Evaluation of Social Media Data for Adverse Events Analysis: New Prospective for Safety of Drugs for Human use

Suman Yadav[1], *Dr. Md. Aftab Alam[2], Dr. Ranjana Patnaik[3]

[1]Research Scholar, Division of Clinical Research, Department of Biosciences, School of Basic and Applied Sciences, Galgotias University, Greater Noida, Uttar Pradesh, India.

[2] Department of Pharmacy, School of Medical & Allied Sciences, Galgotias University, Greater Noida, Uttar Pradesh, India.

[3] Division of Clinical Research, Department of Biosciences, School of Basic and Applied Sciences, Galgotias University, Greater Noida, Uttar Pradesh, India

*Corresponding Author: Dr. Md. Aftab Alam, Associate Professor & HOD (M.Pharm-Pharmaceutics), School of Medical and Allied Sciences, Galgotias University, Greater Noida,
Email ID: aftab.alam@galgotiasuniversity.edu.in

ABSTRACT:

Introduction: For the monitoring of adverse events related to medicinal products, social media can be used as potential source of information under pharmacovigilance domain. It also provides the insight about the consumer/patient perceptions on adverse events experienced. The objective of this research is to find out, if data available on social media can help in the determination and assessment of new adverse events, signals, already known risks and signals, consumers or patients opinion on the side effects and any product quality related issues faced. It also includes, number of 'posts/messages found with similarity to adverse drug reaction/adverse events along with the product classification.

Method: For the purpose of this research work, data has been collected from various digital social media platforms which include Twitter, Facebook, YouTube, Reddit etc. A supplementary work also involved literature search using internet search engine, field expert opinion, newsletters, and various blogs available on internet etc.

Results: 5075 records were found in the internet searches and database explored. AEs measured from Twitter were around 3% and whereas Facebook measured 4% of AEs. These outcomes were further supported by the suggestions from peer reviewers, field experts, reference review and studies published by various authors. Narratives analysis posted was used and then AEs posted on the selected social digital media platforms were calculated.

Conclusion: Monitoring of the digital social media forums/ platforms can help in identifying adverse events related to the medicinal product of interest. The validation and reliability of data needs to be assessed in detail. Heterogeneity has also been found in the frequency and the type of adverse events reported.

Keywords: Pharmacovigilance, Adverse Drug Reaction, Digital Platform, Big Data, Social media, Adverse Events, Side Effects

3. INTRODUCTION:

According to the World Health Organization WHO "*Pharmacovigilance is the science and events concerning to discover, analyze, understanding and preclusion of adverse effects or any other possible drug related problems*" [1]. The developing nations follow this definition as a progressive discipline where medications are significantly beneficial rather than harmful. Pharmacovigilance (PV) has formed like a substantial unit that accompanied by close catastrophic global incidences. In the 15th century, the most primitive confirmed event can be discovered [2]. In the post-marketing phase of a therapeutic product, the thalidomide catastrophe taken in 1961 as the façade that necessitate for continued execution as well as the requisite for healthcare authorities to develop a high index of disbelief [3]. It can recognize the adverse drug reactions and explain such reactions instinctively in a systematic process that will provide the optimum benefits.

In the module I, European Medicine Agency (EMA) depicted the Good Pharmacovigilance Practices (GVP) on a pharmacovigilance system as "*a system utilized by an institute to accomplish its legal responsibilities and tasks with respect to pharmacovigilance as well as developed to observe the safety of authorized medicinal products and identify any change to their risk benefit balance [DIR Art 1(28d)*" [4]. Such aims can be achieved by creating an effective PV system that involves efficient frameworks, processes and outcomes. An operative PV model is needed that consists in the collecting, transfer and documentation of data in any institute. It is based on the Individual Case Safety Reports (ICSRs) [5]. These need the stems from the input information that systematic PV synchronization is necessary to reduce the load that ADRs occurrence imposes on patients and society on large-scale.

Globally, WHO develops an International Drug Monitoring (PIDM) center and works as the consolidating organization for PV actions with their program. They are conducting several roles involving the delivery of risk minimization plans, causality assessments and recognition/coordination/analysis of Adverse Drug Reactions (ADRs). Furthermore, it also assures the efficient announcement of signal detection as well as potential risks. WHO also provides the supervision for the building of national pharmacovigilance systems [6]. The role of several member nations has enhanced the ADRs located in the database of WHO (Vigibase™) that carried the record of 16,720,000 million up to February 2018. Every nation was participated into the PIDM depending on ADR reports submission to the database of WHO database in order to fulfill the essential requirements. Although, pharmacovigilance systems and organization qualities may differ from nation to nation and occasionally province to province. WHO monitoring program have worked in currently 158 nations (131 entirely, 27 allies) [7].

In the developed countries, pharmacovigilance system has mutual characteristic of taking initialization of legislation to undertake the relevant pharmacovigilance events, pertinent set-up. Most of such nation have strategies for efficient signal detection by including efficient reporting systems and electronic databases. These also have the well placed pharmacovigilance communication strategies. The PV systems have some strengths that still places in impulsive ADRs reporting as well as such have the shortcoming as under-reporting [8]. Diverse PV

system models occurs with regionalization of PV centers in few nations [9-10], reporting by several cadres of healthcare experts, compulsory reporting of ADRs by healthcare experts and national PV centers self-governing of the regulatory body. All of these is established based on good quality and number of reports that are in surplus of thousands per annum in some of those nations [8, 11].

The PV system of India is completely underdeveloped, as the country deals with several health challenges as well as socio-economic problems. It has insufficient economic budget for healthcare infrastructure. The medicinal situation is overflowing with the challenges regarding logistic considerations, limited access to medicines because of expenses, poor distribution practices and insufficient production of pharmaceutical units. In addition, the unreasonable medicinal usage, existence of second-rate forged therapeutic products that spot the Indian pharmaceutical area makes it even worse [12-13]. Several Indian nations have in place legislature of fluctuating extent to fight against the above concerns and newly, there have been transfers to form an Indian Medicines Agency to accords the numerous international methods for patient safety and to provide the appropriate drug. Computing the influence of the several PV systems had been the main drawback and such caused in the progress of tools set out to initiate the PV centers that enhance their systems and conduct the self-assessment [14-15].

