Expert Opinion on Drug Safety

ISSN: 1474-0338 (Print) 1744-764X (Online) Journal homepage: http://www.tandfonline.com/loi/ieds20

The usefulness of listening social media for pharmacovigilance purposes: a systematic review

Irma Convertino, Sara Ferraro, Corrando Blandizzi & Marco Tuccori

To cite this article: Irma Convertino, Sara Ferraro, Corrando Blandizzi & Marco Tuccori (2018): The usefulness of listening social media for pharmacovigilance purposes: a systematic review, Expert Opinion on Drug Safety, DOI: 10.1080/14740338.2018.1531847

To link to this article: https://doi.org/10.1080/14740338.2018.1531847

Accepted author version posted online: 04 Oct 2018.

Publisher: Taylor & Francis

Journal: *Expert Opinion on Drug Safety*

DOI: 10.1080/14740338.2018.1531847

Article type: review

The usefulness of listening social media for pharmacovigilance purposes: a systematic review

Irma Convertino[1], Sara Ferraro[1], Corrando Blandizzi[1,2], Marco Tuccori[2]

[1]Unit of Pharmacology and Pharmacovigilance, University of Pisa, Pisa, Italy
[2]Unit of Adverse Drug Reaction Monitoring, University Hospital of Pisa, Pisa, Italy

Corresponding author
Marco Tuccori
Unit of Adverse Drug Reaction Monitoring, University Hospital of Pisa, Medical School, Via Roma 55, 56126, Pisa, Italy
Email: m.tuccori@ao-pisa.toscana.it

Abstract

Introduction: Social media mining could be a possible strategy to retrieve drug safety information. The mining of social media is a complex process under progressive evolution, falling into three broad categories: listening (safety data reporting), engaging (follow-up) and broadcasting (risk communication). This systematic review is aimed at evaluating the usefulness and quality of proto-signals by social media listening.
Areas covered: In this systematic search, performed according to MOOSE and PRISMA statements, we selected studies, published in MEDLINE, EMBASE and Google Scholar until December 31st, 2017, that listened at least one social media to identify proto-adverse drug events and proto-signals.
Expert opinion: The selected thirty-eight studies identified serious and unexpected proto-adverse drug events characterized by poorer information quality as compared with spontaneous reporting databases. This feature allows rarely the evaluation of causal relationships. Proto-signals identified by social media listening had the potential of anticipating pre-specified known signals in only six studies. Moreover, the personal perception of patients reported in social media could be used to implement effective risk communication strategies. However, signal detection in social media cannot be currently recommended for routine pharmacovigilance, due to logistic and technical issues.

Keywords: social media, pharmacovigilance, proto-signal, signal detection

Article highlights

- Social media listening has the potential of identifying serious and unexpected adverse drug events
- Limited evidence suggest that social media can identify some signals earlier than the traditional pharmacovigilance approaches
- The poor quality of information provided in social media comments, rarely allows the evaluation of causal relationships as compared with spontaneous reporting databases. Therefore, drug-event pairs will often need to be corroborated by alternative data sources
- Signal detection in social media cannot be currently recommended for routine pharmacovigilance, due to logistic and technical issues
- In the future, the standardization of methodologies and the development of technologies could make the social media promising sources of information suitable for supporting pharmacovigilance

1. Introduction

The widespread use of social media is likely one of the most revolutionary global phenomenon occurred after 2000. Over those years, social media proliferated and differentiated, creating large people communities sharing passions, needs and interests. The possibility of giving access via web to any kind of personal information, not only to the circle of friends and relatives, but even to unknown peoples, became a sort of appealing highway to popularity. The toll to be payed to take this highway is often the renounce, at least in part, to privacy. As a consequence, the information shared on social media has become a fruitful goldmine for data owners, since they may represent an easy way to achieve personalized advertising. Social media started to be mined initially for commercial purposes and later for other purposes, like political ones[1]. Mining strategies and methodologies were developed and evolved in parallel with the differentiation of these scopes[2].

The Council for International Organizations of Medical Sciences (CIOMS) defined a "signal" as an information that arises from one or multiple sources (including observations and experiments), which suggest a putative causal association, or a new aspect of a known association, between a medical intervention and an event or set of related events, either adverse or beneficial, that is judged to be of sufficient likelihood to justify verification actions[3]. Theoretically, social media could be considered a source of data that might be exploited to retrieve drug safety information[2,4,5]. Several investigations, such as the "WEB-Recognising Adverse Drug Reactions" (Web-RADR), have been then implemented with the aim of providing some directives on 'what and how' to use social media to further proactive pharmacovigilance and protection of public health[6]. The Web-RADR team coined the term "proto-adverse event" (proto-AE) in order to signify "posts with resemblance to AEs', designating posts containing discussion of AEs identified in social media sources" and to distinguish them from the traditional pharmacovigilance regulatory definition[7].

The mining of social media is a complex process under progressive evolution that falls into three broad categories: listening (safety data reporting), engaging (follow-up) and broadcasting (risk communication). Therefore, reviews highlighting deeply the state of the art on the progress of related methodologies[8], ethical aspects[9] and reliability[10] are frequently required. Of note, the usefulness and quality of proto-signals generated by social media listening remain questionable[11], and a systematic review is lacking. In this paper, we reviewed systematically current medical literature, searching for articles that attempted a detection of proto-signals by social media mining, with the aim of evaluating the usefulness of listening social media for pharmacovigilance purposes. This review is aimed at: 1) attempting to summarize different data mining strategies; 2) assessing whether the information quality (i.e. seriousness, notoriety, causality, clinical features) of detected drug-event pairs may be affected by the different scopes of social media; 3) evaluating whether the different methodological approaches used in the studies were able to identify proto-signals that could anticipate the safety warnings issued by Health Authorities.

2. Methods

2.1 Data sources

This systematic review was performed in accordance with PRISMA[12] and MOOSE[13] statements. Studies were selected by using MEDLINE, EMBASE, and Google Scholar.

2.2 Definitions

In this review, social media are intended as computer-based technologies that allow the creation and sharing of information, ideas, interests, and other forms of expression via virtual communities as social networks and forums. Users create service-specific profiles for the website or app that are designed and maintained by the social media organization. The network is populated spontaneously by user-generated contents, and may vary from texts to photos and videos. We further classified social media into two large groups, based on their purposes: medical (e.g. askapatient.com) and non-medical (e.g. Facebook, Twitter, and Google Plus). Each post containing a drug-event pair was analysed as a spontaneous report. In this article, a proto-ADE is defined as a drug-event that emerges in a post published on social media. As well, a proto-signal is identified when an abnormal frequency distribution of a proto-ADE is identified in social media (either by considering a defined threshold or not).

2.3 Search strategy

We conducted a systematic search of studies using the above-mentioned data sources by a combination of the keywords "Pharmacovigilance" and "Social Media". The research was performed in English language without limits of time up to December 31st, 2017. The reference list of selected studies was also checked for additional relevant articles. Duplicates were removed firstly by Mendeley auto-deduplication tool[14], and lastly by manual assessment. Two different reviewers (I.C., S.F.) examined the retrieved papers. The relevance of studies was evaluated by the title and the abstract. If the study eligibility remained unclear, the full text was checked. Any disagreement was resolved by discussion with a senior reviewer (M.T.).

2.4 Study inclusion & exclusion criteria

We included studies performed retrospectively on dataset obtained by at least one identifiable social media (i.e. by name or URL), in accordance with the definition of social media given above. Review articles, sentiment analysis only[15], studies that provided only methodological assessments of text mining strategies, and studies performed on search engines only as data sources were excluded. Notably, when a study used indifferently search engines and social media as data sources, we included it only if data resulting from social media listening were provided separately. Studies in which at least one proto-ADE was not identifiable or for which at least a simple numerical frequency for proto-ADEs was not provided were excluded, as well. Studies performed on proto-ADEs following immunization (vaccines) were not included. Finally, we excluded studies in which the reporting of an ADE was prospectively stimulated by means of questionnaire or by means of *ad hoc* designed forms (active surveys) promoted by the social media managers.

2.5 Data extraction

From each selected study, the following information were collected: authors, year of publication, social media data sources, study population (all users or sub-groups of users and their features), outcomes (identified proto-ADE, clinical features and causality assessment), method of data extraction, aims (survey or signal detection) and conclusion key-points.

2.6 Proto-ADEs classification

Proto-ADEs seriousness, notoriety (expectedness) and causality assessment were recorded when provided by the selected articles. When such information were not reported, seriousness and notoriety were classified according to the European Medicines Agency (EMA) Important Medical Event list of the Medical Dictionary for Regulatory Activities (MedDRA)[16] and the related summary of product characteristics, respectively. Proto-ADEs were clustered according to clinical features identified by means at least of the system organ class (SOC) that was extracted by the original article or attributed by the MedDRA classification. Drugs and drug classes were classified, when necessary, according to their Anatomical Therapeutic Chemical (ATC) classification groups.

2.7 Methodology classification

Study methodologies were analysed and classified by strategy for the selection of the primary data source, and by design (drug-based approach when the identification of proto-ADE begins with the identification of posts containing the name of the drug/drugs of interest, or event-based approach, when the posts are initially selected by the presence of at least one event). A description of their approach to the identification of proto-ADEs is also provided based on the dictionaries, lexicon creation, and strategies for identification of a semantic association. Finally, we classified the studies, based on their approach to signal assessment, in descriptive (the authors provided a simple frequency of the event in the posts of the social media) or analytic (the authors identified proto-signals by using a disproportionality approach such as proportional reporting ratio, or measures of semantic association such as "lift" or "leverage").

3. Results

The selection flowchart is displayed in figure 1. Based on the inclusion and exclusion criteria, thirty-eight studies were identified. Nine studies provided information about the age of users/patients included in

the studies[17–25]. Six of these studies were performed on a population with a mean age or an age range lower than sixty[17–20,23,25]. Eight studies reported information about specific countries or geographical areas where the users/patients were supposed to come from[17,18,22,24,26–29]. Table 1 summarizes the main features of the selected studies. The notoriety of proto-ADEs was provided by the authors in sixteen articles[17,18,21,24,30–41]. Detailed information about proto-ADEs has been displayed in supplementary Table 1.

3.1 Strategies for identification of proto-ADEs and proto-signals

Different social media were investigated to retrieve drug safety information. In eight studies, the authors performed a preliminary selection to identify social media that might represent the optimal source for the data of interest by means of search engines[17,18,24,25,27–29,42]. In twenty-one studies, the rationale and the process of selecting a specific source was not clearly provided[4,19–23,26,34,36–38,41,43–51]. Out of thirty-eight studies, eighteen were performed on a single data source[4,18–21,23,26,30,32,40,41,43–45,47,49,51,52], while the remaining on several ones, ranging from two[25,29,34,35,38,39,46] to twelve social media[28].

There are several approaches to the selection of potential proto-ADEs within a single social media or groups of them. Through the evaluation of the selected studies, we could classify the process of social media listening in a three-step approach, with few exceptions.

In the great majority of studies (n=37), the first identified step was the selection of posts containing the name of the drug or the drugs of interest, either as active ingredient or medical product (drug-based approach)[4,18–53]. The only event-based approach was in the study by Butt et al.[17].

