute	Value	Description
Relevance	fully relevant	the information available in the full text was strictly what was searched for and relevant to the DOI-HOI relation
	partially relevant	the information available from the full text or from a part of the citation (title, and/or abstract) was relevant with our query and could have an interest
	excluded	i.e. articles not about the topic, or articles without enough information
Type	animal study, case report, clinical study, database study, fundamental study, response/comment, review or unknown	

(b)

Attribute	Value	Description
Relevance	fully relevant	the post explicitly refers to the ADR
	partially relevant	the post is somehow relevant to the ADR
	excluded	the post is either irrelevant or refers to missing information (e.g. an embodied Web link is not accessible any more)
Type	potential patient experience/concern, potential healthcare professional experience, reference to published study/announcement, part of a discussion thread, duplicate post	

(c)

Table 2. Data gathered from each data source and their classification for each case study: (a) FAERS reports, (b) PubMed citations and (c) Twitter posts.

	Retrieved reports ³	Excluding duplicates	Sex (M/F/Unknown)	Mean Age	Fatal Cases	Single suspected drug
Clozapine and Myocarditis	441	407	310/83/14	37 [15-77]	67	364
Clozapine and Cardiomyopathy	296	272	197/70/5	43 [15-72]	77	242/269
Haloperidol and Myocarditis	50	34	23/9/2	42.5 [23-66]	14	0
Haloperidol and Cardiomyopathy	66	65	45/20	52 [20-82]	29	0
Apixaban and Cerebral Haemorrhage	76	74	44/25/5	75 [6-92]	29	60

(a)

	Retrieved citations ⁴	Fully relevant	Partially relevant	Excluded
Clozapine and Myocarditis				
All citations	103	41	8	54
Full text access	45	41	4	0
Clozapine and Cardiomyopathy				
All citations	53	27	4	22
Full text access	26	24	2	0
Haloperidol and Myocarditis				
All citations	3	2	1	0
Full text access	2	2	0	0
Haloperidol and Cardiomyopathy				
All citations	6	3	2	1
Full text access	4	3	1	0
Apixaban and Cerebral Haemorrhage				

³ Number of reports obtained from FAERS concerning the DOI-HOI, according to the defined lexicon.

⁴ Results correspond to the number of retrieved citations in which the DOI and the HOI appear in the title or abstract, according to the defined lexicon.

All citations	102	13	26	63
Full text access	49	11	13	25

(b)

Retrieved posts⁵	Fully relevant	Partially relevant	Excluded
Clozapine and Myocarditis			
56	33	10	13
Clozapine and Cardiomyopathy			
9	2	5	2
Haloperidol and Myocarditis			
0	-	-	-
Haloperidol and Cardiomyopathy			
1	0	1	0
Apixaban and Cerebral Haemorrhage			
9	0	8	0

(c)

⁵ Number of posts in Twitter mentioning the DOI and the HOI according to the defined lexicon.

Figures

Figure 1. References across time in the considered data sources for: (a) clozapine and cardiomyopathy, and (b) clozapine and myocarditis.

Figure 2. Three different time windows through interactive data visualisation regarding the clozapine-induced myocarditis case. Navigation in time is enabled by scrolling vertically (zoom-in/out) and at fixed time intervals (moving the graph horizontally). Besides the graph elements, the number of cases/articles/tweets are updated based on the displayed time-window.

Potential healthcare professional experience

Recently saw another patient with #brain hemorrhage due to #Warfarin. New alternatives like Apixaban with less bleed risk are welcome

The Lancet @TheLancet · 20 May 2012
perhaps the most important advantage of apixaban over warfarin: less intracranial bleeding [awe.sm/5roOm](#)

Apixaban and Cerebral Haemorrhage

Figure 3. References on apixaban and cerebral haemorrhage across time in the considered data sources.

Supplementary Material

Figure S1. Example transformation of JSON data obtained from openFDA (formatted to facilitate visualisation) into a spreadsheet.

Listing L1. Lexicon for querying the data sources considered in the study in the scope of the elaborated case studies.

Clozapine: Azaleptine, Cloment, Clonex, Clopin, Clopine, Clopsine, Clorazem, Clozapin, Clozapina, Clozapine, Clozapinum, Clozaril, Fazaclo, Lanolept, Lapenax, Leponex, Lodux, Lozapin, Luften, Mezapin, Refract, Refraxol, Sensipin, Sequax, Sizopin, Sizopril, Syclop, Syzopin, Tanyl, Usphen, Versacloz, Zopen, Zopenia, Zapine, Ziproc, Zopin.