Several research works have been performed to estimate the information value of social media about conventional PV approaches. Such research works also attempted to improve the assessment process of social media data for observing the safety of devices/drugs. The information utilization of social media for tracking the proto-AEs or ‘posts with resemblance to adverse events’ is the emergent domain within pharmacovigilance [3]. Several challenges continue concerning the information value of social media in accompanying the present PV information sources (such as observational data, clinical trial, literature reviews and spontaneous reports) as well as combining to conventional PV approaches (quantitative ratios like proportional reporting rate [PRR], AE reports and safety databases). Identified restrictions of existing PV information sources involve the variable post-marketing quality of AE reports, under-reporting and small clinical trials size of samples. The organizational efforts involve the thorough understanding of safety profile about the products. Also, it consists the social media information benefits present in the capability for prompt authorization to evolving problems from patient-generated information. Social media has the abundant of information, still the significance and extent of posts on definite social media products is neither plentiful nor reliable.

Several research works have analyzed the social media information usage to detect the beneficial safety data. Such research works have illustrated the broad range of consequences, developing a gray zone concerning the outcomes from as well as the practicality of social media information assessment. In recent times, researchers [1] assessed the social media info to detect the untapped safety as well as significant data and come to the conclusion that social media info were appropriate for product benefits and the surveillance of post-marketing safety. Although, a frequent challenge amid others assessing social media information for AEs involves converting the non-text entries like emoji and free text format of posts into clinically significant data in a organized statistics format [1, 2, 5, 6]. The main approach is natural

language processing (NLP) for converting the social media post content into significant data as well as is remaining to emerge as upgrading tool for assessing the post categorization accuracy [6–9]. Furthermore, social media posts categorization in suitable perspective has validated challenging, involving the recognition of authentic AEs versus arbitrary drug mentions [2]. Numerous research works have outlined the social media power in recognition of public health issues as well as severe disease outbreaks that validates the assessment of social media as timelier for reporting on specific products instead of typical public health reporting sources [1, 10, 11].

Other research work utilizing social media info from the patient forums or sites such as Twitter, Facebook etc. focusing on proto-AEs as well as tracking mentions for certain drugs from an enterprise's portfolio [12–15]. Further, experts made the comparison between the US FDA Adverse Event Reporting System (FAERS) and Twitter posts coded as proto-AEs. The sample was too noisy to deliver a significant assessment at the Medical Dictionary for Regulatory Activities (MedDRA) preferred term (PT) level. Nevertheless, experts did obtain concordance at the system organ class (SOC) level [3]. Generally, some research works till now made the comparison between AEs from conventional PV sources (such as FAERS) with proto-AEs from social media info. This paper reveals the assessment reports on social media info to recognize the signals that are not identified by routine pharmacovigilance approaches. It is also validated with the help of several social media info. It was further interest to estimate whether the unique insights could be detected as the social media info were assessed like off-label utilization, patient perspectives and product quality complaints. For further hypothesis assessment, the outcomes depicted that social media may signify a huge untapped info source for safety organizations. The study also required to estimate the amount of proto-AEs occurred in social media and to evaluate the types and characteristics of products that would benefit most from social media data analysis and usage.

2. The characteristics and scenario-burden of pharmacovigilance

According to the reports of WHO, India has a large group of lower-middle income people that is also categorized as a developing country [16]. It is a greatly populated nation of Asia with a diverse cultural groups. It is governmentally divided into 28 states with 8 Union Territory. It also involves several geo-political zones such as North, East, West, South and Central zones. The main area it involves the North and Central zones as these are the highly affected regions. India has the great burden of both non-communicable as well as communicable ailments with the high growth of non-communicable diseases [17-18]. It becomes significant to analytically assess the medicinal safety utilized in these atmosphere and also to evaluate the drugs used in treating illnesses that have great occurrence rates. According to the WHO 2013 statistics in India, communal diseases cause the age adjusted mortality rate that was stated to be nearly 832 per 100,000 with the major contributors as non-communicable diseases, diarrhea, HIV, respiratory infections and malaria. According to the same report, mortality rate is 756.7 per 100,000 [17].

Communicable diseases should be treated by following the rational pharmacotherapy and standard policies, although, self-medications is routine with antibiotics as well as analgesics in

India [18], because of the presence of several analgesics as poor restriction of antibiotics usage/procurement and Over the Counter (OTC) medicines [19-21]. Such medicines have the general usage, it is supposed that it may involve the related great occurrence of ADRs. The load in financial aspects has also not been entirely enumerated. Although, taking the account of unreasonable medication usage practices monitored [22]. It further develops an immense burden. Different zones have different practices across the nation that may alter in line along with ethno-cultural effects of the region and currently, the degree of unreasonable recommending practices in several domains is unidentified. This study also involves the paucity in India setting that have explained the ADRs profile to such often utilized medicinal classes.

India has joined the WHO International drug monitoring programme in 1998 and has built its own PV structure with the outline of relevant strategies [23-26], the formation of consumer reporting, institutionalization of PV in health institutions and PV centers [27]. The pharmacovigilance has the governance in India i.e. located from the National Pharmacovigilance Centre (NPC) and includes the marketing authorization holders, academic institutions and the zonal centers [27]. All such parts have certain purposes and roles that are crucial in the progress of drug safety in India. The NPC is still needed to attain the WHO recommended objective of getting hundred ADR reports per million as this program has nearly 18,000 ADR reports in 2017 placed at the database, although, such numbers are still insufficient. This possibly will be credited to the emerging nature of the Indian PV system. Aspects contributing in this deprived ADRs reporting have earlier preliminarily been discovered mostly at only organizations dispersed across the nation including very few to the east zone and none at the Zonal level of the nation [28-33]. Such research works also focused majorly on the ADRs reporting and physician reporters. However, document of NPC guidelines states that nurses can report and some literature provides the assessment on the insight of nurses in the direction of ADRs reporting in India. Such research work did not discover the other pharmacovigilance facades such as product concerns, scope etc. It becomes significant to assess the perception in this context, training of the healthcare experts in East India towards such challenges as per the latest formation of Central Zonal Pharmacovigilance Centre.

Once more, patient reporting is being motivated globally as well as several approaches are being assessed on the assurance of fulfillment and best quality of ADR reports [34-35]. Certainly in few centers, it has been depicted that the ADR reports of patients are similar with those from the healthcare experts [36]. Thus, it becomes substantial to research the forms of reactions explained by patients to often usage of medicines in this setting to assess the conceivable contributions of patient recording to the PV organization in India. Already a consumer form recording is being motivated via the utilization of the Pharmacovigilance Rapid Alert system for Consumer Reporting (PRASCOR). Such uses text messages delivered directly to the NPC for assessment. Although, it becomes significant to assess the approaches that may be helpful for healthcare experts and so as to assure that satisfactory documentation is prepared and moreover to evade problems that may ascend from an unintentional issue.

