MEDINFO 2021: One World, One Health – Global Partnership for Digital Innovation
P. Otero et al. (Eds.)

doi:10.3233/SHTI220138

Data Veracity of Patients and Health Consumers Reported Adverse Drug Reactions on Twitter: Key Linguistic Features, Twitter Variables, and Association Rules

Tianchu Lyu[a*], Andrew Eidson[b*], Jungmi Jun[c], Xiajie Zhou[d], Xiang Cui[e], Chen Liang[a]

[a] *Department of Health Services Policy and Management, Arnold School of Public Health, University of South Carolina, Columbia, South Carolina, USA*
[b] *School of Medicine, University of South Carolina, Columbia, South Carolina, USA*
[c] *School of Journalism and Mass Communications, College of Information and Communications, University of South Carolina, Columbia, South Carolina, USA*
[d] *Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, Nanjing, Jiangsu, China*
[e] *Department of Epidemiology, Fuwai Hospital, National Center for Cardiovascular Diseases, Chinee Academy of Medical Sciences and Peking Union Medical College, Beijing, China*
* *Equal contribution*

Abstract

As Twitter emerged as an important data source for pharmacovigilance, heterogeneous data veracity becomes a major concern for extracted adverse drug reactions (ADRs). Our objective is to categorize different levels of data veracity and explore linguistic features of tweets and Twitter variables as they may be used for automatic screening high-veracity tweets that contain ADR-related information. We annotated a published Twitter corpus with linguistic features from existing studies and clinical experts. Multinomial logistic regression models found that first-person pronouns, expressing negative sentiment, ADR and drug name being in the same sentence were significantly associated with higher levels of data veracity ($p<0.05$), using medical terminology and fewer indications were associated with good data veracity ($p<0.05$), less drug numbers were marginally associated with good data veracity ($p=0.053$). These findings suggest opportunities for developing machine learning models for automatic screening of ADR-related tweets using key linguistic features, Twitter variables, and association rules.

Keywords:

Drug-Related Side Effects and Adverse Reactions, Pharmacovigilance, Data Mining

Introduction

ADRs have been one of the leading causes of morbidity and mortality in the United States (US) and a significant cost driver, accounting for about USD 30.1 billion of annual health expenditure [23]. The World Health Organization (WHO) defined the ADR as “a response to a drug which is noxious and unintended, and which occurs at doses normally used in man for the prophylaxis, diagnosis, or therapy of disease, or for the modifications of physiological function” [24]. According to the Institute of Medicine, ADRs account for 7,000 deaths among 44,000 to 98,000 deaths caused by medical errors annually [12]. The annual rate of deaths has increased from 1999 to 2006, which is 8 to 12 per 1,000 people [22]. Up to 12% of hospital admissions were associated with ADRs, with higher rates in old patients [16]. Age and the number of medications are risk factors for ADRs in outpatients [25] as older people often need more medications due to coexisting chronic conditions, and incidents often occur post outpatient discharges, collectively suggesting the importance of collecting pharmacovigilance (PV) data from patients and health consumers when they are no longer monitored by clinical systems. The notion of PV refers to the process of monitoring ADRs and other drug-related problems, which has been widely used to detect, assess, and understand ADRs.

Over years, researchers have been seeking ways to improve PV and have identified problems limited by the data sources of PV. Traditional sources of data used for identifying ADRs include clinical trials, pharmaceutical industry reports, and spontaneous reporting systems (SRSs). Electronic health records (EHR) have also been used as a promising data source for PV in recent years. However, these data sources may be limited in terms of capturing the full spectrum of ADRs along the course of health care and social care. For example, clinical trials have limitations in detecting rare ADRs because of typical short durations and small sample sizes. SRSs include information only from standardized reports and, therefore, suffer from under-reporting problems. It is also particularly challenging to acquire ADR data from hard-to-reach patients and health consumers who have experienced ADRs but have not interacted with any sector of the health care system [14].

