BMJ Open

Utility of social media and crowd-sourced data for pharmacovigilance: a scoping review protocol

Andrea C Tricco,[1,2] Wasifa Zarin,[1] Erin Lillie,[1] Ba Pham,[1] Sharon E Straus[1,3]

[1]Li Ka Shing Knowledge Institute of St. Michael's Hospital, Toronto, Ontario, Canada
[2]Epidemiology Division, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada
[3]Faculty of Medicine, Department of Geriatric Medicine, University of Toronto, Toronto, Ontario, Canada

Correspondence to
Dr Andrea C Tricco;
TriccoA@smh.ca

ABSTRACT

Introduction: Adverse events associated with medications are under-reported in postmarketing surveillance systems. A systematic review of published data from 37 studies worldwide (including Canada) found the median under-reporting rate of adverse events to be 94% in spontaneous reporting systems. This scoping review aims to assess the utility of social media and crowd-sourced data to detect and monitor adverse events related to health products including pharmaceuticals, medical devices, biologics and natural health products.

Methods and analysis: Our review conduct will follow the Joanna Briggs Institute scoping review methods manual. Literature searches were conducted in MEDLINE, EMBASE and the Cochrane Library from inception to 13 May 2016. Additional sources included searches of study registries, conference abstracts, dissertations, as well as websites of international regulatory authorities (eg, Food and Drug Administration (FDA), the WHO, European Medicines Agency). Search results will be supplemented by scanning the references of relevant reviews. We will include all publication types including published articles, editorials, websites and book sections that describe use of social media and crowd-sourced data for surveillance of adverse events associated with health products. Two reviewers will perform study selection and data abstraction independently, and discrepancies will be resolved through discussion. Data analysis will involve quantitative (eg, frequencies) and qualitative (eg, content analysis) methods.

Dissemination: The summary of results will be sent to Health Canada, who commissioned the review, and other relevant policymakers involved with the Drug Safety and Effectiveness Network. We will compile and circulate a 1-page policy brief and host a 1-day stakeholder meeting to discuss the implications, key messages and finalise the knowledge translation strategy. Findings from this review will ultimately inform the design and development of a data analytics platform for social media and crowd-sourced data for pharmacovigilance in Canada and internationally.

Registration details: Our protocol was registered prospectively with the Open Science Framework (https://osf.io/kv9hu/).

Strengths and limitations of this study

- We will conduct a comprehensive literature search of multiple electronic databases and sources for difficult to locate and unpublished studies (or grey literature).
- Our scoping review will conform to the methodologically rigorous methods manual by the Joanna Briggs Institute.
- Numerous strategies will be used to disseminate our results widely.
- To increase the feasibility of our scoping review, we will limit to English and have one data abstractor and one verifier.

INTRODUCTION

Social media has gained unprecedented popularity worldwide. Currently, there are over 2.3 billion active social media users, and grows by an estimated 1 million new users every day.[1] Social media platforms such as Twitter, Tumblr and Facebook are increasingly being used to discuss and share health issues. Statistics Canada revealed that over 80% of Canadians were internet users as of 2009,[2] and almost 70% of these individuals were using the internet to search for medical or health-related information.[3] Social media and crowd-sourced data have been used to successfully extract information for surveillance of disease outbreaks,[4 5] health behaviour[6 7] and patient views on health issues.[8]

The use of social media to exchange and discuss health information by the general public generates a large volume of unsolicited and real-time information. Health-related social networks, such as DailyStrength and MedHelp, attract users daily to discuss their health-related experiences, including use of prescription drugs, health products, side effects and treatments. During the 2004–2005 influenza season, social media listening by

means of a Google 'click ad', which appeared on the search page when information seekers typed influenza-specific key words into the Google search engine, closely approximated the incidence of influenza cases.[9] It was revealed that the Google ad click rate correlated more closely with retrospectively confirmed cases of influenza than the Physicians Sentinel Surveillance system for 'influenza-like illness'.[9] Other researchers have also examined the use of social media for influenza outbreaks.[10–12] Similarly, during the Canadian listeriosis outbreak, online search trends related to listeriosis correlated closely with laboratory-confirmed cases determined retrospectively, and preceded official announcements of an epidemic.[13]

Recently, researchers evaluated the types of information[14] including the prevalence of misinformation[15] posted on Twitter and the Sina Weibo Chinese microblog platform related to the 2014–2015 Ebola epidemic. Given the observed predictive power of social media and crowd-sourced data as an information source for public health surveillance, a lot of interest has been generated about its use for surveillance of adverse events to health products, often referred to as pharmacovigilance.