In Asia, the faults revealed in the pharmacovigilance structures and certainly India has mostly the inadequate expertise, lack of resources and deprived infrastructural set up being dedicated

to pharmacovigilance [37]. Other recognizable complications contain second-rate forged medicinal products resulting as development of drug-food, drug-herb interactions, use of herbal medicines, faulty drug distribution systems, adulteration of medicines and therapeutic ineffectiveness [12, 37]. Other problems associating with pharmacovigilance like misuse of medicines, drug abuse, poisoning (chronic and acute) and medication errors have not also been accurately enumerated. Based on the PV indicators prepared by WHO, the utilization of standard indicators in computing the efficiency of a PV structure would be beneficial in estimating the phase of PV actions in that setting [15].

3. Traits of ADRs in Indians:

Social media becomes a trending platform to talk over health related problems. In the last twelve months, 34% have read other's opinion on health or medical problems, while 80% of internet users have explored online for health information [1]. According to such figures, it is expected that social media becomes a trending research tool. A significant literature is available on extracting social media data to observe patient views, health behaviors and disease outbreaks. Several patients select online people to converse the adverse influences of treatments, specifically drug intrusions. Such produces a huge volume of spontaneous as well as latest information. Researchers propose that by observing social media, it takes just 5-7 days to get aware about thalidomide disaster [27].

Even with public accessibility of such information, their suitable role in PV has not been built, nor are they regularly utilized for gathering the information about adverse effects. The social media also needs the comparative value with other reliable data sources (e.g. clinical trials or PV systems). Researchers require to understand that how adverse events info obtained from social media relate with data from other sources as timeliness, frequency, range and type of discovered AEs. Social media data may not be simply accessible from other sources. With several queries that require to address the info retrieval on AEs from social media is predominantly suitable. This paper presents a methodical review that summarizes the literature work on the occurrence, form and frequency of AEs info for healthcare intrusions accessible via social media as well as on the comparative value of social media as a source for AEs info with respect to other data sources.

4. Social media relevance to public health

In recent years, social media becomes a new source where information has grown in an immense proportion as well as continues to develop. In terms of global users, social networks have perceived an exceptional development (such as Twitter has over 645,750,000 handlers on 11th July 2014 and develops by a projected rate of 135,000 handlers every day, producing 9100 tweets per second). In social media, large population of patients are actively engaged in posting/sharing health connected social media data as well as specifically health social networks [12]. Recently, Pew Research Center's conducted a survey [11] that has revealed the social media significance according to modern scenario public health, describing that 34% of healthcare workers as well as 20% of patients observe or read other's online experience/opinion. Furthermore, 6% of patients as well as 11% of caregivers post questions and share experiences online. Health related social networks have been attracting several users

particularly focusing on health related problems. In these social media platforms (such as *e.g.*, DailyMotion, Vkontakate), handlers converse their experiences about health-related issues involving the prescribed drugs usage, treatments and side effects.

Users consistently share their experiences with others who are facing same issues and consequences that makes these social networks exceptional as well as robust sources of data on treatments, drugs and health. Because of the emergence of these social media as well as the plentiful of information from them, exploration on ADR monitoring or public health monitoring, has concentrated on manipulating info from such sources recently. Monitoring public health involves research on smoking cessation configurations on Facebook [18], monitoring malpractice and detecting the user social circles with general clinical experiences such as drug abuse. When several patients suffering from the similar disease or use certain prescription, share data about their drug results, treatments or symptoms, this data can deliver valuable medical insights for both health-related trades and patients that go beyond conventional communication approaches [21].

Monitoring of viral or infectious disease, particularly, can advantage sturdily from social media usage. For instance, conventional systems may miss rare or new events such as the new viral epidemic and deficiency of demographic reach and real time potentials that social media can offer, involving information from people that may not authorize the healthcare via formal channels [22]. However, certain data on a solo user may not be accessible or operational for privacy concerns, several resources are presently available to conduct some demographic assessment with social media info. In addition, last decade witnessed the development in a huge number of social media based surveillance systems that is further reviewed and executed internationally, nationally and locally [23]. Contemporary developments in ADR monitoring systems have seen substantial progresses towards the automatic NLP approaches utilization for mining drugs and connected reactions from social media. Social media posts involve the data on treatment results and deliver the initial authorization to report ADRs that could be valuable for pharmaceutical/health industries. The form and capacity of ADR data that social media makes accessible to the health sectors may not be easily accessible by other means. This contains the ADRs experienced by users having certain conditions like elderly people or patients with comorbidities, pregnant/nursing women and patients with rare diseases, who are typically omitted from medical trials [24].

5. Social media from the perspective of regulatory authorities and the industry

The regulatory authorities' perspective involves the intent of extracting the information from social media in order to get the further data from general public that may be utilized to complement prevailing voluntary data schemes. For instance, in 2013, the Association of the British Pharmaceutical Industry (ABPI) issued a guidance notes that assistance stakeholders/investigators to handle the ADR grievances on digital media [25]. However, the report was developed for data drives only, instead of legal/regulatory information, it delivers the guidelines on how to manage these ADR documents. It evidently outlines a least information set required to report the ADR that involves identifiable reporter, adverse event,

suspect drug and a recognizable patient. The contact details needed for the recognizable reporter are compatible with the domain of social media involving the screen names or emails. It also states that such data should be gathered "if possible," that leaves room for inadequate info [25]. The FDA has not yet issued the clear regulations for social media based PV, but it has published the guidelines for issuing the risk/benefit information and promotional material on social media. ADRs from social media can still be reported to the FDA regardless of the absence of formal guidance. The least info set for an ADR report to the FDA is the same as that for the ABPI. Furthermore, a recent FDA presentation referred that social media ADR reports are reviewed like any other unprompted reporting systems, though recognizing inconsistency in the submitted report quality [28].

Moreover to regulatory specialists, signals recognized from social media could be utilized by pharmaceutical producers, the healthcare researchers/system to accomplish the mandatory reporting necessities. Although the social media mining determined to give early signals, it could capably be utilized by the concerned parties to authenticate or reject signals that have ascended in other reporting schemes. Pharmaceutical producers like AstraZeneca, have deliberated the social media utilization from a business perception: focusing on producers' accountabilities to offer precise and quality data on drugs [29]. Since pharmaceutical producers and regulatory authorities play a role in public safety, both may use the social media to accomplish the safety mission.