Social media is a viable data source for collecting ADRs reported by patients and consumers, serving as an important supplement for traditional PV [21]. Avery and colleagues suggested that the types and adverse effects of drugs reported by patients are different from those reported by health care professionals and patient-reported data contain more details [2]. Social media platforms, such as Twitter, are less costly to access and produce richer data abundance [6]. It was reported that social media data can be utilized for validating pre-existing ADRs as well as detecting signals of new or rare ADRs [8]. Abou Taam et al. examined ‘benfluorex’ related content on three social media websites and found various ADRs such as anxiety, anger, and valvulopathy [1]. Yang et al.’s analysis of social media content (e.g., MedHelp, PatientsLikeMe) and identified drug-ADR pairs, such as ‘Lansoprazole’ and ‘diarrhea’, ‘Prozac’ and ‘depression’, and ‘Luvox’ and ‘heart disease’ [26]. Pierce et al. examined ten safety signals and found that ‘dronedarone–vasculitis’ exhibited in social media prior to FDA signal detection, suggesting that social media listening could contribute to early detection of ADRs [20].

Despite these advantages, uneven data veracity is becoming a major concern, yet systematic studies and clinically meaningful guidelines to obtain high veracity of data is sparse. Among pilot studies, Hoang and colleagues measured data uncertainty and rarity for ADRs extracted from Twitter and proposed authenticity and credibility as two root causes of poor data veracity [10]. Nguyen et al. evaluated trustworthiness to improve the accuracy of social media data [17]. Additionally, data veracity problems were reported to threaten the availability, confidentiality, and integrity of social media data, and data analyses [11]. Building on the existing literature, we summarized three obstacles to good data veracity on ADR extracted from social media. First, most patients and health consumers are laypersons who tend to use their own terms in describing a health issue on social media, which often differ from terminologies that are validated and used by healthcare professionals such as the Unified Medical Language System (UMLS) [4]. Second, social media users may not constantly provide credible information as compared to regulatory individual case safety reports (ICSR) provided by pharmaceutical and biotechnology companies [3]. Third, laypersons may not be able to correctly link an ADR to the corresponding medication. They have difficulties in distinguishing ADRs from comorbidities or indications [9]. These obstacles significantly impede novel PV research using social media, yet the process of identifying different levels of data veracity is labor-intensive because the task requires dissecting improvised social media posts in free text, often complicated by frequent typos, copy-forwarding, and context-dependent language.

Natural language processing (NLP) methods have provided a feasible way to clean and prepare the textual social media data before being used for PV, including spelling correction, lemmatization, lowercasing, annotation, etc. Using linguistic features such as first-person pronoun (i.e., "I", "me", "my") or URL could help improve the accuracy of extracting ADR-related messages [7]. NLP also showed potentials to recognize, extract, and quantify subjective experiences from Twitter users, which is often denoted as sentiment analysis [15]. Despite the fast development of NLP approaches used for identifying ADR-related tweets, linguistic features and methods tailored for data veracity of ADRs on social media are sparse.

While there has been increasing attention devoted to the data veracity issue of ADR-related social media [10], no existing studies have systematically defined and characterized the data veracity of ADR-related social media data to be used for PV. We propose a viable approach for identifying different levels of data veracity for ADRs extracted from Twitter and compared levels of data veracity with different linguistic features and Twitter variables. We integrated existing Twitter variables and nuanced linguistic features extracted from Twitter posts to identify principal Twitter variables and linguistic features that contribute to individual levels of ADR data veracity using multinomial logistic regression models. Linguistic features of Twitter posts are detected by employing an annotation protocol that is clinically plausible for screening ADR events. Anticipated outcomes of this study hold promise to inform the future development of a machine learning model that can automatically screen high-veracity ADRs detected from Twitter, which will greatly facilitate PV research using social media data.

