



# Harnessing social media data for pharmacovigilance: a review of current state of the art, challenges and future directions

Dimitra Pappa<sup>1</sup> · Lampros K. Stergioulas<sup>1</sup>

Received: 29 September 2017 / Accepted: 14 January 2019 / Published online: 12 February 2019  
© The Author(s) 2019

## Abstract

The ever-increasing supply of information combined with the growing knowledge elicitation capabilities of key emerging technologies presents pharmacovigilance with enormous opportunities. Currently, safety monitoring is expanding its evidence base, moving beyond traditional approaches towards sophisticated methods that can identify possible safety signals from multiple information sources, both structured and unstructured. In this context, health information posted online by patients represents a potentially valuable, yet currently left largely unexploited source of post-market safety data that could supplement data from traditional sources of drug safety information. As the use of social media data for pharmacovigilance is still in its infancy, the present paper explores the state of the art in the application of social data to adverse drug reaction detection; provides a thorough review of existing work in the field, highlighting important research efforts and achievements; and finally, discusses the current challenges and promising avenues for future work. Following a literature review methodology, a critical appraisal was conducted of carefully selected work on the use of social data in post-market surveillance, as presented in the recent scientific literature. Out of a sample of more than 1300 articles, which was the result of the literature search, the final selection of articles was made based on their relevance to the applications of social networking sites (SNS) to pharmacovigilance, and a thorough review of this corpus was completed with a total of 100 articles reviewed. The main contributions of this review include the mapping and systematisation of the current knowledge in the field by drawing comparisons of different approaches, types of social data and of relevant sources currently used in the field, and by developing new classifications of social data sources and taxonomies for social data for use in pharmacovigilance, as well as the identification of key challenges and the extraction of new insights in terms of potential for practical applications and future research directions in the area of pharmacovigilance.

## 1 Introduction

Issues related to the safe use of medicines have attracted tremendous attention over recent decades. Pharmaceutical products are used in or on the human body for the prevention, diagnosis or treatment of disease, or for the modification of physiological function [1]. Modern drugs have changed the way in which diseases are managed and controlled. However, adverse reactions to medicines remain a common, yet often preventable, cause of illness, disability and even death

[2,3]. Studies have shown that adverse drug reactions (ADRs) are probably responsible for millions of deaths globally each year (in 2008, 197,000 ADR-related deaths were reported in the EU alone, according to official EC statistics), in both in- and outpatient settings [4]. 6.5% of UK hospital admissions are due to ADRs, and almost 15% of UK patients' experience an ADR during their admission [5]. In France, the estimated annual number of ADR-related hospitalisations was 144,000 in 2007 [6]. A recent study in Spain estimated that the incidence of ADR-related hospitalisations was 7.11%, with fatal ADRs amounting to 1.97% [7]. ADR occurrence in an outpatient setting cannot be fully estimated, as currently such studies are scarce [3]. The overall impact of ADRs is high, accounting for considerable morbidity, mortality, prolonged hospital stays and extra costs. Although many of the suspected drugs have proved benefit, measures need to be taken to reduce the burden of ADRs and therefore further improve the benefit-to-harm ratio of the drugs [8].

---

✉ Dimitra Pappa  
d.pappa@surrey.ac.uk; dimitra@dat.demokritos.gr  
Lampros K. Stergioulas  
l.stergioulas@surrey.ac.uk  
<http://www.surrey.ac.uk/sbs/>

<sup>1</sup> The Surrey Business School, University of Surrey, Guildford, Surrey GU2 7XH, UK

In addition to advances in technological capabilities, today's biggest trend is data. New data are created in novel ways and processed and analysed with the help of new and increasingly intelligent methods. In this context, safety monitoring is expanding its evidence base, moving beyond traditional approaches towards sophisticated methods that can identify possible safety signals from multiple information sources, both structured and unstructured [9]. Health-related information increasingly shared online by patients represents a potentially valuable, yet currently largely unexploited source of post-market safety data that could supplement data from traditional sources of drug safety information.

With the use of social media data for pharmacovigilance being still in its infancy, the present paper aims to explore the state of the art in the field, highlighting important research efforts and achievements, and discussing current research challenges and the way forward. In particular, the thrust of the presented work is to map the state of the art in the application of social data to pharmacovigilance and explore its future potential. Thus, the main aim of this paper is to provide a comprehensive and up-to-date review of existing research in this area, and make significant contributions to the area in terms of generating awareness and systematising the knowledge around social data applications to ADR detection, as well as offering new insights and recommendations for future research and practice in this context. In particular, and in the above context, the objectives (and aimed contributions) of our work are as follows:

1. To conduct a comprehensive literature review of the applications of social media to pharmacovigilance, with critical analysis and comparative assessment of the relevant body of literature
2. To derive a classification of social media sources for use in pharmacovigilance
3. To develop classifications and taxonomies for social data use in pharmacovigilance
4. To derive new insights, key challenges and recommendations for future research and practice

The paper is structured as follows.

The present section (Sect. 1) outlines the plan and scope of our scientific literature review and describes the methodology that the review follows; it also presents a background on pharmacovigilance by reviewing the basic literature in this field, as well as sets out the purpose and rationale of the present study, based on the examined literature; and finally, it summarises preliminary knowledge on social networking sites (SNS), explores the relevance of SNS data to pharmacovigilance and provides definitions and analyses of the basic concepts in this research area.

Section 2 provides an overview of the applications of social data to adverse drug reaction detection and their poten-

tial and presents a new taxonomy of social data sources based on the conducted literature review and a set of key challenges for the future.

Section 3 discusses the identified key challenges, each in separate subsections.

Section 4 draws the conclusions of this research, by summarising its contributions and discussing the gained insights in terms of potential for practical applications and future research directions.

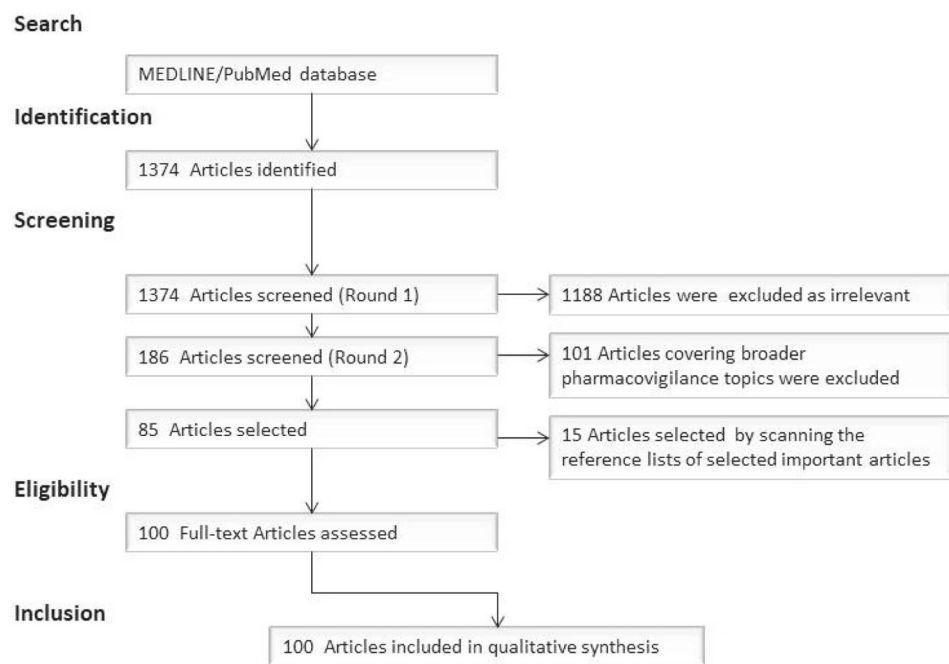
## 1.1 Methodology

In recent years, there has been an increasing number of studies linking social data with ADR detection. To offer a broad overview of this emerging research domain, a review of academic literature was undertaken to examine relevant publications in the MEDLINE/PubMed database. The study methodology applied builds on the PRISMA methodology for systematic reviews. Firstly, the scope of this review was appropriately defined, according to the following specification:

Scope of literature review:

- Literature sources: all corpora included in the MEDLINE/ PubMed database
- Time frame: 2007 early 2018 (covering all eligible literature in last decade)
- Geographic coverage: all inclusive
- Literature selection: the literature search (covering a time window of the last 10 years) used two groups of keywords. The first group included the following terms as approximate synonyms for social data: social media, social networking, forum, Twitter, Facebook, search log and social data. The second group referred to ADR detection and included the terms: adverse drug reaction, side effect and pharmacovigilance. Thus, the literature search query for article selection had the logical form of: Social Data, or equiv. AND Adverse Drug Reaction detection, or equiv.

The initial search resulted in a total of 1374 articles, due to the relatively large set of keywords used. All search results were subsequently scanned, based on the paper title and abstract, to determine whether the respective article should be included in the present literature review. Documents were excluded if they met one or more of the following criteria: (1) irrelevant, (2) not written in English, (3) not a primary or secondary research paper. This first screening resulted in a collection of a total of 186 articles, broadly covering issues related to information technology-enabled post-marketing medicine safety monitoring (traditional methods, consumer behaviour, etc.). Following screening, 101 documents were excluded as they were not directly linked to the specific topic of the present study. As a result of this process, only

**Fig. 1** Flow chart for article review and selection process

85 articles that cover topics relevant to the application of social data to pharmacovigilance were selected for inclusion in the final study corpus. Additional articles were selected, by scanning the reference lists of selected important articles. This screening resulted in a final collection of a total of 100 articles. The flow chart for the article review and selection process is illustrated in Fig. 1.

The most significant insights drawn are outlined in the following sections. Section 2 provides an overview of the application of social data in pharmacovigilance, while Sect. 3 summarises the current challenges and examines the way forward.

## 1.2 Pharmacovigilance background

Post-market surveillance of health and drug products is of paramount importance for the pharmaceutical stakeholders (industry and regulators), since many adverse events are not captured in randomised clinical trials (RCTs) and previously undetected adverse reactions may occur as the drug is exposed to patients and situations not controlled for during the clinical trial. The practice of monitoring the safety of medicines is commonly referred to as **pharmacovigilance**, the origins of which can be traced back to the case of thalidomide [10,11], which highlighted the importance of drug safety and prompted the start of systematic approaches to monitor the safety of marketed medications [1]. Pharmacovigilance is defined by the World Health Organization (WHO) [1] as the science and activities relating to the detection, assessment, understanding and prevention of adverse effects, particularly long-term and short-term side

effect, of medicines. The practice of pharmacovigilance is sometimes called post-market (or post-market) surveillance or post-authorisation monitoring. Within the scope of pharmacovigilance fall the detection, assessment, understanding and prevention of adverse effects or any other possible medication-related problems of herbal, traditional and complementary medicines, blood products, biologicals, medical devices and vaccines.

The term **adverse event** (AE) is used to refer to any untoward medical occurrence that may appear during treatment with a pharmaceutical product but which does not necessarily have a causal relationship with the treatment [12,13]. The WHO [1] describes **adverse drug reaction** (ADR) as a response to a drug which is noxious and unintended and which occurs at doses normally used in man for the prophylaxis, diagnosis or therapy of disease, or for the modification of physiological function, a definition that denotes the existence of a causal relationship between the drug therapy and the observed adverse event. The timely signalling of adverse drug effects is required, in order to promote the safety and quality of drug therapies.

According to the WHO [14], the major tasks of pharmacovigilance are:

- Early detection of unknown adverse reactions and interactions;
- Detection of increases in frequency of known adverse reactions;
- Identification of risk factors and possible mechanisms;
- Estimation of quantitative aspects of benefit/risk analysis and dissemination of information needed to improve medicine prescribing and regulation.

Post-market safety surveillance relies mostly on data from spontaneous reports of adverse events, medical literature and observational databases. Limitations of these data sources include potential under-reporting, lack of geographic diversity, possibility of patients' perspectives being filtered through healthcare professionals and regulatory agencies, and time difference between event occurrence and discovery [15,62]. The need to enhance patient safety calls for a proactive approach to pharmacovigilance, in order to improve patient care and safety in relation to the use of medicines.

Currently, the rapidly increasing supply of information combined with the growing knowledge elicitation capabilities of trending and emerging technologies present epidemiology, pharmacoepidemiology and pharmacovigilance with enormous opportunities [16–18]. Recently, the evidence base of safety monitoring has been expanding, moving beyond traditional approaches towards sophisticated methods that can identify potential safety signals from multiple information sources, both structured and unstructured. This refers to the exploitation of secondary data, i.e. data made available and/or collected for other purposes, namely electronic health records (EHR), social media data, etc. [16,19,155,156]. In the broad context of pharmacotherapy, patient perspectives have always been an essential component of medicine safety monitoring. As efforts directed to patient-centric drug development are intensifying, it becomes increasingly important to incorporate the patients' voice in the pharmacovigilance systems and processes. The Internet has changed our relationship with health care, as people are increasingly sharing online their healthcare experiences [20]. Health stakeholders are adapting to this trend, together with the involved sections of the computer science research community. Mining social media for the extraction of health-related information emerges as a hot topic, particularly in certain areas of health care, for example with regard to health concerns like mental illness [163]. De Choudhury and De [21] studied mental illness communities on Reddit (<http://www.reddit.com/>), a forum-like platform hosting numerous virtual, text-based support groups in which patients openly discuss a variety of concerns related to their condition, benefiting from the dissociative anonymity of the environment.