6. Methods:

Data collection and assessment

Search methods

Sixteen databases covering a range of topic areas, including health and medical research, nursing, information and computer science and grey literature (i.e. literature that is not formally published) were searched. We undertook other supplementary methods which included searching two internet search engines, browsing internet blogs, hand searching journals, newsletters and conference proceedings, reference checking all included articles and related systematic reviews, and contacting experts in the field. This work involves the data collection from the following social sites:

- Twitter
- Facebook
- YouTube
- Tumblr
- DailyMotion
- Flickr
- Vkontakate
- Web
- Reddit

.**Search strategies**

The database search strategies contained two facets – 'social media' and 'adverse events'. A date restriction of 1996 onwards was placed on the searches as blogging first began in 1997. No language restrictions were placed on the searches, although financial and logistical restraints did not allow translation from all languages.

Data extraction

Information was collected on the type of social media used (such as Twitter or Facebook), the adverse events and type of interventions searched for, the primary aim of the study as stated by the authors and the type and frequency of adverse events data identified. Details of comparator sources were noted along with any comparisons of the data collected. Lastly, data were extracted on the conclusions of the original investigators.

Assessment of methodological quality

We did not stipulate any restriction on design of the included studies. As there is no relevant quality assessment checklist for these types of studies, we designed a bespoke tool based on four key areas to reflect potential risks of bias. These four key criteria were:

- Search strategy to identify posts: How were the posts searched for? Were adequate search terms used? Searching social media is difficult due to the unstructured nature of the data. In particular, colloquial expressions/informal speech, mis-spellings, nicknames, the use of non-standard abbreviations, different synonyms and different spellings make a comprehensive search impossible. However, attempts should be made to include numerous synonyms, spellings etc. in a search strategy.
- Selection of relevant posts: What methods were used in selecting relevant posts? For example, were double screening methods used for manual selection? Were computerized methods validated?
- Definition of a report of an adverse event: Was there a clear definition of what constitutes an adverse event report? (for example, the Food and Drug Administration (FDA) minimum criteria of an identifiable reporter, an identifiable patient, a reaction or event, and a suspected medicinal product)
- Duplicate data: Did the researchers measure the amount of duplicate data? Were duplicate reports from the same user excluded?

Analysis

It was anticipated that the included studies would be heterogeneous in nature as methods in this area are still under development. A narrative synthesis was therefore used.

Results

The database and internet searches identified 5045 records, and these results were augmented with studies identified from hand searching, reference checking, contacting experts, peer reviewers suggestions and studies already known by the authors. Altogether 51 studies from 64 publications met the inclusion criteria. There were 180 excluded studies based on the full text papers.

Baseline characteristics of studies

Social media Over 174 different social media sites were represented in the 51 studies. Seventy-one % (36/51) of the studies examined discussion forums, 10 looked at Twitter, five at Facebook, four at blogs, three at YouTube, one at DailyMotion and one at Reddit. Three studies did not report the websites searched and seven studies looked at more than one type of social media. Discussion forums Of the 36 studies that looked at discussion forums six did not specify the forums searched, and one gave an incomplete listing. The most popular named forums were Twitter, Facebook, YouTube, Tumblr, DailyMotion, Flickr, Vkontakate, Web and Reddit. The most popular disease specific patient forums were for cancer (five studies), depression (three studies), heart conditions (two studies) and diabetes (two studies). The number of forums searched ranged from 1 to 24, with an average of four forums searched in each study. Interventions Most studies looked at drug interventions (86%, 44/51), with only three looking at surgery (one with YouTube), two limiting by illness and two looking for a medical device (both with YouTube). Of those studies that included drug interventions most assessed multiple drugs (84%, 37/44) whilst those studies assessing surgery or a medical device only evaluated one intervention.

Table 1 Social Media post assessment to obtain Percentage of adverse event posts from posts related to intervention/illness

Social media	% of AE posts from all posts	% of AEs posts from posts related to intervention/illness
Facebook	4% (n = 1)	0.7%–2% (n = 2)
Blogs, Facebook, Twitter and forums	0.3%–8% (n = 2)	
Twitter	2%–4% (n = 1)	0.02%–11.5% (n = 3)
General forums	0.2%–1.42% (n = 4)	18.2%–35% (n = 4)
General and disease specific forums		12%–58% (n = 3)
Disease specific forums		12%–62% (n = 4)
Disease specific forums and blogs		12.4% (n = 1)
YouTube		40%–78% (n = 3)

n = number of studies; NB. Some studies measured more than one value contained in the above table

Adverse events Ninety % (46/51) of the papers looked for any adverse events whilst only 10% (5/51) specified the adverse events they were looking for (withdrawal symptoms, SJS/TEN, fatal skin reactions/hypersensitivity, pain and sexual dysfunction).

Data Dashboard and Visualization

After the data were processed, they were reviewed using a proprietary interactive tool SPSS. The tool included search and visualization capabilities, facilitating the detection of patterns and multidirectional trends. The tool had options for identifying, displaying, and comparing large

volumes of data, both in aggregate and at the individual de-identified post level, from Twitter, Facebook, and forums. There were dashboard views for the aggregate visualization of social media sites with respect to FAERS, disproportionality scores, and geographic distribution of reports. Users created filters to explore the data according to different areas of interest, such as specific events, comparison of products, posts mentioning health system interactions or severe events, and specific time ranges sorted by drug, PT, and geo-coordinate data.

2.3 Traditional Pharmacovigilance (PV)

Methodologies in Social Media Data Assessment Post-marketing safety surveillance typically relies on data from ‘‘traditional’’ sources, including spontaneous AE reports, clinical and observational databases, and literature. These sources often include reports from healthcare providers that may provide event details, diagnostic details, medical history, and concomitant medications. These reported data are coded and evaluated using quantitative tools to identify events of interest that require qualitative review. Proto-AEs were compared with AEs retrieved from FAERS reports, the global safety database and internal safety signals from the same time period and product labeling. We also sought to evaluate whether previously identified signals identified using one traditional source, the FAERS database, might have been identified sooner through social media data analysis. In addition, the analysis of AE data using traditional PV methods typically incorporates a measure of quantitative analysis such as PRR to assess trends in the data. The usefulness of the PRR (provided by the vendor) was evaluated in assessing proto-AEs from social media data as a possible method for quantitative evaluation.