The second step was the selection of the event or events of interest. This can be achieved by different approaches. In some studies, for instance confirmatory studies and studies where social media were tested for their ability of anticipating safety alerts by regulatory agencies, there were few medical terms that could be combined with the drug or drugs of interest[19,25,29,32,35,37,53]. In these cases, standard dictionaries of medical terminologies can be used[20,29,33,35,46,48,53] or, alternatively, cluster of medical terms created *ad hoc* for the study[4,18,19,23,25,42,45,48]. Sometimes, these clusters of codes were generated using a manually revised sample of the dataset [21,24,25,27,28,42,45,47], and sometimes automatically by a word clouds graph or a similar strategy for posts containing the target drug name[18,36]. In particular, nineteen studies combined drugs with a lexicon of all medical terms (listed in standard dictionaries, often developed to contain even colloquial sentences that may be used to describe medical concepts, and sometimes integrated in automatic extraction tools) for the identification of unexpected proto-ADEs[4,18–20,22,30–32,34,38–41,43,44,49–52]. Proto-ADEs could be identified by any post reporting the drug-event pair of interest[4,17,19,21,22,24,26–40,43,44,47–49,51,52] or by any user that posted a drug-event pair of interest[18,20,23,25,42,45,46,53]. This second step is critical for the identification of proto-ADEs. In particular, when the dataset contains an elevated number of posts, filters can be used to reduce the number of posts to a level that allows a final manual validation (i.e. identifying efficiently proto-ADEs, excluding noises, such as redundant comments of the same post, and avoiding false positive, such as when medical terms are quoted in a post to report a drug indication and not a drug-related event)[18,22,26,35,43]. In this regard, nine studies used strategies of semantic association that included other terms that might have been frequently used by social media users to relate drugs with medical events[4,34,38,39,41,49,50,52,53], for improving medical lexicon accuracy. Sometimes, "lift", "leverage", or similar parameters can be calculated as a measure of the strength of semantic association among the words, by calculating the degree of statistical dependence among them[30–32,34,40,41,49,53]. Even the "distance" in terms of words (or tokens) between the drug and the medical event, or "word mapping" within the posts, can be used to increase precision[20,28,36,39,41,47,50,52]. Tools have been developed to fix spelling errors which can be frequent in social media posts[47,52]. Rating systems to express positive or negative feelings (sentiment analysis) about a drug, either available on websites or attributed by the authors, have been sometimes used to refine the research[4,19,23,25,26,35]. There are not studies comparing the efficiency of different approaches. However, six studies evaluated the efficiency of extraction methods by calculating specificity, sensitivity, positive predictive value (precision) or negative predictive value, as compared to a gold standard sample of manually reviewed cases[4,18,36,39,47,50].

Once the dataset of drug-event pairs had been created, the third step was the detection of proto-signals. Twenty-five studies[4,17,19–21,23–26,29,34–37,39,42,43,45–53] identified proto-ADEs by a descriptive approach, and thirteen studies[18,22,27,28,30–33,38,40,41,44] by an analytic strategy. Four studies attempted also a description of the proto-signal based on the few clinical information provided in the posts[28,35,38,44].

3.2 Qualitative evaluation: the identified proto-ADEs

Out of thirty-eight included articles, eleven studies used comments reported on non-medical social media[21,26,35,37,38,40,43,44,48,49,52], twenty-six studies analysed posts on medical message boards[4,17–20,22–25,27–34,36,39,41,42,45,47,50,51,53], and one study mined comments from both the above mentioned sources[46]. The latter study was aimed at investigating detrimental prescribing cascades, particularly for drugs belonging to the following pharmacological classes: antihypertensive, lipid-lowering medications, antidepressants, anti-inflammatory, antineoplastic, anticoagulants, and antimicrobials. The related expected proto-ADEs were both serious and not serious, and involved musculoskeletal and cardiovascular SOCs[46]. Only Pierce et al. showed that social media can provide sufficient clinical information to attempt a causality assessment for a small set of known proto-ADEs (n=13)[35].

3.2.1 Proto-ADEs in non-medical social media

Three out of eleven studies[38,40,52], performed on non-medical social media, focused on the detection of proto-ADEs related to drugs or drug classes selected among the medications most widely mentioned in the posts, and often corresponding to those broadly used in the real word clinical practice (i.e. antidepressants, antipsychotics, hormone therapies and anti-inflammatory drugs)[40,52]. The remaining eight studies were aimed at the detection of proto-ADEs related to specific drugs, that were pre-specified in the study objectives, and included anti-retroviral medications[26], antidiabetics[37], treatment for chronic inflammatory arthropathies[43,44], antipsychotic drugs[21,35], antidepressants[48] and glucocorticoids[49].

Five studies, evaluating comments on non-medical social media, detected only proto-ADEs classified as expected, according to the respective drug labels[21,35,37,43,48]. Unexpected proto-ADEs were identified in the remaining six studies[26,38,40,44,49,52]. These unexpected proto-ADEs were almost not serious and frequently related to withdrawal syndromes[49,52], hangover effects[52] and lack of efficacy[38,44,49,52].

Nine studies were able to detect serious proto-ADEs[21,35,37,38,40,44,48,49,52]. Two of these studies were conducted on pre-specified serious cardiovascular proto-ADEs (i.e. rosiglitazone and stroke and myocardial infarction, clozapine-induced cardiomyopathy or myocarditis, haloperidol-induced cardiomyopathy or myocarditis, and apixaban-induced cerebral haemorrhage)[21,37]; one focused on several drug-adverse event pairs alerted by FDA[35]. When the proto-ADE of interest was not pre-specified, the most frequently detected proto-signals were related to gastrointestinal, psychiatric (mainly sleep disturbances) and injection site. In two studies, the lack of standardized dictionary for proto-ADE codification did not allow seriousness assessment[26,43].

3.2.2 Proto-ADEs in medical social media

Among the twenty-six studies using only medical forums, nineteen analysed data retrieved in networks discussing general healthcare contents[4,18–20,22,23,27,30–34,39,41,45,47,50,51,53], while two used data collected in disease-specific forums, one dedicated to cancer[28] and the other one to Parkinson disease[29]. Finally, five studies created datasets combining posts from disease-specific social media with threads published in general healthcare discussion forums[17,24,25,36,42].

Among the nineteen studies, performed on medical social media discussing general healthcare contents, ten studies detected proto-ADEs regardless of specific therapeutic areas. Indeed, each of these studies investigated several drug classes used for heterogeneous indications[4,27,31–34,39,41,47,50]. Seven studies were focused on pre-specified drugs or drug classes: five studies were dedicated to lipid-lowering agents (n = 5)[19,20,22,51,53], one study to biologic drugs used for arthritis (n = 1)[18], and one study to antipsychotic medications (n = 1)[23]. Two studies evaluated proto-ADEs associated with particular drug issues, namely interactions[30] and off label drug use[45]. The study by Butt at al.[17] is the only one with an event-based approach, thus selecting posts containing specific ADEs (Stevens-Johnson syndrome and toxic epidermal necrolysis), with the aim of identifying the most frequently mentioned associated drugs.

The seven studies carried out on medical social media, including at least a disease-specific forum[24,25,28,29,36,42], detected proto-ADEs related to drugs belonging to homogeneous therapeutic areas, namely the diseases objects of the forum. Examples include antineoplastic drugs in cancer patients forums[24,28,36], antidepressants and antipsychotics in psychiatric patient forums[25], anti-Parkinson treatments in Parkinson message boards[29] and anticoagulants in cardiologic patients networks[42].

Eleven studies, performed on medical social media, detected only labelled proto-ADEs[4,17,53,18,20,23,25,28,29,45,50]. The remaining fifteen studies mined also unexpected proto-ADEs[19,22,24,27,30–34,36,39,41,42,47,51]. These unexpected proto-ADEs were often not serious and related to mild psychiatric/neurological conditions (e.g. nocturnal sweat with tamoxifen or appetite loss with loratadine)[27,36]. Sometimes, serious unexpected proto-ADEs were detected (amyotrophic lateral sclerosis with statins, Parkinson disease with statins and Alzheimer disease with atorvastatin)[22,51]. Notably, some of these studies provided also information about drug effectiveness by checking posts mentioning each drug with its indication, e.g. breast cancer associated with aromatase inhibitors[36], depression related to antidepressants[41], seizures associated with anticonvulsants[33].

Twenty-one studies detected at least one serious proto-ADE[17–20,22–25,28–34,36,41,42,50,51,53]. The most frequently reported proto-ADEs belonged to the following SOCs: gastrointestinal[18,20,25,29,31–33,39,42,47,50,51,53], neurologic[19,23–25,27–29,34,36,39,47,53], psychiatric[25,29,31,34,36,39,41,45,47,50], musculoskeletal[19,20,22,29,30,39,51,53] and cardiovascular[30–33,41,42]. Three studies detected proto-ADEs related to neoplasms[32,41,53].

3.3 Proto–signal detection: could it anticipate the regulatory action?

Among the thirty-eight studies included in this review, the detection of potential signals appearing on social media prior to their first regulatory notification was at least one of the goals of eight studies[19,31,32,35,37,40,44,53]. Five of these studies were descriptive[19,35,37,40,44] and three were analytic[31,32,53] in nature.

3.3.1 Proto-signal detection: descriptive methodology approach

With regard for the descriptive studies, proto-signals were analysed taking into account the timing of drug-safety notifications issued by the U.S. Food and Drug administration (FDA)[19,35,37,40,44] and, for one paper, also by EMA[37].

In three studies[19,35,40], the proto-signal was able to forecast the identification of the drug-related problem by means of a traditional pharmacovigilance approach. Wu et al.[40] used the trends of mentioning known and unknown proto-ADEs in the social media for drugs involved in four known signals (standard) to create a threshold for the identification of a proto-signal. The results showed that the proto-signals originated in social media would have anticipated the regulatory alert for the four known signals by a period ranging from four to six years. In the study by Duh et al.[19], the Granger causality test was used to evaluate the temporal correlation between the numbers of FAERS reports in a given month and social media postings in prior months. This study used two known signals, namely cardiovascular events associated with sibutramine and muscular events associated with atorvastatin, to explore the ability of social media to forecast the identification of the signals in the FAERS database. The study succeeded in the demonstration of anticipating the potential of social media for the signal associated with sibutramine (the proto-signal was detectable eleven months before the safety alert), but not for that associated with atorvastatin. Pierce et al., focused on warnings of ten safety signals for FDA-regulated medications to evaluate their temporal occurrence in non-medical social media compared with the FAERS. This study highlighted that, in one case, the report occurred in social media just eighteen days prior to signal detection from FAERS, whereas the other case occurred first in FAERS[35].

In the other two studies[37,44] the time of proto-signals identification in social media was concomitant or delayed as compared to the date in which the regulatory authority issued the corresponding alert. Bhattacharya et al. compared the trends of discussion of potential proto-ADEs in non-medical social media for a set of specific drugs with the trends of spontaneous reports in FAERS and the company database, taking into account also the information contained in product labels. In their analysis, social media were not able to anticipate the traditional pharmacovigilance reporting system in the identification of signals[44]. Coloma et al. evaluated posts about rosiglitazone and cardiovascular events (i.e. stroke and myocardial infarction) in non-medical websites, and described a scenario in which social media activity ran in parallel with regulatory notification. Indeed, the majority of posts about the proto-signal of interest was published immediately after issuing the corresponding regulatory communications[37].

3.3.2 Proto-signal detection: an analytic methodology approach

The three studies[31,32,53] carried out on medical forums by an analytic evaluation methodology were based on "lift" measure, and their benchmark regulatory body was the FDA.

Feldman et al. focused their research on lipid lowering drugs and antidepressants, and they were able to find proto-signals for statins and bupropion nine and seven years earlier than the FDA alerts, respectively[53]. Two studies[31,32], performed by the same group, aimed at identifying posts related to five drug-event pairs alerted by the FDA, assessed through a temporal analysis whether these posts appeared in social media prior to the first regulatory communication. In these studies, the trend of the lift value associated with each drug-event pair alerted by the FDA was compared to the average lift value of all drugs for that event year by year. Therefore, if the value calculated for a certain drug was significantly higher than the average value for all drugs, this was considered a proto-signal. The first study[31] showed that social media could detect proto-signals from five to fourteen years earlier than the correspondent FDA alert. In the second study[32], the authors assumed that a proto-signal (defined as described above) could occur merely by chance. They observed that in a year-by-year analysis a proto-signal must be outlined for two and three times in consecutive periods preceding FDA's alert or labelling revision, to strengthen the possibility of identifying a true signal. This finding resulted in a shorter forecasting lag time for social media signal detection and the loss of some proto-signals detected in the first study. Nevertheless, the identified proto-signals appeared in social media from eighteen days to thirteen years earlier than the respective regulatory communications.

4. Expert opinion

The present review investigated the usefulness and reliability of social media in the assessment of drug safety issues. Adequate of pharmacovigilance activities should allow at least: 1) to identify as early as possible (or at least support the early identification of) serious and unexpected adverse drug reactions; or 2) to provide useful information for the clinical characterization of drug safety issues that are poorly known; 3) to support the identification and management of risk minimization strategies.