Pharmacovigilance is defined as 'the science and activities relating to the detection, assessment, understanding and prevention of adverse effects or any other drug-related problem'.[16] It includes drug safety surveillance activities to monitor incidents of adverse effects in real-life conditions. Adverse events, in particular to drug use, are a significant cause of morbidity and mortality, and are the fourth most common cause of death in hospitalised patients.[17] Since many adverse events are not captured in randomised clinical trials, postmarketing surveillance of health and drug products is of paramount importance for drug and health technology industries and regulatory authorities, such as Health Canada, the US Food and Drug Administration (FDA) and European Medicines Agency (EMA). These governmental agencies require clinicians to report all suspected adverse events, but the voluntary nature of the reporting systems most likely contributes to the underreporting of adverse events.[18–20] A systematic review of published data from 37 studies worldwide (including Canada) found the median under-reporting rate of adverse events to be 94% in spontaneous reporting systems.[21] In response to the limitations in the current postmarketing surveillance systems, attention is being directed towards using social media and crowd-sourced data to detect adverse events and to improve consumer safety. Reviews have been conducted assessing social media for pharmacovigilance, such as a systematic review including 51 studies[22] and a scoping review including 24 studies,[23] but this is a rapidly evolving field and an updated scoping review with a comprehensive grey literature search may provide more clarity to the field. In addition, these previous reviews did not summarise pre-existing platforms that exist on this topic, which was requested by our knowledge user, Health Canada.

As such, we aim to assess the utility of social media and crowd-sourced data to monitor and detect adverse events related to health products. For the purpose of this review, health products include pharmaceuticals and drug products, medical devices, biologics, and natural health products. The specific research questions are:

1. What social listening and analytics platforms exist internationally to detect adverse events related to health products using social media and crowd-sourced data? What are their capabilities and characteristics?
2. What is the validity and reliability of user-generated data from social media for surveillance of adverse events to health products?

METHODS
Study design
Our research objectives will be addressed using the scoping review methodology, which is a type of knowledge synthesis approach used to map the concepts underpinning a research area and the main sources and types of evidence available.[24] This scoping review will be conducted in accordance with standard practices used by the Knowledge Synthesis Team within the Knowledge Translation Program of St Michael's Hospital.[25] Our approach will be informed by the methodological framework proposed by Arksey and O'Malley,[24] as well as the methodology manual published by the Joanna Briggs Institute for scoping reviews.[26] This review has been commissioned by the Health Products and Food Branch (HPFB) of Health Canada and funded by the Canadian Institutes of Health Research Drug Safety and Effectiveness Network with a 6-month timeline.

Protocol
Our protocol was drafted using the Preferred Reporting Items for Systematic Reviews and Meta-analysis Protocols (PRISMA-P; see online supplementary appendix A),[27] which was revised by the research team and members of Health Canada, and was disseminated through our programme's Twitter account (@KTCanada) and newsletter to solicit additional feedback. The final protocol was registered prospectively with the Open Science Framework on 6 September 2016 (https://osf.io/kv9hu/).

Eligibility criteria
The PICOS (Population, Intervention, Comparator, Outcome, Study design)[28] eligibility criteria are as follows:

Population
Patients of any age with an adverse event related to health products including pharmaceuticals and drug products, biologics, medical devices, and natural health products.[29] Examples of pharmaceuticals and drug products include both prescription and non-prescription (over-the-counter) medicines, disinfectants and

sanitisers with disinfectant claims. Biologics can include, but are not limited to: vaccines, insulin, serums, blood-derived products, hormones, growth factors and enzymes manufactured in bacterial, yeast or mammalian cell lines; and gene therapy and cell therapy products. Medical devices can include defibrillators, syringes, surgical lasers, hip implants, medical laboratory diagnostic instruments (including X-ray, ultrasound devices), contact lenses and condoms. Natural health products can include vitamins and minerals, herbal remedies, homoeopathic and traditional medicines, probiotics, and other products like amino acids and essential fatty acids. Adverse events, such as addiction and overdose from prescription medical products, are also eligible for inclusion. Adverse events related to programmes of care, health services, organisation of care, public health programmes, health promotion programmes and health education programmes will be excluded.