Presently, there is a growing interest by drug safety stakeholders (pharmaceutical companies and regulators) in exploring the use of social media (**social listening**) to supplement established approaches for pharmacovigilance, by harvesting information on patients' experiences after exposure to pharmaceutical products. Health information posted online by patients is in abundance and is often publicly available, thus representing an untapped source of post-market safety data that could supplement data from existing sources of medicine safety information [22,156]. For example, the study by Gage-Bouchard et al. [23] on cancer information exchanged on personal Facebook Pages revealed that this

information predominantly related to treatment protocols and health services use (35%), followed by information related to side effects and late effects (26%) and medication (16%).

### 1.3 Social networking sites and their relevance to pharmacovigilance

**Social networking sites** (SNS) and applications allow for the exchange of user-generated content whereby people talk/communicate, share information, network and participate in community activities. Boyd and Ellison [24] describe SNS as a web-based service that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection and (3) view and traverse their list of connections and those made by others within the system. The nature and nomenclature of these connections may vary from site to site.

More and more individuals are making use of SNS to communicate and stay in contact with family and friends, to engage in professional networking or to connect around shared interests and ideas [25]. There currently exists a rich and diverse ecology of SNS, which vary in terms of their scope and functionality, and include: general-purpose and specialised community sites (e.g. Facebook and LinkedIn); media sharing sites (e.g. YouTube and Flickr); weblogs (blogs); micro-blogging sites (e.g. Twitter); and question/answer discussion forums, which have continued to be around for decades with undiminished popularity despite relentless Internet evolution. Social media user base has undergone a nearly tenfold increase in the past decade: 65% of adults now use social networking sites [26]. Since 2005, SNS have experienced significant growth in active users, with Facebook and LinkedIn in particular being among the fastest growers. Micro-blogging services such as Twitter and Tumblr are also on a growing trajectory. As a result, social media is creating real-world data at an unprecedented rate, with people using social media to discuss their everyday lives, including their health and their illnesses. The motivation to connect and learn about one another has given rise to niche SNS. Recent years have seen the emergence and proliferation of **SNS dedicated to healthcare communities** (usually consisting of health professionals and/or consumers/patients), which have become particularly popular among patients, with the most common intended use being **self-care** [27], i.e. social media serving as a platform that allows patients to exchange information about their health condition with others who are battling with the same health issues, and receive peer-to-peer support (**online patient communities**) [28,29]. Social support is deemed extremely beneficial in combating health concerns like depression and mental illness [21,33]. Such networks can be classified mainly in terms of two categories:

(A) **Generic SNS**—this category can include:

- **Big public platform SNS**, such as *Facebook*, *Twitter*, *Flicker* and *Tumblr*, which host a plethora of health-related communities/groups, and also contain big volumes of posts by individual users related to health issues.

(B) **Specialised healthcare social networks and forums**—this category can include:

- **Generic health-centred SNS** (generic networking sites on general health topics and disease support, usually requiring user profiles), such as *PatientsLikeMe* ([/www.patientslikeme.com](http://www.patientslikeme.com)), *DailyStrength* ([www.dailystrength.org](http://www.dailystrength.org)), *MedHelp* ([www.medhelp.org](http://www.medhelp.org)), *WebMD* (<https://exchanges.webmd.com/>), and *CureTogether* (<http://curetogether.com/>), where users discuss their health-related experiences, including use of prescription drugs, side effects and treatments,
- **Medicine-focused sharing platforms** (patient forums), like *Ask a Patient* (<http://www.askapatient.com>) and *Medications.com* (<http://www.medications.com/>), which allow patients to share and compare medication experiences. **Disease-specific online health forums** focused on specific diseases, e.g. the *TalkStroke forum* (<https://www.stroke.org.uk/forum>) for stroke survivors and caregivers hosted by UK's Stroke Association [30], Australia's *ReachOut.com* (<https://au.reachout.com/forums>) forum for mental health support, etc.

In the above SNS, users tend to share their views with others facing similar problems/conditions or health outcomes, and this makes such social networks unique and robust sources of information about drugs, health effects and treatments, which can significantly augment the evidence base of research studies and provide additional insight on the needs of specific populations [30]. *CureTogether* specifically promotes patient-driven research, by establishing research partnerships with universities, research organisations and self-experimenters; SNS can further promote medication adherence, enhance the effectiveness of therapies and contribute to secondary prevention against recurrence of disease [31] and chronic pain management [32]. In the case of mental illness, SNS can serve for the identification of signals of mental disorders and users at risk of self-harm [33]. The rapidly growing popularity of such networks and the abundance of data available through them have recently enabled new research on public health monitoring, including ADR monitoring and formal clinical trial procedures [34,35]. Cohort discovery and metadata platforms have also emerged (e.g. the *Dementias Platform UK*, <https://www.dementiasplatform.uk>). The term social data refers to data derived from social

networks. The context of disclosure might be a conversation with friends or in dedicated groups on Facebook or Twitter (where dedicated discussion threads have emerged via hashtags, e.g. #LCSM denoting lung cancer social media posts), or discussions with other patients on online social networks like *PatientsLikeMe*, where they will share diagnoses, treatments, coping mechanisms and outcomes with one another [29,36]. Researchers have considered a range of motivations for disclosure in social network sites. The fact that users publish in SNS a considerable amount of information which is not otherwise available makes social data one of the most important potential sources of knowledge in the pharmacovigilance field. Various new types of diverse data are disclosed via SNS. As noted by the OECD [37], personal data are collected online in different, arguably complementary, ways: (i) data can be voluntarily and explicitly shared by a consumer (e.g. when subscribing to a social network); (ii) data can be observed or recorded (e.g. through cookies monitoring access to a website), with or without consumers knowledge, or explicit consent; (iii) data that can be inferred, also by mixing several sources of data that are, by themselves, anonymous.

Based on the works of Van Alsenoy [38] and Schneider [39], social (networking) data can be categorised according to the context and purposes of data disclosure as:

1. **Service data:** data that users need to give to an SNS in order to use it. This might include the persons legal name, age, credit card number, etc.
2. **Disclosed data:** data that are posted by SNS users on their own pages (e.g. blog entries, photographs, videos, messages, comments, etc.).
3. **Entrusted data:** data that are posted by SNS users on the profile pages of other SNS users (e.g. a wall post, comment). Similarly to disclosed data, entrusted data appear on the users own pages, but they do not have control over the data—someone else does.
4. **Incidental data:** data about an SNS user which has been uploaded by another SNS user (e.g. a picture). Similarly to disclosed and entrusted data, incidental data appear on the users own pages, but they do not have control over it, and they did not create it in the first place.
5. **Derived data:** data that are inferred from (other) SNS data (e.g. membership of group X implies attribute Y).
6. **Behavioural data:** data regarding the activities of SNS users within the SNS (e.g. user habits, who they interact with and how).

A further classification can also be made between these two types of data:

- **Collected (back-office) data:** Data collected by the service provider, which usually include profile and network data explicitly provided by the user, and click history

implicitly provided. Users can assume that everything they do and upload in the browser tab of the service is collected, if the privacy policy of the service does not state it otherwise.

- **Front-end Data:** Data that are knowledgeably shared. This includes: (i) **Public/disclosed data.** Data that are published openly, such as complete name or e-mail. It can be useful for users trying to contact other users. (ii) **Social data.** Data that are openly shared with the users trusted contacts. Unless these contacts are inside their circle of trust, they cannot access it.

Inevitably and quite understandably, SNS data have come to attract the interest of a wide range of actors. SNS allow for the collection of a large amount of information from different sources. Datasets can be continuously monitored in order to identify the emerging trends in the flows of data. This capability is revolutionary and differs from the traditional sampling method, which is based on the extraction of a representative sample from the total statistical population.

Aligned with this trend, the social media participation model for health organisations is evolving along three dimensions: **listening, participating** and **engaging** [40]. With its recognised use as a source of information [41], social media has been used by healthcare stakeholders, in order to distribute information about diseases and their treatment, medicines and announcements [42]. Currently, the potential of SNS as a source of insight is increasingly recognised. Social media mining is becoming an integral part of public health monitoring and surveillance [43], assisted by advances in automated data processing, machine learning and natural language processing (NLP) technologies. Applications include epidemiological investigations, e.g. mining Twitter data for disease topic detection and surveillance [15,44,45,159], and tracking the spread of infectious diseases. However, scholars note that the analysis of health social media content requires further innovations [46]. Social media presents new channels and methods that can enable pharmacovigilance to move away from traditional safety reporting methods towards more patient-centric models for reporting, analysing and monitoring of safety data. Essentially, in terms of safety research, social platforms allow for both **information pull** (social media listening) and **targeted investigations** (direct-to-patient research). Social media activities for pharmacovigilance by pharmaceutical companies fall into three broad categories: **listening** (safety data reporting), **engaging** (follow-up) and **broadcasting** (risk communication)—each with varying degrees of complexity, associated issues and requirements [158,161,162].

Traditionally, pharmacovigilance has mainly relied on post-marketing spontaneous reporting systems (SRSs) [47], such as the EudraVigilance system operated by the European Medicines Agency (EMA) and the Adverse Event

Reporting System (AERS) used by the US Food and Drug Administration (FDA), collecting voluntary reports produced by healthcare professionals, marketing authorisation holders (MAHs) and consumers. However, the reporting rate of such systems is low, causing delays in the detection of ADRs. Basch [48] notes that many adverse reactions are missed due to lack of interest, willingness, availability or awareness of stakeholders to report. Strom [49] considers the systems under-ascertainment (not recognising an event is due to a drug), overascertainment (erroneously ascribing an adverse event to a drug) and under-reporting as its major flaws.

Patient reports have been shown to be of high importance for pharmacovigilance, providing a complementary [50] and independent perspective from those of health professionals [51]. According to Santos [52], the combination of reports from healthcare professionals with first-hand information from patients is of great added value because it increases the chances to identify new safety issues.

With the increasing use of social media and social networks, social data are increasingly recognised as a valid source of real-time information on drug-related adverse events [53], including the assessment of the behaviour and risk perception of consumers [54]. In recent years, many scholars have investigated the availability of adverse event information in social media and appropriate technologies and methods to extract it. Statistical analysis shows that there is value in pharmaceutical companies and regulatory authorities taking a more proactive approach to social media monitoring [55]. **Proactive monitoring** could provide early warning of new adverse events or clinical information that helps guide drug development and avoid preventable litigation. Regulators and pharmaceutical companies are also starting to monitor social media posts for potential ADR signals. While previously much of the interest in social media has been on the marketing front, their potential application to improve drug safety and pharmacovigilance is also recognised, particularly for the identification of signals of unknown ADRs and unknown drug–drug interactions (DDIs) of concomitant medications, which are often linked to unexpected ADRs [56]. As concluded by Sloane et al. [57], the quantity and near-instantaneous nature of social media provide potential opportunities for real-time monitoring of ADRs, greater capture of ADR reports and expedited signal detection if utilised correctly. According to Liu and Chen [71], an advantage of patient social media is that they cover a large and diverse population and contain millions of unsolicited and uncensored discussions about medications. Furthermore, patient reports of adverse events in social media are more sensitive to underlying changes in the patients' physiological state than clinical and spontaneous reports. Today, social media are a formal part of the potential sources of data of interest to pharmacovigilance. Most of the regulatory guidance and hence pharmacovigilance activities involving social

media and Internet are primarily focused on screening of social media sites and follow-up of reported safety data.

## 2 An overview of applications of social data in pharmacovigilance

Nowadays, the pharmaceutical industry is engaging more actively with patients on social media for the collection of direct-from-patient information. Pharmacovigilance is increasingly drawing upon different types of data sources in both solicited and unsolicited ways. Figure 2 depicts the principal sources of social data employed presently for ADR detection—these social data sources are the focal points of the present review. The proposed taxonomy is derived from our literature review, by analysing the existing work on social data applications to ADR detection, in terms of category of SNS and nature of reporting. The *specialised healthcare SNS* category can be further classified into three types of specialised healthcare SNS/forums: *health-centred SNS* (generic networking sites on general health topics, usually requiring user profiles), *disease-specific online health forums/SNS* (focused on specific diseases) and *medicine-focused sharing platforms* (or patient forums), as explained in Sect. 1.3.

Two types of reporting can be distinguished—namely solicited and unsolicited—which can be further analysed in terms of the context and purposes of data disclosure and the area where data are captured according to the classifications proposed in Sect. 1.3.

### Type 1: solicited reporting (use of social media as a reporting channel)

Previous research overwhelmingly suggests that further promotion of patient reporting to the SRSs is justified. Direct patient reporting of suspected ADRs has the potential to add value to pharmacovigilance [51,165]. A study by Avery et al. [59] concluded that patient reporting can (a) contribute types of drugs and reactions different from those reported by healthcare professionals, thus generating new potential signals, and (b) describe suspected ADRs in more detail, thus providing useful information on likely causality and impact on patients' lives.