The present systematic review showed that social media listening has the potential detecting serious and unexpected proto-ADEs. However, in the available studies, the initial identification of a proto-ADE was rarely substantiated by a qualitative evaluation of the elements that may stand in favour of a causal relationship. Indeed, the only study attempting a causality assessment was the one by Pierce et al., in a small set of identified proto-ADEs[35]. This is likely because posts describing proto-ADEs uncommonly provide information supporting a temporal plausibility between drug exposure and the event, dechallenge or rechallenge information and elements that would allow the exclusion of alternative causes (e.g. for lateral amyotrophic sclerosis associated with statins reported by Liu et al.[22], we cannot check the time to onset or the presence of alternative causes or risk factors that could substantiate the causative role of statins). A recent French study, published after the censoring time point of the present review, confirmed that information was poorer in medical forums compared with spontaneous reporting databases. Notably, the only relevant parameter for causality assessment evaluated in this study was temporal plausibility (time to onset) available in 24% of the proto-ADEs in forums vs 70% in the French spontaneous ADR reporting database. Therefore, we must expect that these proto-ADEs will need to be often corroborated by means of alternative data sources. Nevertheless, the related proto-signals could be considered of a certain utility, since they may generate hypotheses that could support the early identification of undisclosed drug safety issues.

Signal detection by social media data has shown to be able to forecast signals detected with traditional pharmacovigilance approaches. This evidence comes from a limited number of studies (n=6)[19,31,32,35,40,53] and deserve to be confirmed, but it seems to suggest that, pending and adequate standardization of methodology, it can be of a certain efficiency. In this respect, our review suggests that both descriptive and analytical strategies for signal detection, either performed on non-medical or medical social media, have the potential of anticipating the identification of pre-specified known signals. Further studies, comparing head-to-head sources and approaches, are needed to establish the advantages (if any) of using one strategy over others. The lag period that may elapse from the proto-signal identification in social media and the communication of the correspondent signal by regulatory authority can be even relevant, ranging from few days [35] to some years[31,32]. The two studies[37,44] that did not succeed in the forecasting of known signals likely failed because of the lack of a sufficient observation period. Indeed,

social media sources, similarly to traditional ones, need to accumulate an appropriate volume of data to be effective in signal detection assessments[37]. The study by Duh et al. highlighted also that signal detection in social media could be more effective for drugs used more frequently in younger populations, since they represent the majority of social media users[19].

In the available studies, social media listening was shown to provide poor clinical information, scarcely useful for the characterization of the identified proto-ADE. Indeed, these are frequently characterized only by the gross category of SOC. However, this limitation is somehow balanced by other information usually not available in spontaneous ADR report databases (or any other healthcare source including for instance the quality of life). Furthermore, these proto-ADEs provide information about clinical events associated with prescription drugs that cannot be easily detected by the traditional pharmacovigilance approaches, such as the lack of efficacy, withdrawal syndromes, hangover and drug-dependence[39,44,49,52]. Notably, non-medical social media were involved as sources for the detection of these proto-ADEs more frequently than medical forums. We can hypothesize that medical social media are more suitable for discussing specific proto-ADEs. On the other hand, medical social media likely represent the preferred source when peculiar events or not easily assessable by other sources must be investigated. These may include drug-related cancer[53] or cardiovascular events[41,42]. Unfortunately, the multifactorial nature and the relatively high prevalence of these events in the population keep representing a hurdle for the identification of an association in social media. These limitations are well known even for the traditional pharmacovigilance strategies and can be overcome only by observational studies performed in healthcare systems' databases. Albeit serious proto-ADEs could be detected by both medical and non-medical social media, the latter ones likely perform better with a pre-specified approach (i.e. the a priori definition of a drug-event pair), as performed for example by Bhattacharya et al.[44]. Notably, social media can be used also to investigate peculiar aspects of drug safety, such as drug-drug interactions[30], off label use of drugs[45], and detrimental prescription cascades[46].

There is information minable in social media that can be helpful in issuing drug safety communication and risk minimization strategies. Firstly, social media have been shown to be not only anticipatory of a signal but also reactionary. News about safety issues can be easily discussed in social media, providing useful information about the perception of drug-related risk in the population[37]. Secondly, social media may reveal personal perception of patients about the effect of treatment and, consequently, inappropriate behaviour that may enhance the risk of adverse events or even low-treatment compliance[44]. This information can be used to prompt effective risk communication strategies, including recommendations for the appropriate use of drugs.

A standard approach could be likely a key-element to perform an effective signal detection in social media. The use of standardized dictionaries, which may include vernacular terms, for the definition of medical events is essential in the phase of data extraction and dataset construction, especially when large non-medical social media are investigated. Heterogeneous approaches to data extraction have been described, ranging from the manual revision of the posts to the use of natural language processing and the subsequent identification of semantic association, suggestive of a relationship between a drug and an event mentioned in the same post. However, there is a lack of studies that have performed head-to-head comparisons across methodologies, and therefore we cannot recommend an approach that could be preferred over others. The drug-based design is largely the most frequently used in these studies with the only exception of Butt et al.[17], in which the authors selected posts related to the events, namely Stevens - Johnson syndrome or toxic epidermal necrolysis (event-based approach). The latter strategy is expected to be more effective when the event is typically drug-induced. In summary, social media have the potential of being a useful source for information on drugs safety. However, the possibility of using these data for practical regulatory purposes remains a hard challenge. Given the high number of drugs to be monitored, a routine practice of broad-ranging social media for signal detection would be logistically impossible even for a limited group of drugs, such as those included in the portfolio of a marketing authorization holder (MAH). Therefore, committing such an activity to MAHs would be unreasonable, since it would require human and technological resources that are not currently available. The current European legislation is somewhat (and likely deliberately, in view of the uncertainties surrounding the effectiveness of the practice) open to interpretation with unclear legal implications for MAHs and regulators, recommending that all individual case safety reports (ICSRs) on digital media being captured, recorded and reported in accordance with the

law[9]. Even in the case proto-ADE, identified by routine performed social media signal detection, could provide the minimum information required to be reported by traditional spontaneous reporting pathways, to the best of our knowledge there is no evidence that this practice may produce any benefit to early identification of drug safety issues. Our review supports the view[54] that the identification of proto-signals could be useful as an integrate activity that must be performed in parallel, and surely not in replacement of traditional signal detection. Notably, given the extreme uncertainty of their nature, proto-signals would require the definition of a specific communication pathway, and it is unlikely that these can be used alone as evidence in decision making processes.
In the future, technological progress will probably allow a periodic routine mining that could be logistically reliable. In the meantime, social media listening cannot be a routine practice, but, more often, could be used to check specific information, for which social media was proven to be a suitable observatory, particularly for those related to risk perception under a condition of crisis.

5. Conclusions

Over the last years, social media have been considered a promising source of drug safety information. Our review shows that social media hold actually the potential of identifying undisclosed drug safety issues for different drug classes. Moreover, social media could be able to forecast the identification of drug safety signals, highlighted by traditional pharmacovigilance approaches (i.e. spontaneous reporting). Unfortunately, the quality of information provided in the posts usually do not allow the causality assessment of proto-ADEs. Therefore, social media cannot be used for signal detection without being integrated by data from other sources, particularly, spontaneous ADR reporting. Signal detection in social media cannot be currently recommended for routine pharmacovigilance practice due to logistic and technical issues (i.e. standardized methodologies). However, social media can be used as a privileged observatory to evaluate risk perception in the population and represent a promising source of information to address effective drug safety communication strategies.

Funding
This paper was not funded.

Declaration of interest
The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

Reviewer disclosures
Peer reviewers on this manuscript have no relevant financial or other relationships to disclose.

References

Papers of special note have been highlighted as either of interest (*) or of considerable interest (**) to readers.

1. Brady JT, Kelly ME, Stein SL. The Trump Effect: With No Peer Review, How Do We Know What to Really Believe on Social Media? *Clin Colon Rectal Surg*. 2017;30(4):270-276. doi:10.1055/s-0037-1604256
2. Lardon J, Abdellaoui R, Bellet F, et al. Adverse drug reaction identification and extraction in social media: A scoping review. *J Med Internet Res*. 2015;17(7):1-16. doi:10.2196/jmir.4304
3. CIOMS. Practical aspects of signal detection in pharmacovigilance. Report of CIOMS Working Group VIII. 2012.
4. Patki A, Sarker A, Pimpalkhute P, et al. Mining Adverse Drug Reaction Signals from Social Media: Going Beyond Extraction. 2014. http://diego.asu.edu/drugstats/Publication_repository/2014/patki_biolinksig/mining adverse drug reaction signals from social media-going beyond extraction.pdf. Accessed June 6, 2018.
5. Golder S, Norman G, Loke YK. Systematic review on the prevalence, frequency and comparative value of adverse events data in social media. *Br J Clin Pharmacol*. 2015;80(4):878-888. doi:10.1111/bcp.12746
6. Ghosh R, Lewis D. Aims and approaches of Web-RADR: a consortium ensuring reliable ADR reporting via mobile devices and new insights from social media. *Expert Opin Drug Saf*. 2015;14(12):1845-1853. doi:10.1517/14740338.2015.1096342
7. Freifeld CC, Brownstein JS, Menone CM, et al. Digital Drug Safety Surveillance: Monitoring Pharmaceutical Products in Twitter. *Drug Saf*. 2014;37(5):343-350. doi:10.1007/s40264-014-0155-x
8. Sarker A, Ginn R, Nikfarjam A, et al. Utilizing social media data for pharmacovigilance: A review. *J Biomed Inform*. 2015;54:202-212. doi:10.1016/j.jbi.2015.02.004
9. Sloane R, Osanlou O, Lewis D, Bollegala D, Maskell S, Pirmohamed M. Social media and pharmacovigilance: A review of the opportunities and challenges. *Br J Clin Pharmacol*. 2015;80(4):910-920. doi:10.1111/bcp.12717
10. Rees S, Mian S, Grabowski N. Therapeutic Advances in Drug Safety. *Ther Adv Drug Saf*. 2018:1-9. doi:10.1177/2042098618789596
11. Tricco AC, Zarin W, Lillie E, Pham B, Straus SE. Utility of social media and crowd-sourced data for pharmacovigilance: A scoping review protocol. *BMJ Open*. 2017;7(1):1-5. doi:10.1136/bmjopen-2016-013474
12. Shamseer L, Moher D, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (prisma-p) 2015: Elaboration and explanation. *BMJ*. 2015;349(January):1-25. doi:10.1136/bmj.g7647
13. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of Observational Studies. 2000. doi:10.1001/jama.283.15.2008
14. Kwon Y, Lemieux M, McTavish J, Wathen N. Identifying and removing duplicate records from systematic review searches. *J Med Libr Assoc*. 2015;103(4):184-188. doi:10.3163/1536-5050.103.4.004
15. Serrano-Guerrero J, Olivas JA, Romero FP, Herrera-Viedma E. Sentiment analysis: A review and comparative analysis of web services. *Inf Sci (Ny)*. 2015;311:18-38. doi:10.1016/J.INS.2015.03.040
16. Brown EG, Wood L, Wood S. The Medical Dictionary for Regulatory Activities (MedDRA). *Drug Saf*. 1999;20(2):109-117. doi:10.2165/00002018-199920020-00002
17. Butt TF, Cox AR, Oyebode JR, Ferner RE. Internet Accounts of Serious Adverse Drug Reactions. *Drug Saf*. 2012;35(12):1159-1170. doi:10.1007/BF03262001
18. Curtis JR, Chen L, Higginbotham P, et al. Social media for arthritis-related comparative effectiveness and safety research and the impact of direct-to-consumer advertising. *Arthritis Res Ther*. 2017;19(1):48. doi:10.1186/s13075-017-1251-y
19. Duh MS, Cremieux P, Audenrode M Van, et al. Can social media data lead to earlier detection of drug-related adverse events? *Pharmacoepidemiol Drug Saf*. 2016;25(12):1425-1433.