2.4 Analyzing Social Media Data to Identify Specific Issues

We investigated whether analysis of social media data could identify specific issues, such as device and drug quality issues, rare and serious events, and off-label use, and provide the patient perspective, especially regarding decisions to change or discontinue treatment. Proto-AEs were manually reviewed for these events and the patient perspective by PV subject matter experts.

Table 2 Articles published on social media mining for ADR detection, their years of publication, data sources, domains of focus, numbers of drugs, size of data, and annotations and public availability of the data.

Year	Source	Subdomain	# Drugs	# Instances	Annotations/ Available
2019	DailyMotion	–	6	10700	Yes/No
2020	DailyMotion	–	4	3600	Yes/No
2019	Yahoo! groups	–	435	3,600,000	No/No
2019	Various forums	Breast cancer	4	3,300,000	No/No
2020	Flickr	Pediatrics	9	4590	Yes/No
2020	MedHelp	–	10	18.744	No/No
2020	Twitter	Cancer	5	44,000,000,000	No/No

2019	American Diabetes Association	Diabetes	–	5,726,524	No/No
2019	Yahoo! groups	–	2	18,600	Yes/No
2020	Twitter	–	5	2431	Yes/No
2020	AskAPatient, Drugs.com, DrugRatingZ	Breast cancer	5	7500	Yes/Yes
2019	PatientsLikeMe, DailyStrength, MediGuard	–	12	38,750	No/No
2020	Twitter	–	23	270,000	No/No
2019	MedHelp	Heart disease	–	8600	Yes/No

Table 1 summarizes crucial characteristics of the studies. In addition to the publication years, it shows the data sources, subdomains of focus (if any), number of drugs involved in the studies, the sizes of the data used, and annotations and availability of the data. The table illustrates some key information regarding what pharmacovigilance using social media research has covered over the duration of January 2020- December 2020, and how research has evolved. The data collection further assessed from the surveying information utilizing the prescription reports. All information was fed into the file of Microsoft excel and further assessed by using SPSS software.

9. Results

9.1 Social Media Data Analysis

Table 3 Mention from Twitter, Facebook, and patient forums for three products from Social tool from 1 January 2020 through 31 December 2020

Product	Twitter	Facebook	Forums	Total mentions
Amlodipine	29,524	11,256	8170	48,950
Metformin	3120	92	1978	5190
Paracetamol	8168	1134	1106	10408
ibuprofen	4243	5254	13	9510
Loratadine	3510	2419	19	7839

a Mention means a social media post containing a discussion that includes the medical product of interest

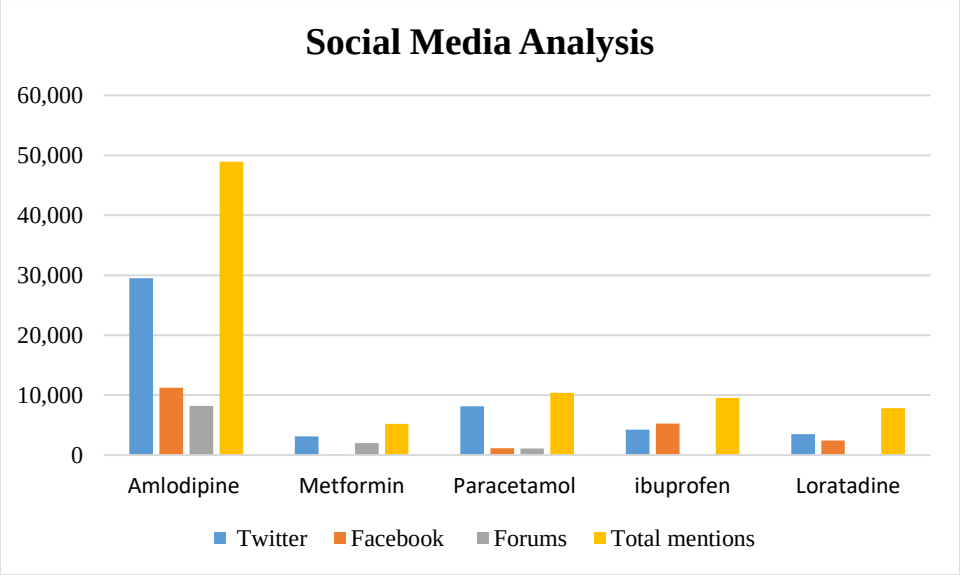
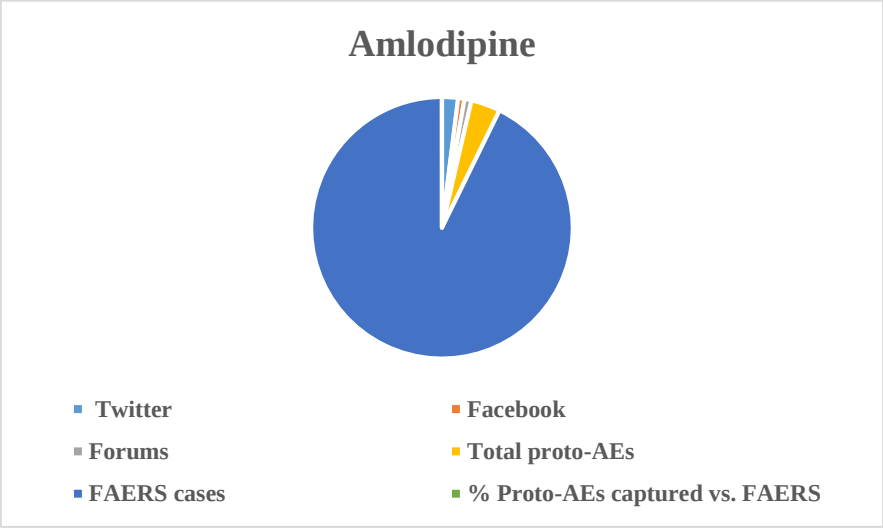
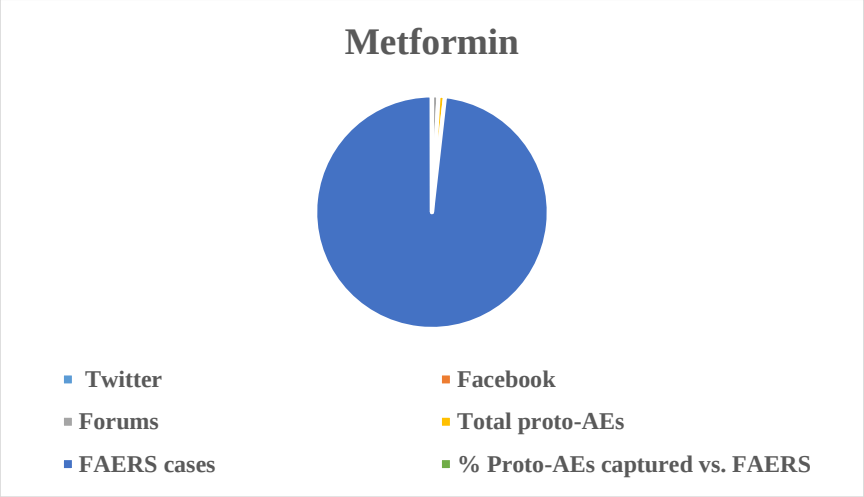