Intervention
Any data analytics or social listening platforms that enable the extraction of user-generated and crowd-sourced data about adverse events to health products from social media are eligible for inclusion. Social media technology is defined as a web-based application that allows for the creation and exchange of user-generated content. This includes, but is not limited to: websites, web pages, blogs, vlogs, social networks, internet forums, chat rooms, wikis and smartphone applications, where users have the ability to generate content (typically by providing posts and comments, often in an anonymous fashion or with limited identifying information) and are able to view/exchange content from and with others in an interactive digital environment.[30] Crowd sourcing is the practice of obtaining needed services, ideas or content by soliciting contributions from a large group of people and especially from the online community rather than from traditional employees or suppliers.[31] Social media listening and data analytics for public health surveillance related to non-communicable (eg, disease prevalence) and communicable diseases (eg, outbreak investigation) will be excluded.

Comparators
Any comparator is relevant for inclusion (eg, studies comparing one form of social media or crowd-sourced data to another or comparing social media with traditional reporting systems). In addition, studies without a comparator are eligible for inclusion.

Outcomes
There are two broad categories of outcomes that are of interest: (1) characteristics of social media listening and analytics platform (eg, data sources, scope of surveillance, capabilities, data extraction, preprocessing data, annotation, text mining methods, computational frameworks, added value to existing surveillance capacities, technical skills required, infrastructure support to implement and sustain, privacy and security of the data); and (2) validity and reliability of user-generated data captured through social media and crowd-sourcing networks (eg, relationship between medications and adverse events, algorithms or processes used to validate the data from social media, and related results of the evaluation).

Study designs
All types of publications including published articles, articles in conference proceedings, editorials, websites and chapters in textbooks are relevant.

Time periods
All periods of time and duration of follow-up are eligible.

Other
Given the 6-month timeline, only publications written in English will be considered for inclusion. If time allows, publications in other languages may be considered.

Information sources and search strategy
Comprehensive literature search strategies were developed by an experienced librarian for the following electronic bibliographic databases: MEDLINE, EMBASE and the Cochrane Library. The search strategy was peer-reviewed by another expert librarian using the PRESS (Peer Review of Electronic Search Strategies) checklist.[32] The final search strategy incorporated feedback from the peer review process and the complete search string for MEDLINE can be found in online supplementary appendix B. The full search terms for the other databases can be obtained by contacting the corresponding author. A trained library technician performed the final searches from inception to May 2016, exported the search results into Endnote and removed all duplicates.

A grey literature search was conducted according to the Canadian Agency for Drugs and Technologies in Health (CADTH) guide.[33] Specifically, we searched 59 sources and websites of 119 relevant regulatory authorities for additional publications or pre-existing platforms of social media listening and data analytics. Examples of such social media listening and analytics platforms include the MedWatcher Social created in collaboration with the US FDA and Web-RADR (Recognising Adverse Drug Reactions) for the European Union regulators.[34 35] See online supplementary appendix C for a full list of grey literature sources that were searched. Literature saturation will be ensured by searching the reference lists of relevant reviews.[22 23 36]

Study selection process
To ensure high inter-rater reliability, a training exercise will be conducted prior to starting the screening process. Using our predefined eligibility criteria, a standardised questionnaire for study selection will be developed and tested on a random sample of 50 titles and

abstracts (ie, level 1 screening) by all team members. The same training exercise will be repeated for screening of full-text articles (ie, level 2 screening). Subsequently, pairs of reviewers will screen citations and full-text articles for inclusion, independently, for level 1 and 2 screening. Inter-rater discrepancies will be resolved by discussion or a third adjudicator. All levels of screening will be conducted using Synthesi.SR, the proprietary online software developed by the Knowledge Synthesis Team.[37]

Data items and data abstraction process

We will abstract data on characteristics of the articles (eg, type of article or study, country of corresponding author), population characteristics (eg, type of patients, type of adverse events, disease condition), intervention characteristics (eg, type of social media or crowd-sourced data used) and outcomes (eg, data analytics/listening platform characteristics, data analytics used, validity and reliability of social media or crowd-sourced data). A standardised data abstraction form will be developed a priori and revised, as needed, after the completion of a training exercise.

Prior to data abstraction, we will complete a training exercise of the data abstraction form on a random sample of five articles. Subsequently, all included studies will be abstracted by pairs of reviewers, independently, with conflicts resolved by a third reviewer. If a large number of studies is identified (>25), we will conduct data abstraction with one reviewer and one verifier.