doi:10.1002/pds.4090
20. Karimi S, Metke-Jimenez A, Kemp M, Wang C. Cadec: A corpus of adverse drug event annotations. *J Biomed Inform*. 2015;55:73-81. doi:10.1016/J.JBI.2015.03.010
21. Koutkias VG, Lillo-Le Louët A, Jaulent M-C. Exploiting heterogeneous publicly available data sources for drug safety surveillance: computational framework and case studies. *Expert Opin Drug Saf*. 2017;16(2):113-124. doi:10.1080/14740338.2017.1257604
22. Liu J, Li A, Seneff S. Automatic Drug Side Effect Discovery from Online Patient-Submitted Reviews: Focus on Statin Drugs. 2011. http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=71E46AE1B1271336F129B9A8811369E5?doi=10.1.1.221.831&rep=rep1&type=pdf. Accessed June 6, 2018.
23. Moncrieff J, Cohen D, Mason JP. The subjective experience of taking antipsychotic medication: a content analysis of Internet data. *Acta Psychiatr Scand*. 2009;120(2):102-111. doi:10.1111/j.1600-0447.2009.01356.x
24. Pages A, Bondon-Guitton E, Montastruc JL, Bagheri H. Undesirable Effects Related to Oral Antineoplastic Drugs: Comparison Between Patients' Internet Narratives and a National Pharmacovigilance Database. *Drug Saf*. 2014;37(8):629-637. doi:10.1007/s40264-014-0203-6
25. Hughes S, Cohen D. Can online consumers contribute to drug knowledge? A mixed-methods comparison of consumer-generated and professionally controlled psychotropic medication information on the internet. *J Med Internet Res*. 2011;13(3):e53. doi:10.2196/jmir.1716
26. Adrover C, Bodnar T, Huang Z, Telenti A, Salathé M. Identifying Adverse Effects of HIV Drug Treatment and Associated Sentiments Using Twitter. *JMIR public Heal Surveill*. 2015;1(2):e7. doi:10.2196/publichealth.4488
27. Hadzi-Puric J, Grmusa J. Automatic Drug Adverse Reaction Discovery from Parenting Websites Using Disproportionality Methods. In: *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. IEEE; 2012:792-797. doi:10.1109/ASONAM.2012.144
28. Mao JJ, Chung A, Benton A, et al. Online discussion of drug side effects and discontinuation among breast cancer survivors. *Pharmacoepidemiol Drug Saf*. 2013;22(3):256-262. doi:10.1002/pds.3365
29. Schröder S, Zöllner YF, Schaefer M. Drug related problems with Antiparkinsonian agents: consumer Internet reports versus published data. *Pharmacoepidemiol Drug Saf*. 2007;16(10):1161-1166. doi:10.1002/pds.1415
30. Yang H, Yang CC. Harnessing Social Media for Drug-Drug Interactions Detection. In: *2013 IEEE International Conference on Healthcare Informatics*. IEEE; 2013:22-29. doi:10.1109/ICHI.2013.10**
31. Yang CC, Yang H, Jiang L. Postmarketing Drug Safety Surveillance Using Publicly Available Health-Consumer-Contributed Content in Social Media. *ACM Trans Manag Inf Syst*. 2014;5(1):1-21. doi:10.1145/2576233**
32. Yang H, Yang CC. Using Health-Consumer-Contributed Data to Detect Adverse Drug Reactions by Association Mining with Temporal Analysis. *ACM Trans Intell Syst Technol*. 2015;6(4):1-27. doi:10.1145/2700482**
33. Yeleswarapu S, Rao A, Joseph T, Saipradeep VG, Srinivasan R. A pipeline to extract drug-adverse event pairs from multiple data sources. *BMC Med Inform Decis Mak*. 2014;14:13. doi:10.1186/1472-6947-14-13
34. Zheng Y, Lan C, Peng H, Li J. Using constrained information entropy to detect rare adverse drug reactions from medical forums. In: *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE; 2016:2460-2463. doi:10.1109/EMBC.2016.7591228
35. Pierce CE, Bouri K, Pamer C, et al. Evaluation of Facebook and Twitter Monitoring to Detect Safety Signals for Medical Products: An Analysis of Recent FDA Safety Alerts. *Drug Saf*. 2017;40(4):317-331. doi:10.1007/s40264-016-0491-0
36. Benton A, Ungar L, Hill S, et al. Identifying potential adverse effects using the web: A new approach to medical hypothesis generation. *J Biomed Inform*. 2011;44(6):989-996. doi:10.1016/J.JBI.2011.07.005
37. Coloma PM, Becker B, Sturkenboom MCJM, van Mulligen EM, Kors JA. Evaluating Social Media Networks in Medicines Safety Surveillance: Two Case Studies. *Drug Saf*. 2015;38(10):921-930.

doi:10.1007/s40264-015-0333-5
38. Powell GE, Seifert HA, Reblin T, et al. Social Media Listening for Routine Post-Marketing Safety Surveillance. *Drug Saf*. 2016;39(5):443-454. doi:10.1007/s40264-015-0385-6
39. Sampathkumar H, Chen X, Luo B. Mining Adverse Drug Reactions from online healthcare forums using Hidden Markov Model. *BMC Med Inform Decis Mak*. 2014;14(1):91. doi:10.1186/1472-6947-14-91
40. Wu H, Hall E, Wu H, Fang H, Stanhope SJ. Exploiting Online Discussions to Discover Unrecognized Drug Side Effects. *Methods Inf Med*. 2013;2(52):152-159. doi:10.3414/ME12-02-0004
41. Yang CC, Haodong Yang D, Jiang L, Tang X. Detecting Signals of Adverse Drug Reactions from Health Consumer Contributed Content in Social Media. 2012. http://wan.poly.edu/KDD2012/forms/workshop/HI-KDD12/doc/paper_14.pdf. Accessed June 6, 2018.**
42. Vaughan Sarrazin MS, Cram P, Mazur A, Ward M, Reisinger HS. Patient perspectives of dabigatran: analysis of online discussion forums. *Patient*. 2014;7(1):47-54. doi:10.1007/s40271-013-0027-y
43. Martínez-López De Castro N, Samartín- Ucha M, Martín-Vila A, Álvarez-Payero M, Piñeiro-Corrales G, Pego-Reigosa JM. Content analysis of Twitter in relation to biological treatments for chronic inflammatory arthropathies: an exploratory study. *Eur J Hosp Pharm*. January 2018:ejhpharm-2017-001402. doi:10.1136/ejhpharm-2017-001402
44. Bhattacharya M, Snyder S, Malin M, et al. Using Social Media Data in Routine Pharmacovigilance: A Pilot Study to Identify Safety Signals and Patient Perspectives. *Pharmaceut Med*. 2017;31(3):167-174. doi:10.1007/s40290-017-0186-6
45. 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*. 2011;13(1):e6. doi:10.2196/jmir.1643
46. Hoang T, Liu J, Pratt N, et al. Detecting signals of detrimental prescribing cascades from social media. *Artif Intell Med*. 2016;71:43-56. doi:10.1016/j.artmed.2016.06.002
47. 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. 2010:117-125. http://delivery.acm.org/10.1145/1870000/1869976/p117-leaman.pdf?ip=131.114.151.148&id=1869976&acc=OPEN&key=296E2ED678667973.A7DB88CBAA293B6B.4D4702B0C3E38B35.6D218144511F3437&__acm__=1528317145_e9b0eecf6b6136bad22288f0709b4e5e. Accessed June 6, 2018.
48. Nguyen T, Larsen ME, O'Dea B, Phung D, Venkatesh S, Christensen H. Estimation of the prevalence of adverse drug reactions from social media. *Int J Med Inform*. 2017;102:130-137. doi:10.1016/J.IJMEDINF.2017.03.013
49. Patel R, Belousov M, Jani M, et al. Frequent discussion of insomnia and weight gain with glucocorticoid therapy: an analysis of Twitter posts. *npj Digit Med*. 2018;1(1):20177. doi:10.1038/s41746-017-0007-z
50. P Tafti A, Badger J, LaRose E, et al. Adverse Drug Event Discovery Using Biomedical Literature: A Big Data Neural Network Adventure. *JMIR Med informatics*. 2017;5(4):e51. doi:10.2196/medinform.9170
51. Topaz M, Lai K, Dhopeshwarkar N, et al. Clinicians' Reports in Electronic Health Records Versus Patients' Concerns in Social Media: A Pilot Study of Adverse Drug Reactions of Aspirin and Atorvastatin. *Drug Saf*. 2016;39(3):241-250. doi:10.1007/s40264-015-0381-x
52. O'Connor K, Pimpalkhute P, Nikfarjam A, Ginn R, Smith KL, Gonzalez G. Pharmacovigilance on twitter? Mining tweets for adverse drug reactions. *AMIA . Annu Symp proceedings AMIA Symp*. 2014;2014:924-933. http://www.ncbi.nlm.nih.gov/pubmed/25954400. Accessed June 6, 2018.
53. Feldman R, Netzer O, Peretz A, Rosenfeld B. Utilizing Text Mining on Online Medical Forums to Predict Label Change due to Adverse Drug Reactions. In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '15*. New York, New York, USA: ACM Press; 2015:1779-1788. doi:10.1145/2783258.2788608**
54. Caster O, Dietrich J, Kürzinger ML, et al. Assessment of the Utility of Social Media for Broad-Ranging Statistical Signal Detection in Pharmacovigilance: Results from the WEB-RADR Project. *Drug Saf*.

