Pharm Med (2017) 31:167–174
DOI 10.1007/s40290-017-0186-6

ORIGINAL RESEARCH ARTICLE

Using Social Media Data in Routine Pharmacovigilance: A Pilot Study to Identify Safety Signals and Patient Perspectives

Mondira Bhattacharya[1] · Scott Snyder[1] · Murray Malin[1] · Melissa M. Truffa[1] · Sandy Marinic[1] · Rachel Engelmann[1] · Ritu R. Raheja[1]

Published online: 17 April 2017

Abstract

Introduction Social media is recognized as a new source of patient perspectives and data on adverse events (AEs) in pharmacovigilance (PV). Questions remain about how social media data can supplement routine PV surveillance.

Objectives The objectives of this pilot were to determine whether analysis of social media data could identify (1) new signals, (2) known signals from routine PV, (3) known signals sooner, and (4) specific issues (i.e., quality issues and patient perspectives). Also of interest was to determine the quantity of 'posts with resemblance to AEs' (proto-AEs) and the types and characteristics of products that would benefit from social media analysis.

Methods AbbVie conducted a study using 26 months of retrospectively collected social media data from Epidemico, Inc., a third-party vendor, for six products. Posts were classified, interpreted, de-identified, and filtered before analysis.

Results Analysis of social media data did not identify new or previously identified safety signals. The use of traditional PV methods to analyze social media data was unsuccessful. However, analysis of social media data did provide insights into medication tolerability, adherence, quality of life, and patient perspectives but not into device and product quality issues. The quantity of proto-AEs and new information gleaned from social media posts was small.

Conclusion The results suggest that, for selected products, social media data analysis cannot identify new safety signals. However, social media can provide unique insight into the patient perspective. Assessment was limited by numerous factors, such as data acquisition, language, and demographics. Further research is necessary to determine the best uses of social media data to augment traditional PV surveillance.

Key Points

Social media as a data source for monitoring drug safety has the potential to be beneficial; however, more research is needed to determine how it should be incorporated into pharmacovigilance (PV) processes or routine surveillance activities.

Analysis of social media data did provide insights into patient perspectives, quality of life, medication tolerability, and adherence.

Traditional PV methods were not appropriate for analysis of social media data for the selected products. Further research is needed to address limitations encountered in this study.

1 Introduction

A variety of studies have sought to determine the value of social media data in the context of traditional pharmacovigilance (PV) methodologies [1–4]. These studies have also tried to advance the analysis of social media data to

Electronic supplementary material The online version of this article (doi:10.1007/s40290-017-0186-6) contains supplementary material, which is available to authorized users.

✉ Mondira Bhattacharya
mondira.bhattacharya@abbvie.com

[1] Pharmacovigilance and Patient Safety, AbbVie Inc., 1 Waukegan Road, North Chicago, IL 60044, USA

improve its use for monitoring the safety of drugs and devices [1–4]. The use of social media data for tracking 'posts with resemblance to adverse events' (proto-AEs) is an evolving science within PV [3]. Many questions remain regarding the value of social media data in supplementing existing sources of PV data (i.e., spontaneous reports, literature reviews, and clinical trial and observational data) and adding to traditional PV methods (safety databases, AE reports, and quantitative ratios such as proportional reporting rate [PRR]). Known limitations of current sources of PV data include small sample sizes in clinical trials, under-reporting, and variable quality of post-marketing AE reports. In the industry's effort to understand the safety profile of products, the advantage of social media data lies in the potential for rapid access to emerging issues from patient-generated data. Social media data are abundant, yet the relevance and quantity of posts about specific products in social media is neither consistent nor plentiful.