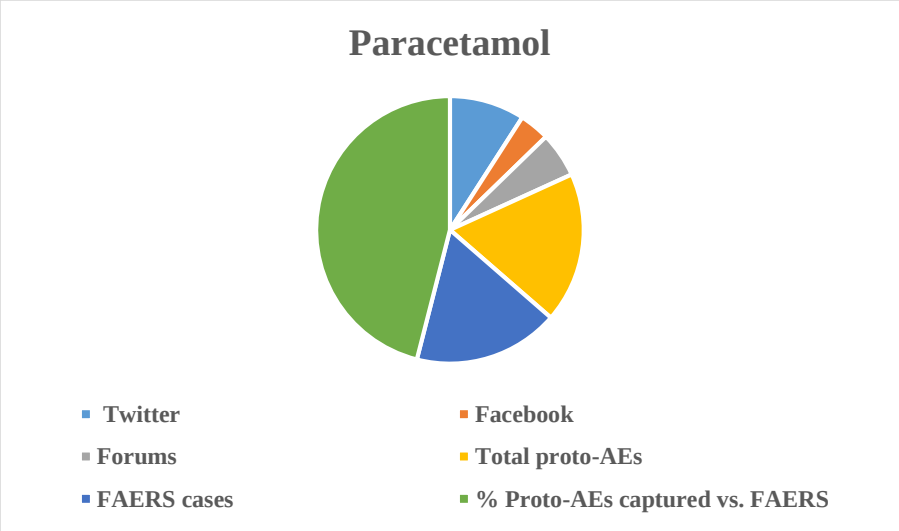
Table 2 Posts with resemblance to adverse events from Twitter, Facebook, and patient forums for three products from the Social Buzz compared with FAERS cases from 1 January 2020 to 31 December 2020

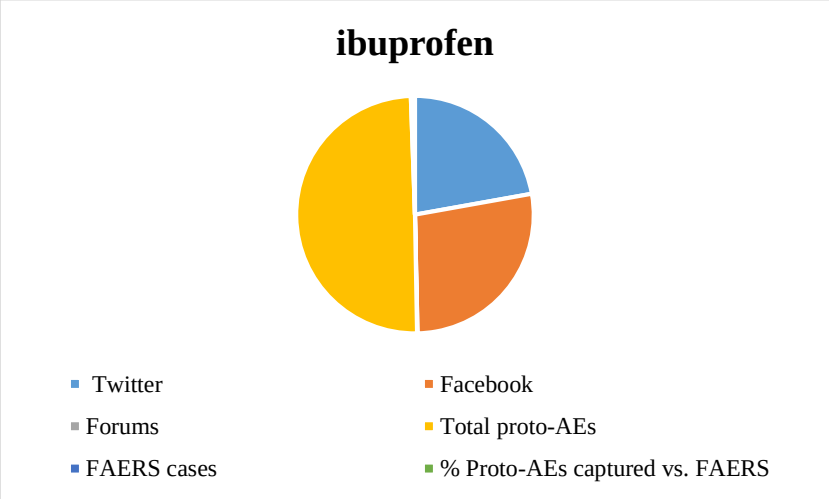
Product	Twitter	Facebook	Forums	Total proto-AEs	FAERS cases	% Proto-AEs captured vs. FAERS
Amlodipine	2130	827	935	3892	99,467	3.23
Metformin	0	0	43	43	4716	0.78
Paracetamol	15	6	9	30	29	75.86
ibuprofen	4243	5254	13	9510	97	10.29
Loratadine	3510	2419	19	7839	45	4.28

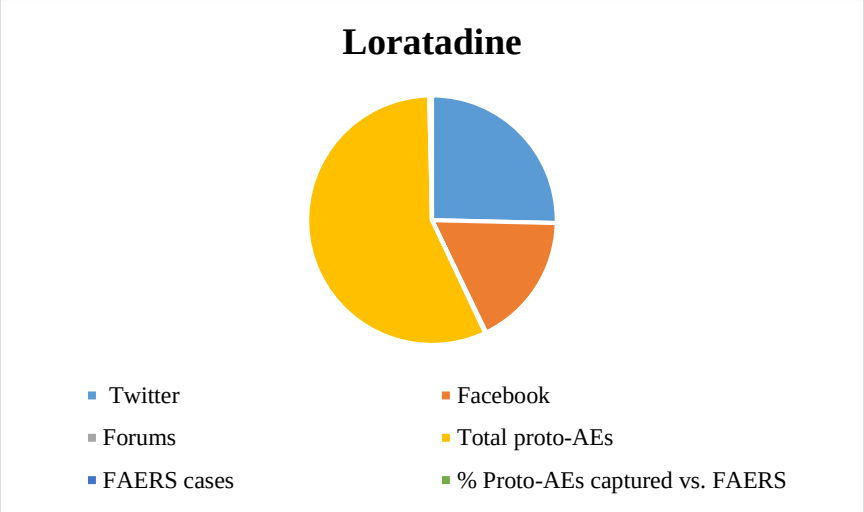
AE adverse event, FAERS US Food and Drug Administration Adverse Event Reporting System, proto-AEs posts with resemblance to Es a This calculation comprises dividing social media cases by FAERS cases, then multiplying by 100 to get the percentage




Metformin
Twitter
Facebook
Forums
Total proto-AEs
FAERS cases
% Proto-AEs captured vs. FAERS