Materials and Methods

Data Source

The data were from a publicly available corpus of tweets containing ADRs [7]. The corpus consists of 10,822 tweets that were annotated with medications and indications.1,217 tweets contain at least one ADR. We retrieved all the available ADR tweets by Twitter IDs via the Twitter API implemented with Python in June of 2020. 766 tweets were no longer retrievable since the corpus was created in 2014 and some tweets or user profiles had been deleted by the time we collected the data. As a result, we retrieved 451 tweets labeled in the corpus as "containing an ADR" for analyses. For validating drug-ADR pairs in the corpus, we used SIDER 4.1 database [13]. Drug-ADR pairs identified in the corpus were compared against the clinically validated drug-ADR pairs documented in SIDER 4.1.

Annotation

Two clinical experts (AE and TL) performed an annotation task to identify levels of data veracity as well as ten critical linguistic features from the tweets. They have received proper training in pharmacology and ADR-related work from medical schools. The operational definition of ADR is "an undesired effect of the drug experienced by the patient" and an indication as "the sign, symptom, syndrome, or disease that is the reason or the purpose for the patient taking the drug or is the desired primary effect of the drug" [19]. Based on the key linguistic features used in prior studies for identifying ADRs on Twitter, clinical experts developed a protocol for the annotation (Figure 1). Each expert followed this protocol and independently performed the annotation task. The interrater reliability was measured by Cohen's Kappa Statistic using data veracity level as representative measurement. The calculated Kappa value is 0.80, indicating high agreement. Disagreement in annotations was resolved during the panel discussion. Below we summarized the key components of the protocol including data veracity levels and key linguistic features.

Veracity Levels

The data veracity of a tweet is defined in three levels including poor, moderate, and good data veracity. To be annotated as good data veracity, the tweets should (1) be correctly identified as an ADR tweet in the corpus, (2) not be a retweet, (3) contain no URL, (4) be the first-person experience, (5) explicitly state the drug name and corresponding ADRs, and (6) correctly include the indications if any. A tweet was classified as moderate data veracity when the tweet did not contain sufficient information required for determining good data veracity. Specifically, moderate data veracity was identified when the expert was unable to decide whether the Twitter user reported an ADR they had experienced, or a single drug-ADR pair could not be identified. For example, tweets with moderate data veracity might contain a potential ADR that has been identified but was not listed in the SIDER database. Tweets with poor data veracity could be expert views, advertisements, case reports, and tweets that contain URLs or indications that were mistakenly identified as ADRs. Table 3 is an operational classification system for data veracity generated based on the annotation protocol (Figure 1).

Key Linguistic Features

- First-person, second-person, and third-person pronouns to improve the accuracy of identification.[5; 10]
- Sentiments, as most ADR tweets are associated with negative sentiment.[5; 7; 10; 15; 18]
- Mentions of drugs and ADRs in the same sentence.
- The presences of ADRs per SIDER and Medical Dictionary for Regulatory Activities (MedDRA).
- The number of drugs.
- The number of indications.

Multinomial Logistic Regression

Multinomial logistic regression was applied to explore the effects of different indicative and contra-indicative factors of interest on data veracity. We removed two variables before model selection: URL and retweet, since the extreme imbalance in the frequencies of values can lead to a quasi-complete separation of data. We used the Backward Selection method to select the final model. The likelihood ratio test was applied to compare the goodness of fit of the full model and the reduced models. Alternatively, we calculated the Akaike Information Criterion (AIC) to test the goodness of fit of the model. No obvious multicollinearity was observed, assessed by variance inflation factors (VIF), tolerance statistics, and eigenvalue and condition index. The analyses were performed via SAS 9.4.