The use of new technology can provide new methods and tools to facilitate direct patient reporting of ADR. In this direction, the WEB-RADR project [60] promotes the utilisation of social media and other technologies for ADR reporting in a convenient, quick and efficient way, also seeking to establish guidelines and a regulatory framework on the use of the technology for such reporting.

### Type 2: unsolicited reporting (social media monitoring)

Social media data are increasingly recognised as a valid source of patient perspectives and data on adverse events

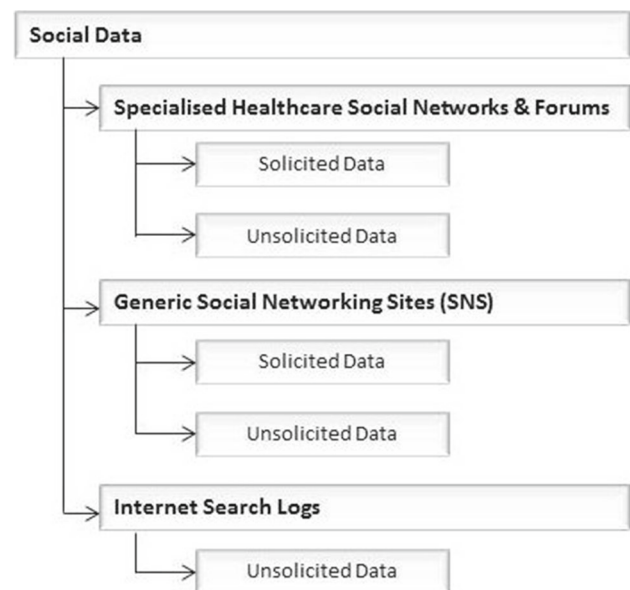


Fig. 2 Taxonomy of social data sources for pharmacovigilance

(AEs) in pharmacovigilance [61]. This information is in abundance and is timely, relevant and often publicly available. Social media have thus the potential to become a new-age tool for monitoring data regarding patients' experience with medications in real time, making and providing early indications of potential safety issues that require further investigation. A typical methodology for signal detection using (in this case, passively collected) social media data includes the following steps, as proposed by Powell et al. [62]:

- Data harvesting: collection of raw data;
- Translation: standardisation of drug names and vernacular symptom/event descriptions;
- Filtering: identification of relevant informative posts and data cleaning (removal of duplicates and noise);
- De-identification: removal of personally identifying information;
- Supplementation: addition of other data sources to facilitate the review process and contextualise the results, in order to assist interpretation (e.g. product label, sales data).

By mining the relationships between drugs and ADRs from data reported by online users on health-related issues, we can speed up the process of detecting and confirming ADRs. This is particularly important in the case of new treatments for rare diseases, which are typically tested only on small patient groups [63]. Social data mining, however, may have limited application to orphan drugs [64].

We shall now further examine in more detail the three most promising sources of social data of use to pharmacovigilance as identified above.

## 2.1 Specialised healthcare social networks and forums

An important source of user-generated content on the Internet is **specialised healthcare social networks and forums**. These platforms allow for the collection of health-related data focused on either a specific disease or multiple disease areas. User **comments** in health-related social networks contain extractable information relevant to pharmacovigilance. Research efforts focused on both general health discussion forums [65–69] and disease-specific discussion forums [46,70,71] have demonstrated that it is possible to extract complex medical concepts, with relatively high performance, from informal, user-generated content.

Recognising the potential of health-related social networks and forums, the FDA has engaged PatientsLikeMe in a research partnership to generate more AE data. PatientsLikeMe claims to have collected more than 110,000 adverse event reports on 1000 different medications, data that the FDA will now be able to access and analyse in addition to its existing sources of information [72].

## 2.2 Generic SNS

General-purpose social platforms (e.g. Twitter, Facebook) are also of value. In a recent article, Pierce et al. [73] concluded that an efficient semi-automated approach to social media monitoring may provide earlier insights into certain adverse events, particularly since patient reports of adverse events in social media are more sensitive to underlying changes in patients' physiological state than traditional spontaneous reports. However, the affordances of SNS vary. SNS privacy policies may hinder the availability of user-generated data for data mining purposes. For example, special dispensation is required for the use of data from Facebook, which is the worlds' largest social network [73]. Twitter content remains publicly available; nonetheless, the noisiness of Twitter data (short sentences, fragmented sentences, use of abbreviations, misspelling errors) can significantly impact the performance of classification methods [74]. Twitter could potentially be an important source of otherwise unreported adverse drug events, but the extracted data are noisy and hard to process [74–76].

Compared to the specialised healthcare social media, generic SNS contain larger volumes of data. The specialised healthcare SNS, however, contain higher proportions of relevant data [55].

Furthermore, the trustworthiness of social data from generic SNS is questioned, since data quality control is lacking and data authenticity cannot be verified. This implies that, while social media mining can reveal early signs of potential ADRs, the information is not sufficient for the proper processing of the identified suspected cases and for the estab-

lishment of causality (signal verification). Domain experts of regulatory authorities still need to employ other instruments in order to assess potential drug safety risks. In this light, Freifeld et al. [77] concluded that while patients reporting AEs on Twitter showed a range of sophistication when describing their experience, the wholesale import of individual social media posts into post-marketing safety databases would not be advisable. Rather, in parallel with other post-marketing sources, such data should be considered for idea generation, and reasonable hypotheses followed up with formal epidemiologic studies. Additional work is needed to improve data acquisition and automation.

## 2.3 Search logs

Search for health information on the web is growing, with an increasing number of people considering the Internet as an important source of knowledge [78]. Anker itet al. [78] note that **health information seeking** represents a purposeful and goal-oriented activity that according to Niederdeppe et al [79] describes active efforts to obtain specific information in response to a relevant event, outside of the normal patterns of exposure to mediated and interpersonal sources that constitute mere information scanning. Search query volume can provide a measure of pharmaceutical utilization in the community [168]. In this light, Bragazzi and Siri [80] demonstrated that online searches for antidepressants reflect the usage pattern recorded and monitored by the Italian Drug Agency.

Tapping on back-office social data, several scholars have demonstrated how **search logs** can be used to detect new ADRs. White et al. [81] demonstrated that anonymised signals on drug interactions can be mined from search logs, using a 2011-reported adverse event (hyperglycaemia) due to a previously unknown interaction between the drugs paroxetine, an antidepressant, and pravastatin, a cholesterol-lowering drug. By **mining search queries** on Google, Bing and Yahoo Search from 2010, White et al. [81] found that people who searched for both drugs were also more likely to search for terms related to the adverse event than those who searched for only one of the drugs. Search information was provided to the researchers anonymously by users who agreed to share their search history. The study was carried out after the ADR had been identified, and using this approach for the identification/detection of unknown ADRs will remain a challenge. Similarly, Chokor et al. [82] mined a variety of Internet data sources and search engines (mainly Google Trends and Google Correlate) for information on reactions associated with the use of two popular major depressive disorder (MDD) drugs: duloxetine and venlafaxine. Yom-Tov and Garilovich [83] noted that web search queries are potentially more suitable for the detection of less acute, later-onset



drug reactions, as acute early-onset ones are more likely to be reported to regulatory agencies.

Sarker et al. [55] reviewed studies describing approaches for ADR detection from social media from the MEDLINE, Embase, Scopus and Web of Science databases, and the Google Scholar search engine. Their review suggests that in terms of sources, both health-related and general social media data have been used for ADR detection, and while health-related sources tend to contain higher proportions of relevant data, the volume of data from general social media websites is significantly higher. Norn [84] further expressed the hypothesis that the study of search logs could help pharmacovigilance systems gain insight on a population segment who would hesitate to share their adverse drug experiences on social media and who would prefer to use search for relevant information on the Internet. In view of this promising outlook, the FDA examined the potential use of search to identify adverse drug reactions in collaboration with Google [85].

The literature review findings were further analysed to identify advantages and limitations of the use of social data in pharmacovigilance and to conduct a comparative assessment of social data sources with respect to a number of key attributes of social big data, including population coverage, usefulness, timeliness, accessibility, quality and processability. These attributes serve as the comparison criteria. Table 1 outlines the advantages and limitations of social media in the context of pharmacovigilance, across the above set of big data attributes, while Table 2 presents the qualitative comparison of the different social data sources in terms of the aforementioned six criteria.

### 3 Current challenges and the way forward

Pharmacovigilance calls for **meaningful, usable and reliable information** [84], i.e. information that is relevant, timely, consistent and complete (so as to support ADR detection and assessment) is available and can be extracted and processed, and is accurate and trustworthy. Experimentation and evidence on the ability of social media to enhance pharmacovigilance methods and processes continue to grow and despite the formal acknowledgement of their potential value by pharmacovigilance managing authorities, many challenges exist at different levels and, as a result, social data remain a largely untapped source of knowledge. There are challenges inherent to big data, namely the high volume of data, the different formats in which data (both structured and unstructured) are captured, as well as other intrinsic, structural and regulatory barriers that impede access and analysis. Risks associated with big data include: extraction of useless data (i.e. extraction of data that do not fit the purpose), extraction of data from sources of unknown quality, or without authorisation, etc. The ADR-PRISM project [86] identified

five major challenges to the operationalisation of social data for pharmacovigilance, namely: (1) variable quality of information on social media, (2) guarantee of data privacy, (3) response to pharmacovigilance expert expectations, (4) identification of relevant information within web pages and (5) robust and evolutionary architecture.

The use of social media for pharmacovigilance represents a knowledge discovery process. The present study examines its challenges in three dimensions:

- Conceptual: Challenges that relate to the purpose and value of social media use in pharmacovigilance (value/utility of social media as knowledge sources for pharmacovigilance),
- Technical: Challenges that relate to the feasibility of the process (information extraction from social media and analysis of social data) and
- Environmental: Challenges that relate to compliance concerns and affect the acceptability of any new pharmacovigilance process proposed (data privacy and regulatory framework).

The major challenges (summarised in Table 3) identified from our literature review are examined in the following subsections, where each challenge is described and critically analysed to produce insights and future directions related to each challenge.

#### 3.1 Value/utility of social media as knowledge sources for pharmacovigilance

Challenges are not solely technical, since the sharing and using the public health data are also conditioned by motivational, economic, political, legal and ethical barriers [87]. Typically, a valid AE that is reportable to a regulatory agency must meet the following four criteria: identifiable reporter, identifiable patient, suspect drug and adverse event. In addition, clinical, pathological and epidemiological information relating to adverse reactions is necessary for a full understanding of the nature of an adverse reaction [13]. Most social media sources fail to provide complete information for case assessment [88].

The credibility of data varies across social media. Concerns relate to the quality, trustworthiness and integrity of information collected from social channels, the volatility and the overall uncertainty of social data.

Social media generate patient-centric data, which is typically **unfiltered** and **unchecked**, and **can use the incorrect terms**, or refer to diagnoses that are based on Internet research rather than confirmed diagnoses from healthcare professionals (risk of misinformation). Social media disclosures are also conditioned by the disclosers gender and cultural background, resulting in differences in linguistic

**Table 1** Advantages and limitations of social data in pharmacovigilance

Attribute	Findings: advantages and limitations	Relevant references
Population coverage	Advantages	[15,25,27–29,32,34,55,90]
	Limitations	[53,57,88]
Usefulness	Advantages	[22,23,25,27,28,30–32,40,41,46,48,49,51,55,56,60,61,64–69,71,73,75–82,84,91,93–95,135,136,149]
	Limitations	[21,57,63,87,89,90]
Timeliness	Advantages	[75–77,93]
	Limitations	[75–77,93]
Accessibility	Advantages	[31,75–77]
	Limitations	[55,57,62,86,87]
Quality	Advantages	[55,57,62,86,87]
	Limitations	[53,63,86,133,135,167]

**Table 1** continued

Attribute	Findings: advantages and limitations	Relevant references
Processability	Advantages	[47,57,60,62,65–67,69–71,73–77,82,83,86,93–102,105,108,112,113,131,136,139–141]
	Advances in big data analytics Availability of advanced NLP and machine learning techniques Advances in the data processing capabilities of machines Trends analysis techniques and tools	
Limitations	Big data analysis issues: e.g. high volume of data	[15,55,90,107]
	Lack of standards and uniformity across different SNS	
	Other structural barriers that impede access and analysis	
	Barriers in analysing data across languages (lack of relevant lexical resources)	