Paracetamol
Twitter
Facebook
Forums
Total proto-AEs
FAERS cases
% Proto-AEs captured vs. FAERS


ibuprofen
Twitter
Facebook
Forums
Total proto-AEs
FAERS cases
% Proto-AEs captured vs. FAERS



9.2 Comparison of Social Media Data vs. Traditional Data Sources

From 1 January 2020 through 31 December 2020, there were 78,289 mentions of the products of interest, and 3944 posts were classified as proto-AEs. The social media proto-AEs were compared with the AE reports from the FAERS database and are summarized in Table 2. Of note, the time interval for the social media data was adjusted to consider proto-AEs from 1 January 2020 to 31 December 2020 to align with the latest available quarterly FAERS update (Table 2). However, when proto-AEs captured in social media data were compared with AEs in the FAERS database, a traditional PV data source, far fewer proto-AEs were identified within social media posts. The relatively small number of proto-AEs found in social media data sources compared with those reported in traditional data sources (e.g., FAERS) limited the ability to identify or validate safety signals using social media data.

10. Discussion

Analysis of Social Media Posts to Identify Other Safety-Related Issues

Several safety-related issues of interest were assessed to explore whether analysis of social media data could provide unique perspectives or insights. We found that analysis of social media posts provided greater insight into medication tolerability, adherence, and quality of life improvement than did events reported in the safety database. Furthermore, social media posts captured patient insights into the impact of the treatment on their quality of life; this was found through reading social media posts for specific types of events related to the drug. For example, posts that described the degree of pain associated with, and misuse/administration errors relating to, injecting medication were helpful in understanding how patients report the use of this medication. Most often, in current PV methodologies, the patient's perspective is lacking in reports by healthcare providers or AE reports in FAERS. Although greater insight was found for certain issues (i.e., quality-of-life improvement and the patient perspective), analysis of social media posts detected few device and drug product quality issues that were previously recognized through spontaneous reports within the safety database.

This pilot study had multiple dimensions: to determine whether analysis of social media data could identify signals not detected through routine traditional PV methods; to determine

whether the signals detected through routine PV methods were also evident and able to be identified sooner within social media data; to determine the quantity of proto-AEs found in social media; and to evaluate the types and characteristics of products that would benefit most from social media data analysis and usage. The results suggest that, for the selected products, analysis of social media data cannot reliably identify new signals through the application of traditional PV methods. Social media did provide unique insight in multiple areas, such as patient perspective, medication tolerability and adherence, and quality of life. However, no unique insights were gained on device and drug product quality issues. Throughout data collection and analysis, a number of limitations were encountered, which are addressed throughout the discussion.

A limitation that impacted on all results was the inherent bias associated with collection of social media data. The following factors introduced bias:

• Data acquisition: Identification and selection of the ‘‘best’’ data sources change over time. As sites that patients use to discuss their disease change, the nature and quantity of data collected from that source also changes. With new and changing data sources, it is difficult to establish a standard for comparison. In addition, a level of uncertainty is associated with who is posting information; one must consider whether the information has been shared by someone multiple times (i.e., re-tweeted) or whether the same person has publicly shared the same data in multiple forums at different times, which are counted as separate events.

There may be an effect within social media (especially in disease state forums) where a microcosm of reports become stimulated by a single or select group of hyperusers’ experiences. These sources of bias require better understanding if reliance on use of social media data for PV purposes is increased.

• Language: The social media data collected and analyzed in this pilot were limited to posts that were in either English or Spanish, and only previoulsy defined emojis were recognized.

• Demographics: Access to and the adoption and usage of the technology required to use social media is not the same for all demographic groups, which impacts on the ability of people to post information using social media.

Social Media Data Assessment

The first step in working with social media data is the collection of the data. All the papers discussed in this review perform data collection from various sources. For health related social networks, such as DailyMotion, the collection of relevant data is generally easy since the data is categorized according to various criteria (e.g., drug name). For generic social networks, such as Twitter, the collection problem is harder. It is possible to collect posts by using drug names as search keywords, but drug names are often misspelt by users. To address this problem recent research has utilized phonetic spelling filters to generate common misspellings for drug names. These recent advances in NLP will aid future data collection processes.

Following data collection, the challenge is to filter data. As explained earlier, data imbalance is an important problem in ADR mining from social media text, which has resulted in various

research tasks on classification of ADR assertive text. With the creation of recent publicly available corpora (e.g., learning algorithms can be trained and optimized to detect ADR assertive instances with high accuracies. Most classification research, however, have only used very basic linguistic features for classification (e.g., bag of words), and only very recent research has focused on exploring deep linguistic and semantic features and advanced machine learning techniques. Effective filtering/classification techniques are likely to aid the process of ADR mention extraction by removing the majority of irrelevant information. We have discussed various ADR extraction approaches in the paper, the most popular being lexicon-based ones. Lexicon-based approaches have benefited from recent expansions and merging of existing lexicons, and the incorporation of colloquial terms. Recent release of publicly available annotated data will inevitably popularize supervised learning approaches for this task.

The last step in the pipeline is to perform statistical analysis on extracted drug-ADR pairs to identify potentially harmful drugs. This step has hardly received any research attention to date, and we only identified two exploratory studies attempting to perform this task on social media data. Progress in ADR extraction and classification research is likely to raise the research focus on the analysis of drug-ADR signals generated from social media data. Considering the rapid growth of social media data, this source of information is likely to have a massive impact on pharmacovigilance research.

11. Conclusion and relevance

This pilot study found that traditional PV methods may not be appropriate when applied to the analysis of social media data for AEs. Social media can provide unique insights into the patient perspective, even with the limitations of data Social Media for Pharmacovigilance acquisition and language, among other limitations. However, these current results suggest that, for the selected products, social media data analysis may not identify new safety signals; therefore, further research is necessary to determine the best sources of and analysis methods for social media data to augment the traditional methods of PV surveillance and signaling. Obviously, social media might be valuable for screening what is going on in the network, however, this sort of information and the nature of the information should likewise be considered while considering any further administrative decision. Since underreporting of ADRs is a common problem in India, it would be appropriate to explore more and more options and prospects for tapping the power of social media in pharmacovigilance activities. However, there could be potential challenges and combined efforts of all key stakeholders that overcome the difficulties. Thus, in the near future social media can be a great tool for ADR reporting and a great platform for consumers and pharma-companies to discuss their opinions and experiences regarding the use of the medicinal product and devices.