Risk of bias assessment or quality appraisal

Since this is a scoping review aiming to map all available evidence, we will not conduct any risk of bias assessment or quality appraisal of included studies. This approach is consistent with the methods manual published by the Joanna Briggs Institute,[26] as well as a database of scoping reviews on health-related topics.[38]

Synthesis of results

The synthesis will focus on providing a description of all social media listening platforms that exist internationally, and the validity and reliability of data from these social listening platforms, when available. This will be achieved by summarising the literature according to the types of participants, interventions, comparators and outcomes identified. Quantitative analysis will be conducted using descriptive statistics (eg, frequencies, measures of central tendency). In addition, we will consider qualitative analysis (eg, content analysis) for open-text data, as necessary. Two reviewers will conduct the initial categorisation coding independently, using NVivo software (NVivo V.10. Australia: International QSR, 2012), and the results will be discussed by the team. These reviewers will subsequently identify, code and chart relevant units of text from the articles using the categorisation code. Discrepancies will be resolved through team discussion.

DISCUSSION
Implications

Findings from this scoping review will inform decision-makers of the types of social listening and analytics platforms that exist to extract user-generated data from social media for surveillance of adverse events to health products. This will inform Health Canada and other regulatory authorities internationally about the potential use of social media and crowd-sourced data for postmarketing surveillance.

Dissemination

The summary of results will be sent to Health Canada and other relevant policymakers and researchers working with the Drug Safety and Effectiveness Network in the form of a one-page policy brief.[39] In addition, a 1-day stakeholder meeting (ie, consultation exercise)[24] will be held to discuss the implications of our scoping review, key messages and to finalise the knowledge translation strategy. All relevant stakeholders will be invited to attend, as recommended by members from the Health Canada HPFB. This meeting will be essential to ensure extensive knowledge translation of our findings and to engage stakeholders and promote our research agenda. We will also present our results at an international conference and publish in an open-access journal. Finally, team members will use their networks to encourage broad dissemination of results.

REFERENCES
1. Smith K. *Marketing: 96 amazing social media statistics and facts for 2016*: Brandwatch Blog, 2016 [updated 7 Mar 2016]. https://www.brandwatch.com/2016/03/96-amazing-social-media-statistics-and-facts-for-2016/
2. Statistics Canada. Internet use by individuals, by selected characteristics 2010 [updated 10 May 2010]. http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/comm35a-eng.htm

3. Statistics Canada. Internet use by individuals, by type of activity (Internet users at home) 2010 [updated 10 May 2010. http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/comm29a-eng.htm

4. Twitter catches the flu: detecting influenza epidemics using Twitter. *Proceedings of the conference on empirical methods in natural language processing. Association for Computational Linguistics.* 2011.

5. Broniatowski DA, Paul MJ, Dredze M. National and local influenza surveillance through Twitter: an analysis of the 2012–2013 influenza epidemic. *PLoS ONE* 2013;8:e83672.

6. Jashinsky J, Burton SH, Hanson CL, *et al.* Tracking suicide risk factors through Twitter in the US. *Crisis* 2014;35:51–9.

7. Scanfeld D, Scanfeld V, Larson EL. Dissemination of health information through social networks: Twitter and antibiotics. *Am J Infect Control* 2010;38:182–8.

8. Nakhasi A, Passarella R, Bell SG, Paul MJ, Dredze M, Pronovost P, editors. Malpractice and malcontent: Analyzing medical complaints in twitter. Association for the Advancement of Artificial Intelligence Fall Symposium Series; 2012.

9. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. *J Med Internet Res* 2009;11:e11.

10. Paul MJ, Dredze M, Broniatowski D. Twitter improves influenza forecasting. *PLoS Curr* 2014;6. doi:10.1371/currents.outbreaks.90b9ed0f59bae4ccaa683a39865d9117

11. Santillana M, Nguyen AT, Dredze M, *et al.* Combining search, social media, and traditional data sources to improve influenza surveillance. *PLoS Comput Biol* 2015;11:e1004513.

12. Sharpe JD, Hopkins RS, Cook RL, *et al.* Evaluating Google, Twitter, and Wikipedia as tools for influenza surveillance using Bayesian change point analysis: a comparative analysis. *JMIR Public Health Surveill* 2016;2:e161.

13. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. *PLoS ONE* 2010;5:e14118.