Various studies have assessed the utility of social media data to identify useful safety information. These studies have presented a range of results, creating a gray area regarding the findings from and usefulness of analysis of social media data. Powell et al. [1] recently analyzed social media data to identify untapped safety and benefit information and found that social media data were suitable for post-marketing safety surveillance and product benefits. However, a recurring challenge amongst others analyzing social media data for AEs includes transforming the free-text format of posts and non-text entries (such as emojis) into clinically meaningful information in a structured data format [1, 2, 5, 6]. Natural language processing (NLP) is the main method for transforming the content of social media posts into meaningful information and is continuing to evolve to improve the accuracy of post categorization [6–9]. Additionally, classification of social media posts to the appropriate context has proven challenging, including identification of actual AEs versus random mentions of the drug [2]. Several studies have highlighted the power of social media in detecting outbreaks of serious diseases and public health issues, which demonstrates that social media analysis is timelier for reporting on certain items than standard public health reporting sources [1, 10, 11].

Other studies using social media data from Facebook, Twitter, and patient forums have focused on tracking mentions and proto-AEs for specific drugs from a company's portfolio [1, 3, 4, 12–15]. Freifeld et al. [3] compared Twitter posts coded as proto-AEs with the US FDA Adverse Event Reporting System (FAERS). At the Medical Dictionary for Regulatory Activities (MedDRA) preferred term (PT) level, the sample was too noisy to provide a meaningful analysis. However, they did find concordance at the system organ class (SOC) level [3]. Overall, few studies to date have compared proto-AEs from social media data against AEs from traditional PV sources (e.g., FAERS).

This pilot study was conducted to ascertain whether social media data could be analyzed to identify signals not detected through routine PV methods and whether signals collected through routine PV methods were also evident and could be identified in social media data. As the social media data were analyzed, it was of additional interest to determine whether unique insights could be identified, such as product quality complaints, patient perspectives, and off-label uses. To further evaluate the hypothesis that social media may represent a large untapped data source for safety organizations, the pilot also sought 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.

2 Methods

Epidemico, Inc. (an authorized third-party vendor) collected publicly accessible posts in English and Spanish from Facebook, Twitter, and patient forums[1] from 1 January 2014 through 29 February 2016[2] using their respective search application programming interfaces (APIs). The sources were prospectively monitored for posts mentioning the six products of interest: AbbVie products Humira (adalimumab), Viekira Pak (dasabuvir-ombitasvir-paritaprevir-ritonavir), Creon (pancrelipase), and Lupron (leuprolide) and non-AbbVie products Harvoni (ledipasvir-sofosbuvir) and Xeljanz (tofacitinib). A first-pass filter, which contained a list of medical product keywords (including misspellings of drugs, generic names, and slang terms) was used to account for the diverse ontology of terms used in colloquial posts.

2.1 Data Preparation Process (Including De-Identification)

Following data collection, Epidemico, Inc. utilized a machine-learning filtering process in which Bayesian classifiers removed duplicates, spam, and irrelevant items on all social media posts collected. This machine-learning algorithm was used to recognize proto-AEs in which a potential AE was discussed within the context of use of a drug. The classifier then translated symptoms (interpretation) described using colloquial language and slang into MedDRA PTs. The classifier is a naive Bayesian

[1] The full listing of patient forums for each product is included in the electronic supplementary material.

[2] In October 2015, Facebook data were withdrawn after changes to its API restricting access to public data to all third parties.

probabilistic model that was developed based on Robinson's approach to eliminating spam emails [17]. Epidemico's vernacular-to-MedDRA processor uses a tree-based, text-matching algorithm to match unstructured text to MedDRA PTs according to a proprietary symptom dictionary that consists of thousands of colloquial phrases that may be used to describe medical concepts.

During Epidemico's classification process, an indicator score between 0 and 1 was automatically given to each post based on the similarity to other proto-AEs in the manually trained classifier dataset. Epidemico determined that an indicator score of 0.7 was required for the post to be included in the data dashboard for further analysis. Posts with higher scores were indicative of proto-AEs and underwent manual curation to improve the automated Bayesian classification process and ensure the ability to recognize correct syntax and slang in order to reduce false positives. Each post was filtered and classified as a proto-AE or mention (i.e., post discusses the drug of interest without a proto-AE).

A data collection and classification process similar to that described in further detail by Powell et al. [1] was followed.