**Table 2** Comparison of social data sources for pharmacovigilance

Attribute	Specialised healthcare social networks and forums	Generic SNS	Search logs
Population coverage	Drug user population and specific cohorts	Large and diverse populations and geographic areas	Large and diverse populations and geographic areas
Data usefulness	Smaller volumes of data but greater salience of drug-specific information. More consistent and complete information regarding demographics, exposure, outcome and possible contextual factors	Incomplete information. Large volumes of data. High noise-to-data ratio	Search frequency as early indication of potential drug–symptom relationship
Data timeliness	Near real time	Near real time	Near real time
Data accessibility	Research partnerships	Open social media and/or sources with conditioned access; privacy concerns	Available
Data quality	Adequate. Availability of structured data possible. Often lengthier postings with greater potential for analysis	Important quality concerns. Lack of quality control. Unstructured data	Adequate for the detection of events and for relating these events to drugs. Inadequate for more rigorous analysis of the ADR, only serves as indicative. Possibility of false positives
Data processability	Text and Data mining, NLP, machine learning methods	Text mining, NLP, machine learning methods	Search analytics

**Table 3** Challenges to the operationalisation of social data for pharmacovigilance

n/n	Challenges	Type (dimension)	Solution timescale (estimated)
1	Value/utility of social media	Conceptual	Short term
2	Information extraction from social media	Technical	Long term
3	Analysis of social data	Technical	Long term
4	Data privacy	Environmental	Short term
5	Regulatory framework	Environmental	Short term

expression, inhibition levels and tendency for social interaction [89], and the risk of bias (in terms of age, gender, ethnicity and physical location) [57]. Furthermore, it is not clear whether patients and clinicians report the same concerns. A study by Topaz et al. [90] compared patients' concerns in social media (*social media data*) with clinicians

reports in electronic health records, discovering significant correlations. However, researchers have identified a different emphasis on the type of adverse events reported on social media compared to SRSs, which suggested that social media may be a better source for symptom-related or less serious (non-life-threatening or not requiring hospitalisation) than

laboratory test abnormalities and serious adverse events. A systematic review conducted by Golder et al. [63], investigating the prevalence, frequency and comparative value of social media-derived information on the adverse events of healthcare interventions, concluded that, although reports of adverse events are identifiable within social media, there is considerable heterogeneity in the frequency and type of events reported, and the reliability or validity of the data is not thoroughly evaluated. Adverse events identified via social media, but not documented elsewhere, also tended to be mild or related to quality of life. In contrast, under-represented adverse events on social media tended to include laboratory abnormalities or effects requiring diagnosis from a healthcare professional. Serious or severe adverse events were also under-represented in social media [91,157].

Hence, pharmacovigilance practice could benefit from empowered, motivated and knowledgeable patients, through increased public awareness of pharmacovigilance, through the closer engagement of pharmacovigilance stakeholders (regulatory authorities, industry) with patients, through the diffusion of scientific knowledge regarding health, illness, therapy and medicines, and through the further strengthening of health support networks and communities.

### 3.2 Information extraction from social media

Social media produce **large amounts of raw data that are challenging to analyse**. Limitations inherent to big data include difficulties in search, the large volume of irrelevant data, issues surrounding the lack of validation, user bias, etc. They are characterised by big volumes of data of high variety and a high frequency of new data generation (Data Velocity).

Pierce et al. [73] stressed the inherent variability across data sources that can change rapidly over time. Norn [84] noted that the different sources of Internet-based data vary in terms of scope and coverage, as well as with regard to the richness of the provided information. This may include limitations caused by website characteristics (e.g. character limitation) [73]. As SNS vary in terms of their core functional building blocks (identity, conversations, sharing, presence, relationships, reputation and groups), a profound understanding of their characteristics is required in order to develop strategies for monitoring, understanding and responding to different social media activities [92,164]. For pharmacovigilance, this signifies that the degree of uncertainty and bias of each SNS needs to be taken into consideration. A careful combination of each of these data sources is required to fully realise their benefits and generate valuable signals [57]]. In this process, social media sources need to be considered separately and methods need to be tailored to each channel individually, as each one carries its own separate challenges. The risk of duplicate reports (e.g. caused by parallel posting on multiple platforms) also exists.

The principal task of signal extraction from social media consists in the identification of drugs and symptoms (entity recognition) and of the relationship between them.

Several challenges exist in extracting ADR information from social media, with the principal ones related to the **named entity recognition** (NER) problem (i.e. the fact that both drug names and reaction terms can be described in a variety of ways) and the **relation detection** problem. According to Sloane et al. [57], technical challenges inherently related to signal extraction from SNS include:

1. Drugs may be described by their brand names, active ingredients, colloquialisms or generic drug terms (e.g. antibiotic);
2. ADRs may be referred to using creative idiomatic expressions or terms not found within existing medical lexicons;
3. The informal nature of social media results in a prevalence of poor grammar, spelling mistakes, abbreviations and slang;
4. The existence of a side effect may be clear, while the specific side effect experienced remains unclear;
5. Discussion of a drug could involve indications, beneficial effects or concerns of an adverse event;
6. Only a small percentage of social media will relate to ADRs.

Scholars have made significant progress on topics revolving around the recognition of drug names, symptoms and ADRs in social media texts using automated or semi-automated methods [93]. Research efforts for the development of appropriate text mining methods and natural language processing (NLP) techniques are ongoing.

One of the principal challenges is the extraction medical entities from noisy patient-generated content. Given the large volume of social media posts, efforts towards the automatic text classification for ADR detection are receiving growing attention [70,94–101]. However, lexicon-based approaches [47] for medical entity recognition and tools like MetaMap [102], developed by the US National Library of Medicine to identify medical concepts into the concept codes from the Unified Medical Language System Metathesaurus (UMLS), are not sufficient, given the informal, colloquial nature of discussions and the non-adherence to standardised terminology used by participants [103]. Spelling variants and machine learning techniques are used to detect drug names and symptoms [75,104]. Text classification approaches for ADR detection have been created [95]. Given the limited amount of annotated data available publicly [55], comparable sets of documents for algorithm training are being developed to further assist research efforts (e.g. TwitMed) [105,106]. Segura-Bedmar et al. [107] stressed that the task of extracting relations between drugs and their effects from social media in languages other than English can also be hindered by the

lack of relevant lexical resources. User-expressed medical concepts are often non-technical, descriptive and challenging to extract. The use of sentiment analysis has also been proposed as a means to improve the performance of ADR identification methods [108–111]. Human curation is also investigated for the purpose of establishing a gold standard by which future, automated classification methods may eventually be compared [112]. The need for manual data labelling is expected to drop considerably with the application of neural network-based tools [113,114]. Abdellaoui et al. [98] apply distance-based filtering in order to distinguish between false positives and true ADR declarations. The framework proposed by Liu and Chen [71] employs a hybrid approach combining statistical machine learning methods and rule-based filtering with information from medical knowledge bases, and report source classification to reduce noise. Advanced machine learning-based NLP techniques, also promoted by Nikfarjam et al. [115], allow for the detection of colloquial expressions of ADRs via sequence labelling approaches, such as CRFs or RNNs [116]. Furthermore, multidimensional analysis of ADRs is required to allow for the discovery of associations between drugs and symptoms [117]. Chowdhury et al. [103] proposed a multitask framework based on the sequence learning model with improved learning efficiency and prediction accuracy for ADR and symptom identification.

A scoping review by Lardon et al. [118] noted that, while there is a multitude of methods for identifying target data, the processes of extracting data and evaluating the quality of medical information from social media are not easily scalable and studies usually failed to accurately assess the completeness, quality and reliability of the data that were analysed from social media. While experimental methods have proved advantageous in identifying previously unknown adverse drug reactions, constant active screening of social media is challenging. The knowledge extraction process is effort-intensive and further scrutiny of Proto-AEs (i.e. posts with a resemblance to an adverse event) by medical regulatory authorities and competent medical professionals is often required in order to extract valid ADR signals. The enormous number and diversity of conversations that can take place in a social media setting mean that there are format and protocol implications for stakeholders seeking to make sense and extract knowledge from them. It is crucial that the processes of managing and interpreting this new information are efficient and effective for sustenance, thoughtful use of resources and valuable return of knowledge [60]. Challenges span the entire knowledge discovery value chain from data extraction to data processing and sense-making, including: collection (selection of data types and data sources, as well as validation of the collected dataset), processing (application of relevant knowledge extraction and processing methods, technologies and tools) and analysis (expert analysis for sense-making).

### 3.3 Analysis of social data

Analysis of social data for medicine safety surveillance is a challenging task. Challenges exist across all phases of the process: signal detection, development of a causality hypothesis and testing of the causality hypothesis.

A recent pilot study by Bhattacharya et al. [61] suggests that the use of traditional pharmacovigilance methods to analyse social media data is ineffective. Mao et al. [119] note that frequency data should not serve as prevalence of the adverse effects/reactions, but as a measure of which symptoms may be the most salient to patients on a day-to-day basis. Pierce et al. [73] underline that further research is needed to develop best practices and methods for determining what constitutes a safety signal in social media. Further research is also needed to investigate the impact of cross-channel information diffusion (data source correlation resulting in information duplication).

While significant progress is being made with regard to signal detection and causality hypothesis generation, determining causation (i.e. ascertaining causality for the identified drug–symptom associations) remains a challenge. The collection of more case data relating to possible causation is needed [120]. Further research is required towards the development of a comprehensive approach to combining evidence from multiple social media, while considering the level of trust in each source [57,86]. To describe the process, Adjeroh et al. [121] employ the term **signal fusion** stating the diversity of social media sources, noise, data redundancy and correlation between sources as its major challenges. Abbasi et al. [122] have proposed the CRUFS framework (an acronym denoting credibility, recency, uniqueness, frequency and salience) as a uniform foundation for critically assessing different data channels in social media analysis of adverse drug events.

### 3.4 Data privacy

In the background of the knowledge discovery process stand **personal data protection** concerns, which makes balancing the respective interests of *patient data protection and medication safety monitoring* challenging [123] and the verification of adverse reaction allegations nearly impossible [124]. It is critical that the capabilities offered by new technologies are harnessed in a way that is ethical, compliant with regulations, respecting data privacy and ensuring responsible use of data [60]. As the number of actors engaging with social media and SNS data increases, so does the risk of potential privacy infringements. The increasing demand for data protection driven by the rapid technological advancement and the necessity to reinforce users' trust in services provided by the public and private sectors are inducing legislators to approve data protection laws or amend the existing regula-

tions in order to adapt them to the technological evolution and new challenges [125]. These new data protection regulations confront researchers with imposing hurdles, ranging from the validity of both the data and how it is sampled to the ethical issues regarding its use [126]. Social networking data will qualify as personal data insofar as they relate to an identified or identifiable individual. Although data from social networks is public, this fact does not deprive it from the protection offered by the data protection legislation. The processing of such data still needs to be fair and lawful. As a result, there needs to be a legitimate ground on the basis of which the data could be processed.

### 3.5 Regulatory framework

While SNS offer a large and often untapped potential to identify safety issues, the appropriate and effective use of social media can be overwhelming. Desai [127] stresses that new **regulatory paradigms** are needed and lists the following important questions that need to be answered:

- What is the limit of the industry’s responsibility in collecting and reviewing social media data?
- How can pharmacovigilance teams confirm the identifiability of the reporter and patient in safety data obtained via social media and establish safeguards against faulty adverse event reporting?
- What will be acceptable practices for following up on potential signals within the context of data privacy?
- What are the protocols for big data integration, analysis and interpretation, and reporting of follow-up results?

Golder et al. [63] note that the methods that can help incorporate these datasets into current pharmacovigilance systems are largely unexplored. Regulatory acceptance of social data might be lower than for traditional sources, due to the uncertainty surrounding the appropriate use of such data from patient privacy point of view and the lack of defined strategies or frameworks in place in order to meet the standards regarding data validity and generalisability [128]. Nonetheless, a need for patient-centricity is increasingly recognised by high-profile institutional drug safety stakeholders (Council for International Organizations of Medical Sciences (CIOMS), European Medicines Agency (the ), Food and Drug Administration (FDA), Innovative Medicines Initiative (IMI), etc.), who have developed a series of regulatory guidances and launched initiatives in this direction across the entire value chain of pharmacovigilance [129]. Experts stress the need for strong and systematic processes for selection, validation and study implementation [130]. As practices and legal requirements vary across countries, the need for a concrete policy framework on the further use of social media as a new valid type of data sources for pharmacovigilance is emphasised

[130]. Smith and Benattia [129] further stress the need for an internal revision of the form and function of pharmacovigilance within the biopharmaceutical industry.

In view of the above points, scepticism lingers within the research community regarding the potential use of social media resources in pharmacovigilance. There are differing views among experts on whether social media should only be considered as a support to signal detection (complementing and enriching primary data sources) or whether it has potential for *de novo* signal detection or assessment and whether the evidence on its utility is required before making these decisions [131]. New methods have had limited success in identifying new drug safety signals. Proof exists that social media may be a source for novel or rare adverse events and mild adverse events and for ascertaining patient perspectives [63,132]. The growing interest in the field is met with a call for more studies to demonstrate and understand the potential of social data and define their role for the purposes of pharmacovigilance. Research regarding the utility of social data is ongoing [133], as is work in the field of *cognitive technologies* (e.g. machine learning, artificial intelligence, etc.). Moving beyond the digital trend, IBM [134] predicts that the future of health is cognitive: through the use of cognitive platforms designed to ingest vast quantities of structured and unstructured information from various sources and to allow researchers to find correlations and connections, in order to identify new patterns and insights to accelerate discoveries, treatments and the delivery of health improvements.