Haloperidol: Aloperidin, Bioperidolo, Brotopon, Dozic, Duraperidol, Einalon S, Eukystol, Haldol, Haloperidol, Halopéridol, Haloperidolum, Halosten, Keselan, Linton, Peluces, Serenace, Seviun, Sigaperidol.

Cardiomyopathy: cardiomyopathy, cardiomyopathy acute, cardiomyopathy alcoholic, cardiomyopathy neonatal, congestive cardiomyopathy, cytotoxic cardiomyopathy, diabetic cardiomyopathy, hypertensive cardiomyopathy, hypertrophic cardiomyopathy, hypertrophic obstructive cardiomyopathy, ischaemic cardiomyopathy, non-obstructive cardiomyopathy, peripartum cardiomyopathy, restrictive cardiomyopathy, stress cardiomyopathy, viral cardiomyopathy.

Myocarditis: allergic myocarditis, autoimmune myocarditis, coxsackie myocarditis, cytomegalovirus myocarditis, eosinophilic myocarditis, lupus myocarditis, myocarditis, myocarditis bacterial, myocarditis infectious, myocarditis mycotic, myocarditis post infection, myocarditis septic, viral myocarditis.

Apixaban: Apixaban, Apixabanum, BMS-562247, BMS 562247-01, BMS-562247-01, Eliquis.

Cerebral Haemorrhage: cerebral hemorrhage, cerebral haemorrhage, cerebral bleeding, intracranial hemorrhage, intracranial haemorrhage, intracranial bleeding, brain hemorrhage, brain haemorrhage, brain bleeding.

Queries Q1. Queries submitted to the data sources considered in the study in the scope of the elaborated case studies.

A) FAERS⁶

Clozapine AND Myocarditis:

(patient.drug.medicinalproduct:Azaleptine+Cloment+Clonex+Clopin+Clopine+Clopsine+Clorazem+Clozapin+Clozapina+Clozapine+Clozapinum+Clozaril+Fazaclo+Lanolept+Lapenax+Leponex+Lodux+Lozapin+Luften+Mezapin+Refract+Refraxol+Sensipin+Sequax+Sizopin+Sizopril+Syclop+Syzipin+Tanyl+Uspen+Versacloz+Zapen+Zapenia+Zapine+Ziproc+Zopin+patient.drug.openfda.brand_name:Azaleptine+Cloment+Clonex+Clopin+Clopine+Clopsine+Clorazem+Clozapin+Clozapina+Clozapine+Clozapinum+Clozaril+Fazaclo+Lanolept+Lapenax+Leponex+Lodux+Lozapin+Luften+Mezapin+Refract+Refraxol+Sensipin+Sequax+Sizopin+Sizopril+Syclop+Syzipin+Tanyl+Uspen+Versacloz+Zapen+Zapenia+Zapine+Ziproc+Zopin+patient.drug.openfda.generic_name:Clozapine+patient.drug.openfda.substance_name:Clozapine)+AND+patient.reaction.reactionmeddrapt:myocarditis

Clozapine AND Cardiomyopathy:

(patient.drug.medicinalproduct:Azaleptine+Cloment+Clonex+Clopin+Clopine+Clopsine+Clorazem+Clozapin+Clozapina+Clozapine+Clozapinum+Clozaril+Fazaclo+Lanolept+Lapenax+Leponex+Lodux+Lozapin+Luften+Mezapin+Refract+Refraxol+Sensipin+Sequax+Sizopin+Sizopril+Syclop+Syzipin+Tanyl+Uspen+Versacloz+Zapen+Zapenia+Zapine+Ziproc+Zopin+patient.drug.openfda.brand_name:Azaleptine+Cloment+Clonex+Clopin+Clopine+Clopsine+Clorazem+Clozapin+Clozapina+Clozapine+Clozapinum+Clozaril+Fazaclo+Lanolept+Lapenax+Leponex+Lodux+Lozapin+Luften+Mezapin+Refract+Refraxol+Sensipin+Sequax+Sizopin+Sizopril+Syclop+Syzipin+Tanyl+Uspen+Versacloz+Zapen+Zapenia+Zapine+Ziproc+Zopin+patient.drug.openfda.generic_name:Clozapine+patient.drug.openfda.substance_name:Clozapine)+AND+(patient.reaction.reactionmeddrapt:cardiomyopathy+myocardiopathy)

Haloperidol AND Myocarditis:

(patient.drug.medicinalproduct:Aloperidin+Bioperidolo+Brotopon+Dozic+Duraperidol+Einalon+Eukystol+Haldol+Haloperidol+Haloperidolum+Halosten+Keselan+Linton+Peluces+Serenace+Sevium+Sigaperidol+patient.drug.openfda.brand_name:Aloperidin+Bioperidolo+Brotopon+Dozic+Duraperidol+Einalon+Eukystol+Haldol+Haloperidol+Haloperidolum+Halosten+Keselan+Linton+Peluces+Serenace+Sevium+Sigaperidol+patient.drug.openfda.generic_name:Haloperidol+patient.drug.openfda.substance_name:Haloperidol)+AND+patient.reaction.reactionmeddrapt:myocarditis