Holding other factors as constants, the odds ratio of using first-person pronouns vs. not using first-person pronouns was 5.80 for tweets in moderate vs. poor data veracity ($p<0.001$). The ratio increased to 7.89 when comparing good data veracity with poor data veracity ($p<0.001$). The odds of tweets in moderate and good data veracity that express negative sentiment were 5.59 and 6.33 times, respectively, as that of tweets in poor data veracity ($p<0.001$ and $p<0.001$). Similarly, the odds of the presence of drug names and ADRs in the same sentence in tweets in good and moderate data veracity were 6.20 and 3.77 times, respectively, as that of tweets in poor data veracity ($p<0.001$ and $p=0.002$). No significant difference was observed when we evaluated the effect of using second-person pronouns, the number of drugs, and the number of indications in tweets between different data veracity levels. However, the differences in the number of drugs and indications became significant at 0.1 and 0.05 significance level when we only compared tweets in poor and good data veracity ($p=0.053$ and $p=0.047$). The odds ratio of tweets with good data veracity was multiplied by 0.50 for every one more drug mentioned in the tweets ($p=0.053$), and multiplied by 0.34 for every one more indication mentioned in the tweets ($p=0.047$) (Tables 2).

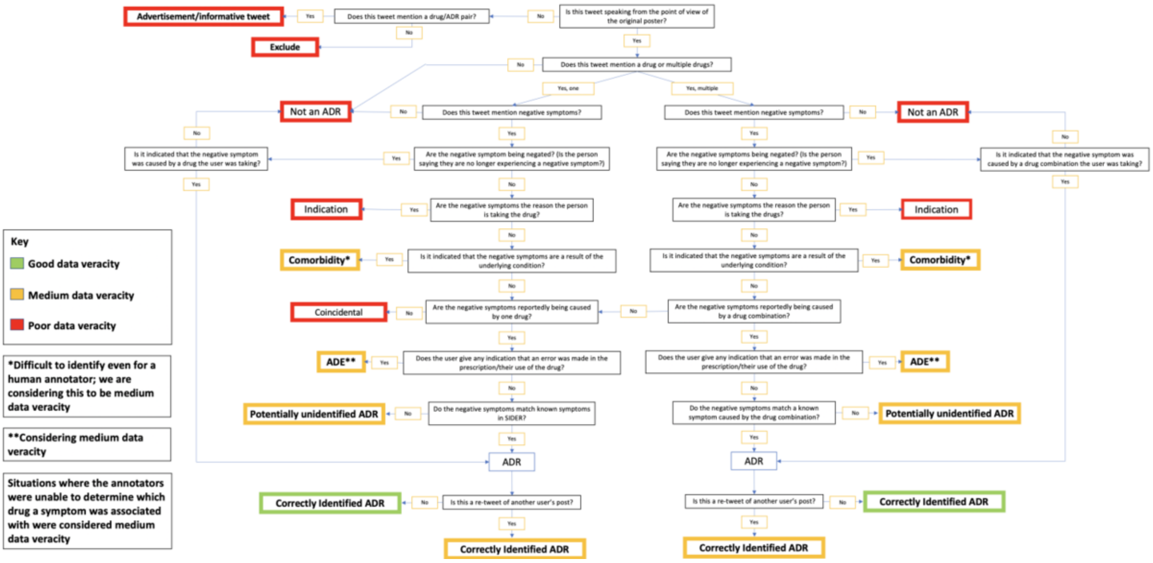

Figure 1 – Annotation Flow Chart.

Results

Among the 451 tweets, 36 (7.98%), 196 (43.46%), and 219 (48.56%) tweets are in poor, moderate, and good data veracity, correspondingly. Compared with tweets of good and moderate data veracity, poor data veracity tweets were less likely to use the first-person pronoun (79.91%, 77.55%, and 41.67%, respectively), more likely to use the second-person pronoun (6.85%, 15.31%, and 16.67%), less likely to present drug and ADR in the same sentence (71.69%, 81.63%, and 44.44%), and less likely to use medical terminology for ADRs (34.70%, 14.80%, and 13.89%). Poor data veracity tweets contain more indications. Specifically, the percentage of tweets in poor data veracity that contain one or more indications was 19.44%, compared with tweets in moderate data veracity (6.12%) and good data veracity (4.57%). The percentage of tweets in poor data veracity that have no drug name mentioned was 16.67%, much higher than moderate (3.96%) and good data veracity (0.00%). Besides, tweets in poor data veracity were more likely to include more than one drug name than tweets in moderate and good data veracity (36.11%, 17.33%, and 13.24%). Equally, the percentage of tweets in poor data veracity containing only one drug name is the lowest across the three data veracity levels (47.22%, 78.71%, and 86.30%) (Table 1).