Benefiting from advances made in the application of knowledge extraction technologies to other scientific domains, some problems that relate to the technical feasibility of social media use in pharmacovigilance seem to be nearly solved, while enhancements, improvements and new promising solutions are being announced frequently. However, for achieving the development of effective instruments for knowledge extraction in real-life situations, there is still a need for an in-depth investigation of the overall feasibility and effectiveness of these basically experimental proof-of-concept methods and of their potential contribution to pharmacovigilance systems.

Social media have the potential to offset limitations of traditional data sources (time difference between event occurrence and discovery, under-reporting, lack of geographic diversity, loss of patients’ perspectives), particularly by means of their volume, broad coverage and timeliness. However, since they too are not without disadvantages, complementarities need to be sought with traditional data sources. Harnessing the complementary strengths of traditional data sources and social media can open new directions and expand the scope of post-market safety monitoring. Powell et al. [62] note that additional value can be created by supplementing social data with other sources of information (e.g. product label, sales data, etc.). Overall, further research is needed

**Table 4** Knowledge extraction from the perspective of pharmacovigilance

Attribute	Challenges and research areas	References
Data sourcing	<p>Challenge :: (near) real-time social listening: Value/utility concerns of social data (quality and credibility of individual social data sources, data redundancy and correlation, etc.) Inherent challenges of big data: big volumes of data of high variety and velocity;</p> <p>Principal research area In-depth investigation and comparative analysis of individual social data sources (i.e. analysis of the number, type and quality disclosures made on each social media, employing natural language processing (NLP), computational linguistics and/or text mining techniques)</p> <p>Complementary research areas Health and medicine knowledge and engagement of citizens, in order to improve the quality and increase the number of solicited/ unsolicited reports</p>	[16,17,19,22,35,36,50,51,55,59,63–65,73,78,84,88–93,118,122,134,142,147,151,160,166,169]
Information extraction	<p>Challenge :: concept/relation extraction and ADR classification: Concept and relation extraction (identification of drugs and symptoms (named entity recognition) and of drug/event relationship)</p> <p>Principal research area Development of automated or semi-automated drug term discovery systems and algorithms. Investigated solutions build on text mining, computational linguistics, natural language processing (NLP) techniques and/or other rule-based approaches (e.g. lexicon-based solutions) and/or on machine learning (e.g. machine learning classifiers) and artificial intelligence;</p> <p>Complementary research areas ADR detection in longitudinal databases and other sources Automated mining of demographic information Sentiment analysis Harnessing behavioural data (e.g. from Internet search logs ) Adherence to medicine therapy Development of algorithm training corpora Linguistic aspects</p>	[18,34,42–47,55,56,60,65,66,69–71,74,75,77,81–83,85,86,93–116,118,121,132,139,143–146,148,150,152–154]
Data analysis	<p>Challenge :: signal detection and causality hypothesis: Identification of previously unsuspected safety signals: quantitative signal detection techniques applied to social data in aggregate (pattern discovery) Detection of valid individual case safety reports (ICSRs) within social data: most social media reports that mention use of a drug and potential adverse event do not meet the basic regulatory definitions of an individual case safety report</p> <p>Principal research area In-depth investigation and comparative analysis of individual social data sources (e.g. analysis of the type and quality of adverse events reported on each SNS) Development of best practices and methods for determining what constitutes a safety signal in social media Development of methods for the joint analysis of multiple data sources (multimodal signal detection or signal fusion)</p>	[16,17,19,41,42,57,60,61,73,77,81,86,88,93,98,117,118,121,122,131,142,149]
Privacy	<p>Challenge :: balancing the interests of patient data protection and medication safety monitoring</p> <p>Principal research area Effective verification of adverse reaction allegations, respecting data privacy and ensuring responsible use of data</p>	[17,60,86,88,123–126,142]
Regulation	<p>Challenge :: regulatory acceptance of social data: New regulatory paradigms for the incorporation of social data into pharmacovigilance systems for de novo signal detection or to complement and enrich primary data sources</p>	[62,63,86,127–130,133]

**Table 4** continued

Attribute	Challenges and research areas	References
	Principal research areas Understanding of the strengths and limitations of social media in post-market safety surveillance and establishment of best practices Investigation of complementarities between social data and traditional data sources Development of regulatory guidances and initiatives for social data incorporation across the entire value chain of pharmacovigilance International harmonisation and regulatory enforcement	

to better understand the strengths and limitations of social media in post-market safety surveillance and establish best practices [62]. With regard to the four points (key tasks, see Sect. 1.2) put forth by WHO to describe the vision of pharmacovigilance in its traditional sense, social media have proved their potential as lead and lag indicators of unknown and known ADRs, respectively. Scholars have successfully exploited social data for the early detection of unknown adverse events and for the detection of increases in frequency of known adverse reactions (although the definition of relevant proportionality measures for social media remains a challenge). In cases where more detailed information is available on the person, their exposure to the medicine, the adverse outcome and any possible contextual factors, which may have affected the observed result, social media can also assist in the identification of risk factors and the formulation of hypotheses regarding possible causal mechanisms between drug and symptom. The estimation of quantitative aspects of benefit/risk analysis remains a challenge for researchers, as the traditional proportionality-based schemes are not applicable in this case. Table 4 provides an overview of the current state of research for knowledge extraction for the purposes pharmacovigilance: the main aspects of knowledge extraction are identified (information extraction, data analysis, privacy, regulation) and elaborated in terms of the specific domain perspective, including a categorised presentation of open challenges, issues and prospects and of current practice in the community (principal and secondary/complementary areas of research).

## 4 Conclusion

As outlined in its research objectives, the present paper makes a number of contributions to the area of pharmacovigilance and particularly to its social data applications. This section summarises these contributions and presents the conclusions and the gained insights and lessons learnt.

### 4.1 Limitations of this study

While there are several strengths to this study, which was conducted in a structured and methodical manner, limitations of the review need to be acknowledged. The study only included evidence from the MEDLINE/PubMed database and was intended as a scoping exercise, aimed at investigating the applications of social media to pharmacovigilance, through the critical analysis and comparative assessment of the relevant body of literature. Although the present research effectively identified the most prominent works in the field, for future studies, it would be interesting to assess evidence from other databases as well. As the body of evidence grows, future studies could also aim to perform an in-depth critical analysis of each specific topic identified in the present work.

### 4.2 Contributions of this study

The presented review was mainly aimed to map the state of the art in applications of social data to pharmacovigilance, to provide a thorough up-to-date literature review, to contribute to the systematisation of current knowledge about the use of social data in this field and to explore the potential of social data for the detection of ADRs.

In particular, the paper makes the following distinct contributions:

#### Literature review of the applications of social media to pharmacovigilance

A comprehensive literature review of the applications of social media to pharmacovigilance was conducted, with a critical analysis of the relevant body of literature, focused on identifying new important dimensions of this topic and the associated key challenges and on gaining new insights. (See review methodology and execution in Sect. 1, while review analysis and its findings span Sects. 1, 2, 3.) A thorough review of recent scientific literature was completed with a total of 100 articles reviewed, and the final selection of articles was made based on relevance to SNS application to pharmacovigilance (Sect. 1.1). The findings of this literature



review were further analysed to formulate new insights, the key challenges and future research directions.

### **Exploration of social media sources for use in pharmacovigilance**

Another contribution is the classification and assessment of social data sources (Sect. 1).

In particular, based on the review findings, the existing work on social data applications to ADR detection was analysed in terms of category of SNS and nature of reporting. A taxonomy of social data sources (i.e. social media including web data) has been proposed in terms of context and purpose. Furthermore, the specialised healthcare SNS category can be further classified into three types of specialised healthcare SNS/forums: health-centred SNS (generic networking sites on general health topics, usually requiring user profiles), disease-specific online health forums/SNS (focused on specific diseases) and medicine-focused sharing platforms (or patient forums), as explained in Sects. 1.2 and 1.3.

Presently, there is a growing interest by institutional drug safety stakeholders in exploring the use of social media (social listening) to supplement established approaches for pharmacovigilance, by harvesting information on patients' experiences after exposure to pharmaceutical products.

This was followed by literature-informed qualitative comparison of social data sources with regard to a set of key attributes of social big data, including population coverage, usefulness, timeliness, accessibility, quality and processability.

### **Systematic examination of social data used in pharmacovigilance**

A third contribution is to do with classifications and taxonomies for social data used in pharmacovigilance (Sect. 1). From the literature review analysis, a new taxonomy of social (networking) data was derived, which also includes two different classifications of social data—(a) according to the context and purposes of data disclosure, as well as (b) in terms of the area where data are captured according to the classifications—front-office or back-office operation—as proposed in Sect. 1.3. Furthermore, two types of reporting were identified (Sect. 2)—namely solicited and unsolicited—which can also be further analysed in terms of the context and purposes of data disclosure and the area where data are captured.

Finally, the advantages and limitations of the use of social data in the context of pharmacovigilance were identified, critically analysed and discussed (Sect. 2).

### **New insights, key challenges and recommendations for future research and practice**

Last but not least, the present study examined the challenges of knowledge discovery using social data in the context of ADR detection in three dimensions—namely conceptual, technical and environmental (Sect. 3). From the

literature review findings, a set of five key challenges related to the use of social data were identified—namely the challenges of value/utility of social media, information extraction from social media, analysis of social data, data privacy and regulatory framework. Each of these challenges was analysed and discussed in the context of social data in pharmacovigilance, along with useful insights extracted from the review findings, and focusing on potential solutions and future research directions.

### **4.3 Key insights and future directions**

While the value potential of social data is increasing, research on social media-based pharmacovigilance is not in a position to supplant more traditional methods [135]. Salathe [15] stresses that data from traditional health systems and patient-generated data have complementary strengths (high veracity in the data from traditional sources and high velocity and variety in patient-generated data) and, when combined, can lead to more robust public health systems. Lazer et al. [135] call for an all data revolution, i.e. methods that employ data from all traditional and new sources. The literature review findings also indicate an agreement among scholars on the potential of social listening to support and supplement pharmacovigilance systems, which currently rely mainly on traditional ADR reporting. According to Incio et al. [136] this can contribute to better decision-making processes in regulatory activities.

Nonetheless, the value of mining social media for ADRs has not yet been realised in practice. Methods exist to reduce noise and make the data suitable for post-market safety surveillance. However, big data cannot be considered a substitute for traditional data collection and analysis, but rather functions as a supplement to existing methods [61,122]. Abbasi et al. [122] stress the importance of developing an understanding of the strengths and limitations of the various social media channels and the capabilities of real-time analytics. Additional research is therefore needed to better understand the strengths, limitations and best practices of social data innovations in the context of pharmacovigilance, to determine which channels are most suitable with respect to various dimensions.

Social media monitoring is expected to become a standard practice in pharmacovigilance in the future. For this purpose, a careful evaluation of the use of social media as a pharmacovigilance instrument is required, along with new data processing techniques and software tools and infrastructure adapted to the volume, velocity, structure and veracity of social media data. Already, marketing authorisation holders (MAHs) are required by European law to establish and maintain a system for pharmacovigilance and record all suspected adverse reactions brought to their attention. This includes recording suspected ADRs from digital social media.

While at present social media monitoring **cannot be used to test hypotheses but just to generate them**, emerging technologies are expected to increase its sense-making capabilities and lead to **actionable evidence** and eventually to the grand vision of achieving **complete digital vigilance**. This implies delving into more challenging questions, such as the detection of drug-to-drug interactions (DDIs) using social media, of which currently limited examples exist [137]. Beyond the early discovery of ADRs (the reduction in false positives, etc.), the ultimate objective of future research should be to enable the development of methods and instruments in order to identify with certitude the existence of a causal relationship between drug and adverse event (causality), and to assess the severity and the preventability of the ADR [138]. This assessment is critical in order for health-care stakeholders to be able to develop strategies and plan interventions to reduce the burden of ADRs.

Research in this field is ongoing, with artificial intelligence (AI)-based web and social listening emerging as a promising solution that may improve levels of accuracy and reliability of human-directed monitoring [139], reduce manual data labelling requirements [113,140] and enable coordinated and efficient systems for developing actionable evidence on medicine safety and effectiveness [141].

Future efforts should be aimed at further developing computational methods for processing large data volumes and natural language processing methods for more detailed and sophisticated data analysis (e.g. to establish **causality** with the help of social media data). Furthermore, there are some early indications that the joint analysis of multiple data sources (**multimodal signal detection**) may lead to improved signal detection [93]. Holistic examination and interpretation of knowledge sources are needed in order to produce timely, reliable and actionable results. According to Harpaz et al. [93], this requires a deeper understanding of the data sources used, additional benchmarks and further research on methods to generate and synthesise signals. Moreover, according to Bate et al. [88], a scientifically robust strategy for measuring the specific value of innovative big data sources is needed before such innovations can be incorporated into formal decision-making processes.