2018. doi:10.1007/s40264-018-0699-2

Figure 1 **Flow chart of the search results**

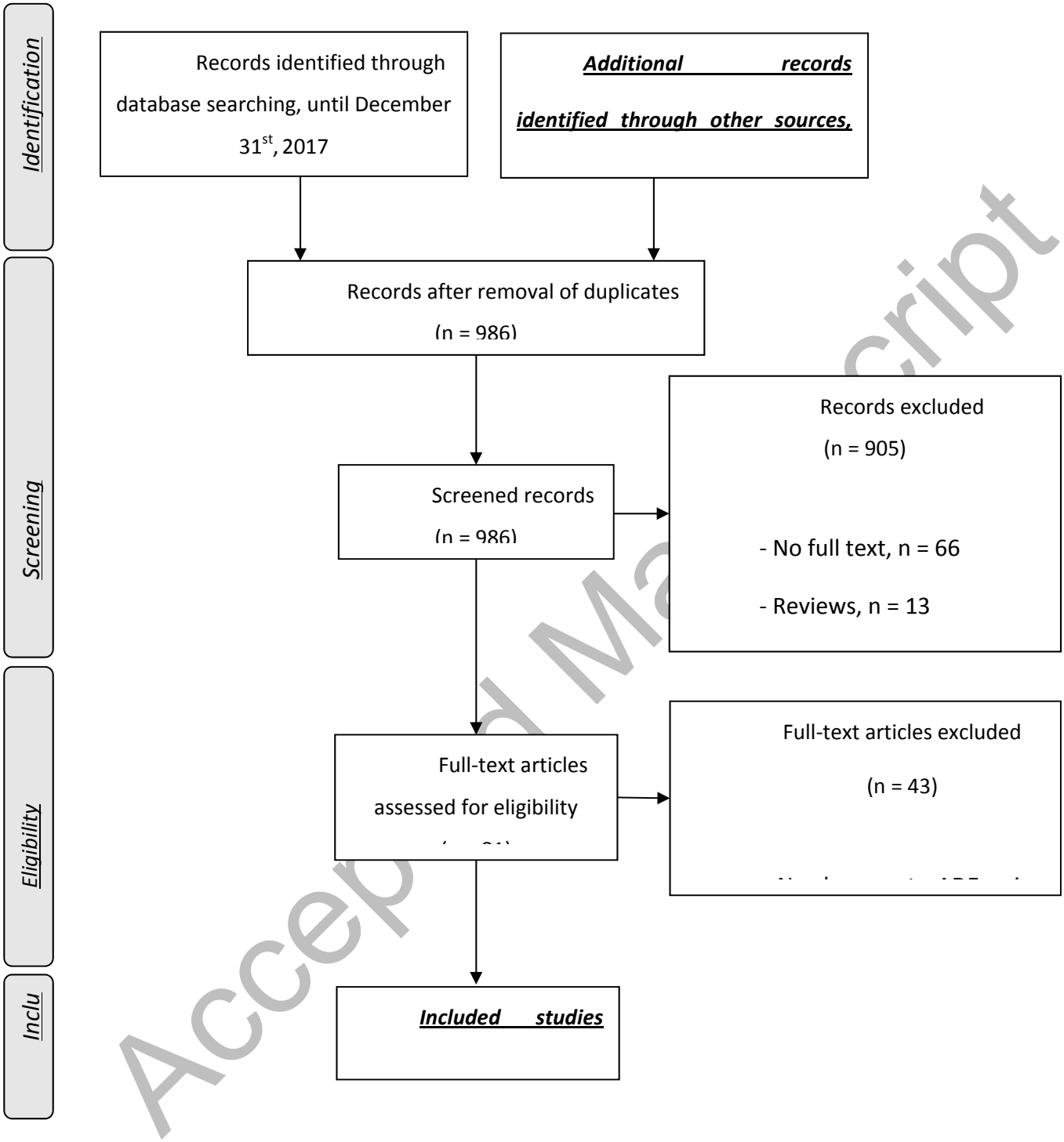

Table 1: Characteristics of the included studies

Author, year	Social media	Post reporting details (sampling keywords, time period, age (Y), country)	Proto-ADEs (notoriety*, seriousness**, clinical features***, suspected drugs****)	Summary of conclusions
Adrover C., 2015[26]	Not-medical (Twitter)	○ anti-HIV drugs ○ 2010-2013 ○ unknown ○ US, UK, South Africa and Canada	○ expected and unexpected ○ not assessable seriousness§ ○ general ○ anti-retroviral drugs	Although the study extrapolated a small size of reports related to anti-HIV treatments, due to the restricted setting, it identified only well-known proto-ADEs.
Benton A., 2011[36]	Medical (Cancer and Health message boards[a])	○ hormonal breast cancer drugs ○ unknown ○ unknown ○ unknown	○ expected and unexpected ○ serious and not serious ○ general ○ tamoxifene, anastrozole, exemestane, letrozole	This approach was able to extract several proto-ADEs from a corpus of breast cancer message board posts, 20% of which were unexpected.
Bhattacharya M., 2017[44]	Not-medical (Epidemico)	○ six pre-specified drugs ○ January 2014-February 2016 ○ unknown ○ unknown	○ expected and unexpected ○ serious and not serious ○ general ○ adalimumab	The use of traditional Pharmacovigilance methods to analyze social media data was unsuccessful. Social media data analysis could not detect new or previously identified safety signals for the selected drugs.
Butt T.F., 2012[17]	Medical (http://www.sjsupport.org; http://www.patient.co.uk; http://www.joeway.co.uk; http://www.milnesjs.com)	○ SJS and TEN ○ 2009-2010 ○ 17.5 (mean) ○ US e UK	○ expected ○ serious ○ SJS and TEN ○ several drugs	Internet descriptions of drug-induced SJS or TEN by patients and their relatives could help to provide health professionals with a deeper insight into patient experience of these serious ADRs.
Coloma P.M., 2015[37]	Not-medical (Facebook, Google +, Twitter)	○ case studies (rosiglitazone-stroke and rosiglitazone-myocardial infarction) ○ 2006-September 2014 ○ unknown ○ unknown	○ expected ○ serious ○ cardiovascular ○ rosiglitazone	Publicly available data from the considered social media networks were sparse and largely untrackable for the purpose of providing early clues of safety concerns regarding the pre-specified case studies.
Curtis J.R., 2017[18]	Medical (Treato)	○ rheumatoid or psoriatic arthritis drugs ○ October 2015 ○ <40 (61% of patients) ○ US (75% of patients)	○ expected ○ serious and not-serious ○ general ○ several drugs	Social media is a challenging yet promising data source that complement traditional approaches for comparative effectiveness research for new medications.
De Castro N.M.L., 2018[43]	Not-medical (Twitter)	○ chronic inflammatory arthropathies drugs ○ May 2015-June 2015 ○ unknown ○ unknown	○ expected ○ not assessable seriousness§ ○ general ○ biological drugs for chronic inflammatory arthropathies	Learning more about subjects dealt with in the tweets will enable us to improve our understanding of the areas of greater interest and concern among patients. This could help caregivers to establish patients focused strategies aimed at addressing their needs.

Author, year	Social media	Post reporting details (sampling keywords, time period, age (Y), country)	Proto-ADEs (notoriety*, seriousness**, clinical features***, suspected drugs****)	Summary of conclusions
Duh M.S., 2016[19]	Medical (Askapatient.com)	o sibutramine and atorvastatin o 2001-December, 2014 o 44-54 o unknown	o expected and unexpected o serious and not serious o general o atorvastatin and sibutramine	Social media proto-ADE reporters were younger and focused on less-serious and fewer types of ADEs than FAERS reporters. The potential for social media to provide earlier proto-ADEs compared with ADEs reported in FAERS is uncertain.
Continued				
Feldman R., 2015[53]	Medical (medhelp.org; exchanges.webmd.com; healthboards.com; ehealthforum.com)	o anti-depressants, cholesterol-lowering drugs o 1999-2013 o unknown o unknown	o expected o serious and not serious o general o anti-depressants and cholesterol-lower drugs	This study was able to predict drug-ADR relations that were unexpected at the time of their mention in the medical forums. Medical forums could forecast safety signals (drug-ADR pairs) reported by FDA.
Frost J., 2011[45]	Medical (PatientsLikeMe.com)	o amitriptyline and modafinil o until May, 2010 o unknown o unknown	o expected o not serious o general o amitriptyline and modafinil	Patient-reported outcomes offered a unique real-time approach to understand utilization and performance of treatments across many conditions. Patients, sharing their data online, could provide relevant and timely information about off-label prescribing and related harmful effects.
Hadzi-Puric J., 2012[27]	Medical (Parenting web sites)	o pre-specified drugs o 2005-2012 o unknown o Serbia	o expected and unexpected o not serious o general o several drugs	Health-related social media could represent an optimal data set for pharmacovigilance. The proposed method could be a valid approach for post marketing surveillance and identification of side-effects.
Hoang T., 2016[46]	Medical and not-medical (http://patient.info/forums, Twitter)	o pre-specified drugs pairs o July 2005-August 2015 o unknown o unknown	o expected o not serious o general o cascade drug prescription	This study showed that social media could be mined to identify detrimental prescribing cascades, including potential unexpected ones.
Hughes S., 2011[25]	Medical (www.askapatient.com; www.crazymeds.us)	o escitalopram and quetiapine o February 2009 o 19-54 (about 43%) o unknown	o expected o serious and not serious o general o escitalopram and quetiapine	Consumer reviews and professional medication descriptions generally reported similar effects but differed in their descriptions and in frequency of reporting.
Karimi S., 2015[20]	Medical (www.askapatient.com)	o several drugs o January 2001-September 2013 o 17-84 (mean 52) o unknown	o expected o serious and not serious o general o diclofenac and atorvastatin	The analysis of post contents provided a corpus of terms able to extract proto-ADEs from layperson reports, automatically.
Koutkias V.G., 2017[21]	Not-medical (Twitter)	o case studies (5 drugs-side effect pairs) o 2006-2016 o 6-92 o unknown	o expected o serious o cardiovascular o clozapine, apixaban and haloperidol	This work contributed in establishing a continuous learning system for drug safety surveillance by exploiting heterogeneous publicly available data sources via appropriate support tools.

Author, year	Social media	Post reporting details (sampling keywords, time period, age (Y), country)	Proto-ADEs (notoriety*, seriousness**, clinical features***, suspected drugs****)	Summary of conclusions
Leaman R., 2010[47]	Medical (Dailystrenght)	○ several drugs ○ 2007-unknown ○ unknown ○ unknown	○ expected and unexpected ○ not serious ○ general ○ olanzapine, carbamazepine, trazodone, ziprasidone, aspirin and ciprofloxacin	This study showed that user comments in health social networks contain relevant information for pharmacovigilance purposes.
Continued				
Liu J., 2011[22]	Medical (AskPatient.com, Medications.com, WebMD.com)	○ statins ○ unknown ○ 40-70 (86% of patients) ○ US	○ expected and unexpected ○ serious and not serious ○ general clinical features ○ statins	Web-based databased could be mined for the association of prescription drugs and proto-ADEs. Many of the findings are supported by the research literature on statins.
Mao J.J., 2013[28]	Medical (cancer message boards [b])	○ aromatase inhibitors ○ 2002-2010 ○ unknown ○ US	○ expected ○ serious and not serious ○ general ○ aromatase inhibitors	Online discussion was often related to drug switching and discontinuation. Physicians should be aware of these discussions and guide patients to effectively manage ADEs and promote optimal adherence.
Moncrieff J., 2009[23]	Medical (Askapatient.com)	○ antipsychotic drugs ○ unknown ○ 34 mean ○ unknown	○ expected ○ serious and not serious ○ nervous system ○ old antipsychotics, risperidone and olanzapine	The generalizability of Internet data is uncertain. However, the data suggested that adverse subjective effects play a central role in the experience of taking antipsychotic drugs and may be related to the drugs' desired benefits.
Nguyen T., 2017[48]	Not-medical (LiveJournal, Reddit, Twitter)	○ most frequently prescribed psychiatric drugs ○ May 1999-August 2015 ○ unknown ○ unknown	○ expected ○ serious and not-serious ○ general ○ psychiatric drugs	This study investigated the potential of applying a novel neural learning framework, words embedding representation, to estimate proto-ADEs for psychiatric drugs from social media, in comparison with conventional methods where a fixed lexicon of ADRs is given.
O'Connor K., 2014[52]	Not-medical (Twitter)	○ several drugs ○ August 2013-February 2014 ○ unknown ○ unknown	○ expected and unexpected ○ serious and not-serious ○ general ○ several drugs	Tweets were about adverse effects experiences related to prescription drugs. Patients mentioned the drug name, along with proto-ADEs, making it possible to automatically extract the drug and proto-ADE relationship.

Author, year	Social media	Post reporting details (sampling keywords, time period, age (Y), country)	Proto-ADEs (notoriety*, seriousness**, clinical features***, suspected drugs****)	Summary of conclusions
Pages A., 2014[24]	Medical (Doctissimo, Sante-medecine, Aufeminin, E-sante, Alte-asso)	o antineoplastic drugs o 2011 (one year) o 21-71 o France	o expected and unexpected o serious and not- serious o general o several antineoplastic drugs	The study suggested how patient websites could be useful to detect proto-ADEs alongside conventional pharmacovigilance.
Patel R., 2018[49]	Not-medical (Twitter)	o prednisone, prednisolone o October 2012 – June 2015 o unknown o unknown	o expected and unexpected o serious and not-serious o general o prednisone and prednisolone	Pharmacovigilance using Twitter data has the potential to be a valuable, supplementary source of drug safety information. In particular, it can illustrate which drug side effects patients discuss most commonly, potentially because of important impacts on quality of life. This information could help clinicians to inform patients about frequent and relevant not-serious side effects as well as more serious side effects.
Continued				
Patki A., 2014[4]	Medical (DailyStrength)	o drugs prescribed for chronic diseases o 2013 o unknown o unknown	o expected o not serious o neurological o several drugs	This paper proposed an approach for classifying drugs into normal and black-box categories, based on the automatic classification of comments extracted from social media. The result obtained while promising with regard for the individual classification of comments as ADRs or no-ADRs, are marginal with respect to the overall classification of the drugs.
Pierce, 2017[35]	Not-medical (Twitter, Facebook)	o several drugs o March 2009 – October 2014 o unknown o unknown	o expected o serious o general o several drugs	An efficient semi-automated approach to social media monitoring may provide earlier insights into certain adverse events. More work is needed to elaborate additional uses for social media data in pharmacovigilance and to determine how they can be applied by regulatory agencies.
Powell G.E., 2016[38]	Not-medical (Twitter, Facebook)	o several drugs o October 2012-October 2014 o unknown o unknown	o expected and unexpected o serious and not serious o general o several drugs	Social media listening was an important tool to improve the post-marketing safety surveillance process.
Sampathkumar H., 2014[39]	Medical (www.medications.com; www.steadyhealth.com)	o several drugs o June 2012-October 2012 o unknown o unknown	o expected and unexpected o not serious o general o several drugs	This study showed that the information extracted from healthcare forums matched those reported in Drug labels. In addition, detected unknown proto-ADEs could act as early indicators for health authorities to help in their efforts towards Pharmacovigilance.