Table 1 - Tweets Characteristics by Data Veracity Levels

	Data Veracity Level			
	Poor	**Moderate**	**Good**	**Total**
1st -person pronoun				
Yes	15 (41.67%)	152 (77.55%)	175 (79.91%)	342
No	21 (58.33%)	44 (22.45%)	44 (20.09%)	109
2nd -person pronoun				
Yes	6 (16.67%)	30 (15.31%)	15 (6.85%)	51
No	30 (83.33%)	166 (84.69%)	204 (93.15%)	400
Negative sentiment				
Yes	7 (19.44%)	114 (58.16%)	132 (60.27%)	253
No	29 (80.56%)	82 (41.84%)	87 (39.73%)	198
Drug and ADR in the same sentence				
Yes	16 (44.44%)	160 (81.63%)	157 (71.69%)	333
No	20 (55.56%)	36 (18.37%)	62 (28.31%)	118
Medical terminology				
Yes	5 (13.89%)	29 (14.80%)	76 (34.70%)	110
No	31 (86.11%)	167 (85.20%)	143 (65.30%)	341
N. of drugs				
0	6 (16.67%)	8 (3.96%)	0 (0.00%)	14
1	17 (47.22%)	159 (78.71%)	189 (86.30%)	365
2	12 (33.33%)	27 (13.37%)	27 (12.33%)	66

3	1 (2.78%)	8 (3.96%)	2 (0.91%)	11
4	0 (0.00%)	0 (0.00%)	1 (0.46%)	1
N. of indications				
0	29 (80.56%)	184 (93.88%)	209 (95.43%)	422
1	7 (19.44%)	9 (4.59%)	9 (4.11%)	25
2	0 (0.00%)	2 (1.02%)	1 (0.46%)	3
3	0 (0.00%)	1 (0.51%)	0 (0.00%)	1

Table 2 - Multinomial Logistic Regression for ADR Data Veracity Levels by Tweets Characteristics

		Maximum Likelihood Estimates				**Odds Ratio Estimates**	
		Data Veracity Level				**Data Veracity Level**	
		Moderate		**Good**		**Moderate**	**Good**
	Poor (ref)	**Coeff. (SE)**	**p**	**Coeff. (SE)**	**p**		
1st -person pronoun							
Yes	-	1.76 (0.44)	0.000	2.07 (0.44)	0.000	5.80 (2.47, 13.63)	7.89 (3.32, 18.76)
No (ref)	-	-	-	-	-	-	-
2nd person pronoun							
Yes	-	0.21 (0.56)	0.709	-0.62 (0.60)	0.298	1.23 (0.41, 3.71)	0.54 (0.17, 1.73)
No (ref)	-	-	-	-	-	-	-
Negative sentiment							
Yes	-	1.72 (0.47)	0.000	1.85 (0.47)	0.000	5.59 (2.21, 14.13)	6.33 (2.50, 16.01)
No (ref)	-	-	-	-	-	-	-
Drug and ADR in the same sentence							
Yes	-	1.83 (0.43)	0.000	1.33 (0.43)	0.002	6.20 (2.67, 14.42)	3.77 (1.64, 8.70)
No (ref)	-	-	-	-	-	-	-
Medical terminology							
Yes	-	0.19 (0.56)	0.738	1.37 (0.55)	0.012	1.21 (0.40, 3.61)	3.93 (1.35, 11.46)
No (ref)	-	-	-	-	-	-	-
N. of drugs		-0.41 (0.35)	0.240	-0.69 (0.36)	0.053	0.66 (0.33, 1.32)	0.50 (0.25, 1.01)
N. of indications		-0.67 (0.48)	0.166	-1.08 (0.55)	0.047	0.51 (0.20, 1.32)	0.34 (0.12, 0.99)