Prior to AbbVie's access and review, metadata, such as geography, timeframe, data source, and personally identifiable information (PII) were removed to de-identify the posts while maintaining non-identifying metadata for geospatial analysis. In this study, PII is any information that may identify the original author of a social media post.

2.2 Data Dashboard and Visualization

After the data were processed, they were reviewed using a proprietary interactive tool, MedWatcher Social©, from Epidemico, Inc. The tool included search and visualization capabilities, facilitating the detection of patterns and multi-directional 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 MedDRA SOC/PTs, frequency, comparison with FAERS, disproportionality scores, and geographic distribution of reports.

Users created filters to explore the data according to different areas of interest, such as specific MedDRA events, comparison of products, posts mentioning health system interactions or severe events, and specific time ranges sorted by drug, MedDRA 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, AbbVie's global safety database and internal safety signals from the same time period (for AbbVie products only), 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.

3 Results

3.1 Comparison of Social Media Data vs. Traditional Data Sources

From 1 January 2014 through 29 February 2016, there were 78,289 mentions of the six products of interest (see Table 1), 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 2014 to 31 December 2015 to align with the latest available quarterly FAERS update (Table 2). Of the six products of interest, Humira ($n = 3213$) and Harvoni ($n = 413$) had the most proto-AEs. 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

Table 1 Mentions[a] from Twitter, Facebook, and patient forums for six products from the MedWatcher Social© tool from 1 January 2014 through 29 February 2016

Product	Twitter	Facebook	Forums	Total mentions
Humira (adalimumab)	24,873	9933	6260	41,066
Viekira (dasabuvir, ombitasvir, paritaprevir and ritonavir)	1822	75	1738	3635
Creon (pancrelipase)	6803	856	984	8643
Lupron (leuprolide acetate)	322	264	40	626
Harvoni (ledipasvir, sofosbuvir)	12,116	499	9242	21,857
Xeljanz (tofacitinib citrate)	1679	738	45	2462

[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 six products from the MedWatcher Social© tool compared with FAERS cases from 1 January 2014 to 31 December 2015

Product	Twitter	Facebook	Forums	Total proto-AEs	FAERS cases	% Proto-AEs captured vs. FAERS[a]
Humira	1920	606	687	3213	99,467	3.23
Viekira	0	0	37	37	4716	0.78
Creon	11	3	8	22	29	75.86
Lupron	5	0	0	5	8585	0.06
Harvoni	19	1	393	413	12,762	3.24
Xeljanz	23	77	9	109	6847	1.59

AE adverse event, *FAERS* US Food and Drug Administration Adverse Event Reporting System, *proto-AEs* posts with resemblance to AEs

[a] This calculation comprises dividing social media cases by FAERS cases, then multiplying by 100 to get the percentage

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.

3.2 Traditional PV Methodologies in Social Media Data Assessment

Traditional PV methodologies were applied to assess trends and rate of occurrence of proto-AEs in social media data compared with traditional data sources (e.g. review of data trends, tabulation, and summary based on MedDRA coding). The frequency and type of PTs in social media posts were not aligned with the known distribution of MedDRA PTs aggregated at the SOC level within traditional data sources (FAERS and company database). Figure 1 illustrates the PTs from proto-AEs for Humira in social media and FAERS. There are differences between the frequencies of PTs between sources. A majority of the PTs from proto-AEs fall within the 'general disorders and administration site reactions' SOC, which contained five times more PTs than any other SOC evaluated through analysis of social media proto-AEs (data not shown).

Other traditional PV methodologies were evaluated in social media data assessment. Proto-AEs identified in social media were evaluated graphically over time, and data peaks were investigated. The results showed that these peaks did not represent safety signals but may be attributable to other biases, including changes in data-collection techniques. The vendor-provided PRR for each event for a given drug and event pair was evaluated; PRR was calculated as follows:

$$\text{PRR} = \frac{a/(a+b)}{c/(c+d)}$$

where a is the number of proto-AEs involving drug X and event Y, b is the number of proto-AEs involving drug X and NOT event Y, c is the number of proto-AEs involving NOT drug X and event Y, d is the number of proto-AEs involving NOT drug X and NOT event Y, and "NOT Drug X" refers to all other products in the MedWatcher Social© system.