In conclusion, substantial future benefits to pharmacovigilance practice are therefore expected to be realised through further advances in data availability and computational methods for mining insights and inferences from large data sets.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

1. World Health Organisation (WHO): Safety of Medicines-A Guide to Detecting and Reporting Adverse Drug Reactions-Why Health Professionals Need to Take Action. World Health Organisation, Geneva (2002)
2. World Health Organisation (WHO): Pharmacovigilance: Ensuring the Safe Use of Medicines, WHO Policy Perspectives 9. World Health Organisation, Geneva (2004)
3. Lazarou, J., Pomeranz, B.H., Corey, P.N.: Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *Jama* **279**(15), 1200–1205 (1998)
4. Bouvy, J.C., De Bruin, M.L., Koopmanschap, M.A.: Epidemiology of adverse drug reactions in Europe: a review of recent observational studies. *Drug Saf.* **38**(5), 437–453 (2015)
5. Davies, E.C., Green, C.F., Mottram, D.R., Rowe, P.H., Pirmohamed, M.: Emergency readmissions to hospital due to adverse drug reactions within 1 year of the index admission. *Br. J. Clin. Pharmacol.* **70**(5), 749–755 (2010)
6. BnardLaribire, A., MiremontSalam, G., PraultPochat, M.C., Noize, P., Haramburu, F.: Incidence of hospital admissions due to adverse drug reactions in France: the EMIR study. *Fundam. Clin. Pharmacol.* **29**(1), 106–111 (2015)
7. Esteban, J., Navarro, C.P., Gonzalez, F.R., Lanuza, F.G., Montesa, C.L.: A study of incidence and clinical characteristics of adverse drug reactions in hospitalized patients. *Rev Esp Salud Pública Revista* **91**, e201712050 (2017)
8. Pirmohamed, M., James, S., Meakin, S., Green, C., Scott, A.K., Walley, T.J., Breckenridge, A.M.: Adverse drug reactions as cause of admission to hospital: prospective analysis of 18 820 patients. *BMJ* **329**(7456), 15–19 (2004)
9. Sonja, M., Ioana, G., Miaoqing, Y., Anna, K.: Understanding value in health data ecosystems: a review of current evidence and ways forward. *Rand Health Q.* **7**(2), 3 (2018)
10. Waller, P.: An Introduction to Pharmacovigilance. Wiley, New York (2011)
11. Caron, J., Rochoy, M., Gaboriau, L., Gautier, S.: The history of pharmacovigilance. *Thrapie* **71**(2), 129–134 (2016)
12. European Medicines Agency: Guideline on Good Pharmacovigilance Practices (GVP). Annex I-Definitions (Rev 4). EMA/876333/2011 Rev 4 (2017)
13. World Health Organization: The Importance of Pharmacovigilance. World Health Organization, Geneva (2002). (ISBN 9241590157)
14. World Health Organization: Safety Monitoring of Medicinal Products: Guidelines for Setting Up and Running a Pharmacovigilance Centre. World Health Organization, Geneva (2000)
15. Salathe, M.: Digital pharmacovigilance and disease surveillance: combining traditional and big-data systems for better public health. *J. Infect. Dis.* **214**(suppl 4), S399–S403 (2016)
16. Harpaz, R., DuMochel, W., Shah, N.H.: Big data and adverse drug reaction detection. *Clin. Pharmacol. Ther.* **99**(3), 268–270 (2016)
17. Bellazzi, R.: Big data and biomedical informatics: a challenging opportunity. *Yearb. Med. Inform.* **9**, 8–13 (2014)
18. Yeleswarapu, S.J., Rao, A., Joseph, T., Saipradeep, V., Srinivasan, R.: A pipeline to extract drug-adverse event pairs from multiple data sources. *BMC Med. Inf. Decis. Mak.* **14**(1), 13 (2014)
19. Ehrenstein, V., Nielsen, H., Pedersen, A.B., Johnsen, S.P., Pedersen, L.: Clinical epidemiology in the era of big data: new opportunities, familiar challenges. *Clin. Epidemiol.* **9**, 245250 (2017)
20. PwC Patient engagement: Pharma's strategy for success in the New Health Economy. Health Research Institute Report (2016)

21. De Choudhury, M., De, S.: Mental health discourse on reddit: self-disclosure, social support, and anonymity. In: Proceedings of ICWSM, pp. 71–80 (2014)
22. Karapetiantz, P., Bellet, F., Audeh, B., Lardon, J., Leprovost, D., Aboukhamis, R., Jaulent, M.C.: Descriptions of adverse drug reactions are less informative in forums than in the french pharmacovigilance database but provide more unexpected reactions. *Front. Pharmacol.* **9**, 439 (2018)
23. Gage-Bouchard, E.A., LaValley, S., Warunek, M., Beaupin, L.K., Mollica, M.: Is cancer information exchanged on social media scientifically accurate? *J. Cancer Educ.* **33**(6), 1328–1332 (2018)
24. Boyd, D., Ellison, N.B.: Social network sites: definition, history, and scholarship. *J. Comput. Med. Commun.* **13**(1), 210–230 (2007)
25. Kuss, D.J., Griffiths, M.D.: Online social networking and addiction—review of the psychological literature. *Int. J. Environ. Res. Public Health* **8**(9), 3528–3552 (2011)
26. Pew Research Center Social Media Usage: 2005–2015. <http://www.pewinternet.org/2015/10/08/social-networking-usage-2005-2015/> (2015). Accessed 20 Sept 2017
27. Hamm, M.P., Chisholm, A., Shulhan, J., Milne, A., Scott, S.D., Given, L.M., Hartling, L.: Social media use among patients and caregivers: a scoping review. *BMJ Open* **3**(5), e002819 (2013)
28. Lamas, E., Salinas, R., Coquedano, C., Simon, M.P., Bousquet, C., Ferrer, M., Zorrilla, S.: The meaning of patient empowerment in the digital age: the role of online patient-communities. *Stud. Health Technol. Inform.* **244**, 43–47 (2017)
29. Housman, L.T.: Im home (screen)!: social media in health care has arrived. *Clin. Ther.* **39**(11), 2189–2195 (2017)
30. De Simoni, A., Shanks, A., Balasooriya-Smeekens, C., Mant, J.: Stroke survivors and their families receive information and support on an individual basis from an online forum: descriptive analysis of a population of 2348 patients and qualitative study of a sample of participants. *BMJ Open* **6**(4), e010501 (2016)
31. Izuka, N.J., Alexander, M.A., Balasooriya-Smeekens, C., Mant, J., De Simoni, A.: How do stroke survivors and their carers use practitioners advice on secondary prevention medications? Qualitative study of an online forum. *Fam. Pract.* **34**(5), 612–620 (2017)
32. Merolli, M., Gray, K., Martin-Sanchez, F., Lopez-Campos, G.: Patient-reported outcomes and therapeutic affordances of social media: findings from a global online survey of people with chronic pain. *J. Med. Internet Res.* **17**(1), e20 (2015)
33. Cohan, A., Young, S., Goharian, N.: Triaging mental health forum posts. In: Proceedings of the 3rd Workshop on Computational Linguistics and Clinical Psychology 2016, pp. 143–147 (2016)
34. Frost, J., Okun, S., Vaughan, T., Heywood, J., Wicks, P.: Patient-reported outcomes as a source of evidence in off-label prescribing: analysis of data from patientslikeme. *J. Med. Internet Res.* **13**(1), e6 (2011)
35. Nakamura, C., Bromberg, M., Bhargava, S., Wicks, P., Zeng-Treitler, Q.: Mining online social network data for biomedical research: a comparison of clinicians and patients perceptions about amyotrophic lateral sclerosis treatments. *J. Med. Internet Res.* **2012** **14**(3), e90 (2012)
36. Hawkins, C.M., DeLaO, A.J., Hung, C.: Social media and the patient experience. *J. Am. Coll. Radiol.* **13**(12), 1615–1621 (2016)
37. OECD: Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value, OECD Digital Economy Papers, No. 220, OECD Publishing (2013)
38. Van Alsenoy, B. Rights and obligations of actors in social networking sites, deliverable 6.2 of the SPION project. <https://lirias.kuleuven.be/handle/123456789/453689> (2014). Accessed 20 Sept 2017
39. Schneider, B.: A taxonomy of social networking data. *IEEE Secur. Priv.* **8**(4), 88–88 (2010)
40. PwC: Social media likes healthcare. From marketing to social business. Health Research Institute Report (2012)
41. Vance, K., Howe, W., Dellavalle, R.: Social internet sites as a source of public health information. *Dermatol. Clin.* **27**(2), 133–136 (2009)
42. Choudhury, M., Morris, M.R., White, R.W.: Seeking and sharing health information online: comparing search engines and social media. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 2014, pp. 1365–1376 (2014)
43. Paul, M., Sarker, A., Brownstein, J., Nikfarjam, A., Scotch, M., Smith K., Gonzalez, G.: Social media mining for public health monitoring and surveillance. In: Biocomputing 2016: Proceedings of the Pacific Symposium, pp. 468–479 (2016)
44. Paul, M., Dredze, M.: A model for mining public health topics from Twitter. *Health* **11**, 16–6 (2011)
45. Byrd, K., Mansurov, A., Baysal, O.: Mining Twitter data for influenza detection and surveillance. In: Proceedings of the International Workshop on Software Engineering in Healthcare Systems, pp. 43–49. ACM (2016)
46. Benton, A., Ungar, L., Hill, S., Hennessy, S., Mao, J., Chung, A., Leonard, C.E., Holmes, J.H.: Identifying potential adverse effects using the web: a new approach to medical hypothesis generation. *J. Biomed. Inform.* **44**(6), 989996 (2011)
47. World Health Organisation (WHO): The Safety of Medicines in Public Health Programmes: Pharmacovigilance an Essential Tool. World Health Organisation, Geneva (2006)
48. Basch, E.: Systematic collection of patient-reported adverse drug reactions: a path to patient-centred pharmacovigilance. *Drug Saf.* **36**(4), 277–278 (2013)
49. Strom, B.L.: How the US drug safety system should be changed. *Jama* **295**(17), 2072–2075 (2006)
50. de Langen, J., van Hunsel, F., Passier, A., den Berg, L.-V., van Grootheest, K.: Adverse drug reaction reporting by patients in The Netherlands: three years of experience. *Drug Saf.* **31**(6), 515524 (2008)
51. Anderson, C., Krska, J., Murphy, E., Avery, A.: The importance of direct patient reporting of suspected adverse drug reactions: a patient perspective. *Br. J. Clin. Pharmacol.* **72**(5), 806–822 (2011)
52. Santos, A.: Direct patient reporting in the European Union. A snapshot of reporting systems in seven member states. Health Action International. <http://haiweb.org/wp-content/uploads/2015/09/Direct-Patient-Reporting-in-the-EU.pdf> (2015). Accessed 20 Sept 2017
53. Berrewaerts, J., Delbecque, L., Orban, P., Desseilles, M.: Patient participation and the use of ehealth tools for pharmacovigilance. *Front. Pharmacol.* **7**, 90 (2016)
54. Abou Taam, M., Rossard, C., Cantaloube, L., Bouscaren, N., Roche, G., Pochard, L., Bagheri, H.: Analysis of patients’ narratives posted on social media websites on benfluorex’s (Mediator) withdrawal in France. *J. Clin. Pharm. Ther.* **39**(1), 53–55 (2014)
55. Sarker, A., Ginn, R.E., Nikfarjam, A., O’Connor, K., Smith, K., Jayaraman, S., Tejaswi, Upadhaya, T., Gonzalez, G.: Utilizing social media data for pharmacovigilance: a review. *J. Biomed. Inform.* **54**, 202–212 (2015)
56. Yang, H., Yang, C.C.: Harnessing social media for drug-drug interactions detection. In: Proceedings of the 2013 IEEE International Conference on Healthcare Informatics, p. 2229 (2013)
57. 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.* **80**(4), 910–920 (2015)
58. Liu, X., Chen, H.: Identifying adverse drug events from patient social media: a case study for diabetes. *IEEE Intell. Syst.* **30**(3), 44–51 (2015)
59. Avery, A.J., Anderson, C., Bond, C.M., Fortnum, H., Gifford, A., Hannaford, P.C., Murphy, E.: Evaluation of patient reporting of