Discussion

Accurate identification and categorization of key linguistic features in Twitter posts are critical for using patients- and health consumers-reported ADRs for PV research. We found several clues of improving data veracity during the progress of annotation. First, drug names or keywords, when they are out of context, become causes of poor data veracity. For example, "lozenge" was identified in the dataset as a drug, but it was never specified what active ingredients the lozenge contained. When "Nicotine" was identified as a drug with ADRs in tweets, the tweets were most discussing over smoking cessation and not the use of nicotine in a medically prescribed way. Tweets often contain mentions of recreational drugs (e.g., heroin), but such drugs and linked adverse reactions are out of the context of ADR. Second, linguistic features could be mistakenly categorized since the information contained may be incomplete and limited due to the low text limits. For example, we found instances of a comorbidity being identified as an ADR. Such cases would be more of a challenge since even medical experts might not be able to distinguish between comorbidity and ADR based on limited information. Another situation is that vague descriptions of ADRs (e.g., feeling sick, aching, etc.) cannot be mapped onto drug-ADR pairs in the SIDER database. The causes of insufficient information vary and could be better understood if we could tease out the medium level of data veracity further because tweets with insufficient information tend to be categorized into medium data veracity. Third, idiomatic language and metaphors can also impede the correct recognition of linguistic features, leading to the difficulty in identifying either the mention of ADRs or specific ADRs. For example, "it feels like my brain is melting …" can be a metaphor for drowsiness, dizziness, sluggishness, and other possibilities. To precisely decipher the information in these tweets, analyzing longitudinal patterns and trends of tweets could be of great help. Fourth, NLP-based sentiment analysis can reveal the attitudes and feelings of the users who composed the Twitter posts. However, it lacks accuracy when modeling some special situations. For example, sentiment analysis may generate an opposite result of the true feeling when consumers use sarcasm in their messages, which also is a cause of low data veracity. Future NLP techniques should be focused on the pragmatics of the sentence instead of semantics solely. Specifically, studies should pay specific attention to the logical conjunctions and take into account the temporal patterns and trends of semantics across tweets posted by the same users and generated at different times.

Table 3. Summary of Key Linguistic Features

Features	Previous Studies	Implications
1st person pronoun	Bian et al, 2012 Ginn et al, 2014 Lim et al, 2017 Hoang et al, 2018	The first-person pronoun helps determine if the tweet regards "personal experiences".
Negative sentiment	Bian et al, 2012 Ginn et al, 2014 Lim et al, 2017 Hoang et al, 2018 Nikfarjam et al, 2015	Consumers tend to report ADRs with negative sentiment.
Drug and ADR in the same sentence *	N/A.	The drug and the ADR are less likely to be a valid "pair" if they are not mentioned in the same sentence.
Medical terminology*	N/A.	Consumers who use medical terminologies tend to provide more reliable information.
N. of drugs *	N/A.	Increased number of different drugs makes it more difficult to identify valid drug-ADR pairs.
N. of indications *	N/A.	Increased number of different indications make it more difficult to distinguish ADRs and indications.

* No previous study examined.

These identified linguistic features, Twitter variables, and association rules are key to identifying different data veracity, yet screening Twitter data in the real world is costly considering that Twitter data consist of more than 50 variables including the Twitter posts in free text and the high volume of dataset needed for PV research. To bridge this gap, our findings suggest great potentials for developing an efficient ML model to automatically detect ADR tweets at different levels of data veracity. This model can greatly reduce human labor when used for Twitter data selection. Our findings shed light on a couple of key designing components for developing ML models. First, the numbers of annotated drugs and indications could be used in the ML models as our statistical results showed that they are strong predictors for good data veracity tweets. Second, tweets with the first-person pronoun, the negative sentiment of ADR sentence, drug name and ADR name in the same sentence, mentions of the drug name, and no mention of the indication tend to be of good data veracity. Thus, these linguistic features should be included in the ML models as well. Third, annotation of words and phrases using terminologies tailored for layperson' language (e.g., Consumer Health Vocabulary) could be a strict filter for tweets with low data veracity. Precise distinguishing between indications and ADRs is also important for ML models but is not fully addressed in this study.