REFERENCES:

1. Meyboom RH, Lindquist M, Egberts AC. An ABC of drug-related problems. *Drug Saf* 2000; 22: 415–423.
2. World Health Organization. *The importance of pharmacovigilance- Safety monitoring of medicinal products.* Geneva: World Health Organisation, 2002.
3. WHO-UMC. *Uppsala Reports, Issue 74, Antimicrobial resistance- An overlooked adverse event*. Uppsala, 2017.

4. World Health Organization. *Promoting rational use of medicines: Core components- WHO Policy Perspectives on Medicines No.5.* Geneva, 2002.
5. Ofori-Asenso R, Agyeman A. Irrational Use of Medicines—A Summary of Key Concepts. *Pharmacy* 2016; 4: 35.
6. Hassan A, Abdullahi GA, Ahmed AA, et al. Assessing prescriber's awareness of essential medicine list, hospital drug formulary and utilization of standard treatment guidelines in a tertiary healthcare facility in North-Central Nigeria. *Alexandria J Med* 2018; 54: 81–84.
7. Lee D, Bergman U. Studies of Drug Utilization. In: *Pharmacoepidemiology*. Wiley-Blackwell, 2012, pp. 377–401.
8. World Health Organisation. *WHO drug use in health facilites.* Geneva, 1993.
9. Isah AO, Ohaju-Obodo J, Isah EC, et al. Drug use profile in a Nigerian city hospital. *Pharmacoepidemiol Drug Saf* 1997; 6: 319–324.
10. Adebayo ET, Hussain NA. Pattern of prescription drug use in Nigerian army hospitals. *Ann Afr Med* 2010; 9: 152–8.
11. Isah A, Laing R, Quick J, et al. The Development of Reference Values for the WHO Health Facility Core Prescribing Indicators. *West Afr J Pharmacol Drug Res* 2001; 18: 6–11.
12. Akoria OA, Isah A. An evaluation of doctors' prescribing performance in Nigeria. *Pakistan J Med Sci* 2009; 25: 533–538.
13. Desalegn AA. Assessment of drug use pattern using WHO prescribing indicators at Hawassa University teaching and referral hospital, south Ethiopia: a cross-sectional study. *BMC Health Serv Res* 2013; 13: 170.
14. Okoli RI, Isah AO, Ozolua RI, et al. Drug prescribing pattern in three levels of health care facilities in the north and central senatorial districts of Edo State, Nigeria. *Int J Herbs Pharmacol Res* 2015; 4: 46–57.
15. Adisa R, Fakeye TO, Aindero VO. Evaluation of prescription pattern and patients ' opinion on healthcare practices in selected primary healthcare facilities in Ibadan , South-Western Nigeria. *Afri Heal Sci* 2015; 15: 1318–1329.
16. WHO Collaborating Centre for Drug Statistics Methodology. *Guidelines for ATC classification and DDD assignment 2018*. Oslo, 2017.
17. Federal Ministry of Health. *Essential Medicines List*. Abuja, Nigeria, 2010.
18. Ibrahim M. Physicians' prescribing behaviour in two tertiary health care facilities in north-western Nigeria - Analysis of 518 prescriptions. *Sahel Med J* 2004; 7: 115–118.
19. Jimoh AO, Omar I, Adebisi IM, et al. Review of Morbidity Profiles and Drug Prescribing Patterns of a University Clinic in North-Western Nigeria. *Int J Med Med Sci* 2014; 1: 107–115.
20. Enato EFO, Chima IE. Evaluation of drug utilization patterns and patient care practices. *West African J Pharm* 2011; 22: 36–41.
21. Sisay M, Mengistu G, Molla B, et al. Evaluation of rational drug use based on World Health Organization core drug use indicators in selected public hospitals of eastern Ethiopia: a cross sectional study. *BMC Health Serv Res* 2017; 17: 161.
22. Ijoma U, Onwuekwe I, Onodugo O, et al. Effect of promotional strategies of pharmaceutical companies on doctors' prescription pattern in South East Nigeria. *Turk Silahl Kuvvetleri, Koruyucu Hekim Bul* 2010; 9: 1–6.
23. Velo GP, Minuz P. Medication errors: prescribing faults and prescription errors. *Br J Clin Pharmacol* 2009; 67: 624–628.
24. Tamuno I, Fadare JO. Drug prescription pattern in a Nigerian Tertiary hospital. *Trop J Pharm Res* 2012; 11: 146–152.
25. Ndukwe Henry C, Ogaji J Ikoni SCN. Drug Use pattern with standard indicators in Jos University Hospital Nigeria. *West African J Pharm* 2013; 24: 88–93.
26. Ragam A, Acharya S, Holla R. Assessment of drug use pattern using World Health Organization prescribing indicators in a tertiary care hospital in Mangalore: A cross-sectional study. *Natl J Physiol Pharm Pharmacol* 2017; 7: 1.
27. Maiyaki MB, Garbati MA. The burden of non-communicable diseases in Nigeria; in the context of globalization. *Ann Afr Med* 2014; 13: 1–10.
28. Federal Ministry of Health. *National antimalarial treatment policy*. Abuja, 2005.

29. Akoria OA, Olowofela A. Impact of an educational intervention on the prescribing of antimalarials at the University of Benin Teaching Hospital. *Int J Heal Res* 2011; 4: 117–122.
30. Maryah P, Denue B, Ikunaiye N, et al. Evaluation of antimalarial drugs usage according to WHO prescribing indicators in a tertiary health facility in Maiduguri, Northeastern Nigeria. *Port Harcourt Med J* 2017; 11: 67.
31. Federal Republic of Nigeria. National Drug Formulary and Essential drug list act. Decree 43, Nigeria, 1989.
32. Atif M, Scahill S, Azeem M, et al. Drug utilization patterns in the global context: A systematic review. *Heal Policy Technol* 2017; 6: 457–470.
33. Gulmez SE, Larrey D, Pageaux G-P, et al. Transplantation for Acute Liver Failure in Patients Exposed to NSAIDs or Paracetamol (Acetaminophen). *Drug Saf* 2013; 36: 135–144.
34. Ekman E, Bäckström M. Attitudes among hospital physicians to the reporting of adverse drug reactions in Sweden. *Eur J Clin Pharmacol* 2009; 65: 43–46.
35. Eland IA, Belton KJ, van Grootheest AC, et al. Attitudinal survey of voluntary reporting of adverse drug reactions. *Br J Clin Pharmacol* 1999; 48: 623–7.