14. Fung IC, Hao Y, Cai J, *et al.* Chinese social media reaction to information about 42 notifiable infectious diseases. *PLoS ONE* 2015;10:e0126092.

15. Fung IC, Fu KW, Chan CH, *et al.* Social media's initial reaction to information and misinformation on Ebola, August 2014: facts and rumors. *Public Health Rep* 2016;131:461–73.

16. World Health Organization. *Essential medicines and health products: pharmacovigilance.* 2016. http://www.who.int/medicines/areas/quality_safety/safety_efficacy/pharmvigi/en/

17. Wester K, Jönsson AK, Spigset O, *et al.* Incidence of fatal adverse drug reactions: a population based study. *Br J Clin Pharmacol* 2008;65:573–9.

18. van Der Hooft CS, Sturkenboom MC, van Grootheest K, *et al.* Adverse drug reaction-related hospitalisations: a nationwide study in The Netherlands. *Drug Saf* 2006;29:161–8.

19. Bates DW, Evans RS, Murff H, *et al.* Detecting adverse events using information technology. *J Am Med Inform Assoc* 2003;10:115–28.

20. McClellan M. Drug safety reform at the FDA—pendulum swing or systematic improvement? *N Engl J Med* 2007;356:1700–2.

21. Mittmann N, Knowles SR, Gomez M, *et al.* Evaluation of the extent of under-reporting of serious adverse drug reactions: the case of toxic epidermal necrolysis. *Drug Saf* 2004;27:477–87.

22. 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;80:878–88.

23. Lardon J, Abdellaoui R, Bellet F, *et al.* Adverse drug reaction identification and extraction in social media: a scoping review. *J Med Internet Res* 2015;17:e171.

24. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol* 2005;8:19–32.

25. Knowledge Translation Program. Canada: Knowledge Translation Program; 2016. http://knowledgetranslation.net/

26. Peters M, Godfrey C, McInerney P, Soares C, Hanan K, Parker D. The Joanna Briggs Institute Reviewers' Manual 2015: Methodology for JBI Scoping Reviews. 2015. Available from: http://joannabriggs.org/assets/docs/sumari/Reviewers-Manual_Methodology-for-JBI-Scoping-Reviews_2015_v2.pdf.

27. Shamseer L, Moher D, Clarke M, *et al.* Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ* 2015;349:g7647.

28. Stone PW. Popping the (PICO) question in research and evidence-based practice. *Appl Nurs Res* 2002;15:197–8.

29. Health Canada. Drugs and Health Products. 2016 [updated 30 June 2016]. http://www.hc-sc.gc.ca/dhp-mps/index-eng.php

30. Naik P, Umrath T, van Stekelenborg J, *et al.* Regulatory definitions and good pharmacovigilance practices in social media: challenges and recommendations. *Ther Innov Regul Sci* 2015;49:840–51.

31. Merriam-Webster. Crowdsourcing. http://www.merriam-webster.com/dictionary/crowdsourcing

32. McGowan J, Sampson M, Salzwedel DM, *et al.* PRESS Peer Review of Electronic Search Strategies: 2015 guideline statement. *J Clin Epidemiol* 2016;75:40–6.

33. CADTH. Grey matters: a practical tool for searching health-related grey literature: Canadian Agency for Drugs and Technologies in Health (CADTH). 2015. https://www.cadth.ca/resources/finding-evidence/grey-matters

34. MedWatcher. MedWatcher for Enterprise 2016. https://medwatcher.org/about

35. WEB-RADR. WEB-RADR: Recognising Adverse Drug Reactions 2016. https://web-radr.eu/

36. Sinnenberg L, Buttenheim AM, Padrez K, *et al.* Twitter as a tool for health research: a systematic review. *Am J Public Health* 2017;107:e1–8.

37. Newton D. *Synthesi.SR.* Toronto, Canada: Knowledge Translation Program, St. Michael's Hospital, 2012.

38. Tricco AC, Lillie E, Zarin W, *et al.* A scoping review on the conduct and reporting of scoping reviews. *BMC Med Res Methodol* 2016;16:15.

39. Tricco AC, Soobiah C, Berliner S, Ho J, Ng C, Ashoor HS, SE. Efficacy and Safety of Cognitive Enhancers for Mild Cognitive Impairment: A Systematic Review and Meta-analysis; 2013. Available from: http://www.cihr-irsc.gc.ca/e/documents/dsen-abstract-en.pdf