- adverse drug reactions to the UK Yellow Card Scheme: literature review, descriptive and qualitative analyses, and questionnaire surveys. *Health Technol. Assess.* **15**(20), 1–234 (2011)
60. 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 on Drug Saf.* **14**(12), 1845–1853 (2015)
  61. Bhattacharya, M., Snyder, S., Malin, M., Truffa, M.M., Marinic, S., Engelmann, R., Raheja, R.R.: Using social media data in routine pharmacovigilance: a pilot study to identify safety signals and patient perspectives. *Pharm. Med.* **31**(3), 167–174 (2017)
  62. Powell, G., Seifert, H., Reblin, T., Burstein, P., Blowers, J., Menius, J., Painter, J., Thomas, M., Pierce, C., Rodriguez, H., Brownstein, J., Freifeld, C., Bell, H., Dasgupta, N.: Social media listening for routine post-market safety surveillance. *Drug Saf.* **39**(5), 443–454 (2016)
  63. Golder, S., Norman, G., Loke, Y.K.: Systematic review on the prevalence, frequency and comparative value of adverse events data in social media. *Br. J. Clin. Pharmacol.* **80**(4), 878–888 (2015)
  64. Price, J.: What can big data offer the pharmacovigilance of orphan drugs? *Clin. Ther.* **38**(12), 2533–2545 (2016)
  65. Leaman, R., Wojtulewicz, L., Sullivan, R., Skariah, A., Yang, J., Gonzalez, G.: Towards internet-age pharmacovigilance: extracting adverse drug reactions from user posts to health-related social networks. In: Proceedings of the 2010 Workshop on Biomedical Natural Language Processing, Association for Computational Linguistics, 2010, p. 117125 (2010)
  66. Nikfarjam, A., Gonzalez, G.H.: Pattern mining for extraction of mentions of adverse drug reactions from user comments. In: AMIA Annual Symposium Proceedings, vol. 2011, pp. 1019–1026. American Medical Informatics Association (2011)
  67. Hadzi-Puric, J., Grmusa, J.: Automatic drug adverse reaction discovery from parenting websites using disproportionality methods. In: Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012) IEEE Computer Society, pp. 792–797 (2012)
  68. Matsuda, S., Aoki, K., Tomizawa, S., Sone, M., Tanaka, R., Kuriki, H., Takahashi, Y.: Analysis of patient narratives in disease blogs on the internet: an exploratory study of social pharmacovigilance. *JMIR Pub. Health Surveill.* **3**(1), e10 (2017)
  69. Sampathkumar, H., Chen, X.W., Luo, B.: Mining adverse drug reactions from online healthcare forums using hidden Markov model. *BMC Med. Inform. Decis. Mak.* **14**(1), 91 (2014)
  70. Yates, A., Goharian, N.: ADRTrace: detecting expected and unexpected adverse drug reactions from user reviews on social media sites. In: European Conference on Information Retrieval, pp. 816–819. Springer, Berlin, Heidelberg (2013)
  71. Liu, X., Chen, H.: A research framework for pharmacovigilance in health social media: identification and evaluation of patient adverse drug event reports. *J. Biomed. Inform.* **58**, 268–279 (2015)
  72. Comstock, J.: FDA taps PatientsLikeMe to test the waters of social media adverse event reporting. *MobileHealthNews*. <http://www.mobihealthnews.com/44366/fda-taps-patientslikeme-to-test-the-waters-of-social-media-adverse-event-reporting> (2015). Accessed 20 Sept 2017
  73. Pierce, C.E., Bouri, K., Pamer, C., Proestel, S., Rodriguez, H.W., Le Van, H., Dasgupta, N.: Evaluation of Facebook and Twitter monitoring to detect safety signals for medical products: an analysis of recent FDA safety alerts. *Drug Saf.* **40**(4), 317–331 (2017)
  74. Bian, J., Topaloglu, U., Yu, F.: Towards large-scale twitter mining for drug-related adverse events. Association for Computing Machinery. In: Proceedings of the 2012 International workshop on Smart health and wellbeing, p. 2532 (2012)
  75. O'Connor, K., Pimpalkhute, P., Nikfarjam, A., Ginn, R., Smith, K.L., Gonzalez, G.: Pharmacovigilance on Twitter? Mining Tweets for adverse drug reactions. In: AMIA Annual Symposium Proceedings, vol. 2014, p. 924933. American Medical Informatics Association (2014)
  76. Carbonell, P., Mayer, M.A., Bravo: Exploring brand-name drug mentions on Twitter for pharmacovigilance. In: Digital Healthcare Empowering Europeans: Proceedings of MIE2015, pp. 55–59 (2015)
  77. Freifeld, C., Brownstein, J., Menone, C., Bao, W., Filice, R., Kass-Hout, T., Dasgupta, N.: Digital drug safety surveillance: monitoring pharmaceutical products in Twitter. *Drug Saf.* **37**(5), 343–350 (2014)
  78. Anker, A.E., Reinhart, A.M., Feeley, T.H.: Health information seeking: a review of measures and methods. *Patient Educ. Couns.* **82**(3), 346–354 (2011)
  79. Niederdeppe, J., Hornik, R.C., Kelly, B.J., Frosch, D.L., Roman-tan, A., Stevens, R.S., Schwartz, J.S.: Examining the dimensions of cancer-related information seeking and scanning behavior. *Health Commun.* **22**(2), 153–167 (2007)
  80. Bragazzi, N., Siri, A.: Google trends-enabled digital pharmacovigilance: monitoring interest towards antidepressants and their usage patterns in Italy. *Eur. Psychiatry* **33**, S281 (2016)
  81. White, R., Tatonetti, N., Shah, N., Altman, R., Horvitz, E.: Web-scale pharmacovigilance: listening to signals from the crowd. *J. Am. Med. Inform. Assoc.* **20**(3), 404–408 (2013)
  82. Chokor, A., Sarker, A., Gonzalez, G.: Mining the web for pharmacovigilance: the case study of duloxetine and venlafaxine. *arXiv preprint arXiv:1610.02567* (2016)
  83. Yom-Tov, E., Gabrilovich, E.: Postmarket drug surveillance without trial costs: discovery of adverse drug reactions through large-scale analysis of web search queries. *J. Med. Internet Res.* **15**(6), e124 (2013)
  84. Norn, G.N.: Pharmacovigilance for a revolving world: prospects of patient-generated data on the internet. *Drug Saf.* **37**(10), 761764 (2014)
  85. Comstock, J.: FDA Google met to discuss use of search to find adverse drug reactions. *MobileHealthNews*. <http://www.mobihealthnews.com/45411/fda-google-met-to-discuss-use-of-search-to-find-adverse-drug-reactions> (2015). Accessed 20 September 2017
  86. Bousquet, C., Dahamna, B., Guillemin-Lanne, S., Darmoni, S.J., Faviez, C., Huot, C., Katsahian, S., Leroux, V., Pereira, S., Richard, C., Schck, S., Souvignat, J., Lillo-Le Lout, A., Texier, N.: The adverse drug reactions from patient reports in social media project: five major challenges to overcome to operationalize analysis and efficiently support pharmacovigilance process. *JMIR Res. Protoc.* **6**(9), e179 (2017)
  87. van Panhuis, W.G., Proma, P., Emerson, C., Grefenstette, J., Wilder, R., Herbst, A., Heymann, D., Burke, D.: A systematic review of barriers to data sharing in public health. *BMC Pub. Health* **14**(1), 1144 (2014)
  88. Bate, A., Reynolds, R.F., Caubel, P.: The hope, hype and reality of big data for pharmacovigilance. *Ther. Adv. Drug Saf.* **9**(1), 5–11 (2018)
  89. De Choudhury, M., Sharma, S., Logar, T., Eekhout, W., Nielsen, R.: Gender and cross-cultural differences in social media disclosures of mental illness. In: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, pp. 353–369. ACM (2017)
  90. Topaz, M., Lai, K., Dhopeswarkar, N., Seger, D.L., Saadon, R., Goss, F., Zhou, L.: 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.* **39**(3), 241–250 (2016)
  91. Duh, M.S., Cremieux, P., Audenrode, M.V., Vekeman, F., Karner, P., Zhang, H., Greenberg, P.: Can social media data lead to earlier detection of drug-related adverse events? *Pharmacoepidemiol. Drug Saf.* **25**(12), 1425–1433 (2016)

92. Kietzmann, J., Hermkens, K., McCarthy, I., Silvestre, B., Kietzmann, J.: Social media? Get serious! Understanding the functional building blocks of social media. *Bus. Horiz.* **54**(3), 241–251 (2011)
93. Harpaz, R., DuMouchel, W., Schuemie, M., Bodenreider, O., Friedman, C., Horvitz, E., Shah, N.H.: Toward multimodal signal detection of adverse drug reactions. *J. Biomed. Inform.* **76**, 41–49 (2017)
94. Chee, B., Berlin, R., Schatz, B.: Predicting adverse drug events from personal health messages. In: *AMIA Annual Symposium Proceedings*. American Medical Informatics Association, 2011, pp. 217–226 (2011)
95. Sarker, A., Gonzalez, G.: Portable automatic text classification for adverse drug reaction detection via multi-corpus training. *J. Biomed. Inform.* **53**, 196–207 (2015)
96. Yang, M., Kiang, M.Y., Shang, W.: Filtering big data from social media—building an early warning system for adverse drug reactions. *J. Biomed. Inform.* **54**, 230–240 (2015)
97. Zhang, Z., Nie, J., Zhang, X.: An ensemble method for binary classification of adverse drug reactions from social media. In: *Proceedings of the Social Media Mining Shared Task Workshop at the Pacific Symposium on Biocomputing 2016* (2016)
98. Abdellaoui, R., Schck, S., Texier, N., Burgun, A.: Filtering entities to optimize identification of adverse drug reaction from social media: How can the number of words between entities in the messages help? *JMIR Pub. Health Surveill.* **3**(2), e36 (2017)
99. Nguyen, T., Larsen, M.E., ODea, B., Phung, D., Venkatesh, S., Christensen, H.: Estimation of the prevalence of adverse drug reactions from social media. *Int. J. Med. Inform.* **102**, 130–137 (2017)
100. Audeh, B., Beigbeder, M., Zimmermann, A., Jaillon, P., Bousquet, C.: Vigi4Med scraper: a framework for web forum structured data extraction and semantic representation. *PloS ONE* **12**(1), e0169658 (2017)
101. Chen, X., Deldossi, M., Aboukhamis, R., Faviez, C., Dahamna, B., Karapetiantz, P., Guenegou-Arnoux, A., Girardeau, Y., Guillemain-Lanne, S., Lillo-Le-Lout, A., Texier, N., Burgun, A., Katsahian, S.: Mining adverse drug reactions in social media with named entity recognition and semantic methods. *Stud. Health Technol. Inform.* **245**, 322–326 (2017)
102. Aronson, A.R.: Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. In: *Proceedings of the 2001 AMIA Symposium* (p. 17), American Medical Informatics Association, pp. 17–21 (2001)
103. Chowdhury, S., Zhang, C., Yu, P.S.: Multi-task pharmacovigilance mining from social media posts. *CoRR*, arXiv preprint [arXiv:1801.06294](https://arxiv.org/abs/1801.06294) (2018)
104. Tuarob, S., Tucker, C.S., Salath, M., Ram, N.: An ensemble heterogeneous classification methodology for discovering health-related knowledge in social media messages. *J. Biomed. Inform.* **49**, 255–68 (2014)
105. Alvaro, N., Miyao, Y., Collier, N.: TwiMed: twitter and PubMed comparable corpus of drugs, diseases, symptoms, and their relations. *JMIR Pub. Health Surveill.* **3**(2), e24 (2017)
106. Sarker, A., Gonzalez, G.: A corpus for mining drug-related knowledge from Twitter chatter: language models and their utilities. *Data Brief* **10**, 122–131 (2017)
107. Segura-Bedmar, I., Martinez, P., Revert, R., Moreno-Schneider, J.: Exploring Spanish health social media for detecting drug effects. *BMC Med. Inform. Decis. Mak.* **15**(2), S6 (2015)
108. Korkontzelos, I., Nikfarjam, A., Shardlow, M., Sarker, A., Ananiadou, S., Gonzalez, G.H.: Analysis of the effect of sentiment analysis on extracting adverse drug reactions from tweets and forum posts. *J. Biomed. Inform.* **62**, 148–158 (2016)
109. Isah, H., Trundle, P.R., Neagu, D.: Social media analysis for product safety using text mining and sentiment analysis. In: *2014 14th UK Workshop on Computational Intelligence (UKCI)*, pp. 1–7. IEEE (2014)
110. Mishra, A., Malviya, A., Aggarwal, S.: Towards automatic pharmacovigilance: analysing patient reviews and sentiment on oncological drugs. In: *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, pp. 1402–1409. IEEE (2015)
111. Ji, X., Chun, S. A., Geller, J.: Monitoring public health concerns using twitter sentiment classifications. In: *2013 IEEE International Conference on Healthcare Informatics (ICHI)*, pp. 335–344. IEEE (2013)
112. Casperson, T.A., Painter, J.L., Dietrich, J.: Strategies for distributed curation of social media data for safety and pharmacovigilance. In: *Proceedings of the International Conference on Data Mining (DMIN) 2016*, p. 118. The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp) (2016)
113. Cocos, A., Fiks, A.G., Masino, A.J.: Deep learning for pharmacovigilance: recurrent neural network architectures for labeling adverse drug reactions in Twitter posts. *J. Am. Med. Inform. Assoc.* **24**(4), 813–821 (2017)
114. Miftahutdinov, Z., Tutubalina, E.: End-to-end deep framework for disease named entity recognition using social media data. In: *2017 IEEE 30th Neumann Colloquium (NC)*, pp. 47–52 (2017)
115. Nikfarjam, A., Sarker, A., OConnor, K., Ginn, R., Gonzalez, G.: Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. *J. Am. Med. Inform. Assoc.* **22**(3), 671–681 (2015)
116. Xu, B., Lin, H., Zhao, M., Yang, Z., Wang, J., Zhang, S.: Detecting potential adverse drug reactions from health-related social networks. In: *Lin C.Y., Xue, N., Zhao, D., Huang, X., Feng, Y. (eds.) Natural Language Understanding and Intelligent Applications. ICCPOL 2016, NLPCC 2016. Lecture Notes in Computer Science*, vol. 10102, pp. 523–530. Springer, Cham (2016)
117. Lin, W.Y., Li, H.Y., Du, J.W., Feng, W.Y., Lo, C.F., Soo, V.W.: iADR: towards online adverse drug reaction analysis. *Springer-Plus* **1**(1), 72 (2012)
118. Lardon, J., Abdellaoui, R., Bellet, F., Asfari, H., Souvignet, J., Texier, N., Bousquet, C.: Adverse drug reaction identification and extraction in social media: a scoping review. *J. Med Internet Res.* **17**(7), e171 (2015)
119. Mao, J.J., Chung, A., Benton, A., Hill, S., Ungar, L., Leonard, C.E., Hennessy, S., Holmes, J.H.: Online discussion of drug side effects and discontinuation among breast cancer survivors. *Pharmacoepidemiol. Drug Saf.* **22**(3), 256–262 (2013)
120. Edwards, I.R.: Causality assessment in pharmacovigilance: still a challenge. *Drug Saf.* **40**(5), 365–372 (2017)
121. Adjeroh, D., Beal, R., Abbasi, A., Zheng, W., Abate, M., Ross, A.: Signal fusion for social media analysis of adverse drug events. *IEEE Intell. Syst.* **29**(2), 74–80 (2014)
122. Abbasi, A., Adjeroh, D., Dredze, M., Paul, M.J., Zahedi, F.M., Zhao, H., Huesch, M.D.: Social media analytics for smart health. *IEEE Intell. Syst.* **29**(2), 60–80 (2014)
123. Dreyer, N.A., Blackburn, S., Hliva, V., Mt-Isa, S., Richardson, J., Jamry-Dziurla, A., Bourke, A., Johnson, R.: Balancing the interests of patient data protection and medication safety monitoring in a public-private partnership. *JMIR Med. Inform.* **3**(2), e18 (2015)
124. Coloma, P.M., Becker, B., Sturkenboom, M.C., Van Mulligen, E.M., Kors, J.A.: Evaluating social media networks in medicines safety surveillance: two case studies. *Drug Saf.* **38**(10), 921–930 (2015)
125. Greenleaf, G.: *Global Data Privacy Laws: 89 Countries, and Accelerating. 115 Privacy Laws & Business International Report, Special Supplement* (2012)