Haloperidol AND Cardiomyopathy:

(patient.drug.medicinalproduct:Aloperidin+Bioperidolo+Brotopon+Dozic+Duraperidol+Einalon+Eukystol+Haldol+Haloperidol+Haloperidolum+Halosten+Keselan+Linton+Peluces+Serenace+Sevium+Sigaperidol+patient.drug.openfda.brand_name:Aloperidin+Bioperidolo+Brotopon+Dozic+Duraperidol+Einalon+Eukystol+Haldol+Haloperidol+Haloperidolum+Halosten+Keselan+Linton+Peluces+Serenace+Sevium+Sigaperidol+patient.drug.openfda.generic_name:Haloperidol+patient.drug.openfda.substance_name:Haloperidol)+AND+(patient.reaction.reactionmeddrapt:cardiomyopathy+myocardiopathy)

⁶ For the syntax of the queries the reader is referred to the relevant openFDA documentation:

<https://open.fda.gov/api/reference/> [Last accessed 20 April 2016].

m+Sigaperidol+patient.drug.openfda.generic_name:Haloperidol+patient.drug.openfda.substance_name:Haloperidol)+AND+(patient.reaction.reactionmeddrapt:cardiomyopathy+myocardiopathy)

Apixaban AND Cerebral Haemorrhage:

(patient.drug.medicinalproduct:Apixaban+Apixabanum+Eliquis)+AND+patient.reaction.reactionmeddrapt:"cerebral+haemorrhage"

B) PubMed⁷

Clozapine AND Myocarditis:

(Azaleptine[tiab] OR Cloment[tiab] OR Clonex[tiab] OR Clopin[tiab] OR Clopine[tiab] OR Clopsine[tiab] OR Clorazem[tiab] OR Clozapin[tiab] OR Clozapina[tiab] OR Clozapine[tiab] OR Clozapinum[tiab] OR Clozaril[tiab] OR Fazaclo[tiab] OR Lanolept[tiab] OR Lapenax[tiab] OR Leponex[tiab] OR Lodux[tiab] OR Lozapin[tiab] OR Luften[tiab] OR Mezapin[tiab] OR Refract[tiab] OR Refraxol[tiab] OR Sensipin[tiab] OR Sequax[tiab] OR Sizopin[tiab] OR Sizopril[tiab] OR Syclop[tiab] OR Syzopin[tiab] OR Tanyl[tiab] OR Uspen[tiab] OR Versacloz[tiab] OR Zapen[tiab] OR Zapenia[tiab] OR Zapine[tiab] OR Ziproc[tiab] OR Zopin[tiab]) AND myocarditis[tiab]

Clozapine AND Cardiomyopathy:

(Azaleptine[tiab] OR Cloment[tiab] OR Clonex[tiab] OR Clopin[tiab] OR Clopine[tiab] OR Clopsine[tiab] OR Clorazem[tiab] OR Clozapin[tiab] OR Clozapina[tiab] OR Clozapine[tiab] OR Clozapinum[tiab] OR Clozaril[tiab] OR Fazaclo[tiab] OR Lanolept[tiab] OR Lapenax[tiab] OR Leponex[tiab] OR Lodux[tiab] OR Lozapin[tiab] OR Luften[tiab] OR Mezapin[tiab] OR Refract[tiab] OR Refraxol[tiab] OR Sensipin[tiab] OR Sequax[tiab] OR Sizopin[tiab] OR Sizopril[tiab] OR Syclop[tiab] OR Syzopin[tiab] OR Tanyl[tiab] OR Uspen[tiab] OR Versacloz[tiab] OR Zapen[tiab] OR Zapenia[tiab] OR Zapine[tiab] OR Ziproc[tiab] OR Zopin[tiab]) AND (cardiomyopathy[tiab] OR myocardiopathy[tiab])

Haloperidol AND Myocarditis:

(Aloperidin[tiab] OR Bioperidolo[tiab] OR Brotopon[tiab] OR Dozic[tiab] OR Duraperidol[tiab] OR "Einalon S"[tiab] OR Eukystol[tiab] OR Haldol[tiab] OR Haloperidol[tiab] OR Halopéridol[tiab] OR Haloperidolum[tiab] OR Halosten[tiab] OR Keselan[tiab] OR Linton[tiab] OR Peluces[tiab] OR Serenace[tiab] OR Sevium[tiab] OR Sigaperidol[tiab]) AND myocarditis[tiab]