Author, year	Social media	Post reporting details (sampling keywords, time period, age (Y), country)	Proto-ADEs (notoriety*, seriousness**, clinical features***, suspected drugs****)	Summary of conclusions
Schröder S., 2007[29]	Medical (www.parkinsonline-forum.de, www.parkinson-web.de)	o anti-Parkinson drugs o 2004-2005 o unknown o Germany	o expected o serious and not serious o general o anti-Parkinson drugs	Online forums may be considered as a suitable source of observational information to complement data from randomized clinical trials.
Tafti A.P., 2017[50]	Medical (MedHelp, Patient, WebMD)	o several drugs o unknown o unknown o unknown	o expected o serious and not serious o general o several drugs	The study suggested that clinical data and patients' concerns from social media might be used to gain insights on similarities and differences between patients' and providers' perceptions of drug-related risk.
Topaz M., 2016[51]	Medical (Treato)	o atorvastatin and aspirin o unknown o unknown o unknown	o expected and unexpected o serious and not serious o general o aspirin and atorvastatin	The study suggested that clinical data and patients' concerns from social media might be used to gain insights on similarities and differences between patients' and providers' perceptions.
Vaughan Sarrazin M.S., 2014[42]	Medical (patients message boards[c])	o dabigatran o January 2011-September 2012 o unknown o unknown	o expected and unexpected o serious and not serious o general o dabigatran	Online communities could provide information about potential topics, but may not be discernible in clinical trials and are not a priority to researchers. Nevertheless, these topics can be a relevant concern to patients.
Continued				
Wu H., 2013[40]	Not-medical (Google Discussion Search)	o several drugs o unknown o unknown o unknown	o expected and unexpected o serious and not serious o general o several drugs	This approach can serve as a complementary tool for drug companies and regulatory agencies to receive feedback from patients and it has the potential for expedite the discovery process of unrecognized drug side effects.
Yang C.C., 2012[41]	Medical (MedHelp)	o case studies (five drug-side effect pairs) o unknown o unknown o unknown	o expected and unexpected o serious and not serious o general o biaxin, lansoprazole, luvox, prozac, tacrolimus	This approach was able to effectively detect ADRs that were already alerted by FDA. However, adverse drug reaction such as cancer cannot be detected effectively.
Yang C.C., 2014[31]	Medical (MedHelp)	o several drug o 1995-2012 o unknown o unknown	o expected and unexpected o serious and not serious o general o several drugs	The result shows that the performance of harnessing health related social media contents to detect adverse drug reaction is good and promising.
Yang H., 2013[30]	Medical (MedHelp)	o several drug pairs o unknown o unknown o unknown	o expected and unexpected o serious and not serious o musculoskeletal and cardiovascular o potential interacting drugs.	This method was able to effectively detect drug-drug interactions reported by DrugBank, and it could be exploited in the future for mining unknown drug-drug interactions.

Author, year	Social media	Post reporting details (sampling keywords, time period, age (Y), country)	Proto-ADEs (notoriety*, seriousness**, clinical features***, suspected drugs****)	Summary of conclusions
Yang H., 2015[32]	Medical (MedHelp)	○ ten drugs ○ 1997-2012 ○ unknown ○ unknown	○ expected and unexpected ○ serious and not serious ○ general ○ several drugs	This study showed that health-related social media are a promising source for ADRs detection, and an effective way to identify early ADRs signals.
Yeleswarapu S., 2014[33]	Medical (PatientsLikeMe; DailyStrength; MediGuard)	○ several drugs ○ 2008-2012 ○ unknown ○ unknown	○ expected and unexpected ○ serious and not serious ○ general ○ several drugs	Mining social media cannot be themselves an alternative for ADEs detection, but this approach could help to substantiate the adverse event databases, by detecting not only known ADEs, but also unknown and unreported proto-ADEs, which could be analyzed further.
Zheng Y., 2016[34]	Medical (www.steadyhealth.com Post; www.medhelp.org)	○ several drugs ○ unknown ○ unknown ○ unknown	○ expected and unexpected ○ serious and not serious ○ general ○ several drugs	This approach was able to filter out irrelevant drug-ADRs pairs (including beneficial reactions) and detect rare ADRs which were hard to identify by co-occurrence-based method.

anti-HIV: anti - human immunodeficiency virus; proto-ADE: proto-adverse drug event; SJS: Stevens Johnson Syndrome; TEN: Toxic Epidermal Necrolysis; UK: United Kingdom; US: United States

* reported in drug labels; ** evaluated in accordance with important medical events lists;*** identified by means of the System Organ Class;**** identified by means of the Anatomical Therapeutic Chemical classification groups; § not assessable seriousness: the resulted proto-ADEs were reported as system organ class or as vernacular terms.

[a] breastcancer.org; komen.org; csn.cancer.org; bcsupport.org; healthboards.com; cancercompass.com; webmd.com; dailystrength.org; revolutionhealth.com; ehealthforum.com; oprah.com.

[b] http://community.breastcancer.org/; http://apps.komen.org/forums/; http://csn.cancer.org/forum/127; http://bcsupport.org/; http://www.healthboards.com/boards/forumdisplay.php?f=23; http://www.cancercompass.com/message-board/cancers/breast-cancer/1,0,119,1.htm; http://boards.webmd.com/webx/topics/hd/cancer/breast-cancer-friend-to-friend/; http://www.dailystrength.org/c/breast-cancer/forum; http://www.revolutionhealth.com/forums/cancer/breast-cancer; http://ehealthforum.com/health/breast_cancer.html; http://www.oprah.com/community/community/health/cancer; http://boards.webmd.com/webx/topics/hd/cancer/breast-; cancer-living-with-metastatic-breast-cancer/

[c] www.drugs.com; www.webmd.com; www.medhelp.org; www.healthboards.com; www.dailystrength.org; www.patientsville.com; health.groups.yahoo.com/group/afibsupport; www.mdjunction.com/; www.afibbers.net; www.mdjunction.com/forums/atrial-fibrillation-discussions

Supplementary Table 1: Details of the included studies

Author, year	*Aim*	*Drug - proto-ADE pairs* *
Adrover C., 2015[26]	To assess proto-ADEs and associated personal feeling related to HIV-drug treatments by using publicly available data from social media.	o efavirenz: sleep, psychological, neurological, gastrointestinal, liver and altered libido symptoms o efavirenz-emtricitabina-tenofovir-disoproxil: sleep, psychological, neurological, gastrointestinal, liver, altered libido, rash, allergy, kidney symptoms o tenofovir-emtricitabina: sleep, psychological, neurological, gastrointestinal, liver, altered libido, kidney symptoms
Benton A., 2011[36]	To identify proto-ADEs from online media related to four hormonal medications commonly used in the treatment of breast cancer.	o tamoxifene: hot flashes, breast cancer, menopause, pain, weight gain, joint pain, uterine cancer, fatigue, night sweat, depression
Bhattacharya M., 2017[44]	To identify new signals, known signals, signals sooner than notification, and specific issues (i.e., quality issues and patient perspectives) in social media and to determine the quantity of proto-ADEs and the type of drugs that would benefit from social media analysis.	o adalimumab: nausea, infection, injection site reaction, headache, burning sensation, antibody test abnormal, inflammation, drug tolerance, surgery, therapy naive, condition aggravated, abdominal symptom, malaise, fatigue, injection site pain, therapy change, drug ineffective, pain
Butt T.F., 2012[17]	To describe Internet narratives of drug-induced Stevens-Johnson syndrome and toxic epidermal necrolysis and to compare results with a previous study conducted by face-to-face interview.	o Stevens-Johnson Syndrome and toxic epidermal necrolysis: sulfonamides, penicillins, ibuprofen, carbamazepine, macrolides, cephalosporins, phenytoin, lamotrigine, tetracyclines
Coloma P.M., 2015[37]	To evaluate the potential contribution of mining social media networks for drug safety surveillance by using the following case study: rosiglitazone and cardiovascular events (i.e. stroke and myocardial infarction).	o rosiglitazione: cardiovascular events
Curtis J.R., 2017[18]	To characterize the demographics of people using social media to discuss rheumatoid arthritis and psoriatic arthritis and psoriasis, to evaluate the suitability of social media as a data source for drug safety research and to classify the content and timing of the posts with a particular focus on newer	o tofacitinib, infliximbam golimumab, rituximab, adalimumab and etanercept: *Herpes zoster* o tocilizumab, rituximab, adalimumab, infliximab, abatacept, certolizumab: Gastrointestinal perforation

Author, year	Aim	Drug - proto-ADE pairs*
	biologic drugs and small molecules in relation to the launch dates.	
Continued		
De Castro N.M.L., 2018[43]	To analyse the volume and content of Tweets related to biological treatments for rheumatoid arthritis, ankylosing spondylitis and psoriatic arthritis.	○ abatacept: administration adverse reactions, infections ○ adalimumab: administration adverse reactions, general adverse effects, infections, skin reactions, allergic reactions, immunosuppression, neurologic adverse effects, gastrointestinal side effects, lupus, fatigue, pulmonary toxicity, onco-haematologic diseases, death, liver damage ○ certolizumab: administration adverse reactions, general adverse effects, infections, allergic reactions, lupus ○ etanercept: administration adverse reactions, general adverse effects, infections, skin reactions, allergic reactions, immunosuppression, fatigue, hair loss, cardiovascular side effect, genitourinary toxicity ○ golimumab: allergic reactions, gastrointestinal side effects, headache, pulmonary toxicity ○ infliximab: administration adverse reactions, general adverse effects, infections, allergic reactions, immunosuppression, neurologic adverse effects, gastrointestinal side effects, lupus, fatigue, headache, onco-haematologic diseases, death, hair loss, genitourinary toxicity
Duh M.S., 2016[19]	To assess in social media proto-ADEs related to atorvastatin and sibutramine as compared to those reported in FDA FAERS and to determine whether social media posts can be useful in accelerating the detection of drug-related ADEs as compared to FAERS reports.	○ atorvastatin: Muscle or bone pain, joint pain, low energy, mental fog or memory loss, cramps, stomach or bowel issues, weakness, depression, insomnia, numbness, headache, vertigo, heart issue or chest pain, blurry vision, swelling, skin issue, anxiety, urinary issue, flu symptoms, mood swings ○ sibutramine: dry mouth, headaches, insomnia, constipation, cardiac symptoms, anxiety or irritability, lack of efficacy, nausea, indigestion, excessive thirst, low energy, muscle or joint pain, hypertension, bad breath, depression, sexual dysfunction, sweating, skins lesions, shortness of breath

Author, year	*Aim*	*Drug - proto-ADE pairs*[*]

Continued

Author, year	*Aim*	*Drug - proto-ADE pairs*[*]
Feldman R., 2015[53]	To evaluate a text mining methodology for the detection of unreported ADEs and to examine whether signals reported by the FDA in the post marketing appear in health forum prior to the first notification.	○ colesevelam hydrochloride, cholestyramine, colestipol hydrochloride: pain, muscle pain, flushing, heart attack, muscle damage, feeling weak, allergic reaction, liver failure, diabetes, cognitive impairment, leg pain, muscle problems, infection, leg cramps, cancer, head pain, stroke, burning sensation ○ ezetimibe: pain, muscle damage, allergic reaction, diabetes, muscle problems, leg cramps, cancer, head pain ○ fenofibrate, gemfibrozil, choline fenofibrate, colfibrate: pain, muscle pain, heart attack, muscle damage, feeling weak, allergic reaction, liver failure, diabetes, cognitive impairment, leg pain, muscle problems, infection, leg cramps, muscle weakness, head pain, heart problems, stroke ○ niacin, icosapent ethyl, dextrothyroxine sodium, lomitapide mesylate, mipomersen sodium: pain, muscle pain, flushing, heart attack, muscle damage, feeling weak, allergic reaction, diabetes, cognitive impairment, leg pain, muscle problems, infection, leg cramps, muscle weakness, cancer, head pain, stroke ○ ezetimibe/simvastatin, lovastatin/nicotinic acid, nicotinic acid/simvastatin, amlodipine besylate/atorvastatin calcium, aspirin and pravastatin, simvastatin/sitagliptin phosphate, atorvastatin calcium/ezetimibe: pain, muscle pain, flushing, heart attack, muscle damage, feeling weak, allergic reaction, liver failure, diabetes, cognitive impairment, leg pain, muscle problems, infection, leg cramps, cancer, head pain, heart problem, stroke, burning sensation ○ lovastatin, rosuvastatin calcium, fluvastatin sodium, atorvastatin calcium, lovastatin, pravastatin sodium, simvastatin, pivastatin, cerivastatin sodium: pain, muscle pain, flushing, heart attack, muscle damage, feeling weak, allergic reaction, liver failure, diabetes, cognitive impairment, leg pain, muscle problems, infection, leg cramps, muscle weakness, cancer, head pain, heart problem, stroke, burning sensation ○ citalopram hydrobromide: anxiety, weight gain, head pain, panic state, sleep disorder, allergic reaction, feeling weak, pain, tremors, agitation, nausea, sweating, seizure, dizziness, suicidality, sexual dysfunction, cognitive impairment, weight loss, mood swing, sleepiness ○ venlafaxine: anxiety, weight gain, head pain, panic state, sleep disorder, allergic reaction, feeling weak, pain, tremors, agitation, nausea, sweating, seizure, dizziness, suicidality, sexual dysfunction, cognitive impairment, weight loss, mood swing, sleepiness ○ desvenlafaxine succinate: anxiety, weight gain, head pain, sleep disorder, feeling weak, nausea, sweating