While we demonstrated a pathway to improving data veracity of ADRs, our study is subject to limitations. The dataset we used is not longitudinal. We were unable to assess whether the user-deleted ADR tweets would have any systemic impact on data veracity. Follow-up studies should employ multi-source datasets for cross-validation. Despite limitations, our study is among the first that incorporated the rule-based logic flow generated by medical experts in defining and identifying data veracity levels of ADR tweets, which has the potential of being generalized for other consumer-reported data.

Conclusions

Intricate linguistic features of Twitter posts and Twitter variables, when incorported with clinical domain knowledge, can be used to examine diverse data veracity of ADR-related tweets. Key linguistic features were found to be associated with specific data veracity levels including two reported features (1st person pronouns, negative sentiment) and four newly found features (drug and ADR in the same sentence, medical terminology, number of drugs, number of indications). These findings suggest that veracity of Twitter data could be further improved if researchers consider using these influential linguistic features to screen ADR-related tweets before downstream analysis in PV research.

Acknowledgements

This study is supported by a seed grant by the Arnold School of Public Health, University of South Carolina.

References

[1] M. Abou Taam, C. Rossard, L. Cantaloube, N. Bouscaren, G. Roche, L. Pochard, F. Montastruc, A. Herxheimer, J.L. Montastruc, and H. Bagheri, Analysis of patients' narratives posted on social media websites on benfluorex's (Mediator®) withdrawal in France, **39** (2014), 53-55.

[2] A.J. Avery, C. Anderson, C. Bond, H. Fortnum, A. Gifford, P.C. Hannaford, L. Hazell, J. Krska, A. Lee, and D.J. Mclernon, Evaluation of patient reporting of adverse drug reactions to the UK ‘Yellow Card Scheme’: literature review, descriptive and qualitative analyses, and questionnaire surveys, (2011).

[3] A. Bate, R.F. Reynolds, and P. Caubel, The hope, hype and reality of Big Data for pharmacovigilance, **9** (2018), 5-11.

[4] A. Benton, L. Ungar, S. Hill, S. Hennessy, J. Mao, A. Chung, C.E. Leonard, and J.H. Holmes, Identifying potential adverse effects using the web: A new approach to medical hypothesis generation, *Journal of Biomedical Informatics* **44** (2011), 989-996.

[5] J. Bian, U. Topaloglu, and F. Yu, Towards Large-scale Twitter Mining for Drug-related Adverse Events, *Shb12 (2012)* **2012** (2012), 25-32.

[6] C.C. Freifeld, J.S. Brownstein, C.M. Menone, W. Bao, R. Filice, T. Kass-Hout, and N. Dasgupta, Digital drug safety surveillance: monitoring pharmaceutical products in twitter, *Drug Safety* **37** (2014), 343-350.

[7] R. Ginn, P. Pimpalkhute, A. Nikfarjam, A. Patki, K. O’Connor, A. Sarker, K. Smith, and G. Gonzalez, Mining Twitter for adverse drug reaction mentions: a corpus and classification benchmark, in: *Proceedings of the fourth workshop on building and evaluating resources for health and biomedical text processing*, Citeseer, 2014, pp. 1-8.

[8] S. Golder, G. Norman, and Y.K. Loke, Systematic review on the prevalence, frequency and comparative value of adverse events data in social media, **80** (2015), 878-888.

[9] T. Hoang, J. Liu, N. Pratt, V.W. Zheng, K.C. Chang, E. Roughead, and J. Li, Detecting signals of detrimental prescribing cascades from social media, *Artificial Intelligence in Medicine* **71** (2016), 43-56.

[10] T. Hoang, J. Liu, N. Pratt, V.W. Zheng, K.C. Chang, E. Roughead, and J. Li, Authenticity and credibility aware detection of adverse drug events from social media, *International Journal of Medical Informatics* **120** (2018), 157-171.