Haloperidol AND Cardiomyopathy:

(Aloperidin[tiab] OR Bioperidolo[tiab] OR Brotopon[tiab] OR Dozic[tiab] OR Duraperidol[tiab] OR "Einalon S"[tiab] OR Eukystol[tiab] OR Haldol[tiab] OR Haloperidol[tiab] OR Halopéridol[tiab] OR Haloperidolum[tiab] OR Halosten[tiab] OR Keselan[tiab] OR Linton[tiab] OR Peluces[tiab] OR Serenace[tiab] OR Sevium[tiab] OR Sigaperidol[tiab]) AND (cardiomyopathy[tiab] OR myocardiopathy[tiab])

Apixaban AND Cerebral Haemorrhage:

⁷ For the syntax of the queries the reader is referred to the relevant PubMed documentation:

http://www.ncbi.nlm.nih.gov/books/NBK3827/#pubmedhelp.Search_Field_Descriptions_and [Last accessed 20 April 2016].

(apixaban[tiab] or Apixabanum[tiab] or "BMS-562247"[tiab] or "BMS 562247-01"[tiab] or "BMS-562247-01"[tiab] or Eliquis[tiab]) and ("Cerebral haemorrhage"[tiab] or "Cerebral hemorrhage"[tiab] or "intracranial hemorrhage"[tiab] or "intracranial haemorrhage"[tiab] or "brain hemorrhage"[tiab] or "brain haemorrhage"[tiab] or "brain bleeding"[tiab] or "cerebral bleeding"[tiab] or "intracranial bleeding"[tiab])

C) Twitter⁸

Clozapine AND Myocarditis:

(Clozapine OR Azaleptine OR Cloment OR Clonex OR Clopin OR Clopine OR Clopsine OR Clorazem OR Clozapin OR Clozapina OR Clozapine OR Clozapinum OR Clozaril OR Fazaclo OR Lanolept OR Lapenax OR Leponex OR Lodux OR Lozapin OR Luften OR Mezapin OR Refract OR Refraxol OR Sensipin OR Sequax OR Sizopin OR Sizopril OR Syclop OR Syzopin OR Tanyl OR Uspen OR Versacloz OR Zapen OR Zapenia OR Zapine OR Ziproc OR Zopin)

AND Myocarditis since:2006-03-21

Clozapine AND Cardiomyopathy:

(Clozapine OR Azaleptine OR Cloment OR Clonex OR Clopin OR Clopine OR Clopsine OR Clorazem OR Clozapin OR Clozapina OR Clozapine OR Clozapinum OR Clozaril OR Fazaclo OR Lanolept OR Lapenax OR Leponex OR Lodux OR Lozapin OR Luften OR Mezapin OR Refract OR Refraxol OR Sensipin OR Sequax OR Sizopin OR Sizopril OR Syclop OR Syzopin OR Tanyl OR Uspen OR Versacloz OR Zapen OR Zapenia OR Zapine OR Ziproc OR Zopin)

AND (Cardiomyopathy OR Myocardiopathy) since:2006-03-21

Haloperidol AND Myocarditis:

(Haloperidol OR Aloperidin OR Bioperidolo OR Brotopon OR Dozic OR Duraperidol OR Einalon SCOR Eukystol OR Haldol OR Haloperidol OR Halopéridol OR Haloperidolum OR Halosten OR Keselan OR Linton OR Peluces OR Serenace OR Sevium OR Sigaperidol)

AND Myocarditis since:2006-03-21

Haloperidol AND Cardiomyopathy:

(Haloperidol OR Aloperidin OR Bioperidolo OR Brotopon OR Dozic OR Duraperidol OR Einalon S OR Eukystol OR Haldol OR Haloperidol OR Halopéridol OR Haloperidolum OR Halosten OR Keselan OR Linton OR Peluces OR Serenace OR Sevium OR Sigaperidol)

AND (Cardiomyopathy OR Myocardiopathy) since:2006-03-21

Apixaban AND Cerebral Haemorrhage:

(Apixaban OR Apixabanum OR "BMS-562247" OR "BMS 562247-01" OR "BMS-562247-01" OR Eliquis) AND ("intracranial hemorrhage" OR "intracranial haemorrhage" OR "brain hemorrhage" OR "brain haemorrhage" OR "intracranial bleeding" OR "brain bleeding" OR "cerebral hemorrhage" OR "cerebral haemorrhage") since:2006-03-21

⁸ For the syntax of the queries the reader is referred to the relevant Twitter documentation:

<https://dev.twitter.com/rest/public/search> [Last accessed 20 April 2016].