Continued

Author, year	Aim	Drug - proto-ADE pairs*
		○ bupropion hydrochloride: anxiety, weight gain, head pain, panic state, sleep disorder, allergic reaction, feeling weak, pain, tremors, agitation, nausea, sweating, seizure, dizziness, suicidality, sexual dysfunction, cognitive impairment, weight loss, mood swing, sleepiness ○ alprazolam: anxiety, weight gain, head pain, panic state, sleep disorder, allergic reaction, feeling weak, pain, tremors, agitation, nausea, sweating, seizure, dizziness, suicidality, sexual dysfunction, cognitive impairment, weight loss, mood swing, sleepiness ○ sertraline hydrochloride: anxiety, weight gain, head pain, panic state, sleep disorder, allergic reaction, feeling weak, pain, tremors, agitation, nausea, sweating, seizure, dizziness, suicidality, sexual dysfunction, cognitive impairment, weight loss, mood swing, sleepiness
Frost J., 2011[45]	To examine the prevalence of on-label versus off-label use, dosing, perceived effectiveness and side effects for amitriptyline and modafinil.	○ modafinil: jittery feeling, dry mouth, anxiety ○ amitriptyline: feeling sleepy, dry mouth, weight gain
Hadzi-Puric J., 2012[27]	To detect drug proto-ADE pairs from parenting websites using statistical methods based on different measures of disproportionality.	○ loratadine: headache, somnolence, dry mouth, fatigue, appetite loss
Hoang T., 2016[46]	To detect cascades of drug prescriptions and proto-ADEs, defined as detrimental prescribing cascades, from social media.	○ metoprolol - stroke → simvastatin - arrhythmia ○ venlafaxine - stroke → simvastatin - hemorrhage ○ celecoxib - hypertension →lisinopril - depression ○ venlafaxine - arthritis → meloxicam - hypertension ○ trazodone - hypertension → prazosin - anxiety ○ doxorubicin - pulmonary embolism → warfarin - myalgia ○ lisinopril - pulmonary embolism → warfarin - myalgia ○ ciprofloxacin - pulmonary embolism → warfarin - myalgia ○ citalopram - pulmonary embolism → warfarin - myalgia ○ clopidogrel - pulmonary embolism → warfarin - myalgia
Hughes S., 2011[25]	To describe the most frequently reported proto-ADEs of escitalopram and quetiapine in online consumer reviews, to compare them with those described in professionally controlled commercial health websites, and to gauge the usability of online	○ escitalopram: somnolence, new/worsened neurologic effects, weight gain, insomnia, nausea, sick stomach, vomiting, agitation, restless, dizziness, suicidal thoughts, hematologic effects ○ quetiapine: somnolence, weight gain, brain fog, abnormal movements, dizziness, vision changes, suicidal thoughts, weakness, sexual dysfunction

Author, year	Aim	Drug - proto-ADE pairs*
	consumer medication reviews.	
Continued		
Karimi S., 2015[20]	To detect possible proto-ADEs from patient comments in social media.	○ diclofenac: diarrhea, nausea, vaginal bleeding, cramps, dizziness ○ atorvastatin calcium: pain, fatigue, depression, muscle pain, memory loss
Koutkias V.G., 2017[21]	To validate a computational framework on specific case studies (clozapine-induced cardiomyopathy or myocarditis versus haloperidol-induced cardiomyopathy or myocarditis, and apixaban-induced cerebral haemorrhage) from data of several publicly sources (i.e. FAERS, PubMed, Twitter).	○ clozapine: myocarditis, cardiomyopathy ○ apixaban: cerebral hemorrhage ○ haloperidol: cardiomyopathy
Leaman R., 2010[47]	To mine the association between drugs and proto-ADEs reported by patients in comments of health-related websites and to compare the frequency of these proto-ADEs to that documented in labels.	○ carbamazepine: somnolence or fatigue, allergy, weight gain, rash, depression, dizziness, tremor/spasm, headache, appetite increased, nausea ○ olanzapine: weight gain, somnolence or fatigue, appetite increased, depression, tremor, diabetes, mania, anxiety, hallucination, edema ○ trazodone hydrochloride: somnolence or fatigue, nightmares, insomnia, addiction, headache, depression, hangover, anxiety attack, panic reaction, dizziness ○ ziprasidone hydrochloride or ziprasidone mesylate: somnolence or fatigue, dyskinesia, mania, anxiety attack, weight gain, depression, allergic reaction, dizziness, panic reaction ○ aspirin: ulcers, sensitivity, stroke, bleeding time increased, somnolence or fatigue, malaise, weakness, numbness, bleeding , tinnitus ○ ciprofloxacin: abdominal pain, malaise, nausea, allergy, somnolence or fatigue, dizziness, weakness, tolerance, rash, yeast infection
Liu J., 2011[22]	To detect the association between statins and proto-ADEs from patient-provided drug reviews on health-related web sites.	○ statins: muscle pain, pain, pain in legs, shoulder pain, back pain, neck pain, pain in arms, muscle cramps, general weakness, muscle weakness, difficulty walking, loss of muscle mass, general numbness, muscle spasms, rhabdomyolysis, tendinitis, balance problems, ALS (Amyotrophic Lateral Sclerosis), memory problems, Parkinson's disease, neuropathy, dementia, heart attack, liver damage, diabetes, stroke, arthritis, raised liver enzymes, heart failure, kidney failure, kidney damage, muscle problems, mobility problems, liver problems, pain, nerve problems, arthritis
Mao J.J., 2013[28]	To evaluate the volume and frequency of proto-ADEs related to aromatase inhibitors from internet message boards	○ aromatase Inhibitors: joint pain, bone pain, muscle pain, osteoporosis, weight gain, hair loss, mental depression, sleeplessness, headache, thyroid issues, dizziness, back pain

Author, year	*Aim*	*Drug - proto-ADE pairs*[*]
	and to focus on the association between aromatase inhibitors and arthralgia, that is one of the most common side effect.	
Continued		
Moncrieff J., 2009[23]	To analyse posts about proto-ADEs associated to olanzapine, risperidone and older neuroleptics.	o older antipsychotics: Sedative effects, cognitive effects, emotional effects, akathisia, anxiety, depression, suicidal thoughts, euphoria, relaxation, calmness, sexual impairment, weight gain o risperidone: Sedative effects, cognitive effects, emotional effects, akathisia, anxiety, depression, suicidal thoughts, euphoria, relaxation, calmness, sexual impairment, hormonal effects, weight gain, food craving o olanzapine: Sedative effects, cognitive effects, emotional effects, akathisia, anxiety, depression, suicidal thoughts, euphoria, relaxation, calmness, sexual impairment, weight gain, food craving
Nguyen T., 2017[48]	To assess whether the rate of proto-ADEs detected in social media is comparable to the rate of known ADRs and to evaluate if the identification of additional terms related to proto-ADEs could improve the rate of the detected proto-ADEs.	o lorazepam, alprazolam, sertraline hydrochloride: asthenia, convulsion, diarrhea, drowsiness, headache, hypotension, muscle rigidity, somnolence, sweating, yawning o citalopram hydrobromide: asthenia, convulsion, diarrhea, drowsiness, headache, hypotension, somnolence, sweating, yawning o duloxetine hydrochloride: asthenia, convulsion, diarrhea, drowsiness, headache, hypotension, muscle rigidity, sweating, yawning o trazodone hydrochloride: hypotension, muscle rigidity, somnolence o venlafaxine hydrochloride: asthenia, convulsion, diarrhea, drowsiness, headache, hypotension, muscle rigidity, somnolence, sweating, yawning o escitalopram oxalate, fluoxetine hydrochloride, bupropion hydrochloride: asthenia, diarrhea, drowsiness, headache, hypotension, muscle rigidity, somnolence, sweating, yawning
Continued		

Author, year	*Aim*	*Drug - proto-ADE pairs*[*]
O'Connor K., 2014[52]	To detect the association between drugs and proto-ADEs reported in Twitter.	○ quetiapine fumarate: somnolence, abnormal dreams, feel like a zombie, weight gain, restless leg syndrome, increased appetite, sleep paralysis, dizziness, psychosis, tremors ○ venlafaxine hydrochloride: withdrawal syndrome, insomnia, headache, malaise, abnormal dreams, nausea, shaking, fatigue ○ lisdexamfetamine dimesylate: insomnia, obsessive compulsive disorder, anger, heart racing, depression, psychosis, headache, feel weird ○ paroxetine hydrochloride: withdrawal syndrome, weight gain, depression, headache, somnolence, allergic, feel sick, emotional ○ fluoxetine hydrochloride: somnolence, withdrawal syndrome, feeling ill, abnormal dreams, suicidal thoughts, tremors, allergic reaction ○ lamotrigine: insomnia, rash, lethargy, joint pain, feel like a zombie, feel sick ○ olanzapine: weight gain, somnolence, increased appetite, dependence ○ adalimumab: somnolence, feel sick, palpitations, ache/pains, joint pain, headache, rash, respiratory disorder ○ duloxetine hydrochloride: withdrawal syndrome, fatigue, somnolence, dizziness, dry mouth, depression, rash, migraine ○ trazodone: somnolence, abnormal dreams, hangover effect, headache, insomnia

Continued

Author, year	*Aim*	*Drug - proto-ADE pairs*[*]
Pages A., 2014[24]	To describe proto-ADEs related to oral antineoplastic agents reported in online discussions and to compare these with reports recorded by health professionals in the French pharmacovigilance database.	○ oral antineoplastic protein kinase inhibitors: vascular disorders, skin and subcutaneous tissue disorders, respiratory disorders, thoracic disorders, mediastinal disorders, reproductive system disorders, breast disorders, nervous system disorder, musculoskeletal disorders, connective tissue disorders, metabolism and nutrition disorders, investigations, administration site conditions, gastrointestinal disorders, eye disorders, cardiac disorders, blood and lymphatic system disorders ○ oral antineoplastic hormone antagonists: vascular disorders, skin and subcutaneous tissue disorders, reproductive system disorders, breast disorders, psychiatric disorders, nervous system disorder, musculoskeletal disorders, connective tissue disorders, investigations, administration site conditions, gastrointestinal disorders, eye disorders, blood and lymphatic system disorders ○ anagrelide: chest pain, mitral valve disease ○ anastrozole: dry eye, libido decreased, weight increased ○ chlorambucil: dizziness, headache, petit mal epilepsy ○ erlotinib: agueusia, gingival pain, haemoptysis, toothache ○ everolismus: hypomagnesaemia ○ exemestane: colitis, hand deformity, libido decreased, memory impairment, mood disorder, weight increased, ○ hydroxycarbamide: augeusia ○ imatinib: bone pain, osteoporosis, pelvic fluid collection, tendinitis, tooth fracture ○ lenalidomide: hyperhidrosis

Author, year	*Aim*	*Drug - proto-ADE pairs*[*]
		○ letrozole: dermatitis bullous ○ tamoxifene: ocular hyperaemia, weight increased ○ temozolomide: oesophageal irritation