The utility of the PRR provided was limited as PRR has not been validated against currently accepted PV metrics. Further research into other possible methods of quantitative analysis for social media data did not yield additional results. No safety signals from social media data were

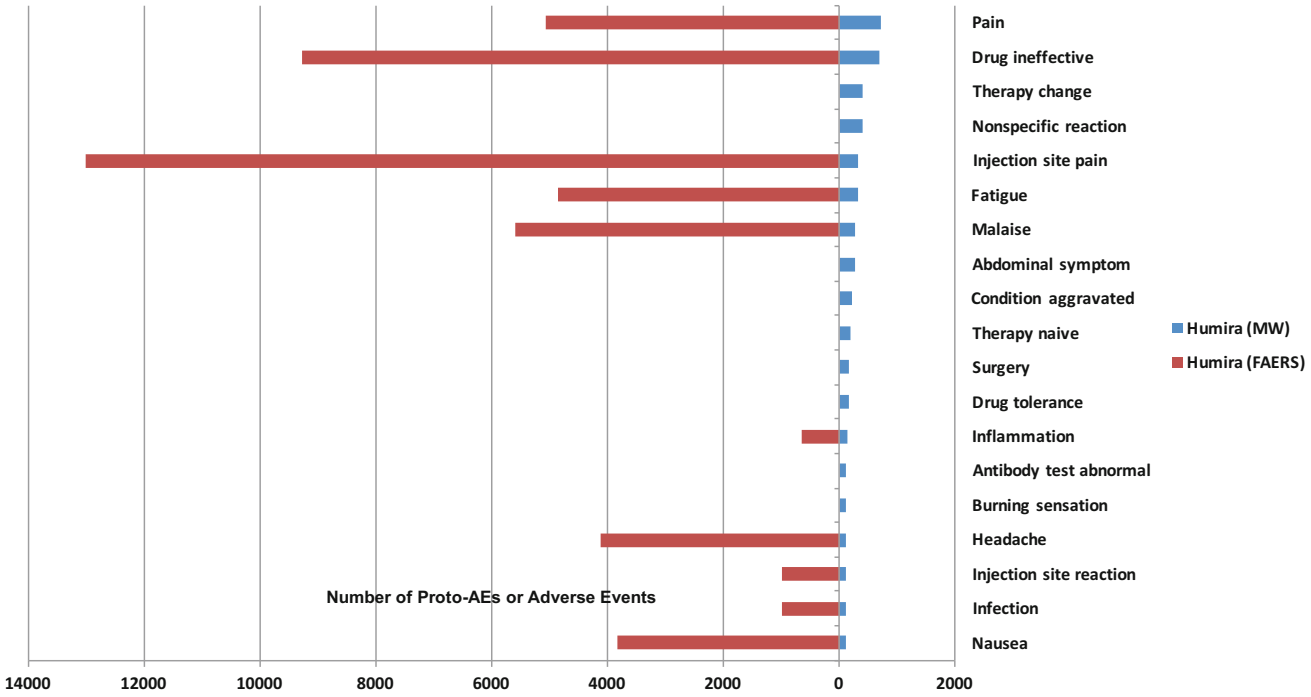

Fig. 1 Comparison of number of proto-AEs PTs in MedWatcher Social© tool to number of AEs in FAERS, from January 01, 2014 through February 29, 2016 for Humira (adalimumab). *AEs* adverse events, *FAERS* Food and Drug Administration Adverse Event Reporting System, *MW* MedWatcher© tool

identified for AbbVie products based on the evaluation of SOC and PT tabulations, comparison with FAERS and AbbVie's internal data (for AbbVie products), time series analysis, and evaluations of PRR.