[11] J. Kepner, V. Gadepally, P. Michaleas, N. Schear, M. Varia, A. Yerukhimovich, and R.K. Cunningham, Computing on masked data: a high performance method for improving big data veracity, in: *2014 IEEE High Performance Extreme Computing Conference (HPEC)*, 2014, pp. 1-6.

[12] L.T. Kohn, J.M. Corrigan, and M.S. Donaldson, *To Err Is Human: Building a Safer Health System*, The National Academies Press, Washington, DC, 2000.

[13] M. Kuhn, I. Letunic, L.J. Jensen, and P. Bork, The SIDER database of drugs and side effects, *Nucleic acids research* **44** (2016), D1075-D1079.

[14] J. Lazarou, B.H. Pomeranz, and P.N. Corey, Incidence of Adverse Drug Reactions in Hospitalized PatientsA Meta-analysis of Prospective Studies, *JAMA* **279** (1998), 1200-1205.

[15] S. Lim, C.S. Tucker, and S. Kumara, An unsupervised machine learning model for discovering latent infectious diseases using social media data, *Journal of Biomedical Informatics* **66** (2017), 82-94.

[16] T. Mjörndal, M.D. Boman, S. Hägg, M. Bäckström, B.-E. Wiholm, A. Wahlin, and R. Dahlqvist, Adverse drug reactions as a cause for admissions to a department of internal medicine, **11** (2002), 65-72.

[17] V.H. Nguyen, K. Sugiyama, M.-Y. Kan, and K. Halder, Treatment Side Effect Prediction from Online User-generated Content, in: *Proceedings of the Ninth International Workshop on Health Text Mining and Information Analysis*, 2018, pp. 12-21.

[18] A. Nikfarjam, A. Sarker, K. O'Connor, R. Ginn, and G. Gonzalez, Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features, *J Am Med Inform Assoc* **22** (2015), 671-681.

[19] K. O'Connor, P. Pimpalkhute, A. Nikfarjam, R. Ginn, K.L. Smith, and G. Gonzalez, Pharmacovigilance on twitter? Mining tweets for adverse drug reactions, *AMIA ... Annual Symposium proceedings. AMIA Symposium* **2014** (2014), 924-933.

[20] C.E. Pierce, K. Bouri, C. Pamer, S. Proestel, H.W. Rodriguez, H. Van Le, C.C. Freifeld, J.S. Brownstein, M. Walderhaug, and I.R.J.D.s. Edwards, Evaluation of Facebook and Twitter monitoring to detect safety signals for medical products: an analysis of recent FDA safety alerts, **40** (2017), 317-331.

[21] A. Sarker, R. Ginn, A. Nikfarjam, K. O’Connor, K. Smith, S. Jayaraman, T. Upadhaya, and G. Gonzalez, Utilizing social media data for pharmacovigilance: A review, *Journal of Biomedical Informatics* **54** (2015), 202-212.

[22] G. Shepherd, P. Mohorn, K. Yacoub, and D.W. May, Adverse Drug Reaction Deaths Reported in United States Vital Statistics, 1999-2006, **46** (2012), 169-175.

[23] J. Sultana, P. Cutroneo, and G. Trifirò, Clinical and economic burden of adverse drug reactions, *Journal of pharmacology & pharmacotherapeutics* **4** (2013), S73-S77.

[24] WHO, Training Courses - Definition, in, World Health Organization, World Health Organization, 1972.

[25] W.K. Wu and N. Pantaleo, Evaluation of outpatient adverse drug reactions leading to hospitalization, *American Journal of Health-System Pharmacy* **60** (2003), 253-259.

[26] C.C. Yang, H. Yang, and L. Jiang, Postmarketing Drug Safety Surveillance Using Publicly Available Health-Consumer-Contributed Content in Social Media, **5** (2014), Article 2.

Address for correspondence

Chen Liang, Ph.D., cliang@mailbox.sc.edu