126. Golder, S.A., Macy, M.W.: Digital footprints: Opportunities and challenges for online social research. *Annu. Rev. Sociol.* **40**, 129–152 (2014)
127. Desai, S.: The impact of social media on drug safety. *Safety & Risk Management Blog*. Posted 15 April 15 2015 (2015)
128. The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP): Guide on Methodological Standards in Pharmacoepidemiology (Revision 6). EMA/95098/2010 (2017)
129. Smith, M.Y., Benattia, I.: The patients voice in pharmacovigilance: pragmatic approaches to building a patient-centric drug safety organization. *Drug Saf.* **39**(9), 779–785 (2016)
130. Lengsavath, M., Dal Pra, A., de Ferran, A.M., Brosch, S., Hrmak, L., Newbould, V., Goncalves, S.: Social media monitoring and adverse drug reaction reporting in pharmacovigilance: an overview of the regulatory landscape. *Ther. Innov. Regul. Sci.* **51**(1), 125–131 (2017)
131. European Medicines Agency WEB-RADR Workshop Report: Mobile Technologies and Social Media as New Tools in Pharmacovigilance. WEB-RADR project. Innovative Medicines Initiative (2016). [https://www.ema.europa.eu/documents/report/innovative-medicines-initiative-web-radr-workshop-report-mobile-technologies-social-media-new-tools\\_en.pdf](https://www.ema.europa.eu/documents/report/innovative-medicines-initiative-web-radr-workshop-report-mobile-technologies-social-media-new-tools_en.pdf). Accessed 20 Sept 2017
132. Zheng, Y., Lan, C., Peng, H., Li, J.: Using constrained information entropy to detect rare adverse drug reactions from medical forums. In: 2016 IEEE 38th Annual International Conference of the Engineering in Medicine and Biology Society (EMBC), pp. 2460–2463. IEEE (2016)
133. Tricco, A.C., Zarin, W., Lillie, E., Pham, B., Straus, S.E.: Utility of social media and crowd-sourced data for pharmacovigilance: a scoping review protocol. *BMJ Open* **7**(1), e013474 (2017)
134. IBM. The future of health is cognitive. Point of view—IBM healthcare and life sciences. <http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=HLW03025USEN> (2016). Accessed 20 Dec 2017
135. Lazer, D., Kennedy, R., King, G., Vespignani, A.: Big data. The parable of Google Flu: traps in big data analysis. *Science* **343**(6176), 1203–1205 (2014)
136. Incio, P., Cavaco, A., Airaksinen, M.: The value of patient reporting to the pharmacovigilance system: a systematic review. *Br. J. Clin. Pharmacol.* **83**(2), 227–246 (2017)
137. Vilar, S., Friedman, C., Hripcsak, G.: Detection of drug-drug interactions through data mining studies using clinical sources, scientific literature and social media. *Brief. Bioinform.* **19**(5), 863–877 (2018)
138. Ithnin, M., Rani, M.D.M., Latif, Z.A., Kani, P., Syaiful, A., Aripin, K.N.N., Mohd, T.A.M.T.: Mobile app design, development, and publication for adverse drug reaction assessments of causality, severity, and preventability. *JMIR mHealth uHealth* **5**(5), e78 (2017)
139. Sherlock, A., Rudolf, C.: Artificial Intelligence as an Aid to Pharmacovigilance. *Pharm Exec Magazine*. Posted on May 12, 2017 at <http://www.pharmexec.com/artificial-intelligence-aid-pharmacovigilance> (2017). Accessed 20 Sept 2017
140. Zorzi, M., Combi, C., Pozzani, G., Arzenton, E., Moretti, U.: A co-occurrence based MedDRA terminology generation: some preliminary results. In: Conference on Artificial Intelligence in Medicine in Europe, pp. 215–220. Springer, Cham (2017)
141. Pitts, P.J.: 21st Century pharmacovigilance: intuition, science, and the role of artificial intelligence. *J. Commer. Biotechnol.* **23**(1), 3–6 (2017)
142. Knowledgegent. Big data enabling better pharmacovigilance. Knowledgegent Whitepaper (2015)
143. Comfort, S., Perera, S., Hudson, Z., Dorrell, D., Meireis, S., Nagarajan, M., Fine, J.: Sorting through the safety data haystack: using machine learning to identify individual case safety reports in social-digital media. *Drug Saf.* **41**(6), 579–590 (2018)
144. Ilimova I., Tutubalina E.: Automated detection of adverse drug reactions from social media posts with machine learning. In: van der Aalst, W. et al. (eds.) *Analysis of Images, Social Networks and Texts*. AIST 2017. Lecture Notes in Computer Science, p. 10716. Springer, Cham (2018)
145. Tutubalina, E., Nikolenko, S.: Exploring convolutional neural networks and topic models for user profiling from drug reviews. *Multimed. Tools Appl.* **77**(4), 4791–4809 (2018)
146. Banerjee, R., Ramakrishnan, I.V., Henry, M., Perciavalle, M.: Patient centered identification, attribution, and ranking of adverse drug events. In: 2015 International Conference on Healthcare Informatics (ICHI), pp. 18–27. IEEE (2015)
147. Hsu, D., Moh, M., Moh, T.: Mining frequency of drug side effects over a large twitter dataset using apache spark. In: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, pp. 915–924. ACM (2017)
148. Jouanous, E., Mallaret, M.P., Micallef, J., Pont, C., Roussin, A., Lapeyre-Mestre, M.: Comment on: “social media mining for toxicovigilance: automatic monitoring of prescription medication abuse from Twitter”. *Drug Saf.* **40**, 183–185 (2017)
149. Renuka, K., Jeetha, B.R., Hirose, H.: A survey and analysis of various health-related knowledge mining techniques in social media. *Int. J. Comput. Appl.* **158**(1), 5–10 (2017)
150. Dupuch, M., Hamon, T., Grabar, N.: Cross-language detection of linguistic and semantic regularities in pharmacovigilance terms. In: 4th International Louhi Workshop on Health Document Text Mining and Information Analysis (2013)
151. Sokolova, M.: Big text advantages and challenges: classification perspective. *Int. J. Data Sci. Anal.* **5**(1), 1–10 (2018)
152. Karimi, S., Metke-Jimenez, A., Kemp, M., Wang, C.: Cadec: a corpus of adverse drug event annotations. *J. Biomed. Inform.* **55**, 73–81 (2015)
153. Akhtyamova, L., Alexandrov, M., Cardiff, J.: Adverse drug extraction in twitter data using convolutional neural network. In: 2017 28th International Workshop on Database and Expert Systems Applications (DEXA), pp. 88–92. IEEE (2017)
154. Liu, Y., Shi, J., Chen, Y.: Patientcentered and experienceaware mining for effective adverse drug reaction discovery in online health forums. *J. Assoc. Inf. Sci. Technol.* **69**(2), 215–228 (2018)
155. Ravoire, S., Lang, M., Perrin, E., Audry, A., Bilbault, P., Chekroun, M., Malbezin, M.: Advantages and limitations of online communities of patients for research on health products. *Therapie* **72**(1), 135–143 (2017)
156. PatientsLikeMe. Research manuscripts bibliography. (2018) <https://patientslikeme-bibliography.s3.amazonaws.com/PLM%20Research%20Manuscripts%20Bibliography.pdf>. Accessed 18 June 2018
157. Kheloufi, F., Default, A., Blin, O., Micallef, J.: Investigating patient narratives posted on Internet and their informativeness level for pharmacovigilance purpose: the example of comments about statins. *Therapie* **72**(4), 483–490 (2017)
158. Sinha, M.S., Freifeld, C.C., Brownstein, J.S., Donneyong, M.M., Rausch, P., Lappin, B.M., Avorn, J.: Social media impact of the food and drug administration’s drug safety communication messaging about zolpidem: Mixed-methods analysis. *JMIR Pub. Health Surveill.* **4**(1) (2018)
159. Koutkias, V.G., Lillo-Le Lout, A., Jaulent, M.C.: Exploiting heterogeneous publicly available data sources for drug safety surveillance: computational framework and case studies. *Expert Opin. Drug Saf.* **16**(2), 113–124 (2017)
160. Amoozegar, J.B., Rupert, D.J., Sullivan, H.W., ODonoghue, A.C.: Consumer confusion between prescription drug precautions and side effects. *Patient Educ. Couns.* **100**(6), 1111–1119 (2017)

161. Park, H., Rodgers, S., Stemmler, J.: Analyzing health organizations' use of Twitter for promoting health literacy. *J. Health Commun.* **18**(4), 410–425 (2013)
162. Claiborne, A.B., English, R.A., Caruso, D. (eds.): *Characterizing and Communicating Uncertainty in the Assessment of Benefits and Risks of Pharmaceutical Products: Workshop Summary*. National Academies Press, Washington (2014)
163. Martin-Sanchez, F., Verspoor, K.: Big data in medicine is driving big changes. *Yearb. Med. Inform.* **9**(1), 14 (2014)
164. Wiley, M.T., Jin, C., Hristidis, V., Esterling, K.M.: Pharmaceutical drugs chatter on online social networks. *J. Biomed. Inform.* **49**, 245–254 (2014)
165. Banerjee, A.K., Okun, S., Edwards, I.R., Wicks, P., Smith, M.Y., Mayall, S.J., Basch, E.: Patient-reported outcome measures in safety event reporting: PROSPER consortium guidance. *Drug Saf.* **36**(12), 1129–1149 (2013)
166. Moore, N.: The past, present and perhaps future of pharmacovigilance: homage to Folke Sjoqvist. *Eur. J. Clin. Pharmacol.* **69**(1), 33–41 (2013)
167. Zhao, Y.Q., Ma, W.J.: A review on the advancement of internet-based public health surveillance program. *Zhonghua liu xing bing xue za zhi Zhonghua liuxingbingxue zazhi* **38**(2), 272–276 (2017)
168. Simmering, J.E., Polgreen, L.A., Polgreen, P.M.: Web search query volume as a measure of pharmaceutical utilization and changes in prescribing patterns. *Res. Soc. Adm. Pharm.* **10**(6), 896–903 (2014)
169. Alshakka, M.A., Ibrahim, M.I.M., Hassali, M.A.A.: Do health professionals have positive perception towards consumer reporting of adverse drug reactions? *J. Clin. Diagn. Res. JCDR* **7**(10), 2181 (2013)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.