3.3 Retrospective Comparison of Social Media Data to Internally Identified Safety Signals

Social media data were evaluated to retrospectively identify known safety signals from traditional PV sources. Specific PTs or distinct medical concepts identified as signals for these AbbVie products were not found within posts from social media for the same products. The social media data were not precise enough to map to the distinct medical concepts that characterized the safety signals. Therefore, this assessment concluded that the methodology and sources of social media data (e.g., hepatic decompensation in patients treated with Viekira) used in this analyses was inadequate for identifying signals that were observed through routine PV. It was assessed that social media was not able to support timely surveillance and evaluation of safety signaling.

3.4 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 AbbVie safety database (Fig. 2). 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 (Fig. 2). For example, posts that described the degree of pain associated with, and misuse/administration errors relating to, injecting Humira 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 AbbVie safety database.

4 Discussion

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

Example 1: Thanks for sharing, I did not realize that this was a side effect. The side effects went away for me (fatigue, nausea and insomnia) and then came back.

Example 2: I have a stressful job. Fatigue is certainly a side effect. Is anyone having difficulty concentrating? I can't seem to stay on track during the day.

Example 3: Joint pain has been the worst for me. The nurse told me this was uncommon, which led me to this site. Thanks for sharing your experiences!

Fig. 2 Sample Social Media posts that capture patient insights into the impact of treatment on their quality of life from MedWatcher Social© tool

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 hyper-users' 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.
- Public discussion of health-related issues: Patients may discuss medically complex or sensitive issues more frequently within closed forums than more public forms of social media. It was observed that events related to tolerability and/or quality of life were discussed more often than other types of AEs. As seen in this study with Viekira, patients with hepatitis C tend to share information only in private forums and not publicly on Facebook or Twitter (Table 2). There most likely are other diseases for which the stigma associated with the disease will influence where patients discuss treatment options and side effects of current treatments.
- Product market exposure: The market exposure and length of time marketed impacts on the number of user posts and AEs [16]. Although older products would be expected to have more posts than a similar "newer" product, products introduced into the market prior to the advent of social media may not have a significant social media presence as patients understood the safety profile prior to the availability of social media [16].

The initial hypothesis was that social media data may represent a large, untapped data source providing a unique perspective to a safety science organization. As the results from this pilot study suggest, based on the methodology utilized for the aggregation and assessment of the data, social media data does not have the magnitude of proto-AEs that may have been expected relative to traditional PV data sources (e.g., FAERS). The value of social media proto-AEs was low in two ways: first, the data were relatively limited and not very relevant and, second, the number of posts was much lower than in traditional data sources (i.e. FAERS and company database).

The nature of the disease, demographics of the indicated population, and the geographic distribution of the patient population may affect the number of proto-AEs for the selected products. For example, patients receiving Creon proportionately had more proto-AEs than those receiving other products with respect to the FAERS database (Table 2). This may have been because Creon dosing is adjusted by patients based on clinical symptoms and other individual factors, notably dietary fat intake, resulting in these patients conversing more frequently while seeking advice from or providing advice to others. One could also

postulate that patients may be more likely to discuss certain diseases more openly than others.

Humira had a disproportionately low number of proto-AEs compared with the FAERS database (see Table 2). This low proto-AE rate may be due to the age of the product and the robust understanding of the safety profile among patients based on the materials provided when initiating therapy. Although Lupron is primarily used by men with prostate cancer, most proto-AEs were reported by females (in whom the medicine is indicated for uterine fibroids and endometriosis), suggesting this bias is a function of demographics.

The application of traditional PV methodologies and analysis techniques to social media data was initially considered a pragmatic approach. However, limitations become evident when applying traditional PV methodologies (including the aforementioned biases).

Coding of social media posts to a standard medical dictionary such as MedDRA is difficult and resulted in inconsistencies. The manner in which patients discuss their health on social media is difficult to compare with standard MedDRA dictionary coding (i.e., patients discuss liver issues rather than hepatic decompensation or non-serious transaminase elevations). This is further demonstrated by the data in Fig. 1 as patients discussed proto-AEs that affected their quality of life more often than other proto-AEs (i.e., pain and fatigue vs. antibody test abnormal and nonspecific reaction). The meaning and clinical context of what the user was referring to in their post was often lost. Translating and providing structure to social media posts has been a common challenge in using social media data to inform PV practices [2, 9]. Although NLP algorithms for extracting and interpreting social media data are becoming more sophisticated, manual curation is still required [6–9]. Given the increasing quantity of social media sources and data, a significant investment in resources would be needed to make the data useful.