Continued

Patel R., 2018[49]	To detect and quantify proto-ADEs related to glucocorticoids from Twitter posts through an automatic detection system and to compare the frequency of proto-ADEs reported in Twitter to the frequency of ADR reports recorded in the UK spontaneous reporting system.	○ prednisone and prednisolone: insomnia, weight increased, not-specific reaction, increased appetite, malaise, drug ineffective, swelling, alerted state of consciousness, fatigue, affect lability, restlessness, swelling face, anger, withdrawal syndrome, condition aggravated, irritability, weight decreased, anxiety, abdominal pain, somnolence, hyperhidrosis, abnormal dreams, abdominal distention, skin discomfort, depression, pyrexia, vomiting, death, diarrhea, diabetes mellitus, nausea, dizziness, rash, confusional state
Patki A., 2014[4]	To develop automatic classification techniques to identify proto-ADEs from health-related social media data and to validate this approach by evaluating if the probabilities estimated for the reported proto-ADEs can be useful for categorizing drugs.	○ pregabalin: dizziness

Author, year	*Aim*	*Drug - proto-ADE pairs*[*]
Pierce, 2017[35]	To examine whether specific product-adverse event pairs were reported via social media before being reported to FAERS	○ dronedarone – vasculitis ○ ziprasidone – DRESS ○ methylphenidate – priapism ○ dimethyl fumarate – PML
Powell G.E., 2016[38]	To describe an approach of social media listening for pharmacovigilance purposes.	○ salbutamol: tremor, chronic obstructive pulmonary disease, wheezing, bronchitis, pallor, pneumonia, dysesthesia, cough, restlessness, mobility decreased, lung disorders, palpitations, hemorrhoids, muscle twitching, infection, hear rate increased

Continued

Author, year	*Aim*	*Drug - proto-ADE pairs*[*]
Sampathkumar H., 2014[39]	To extract reports of proto-ADEs from messages in online healthcare forums and to match them with those reported in drug labels.	o lisinopril: cough, dizziness, headache, fatigue, cramps, diarrhea, nausea, rash, hearing loss, hair loss, shingles, fits o prednisone: anxiety, insomnia, depression, dizziness, mood swings, weight gain, nausea, moon face, hives, acid reflux, avascular necrosis, dry mouth o montelukast sodium: headache, infection, cough, fever, diarrhea, sinusitis, inflammation, Seizure, depression, nightmares, aggression, mood swings, suicide, suicidal thoughts o topiramate: tingling, weight loss, memory loss, numbness, dizziness, tired, sleepy, hair loss , depression, stress, aches, anxiety, diarrhea, dry mouth, itching o dextroamphetamine/amphetamine: depression, weight loss, headache, dizziness, dry mouth, insomnia, constipation, loss of appetite, death, seizures, high blood pressure, restlessness, anxiety, fatigue, addiction, mood swings, vomiting, nausea, hallucinations o cortisone: headache, allergies, nausea, weight gain, depression, insomnia, high blood pressure, acne, atrophy, rash, anxiety, cramps, bleeding, back pain, dizziness, numbness, diarrhea o venlafaxine hydrochloride: dizziness, headache, dizziness, headache, insomnia, vomiting, chills, diarrhea, tachycardia, weight gain, acne, shocks, hives, mood swings o buprenorphine/naloxone: pain, insomnia, depression, chronic pain, sweats, headaches, anxiety, tired, restlessness, chills, weight gain, runny nose
Schröder S., 2007[29]	To identify proto-ADEs related to Parkinson's disease treatments through the analysis of online outpatient forums.	o antiparkinsonian agents - pramipexole, ropinirole, pergolide, cabergoline, levodopa, entacapone, tolcapone, carbidopa, rasagiline, selegiline and amantadine: dizziness, headache, migraine, insomnia, vivid dreams, sleepiness, sleep attacks, fatigue, depression, impaired memory/impaired concentration, hallucinations, aggressiveness, restlessness, hypersexuality, gambling, orthostatic hypertension, sweating , increased urge to urinate, dry mouth, musculoskeletal effects, diarrhea, constipation, increased appetite/weight gain, nausea, general eyesight problems/impairment of vision, loss of visual acuteness, electrocardiography changes, valvular changes/fibrosis, cardiac dysrhythmia, cardiac palpitations, breathing trouble, hacking cough, allergic skin reactions, oedema, wound healing problems, pigment disorders, loss of hair

Continued

Tafti A.P., 2017[50]	To develop a big data analytics strategy that mines the	o antihistamine: nausea o antipsychotic: weight gain

Author, year	*Aim*	*Drug - proto-ADE pairs*[*]
	content of scientific articles and health-related web-based social media for detecting and identifying proto-ADEs.	○ aspirin: bleeding, constipation, panic, rash ○ atenolol: dizziness, hypotension, tiredness, vomiting, bradycardia ○ dexamethasone: nausea, weight gain, vomiting ○ diazepam: drowsiness, nausea ○ dopamine: weakness, sleep problems, hypertension ○ ephedrine: anxiety, hypertension ○ gabapentin: diarrhea, constipation ○ heparin: anemia ○ ibuprofen: constipation ○ lamotrigine: dizziness, vertigo ○ lorazepam: dizziness, insomnia, amnesia ○ melatonin: depression ○ metformin: nausea, diarrhea, vomiting dizziness, abdominal pain ○ methylphenidate: nervous feeling ○ sildenafil: chest pain, myocardial infarction, sweating, nausea ○ statins: fatigue, rhabdomyolysis ○ warfarin: bleeding ○ bupropion hydrochloride: dry mouth, sweating, nausea
Topaz M., 2016[51]	To compare data between electronic health records and social media about proto-ADEs related to aspirin and atorvastatin.	○ aspirin: hives or other rash, bleeding, swelling, anaphylaxis, angioedema, bronchospasm or wheezing, shortness of breath, nose bleeds, itching, asthma, thrombocytopenia, anemia, tinnitus, arrhythmia, flushing, hypotension, Reye's syndrome, hypoglycemia ○ atorvastatin: musculoskeletal pain, cramps, hives or other rash, gastrointestinal upset, hepatotoxicity, weakness, tiredness, headaches, myositis, nausea or vomiting, diarrhea, swelling, itching, memory loss, cough, hair loss, shortness of breath, anaphylaxis, angioedema, bronchospasm or wheezing, renal toxicity, flushing, depression, thrombocytopenia, anemia, fever, arthritis, seizures, hypotension, Alzheimer
Continued		
Vaughan Sarrazin	To examine perceptions and experiences about	○ dabigratan: stomach discomfort, upper gastrointestinal discomfort, lower gastrointestinal discomfort, esophagus discomfort, new/worsening hemorrhoids,

Author, year	*Aim*	*Drug - proto-ADE pairs*[*]
M.S., 2014[42]	dabigatran used for atrial fibrillation in online comments reported by patients and caregivers.	loss of appetite/weight loss, metallic taste in mouth, fatigue, headaches, dizziness, excessive sweating, feeling unusually cold, feeling usually warm, change in blood pressure, weight, gain, blood rushing to head, muscle or joint pain, swelling of feet/limbs, major or serious bleed, minor bleeding, fluid retention, difficulty urinating, kidney damage, confusion/disorientation, depression, psychiatric disorders, skin rash/itching, change in blood sugar level, blurry vision, chest pain/angina, abnormal liver tests, dry mouth, stroke/transient ischemic attack, shortness of breath/ difficulty breathing
Wu H., 2013[40]	To investigate the feasibility of exploiting social media discussions to discover unrecognized drug side effects.	o dextropropoxyphene/paracetamol: abnormal heart rhythm o drospirenone/ethinyl estradiol: blood clot o drospirenone/ethinyl estradiol: blood clot
Yang C.C., 2012[41]	To test the effectiveness of using association rule mining to extract proto-ADEs caused by certain drugs from online healthcare communities.	o clarithromycin: heart disease , diarrhea, cancer o lansoprazole: diarrhea, heart disease, cancer o fluvoxamine: heart condition, suicide, depression o fluoxetine: suicidal, depression
Yang C.C., 2014[31]	To detect signals of proto-ADEs from online health communities.	o clarithromycin: diarrhea, heart disease, kidney disease o lansoprazole: diarrhea, heart disease o fluvoxamine: depression, suicide o fluoxetine: depression o tacrolimus: diarrhea, kidney disease o adenosine: heart disease o tadalafil: heart disease, depression o pimecrolimus: diarrhea, depression o insulin glargine: diarrhea, heart disease, kidney disease o lisdexamfetamine dimesylate: depression, suicide, kidney disease o methylphenidate hydrochloride: depression, suicide o epoetin alfa: kidney disease o gadolinium: kidney disease o ziprasidone hydrochloride: depression, suicide o heparin: diarrhea, heart disease, kidney disease o eszopiclone: depression o risperidone: depression, suicide o simvastatin: diarrhea, heart disease, kidney disease o simvastatin: diarrhea, heart disease, kidney disease o olanzapine: depression, suicide
Continued		
Yang H., 2013[30]	To detect drug-drug interaction signals from consumer contributed contents in online healthcare communities by using	o quinidine – clarithromycin: arrhythmias o quinidine – tacrolimus: arrhythmias o quinidine – simvastatin: arrhythmias

Author, year	Aim	Drug - proto-ADE pairs*
	associations mining.	○ quinidine – simvastatin: arrhythmias ○ ticlopidine – heparin: bleeding ○ ticlopidine – luvox: bleeding ○ gemfibrozil – simvastatin: myopathy ○ gemfibrozil – simvastatin: myopathy
Yang H., 2015[32]	To detect the association between drugs and proto-ADEs and to perform a temporal analysis for detecting drug safety signals at the early stage.	○ lansoprazole: diarrhea ○ fluvoxamine: heart disease, suicidal thoughts ○ fluoxetine: depression, suicidal thoughts ○ simvastatin: kidney disease ○ simvastatin: kidney disease ○ tadalafil: stroke ○ methylphenidate hydrochloride: blurred vision, hypertension ○ heparin: diarrhea ○ pimecrolimus: skin discoloration, cancer ○ tacrolimus: cancer
Yeleswarapu S., 2014[33]	To detect potential proto-ADEs from FAERS, health-related websites and MEDLINE, to compare these with the label information of drugs and to evaluate by a statistical method the significance level of drug – proto-ADEs pairs.	○ aspirin: hemorrhage, asthma, ulcer ○ bupropion: weight loss ○ carbamazepine: seizures, exanthema, lymphoma ○ ibuprofen: pain, osteoarthritis, stomach ulcer ○ morphine: Hashimoto disease, breathlessness, violent behavior ○ olanzapine, ciprofloxacin: diverticulitis, acne vulgaris ○ paroxetine, rosiglitazone: heart diseases, diabetes mellitus, coronary artery disease ○ trazodone: back pain, sleep initiation and maintenance disorders, condylomata acuminata ○ warfarin: international normalized ratio, hemorrhage, stroke ○ ziprasidone: vomiting, tremor, psychotic disorders
Continued		
Zheng Y., 2016[34]	To test a novel constrained information entropy approach to detect proto-ADEs in medical forum.	○ norepinephrine: anxiety ○ topiramate: migraine ○ arthritis pain relief drugs: pain ○ ibuprofen: headache ○ lithium: pain ○ gabapentin: pain ○ azithromycin: bronchitis

Author, year	*Aim*	*Drug - proto-ADE pairs*[*]
		○ methadone: pain ○ spironolactone: acne ○ etanercept: psoriasis ○ naproxen: pain ○ pregabalin: pain ○ clindamycin: acne ○ moxifloxacin hydrochloride: sinusitis ○ acetaminophen: pain ○ cortisone: coccydynia ○ analgesic: pain ○ aripiprazole: polyuria ○ antibacterial: infection ○ modafinil: excessive daytime sleepiness ○ ibuprofen: pain ○ diphenhydramine hydrochloride: erectile dysfunction ○ compazine: nausea ○ vinorelbine tartrate: pain ○ flecainide acetate: stress ○ tramadol: dependence ○ epinephrine: overdose ○ celecoxib: arthritis ○ moxifloxacin hydrochloride: bronchitis

[*]Only drug - proto-ADE pairs reported as frequency or risk value were extracted

DRESS: Drug reaction with eosinophilia and systemic symptoms; FAERS: FDA Adverse Event Reporting System; FDA: Food and Drug Administration; HIV-drug: Human Immunodeficiency Virus drug; PML: progressive multifocal leukoencephalopathy