Comparison of FAERS events (i.e., PTs) with social media proto-AEs is difficult because of bias, including limitations mapping social media posts to MedDRA PTs and differing patient populations reporting each channel. Freifeld et al. [3] compared Twitter proto-AEs with FAERS events but the excessive noise in the data prevented any solid conclusion; our analysis yielded similarly inconclusive results. Topics discussed on social media, particularly on Twitter and Facebook, tended to be more general. Golder et al. [5] found that social media tended to over-represent mild and AE-related symptoms, whereas laboratory test abnormalities and serious AEs were underrepresented compared with other data sources. In this pilot study, it was often challenging to interpret the posts in the absence of medical history, comorbidities, and other demographic information, all of which were rarely available. Social media data are better for measuring awareness and public sentiment for emerging safety issues [14].

A common quantitative tool used in traditional PV methods is disproportionality analysis, including evaluation of PRR. However, this is not yet a meaningful method for analysis of social media data. The size and scope of the denominator is constantly changing within the realm of social media, making interpretation of the PRR challenging. The PRR would also be subject to the same bias of the sources used to collect the data. To more easily quantitatively interpret social media data, a method will need to be developed to analyze disproportionality. While this pilot did not find social media data analysis to be useful in the identification of safety signals for the selected products, social media data may contain unique insights into issues such as patient perspectives, medication adherence, tolerability, and off-label use. Social media data analysis may also support risk-minimization activities (risk-effectiveness studies), benefit–risk assessments, and real-time safety profile assessment.

Throughout this study, a number of limitations and biases were encountered that led to recommendations for future actions to improve the utility of social media data analysis for supplementing routine PV surveillance. First, the most useful data sources for specific products could be identified through improved understanding of the stigma associated with each disease and how comfortable the patients feel discussing their treatments in forums or on public sites. Second, analysis of social media data has progressed over the last 5 years because of advancements in NLP and improved language detection. Future research on these techniques will likely lead to new developments in the field. With new social media sites and technologies continuing to emerge and evolve, additional data sources and tools will become available to improve NLP. Third, as social media post structure has caused difficulties in translating information into medical events and terminologies, there is reason to consider other methods of collecting relevant information about drug safety and AEs from social media until the technology improves. There is great potential for growth in using social media for monitoring drug safety, and only time will tell how it will be incorporated into PV processes or routine surveillance activities.

5 Conclusions

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

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.

Acknowledgements Epidemico Inc., and staff provided access to data and training but did not influence the data analysis, results, conclusions, or content of this manuscript. Epidemico Inc. had an opportunity to review drafts of this manuscript. The authors acknowledge the following groups within AbbVie that supported the review of the social media data and tool used in this analysis: Medical Safety Operations, Medical Safety Assessment, Medical Safety Evaluation, Epidemiology, Contact Center, Aggregate Safety, Safety Technology Solutions, and Data Management. Contractual medical writing support was provided by Mia DeFino, MS, PharmaStart, LLC, Northbrook, IL, USA. Her role was to organize the work by providing an outline to the multiple authors (each contributing to specific sections), compile the various drafts to allow for ease of review by all authors and edit and finalize the supportive tables.

Compliance with Ethical Standards

Funding AbbVie Inc. paid for access to data through the Epidemico Inc. platform, including time from Epidemico Inc. staff to train AbbVie personnel on data review. Mia DeFino, MS, received compensation for medical writing support.

Conflict of interest Mondira Bhattacharya, Scott Snyder, Murray Malin, Melissa M. Truffa, Sandy Marinic, Rachel Engelmann, and Ritu R. Raheja are employees of AbbVie Inc. and are shareholders of AbbVie stock.

Ethical standards Internal authorship standards were followed in that only individuals who contributed to the design and conduct and analysis of the research were co-authors of the manuscript. Data were de-identified (Sect. 2.1).

References

1. Powell GE, Seifert HA, Reblin T, Burstein PJ, Blowers J, Menius JA, et al. Social media listening for routine post-marketing safety surveillance. Drug Saf. 2016;39(5):443–54. doi:10.1007/s40264-015-0385-6.
2. Lardon J, Abdellaoui R, Bellet F, Asfari H, Souvignet J, Texier N, et al. adverse drug reaction identification and extraction in social media: a scoping review. J Med Internet Res. 2015;17(7):e171. doi:10.2196/jmir.4304.
3. Freifeld CC, Brownstein JS, Menone CM, Bao W, Filice R, Kass-Hout T, et al. Digital drug safety surveillance: monitoring pharmaceutical products in Twitter. Drug Saf. 2014;37(5):343–50. doi:10.1007/s40264-014-0155-x.
4. Carbonell P, Mayer MA, Bravo A. Exploring brand-name drug mentions on Twitter for pharmacovigilance. Stud Health Technol Inform. 2015;210:55–9. doi:10.3233/978-1-61499-512-8-55.
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):878–88. doi:10.1111/bcp.12746.
6. Leaman R, Wojtulewicz L, Sullivan R, Skariah A, Yang J, Gonzalez G, editors. Towards internet-age pharmacovigilance: extracting adverse drug reactions from user posts to health-related social networks. In: Proceedings of the workshop on biomedical natural language processing; 2010 July 15; Uppsala, Sweden: Association for Computational Linguistics.
7. Sarker A, Ginn R, Nikfarjam A, O'Connor K, Smith K, Jayaraman S, et al. Utilizing social media data for pharmacovigilance: a review. J Biomed Inform. 2015;54:202–12. doi:10.1016/j.jbi.2015.02.004.
8. O'Connor K, Pimpalkhute P, Nikfarjam A, Ginn R, Smith K, Gonzalez G, editors. Pharmacovigilance on twitter? Mining tweets for adverse drug reactions. In: AMIA Annual Symposium; 2014.
9. Liu X, Chen H. A research framework for pharmacovigilance in health social media: identification and evaluation of patient adverse drug event reports. J Biomed Inform. 2015;58:268–79. doi:10.1016/j.jbi.2015.10.011.
10. Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection–harnessing the web for public health surveillance. N Engl J Med. 2009;360(21):2153–7. doi:10.1056/NEJMp0900702.
11. Chowell G, Nishiura H. Transmission dynamics and control of Ebola virus disease (EVD): a review. BMC Med. 2014;12:196. doi:10.1186/s12916-014-0196-0.
12. Wu H, Fang H, Stanhope SJ. Exploiting online discussions to discover unrecognized drug side effects. Methods Inf Med. 2013;52(2):152–9. doi:10.3414/ME12-02-0004.
13. Harpaz R, Callahan A, Tamang S, Low Y, Odgers D, Finlayson S, et al. Text mining for adverse drug events: the promise, challenges, and state of the art. Drug Saf. 2014;37(10):777–90. doi:10.1007/s40264-014-0218-z.
14. Coloma P, Becker B, Sturkenboom M, van Mulligen E, Kors J. What can social media networks contribute to medicines safety surveillance? Pharmacoepidemiol Drug Saf. 2015; 24(Suppl. 1):467–8 [abstract 818].
15. Pages A, Palosse L, Guitton E, Lacroix I, Damase-Michel C, Herxheimer A, et al. PharmacoCyberVigilance: new tool to improve adverse drug detection? J Biomed Inform. 2015;54:202–12.
16. Weber J. Epidemiology of adverse reactions to nonsteroidal anti-inflammatory drugs. Adv Inflamm Res. 1984;6:1–7.
17. Robinson G. A statistical approach to the spam problem. Linux J. 2003;107:3.