BMC Medicine

OPINION

Open Access

CrossMark

Real-world data and the patient perspective: the PROmise of social media?

Laura McDonald[1], Bill Malcolm[2], Sreeram Ramagopalan[1*] and Hayley Syrad[3]

Abstract

Understanding the patient perspective is fundamental to delivering patient-centred care. In most healthcare systems, however, patient-reported outcomes are not regularly collected or recorded as part of routine clinical care, despite evidence that doing so can have tangible clinical benefit. In the absence of the routine collection of these data, research is beginning to turn to social media as a novel means to capture the patient voice. Publicly available social media data can now be analysed with relative ease, bypassing many logistical hurdles associated with traditional approaches and allowing for accelerated and cost-effective data collection. Existing work has shown these data can offer credible insight into the patient experience, although more work is needed to understand limitations with respect to patient representativeness and nuances of captured experience. Nevertheless, linking social media to electronic medical records offers a significant opportunity for patient views to be systematically collected for health services research and ultimately to improve patient care.

Keywords: Social media, Patient-reported outcomes, Epidemiology, Real-world data, Patient-centricity

Patient-centricity in real-world research: prioritising the patient perspective

Real-world data (RWD) are those data collected outside conventional randomised clinical trials to evaluate what is happening in routine clinical practice. These data are increasingly used to support regulatory decision-making and to guide clinical practice in real-world populations [1]. While the focus of evidence generation using RWD has traditionally been on clinical endpoints (safety and effectiveness outcomes), in order to provide a more holistic view of disease and well-being there is a need for RWD that capture the patient perspective.

A patient-centric paradigm shift has already occurred in the clinical trial domain, where patient-reported outcomes (PROs) are routinely integrated into trial design [2]. These data provide assessments of how a patient feels and functions at a given point; they are measured using standardised direct-to-patient questionnaires. Particularly in fields such as oncology, these data can be pivotal in helping to differentiate interventions in which clinical outcomes (such as survival) may appear comparable, and to provide

additional data on the impact of a treatment beyond that which can be obtained from traditional endpoints (e.g. an assessment of tolerability). Furthermore, it is well-documented that a patient's and clinician's view of disease and well-being can differ substantially [3, 4], so these data provide valuable insight into patient experiences that might not otherwise be reported to or recorded by treating physicians, but that may have a meaningful impact on clinical outcomes [5–7].

Outside clinical trials, PRO measures can be incorporated into sources of RWD that are designed for research purposes, such as patient registries. The integration of PROs into prospective data capture is, however, resource intensive and maintaining patient engagement can be challenging, particularly in groups of patients who are older, sicker and of lower socioeconomic status [8]. Distinct from structured PROs, unstructured patient-generated health data (PGHD) are those data captured or recorded spontaneously by patients or their carers [9]. These data can be collected from a variety of sources including patient-powered research networks and smart wearable devices, as well as social media. Leveraging PGHD to generate insight into patient-experienced outcomes in the real world offers an exciting area for research, and is gaining attention from scientists, industry

* Correspondence: sreeram.ramagopalan@bms.com
[1]Centre for Observational Research and Data Sciences, Bristol-Myers Squibb, Uxbridge, UK
Full list of author information is available at the end of the article

and regulators. Indeed, the US Food and Drug Administration (FDA) has recently encouraged the exploration of social media for this purpose [10]. The goal of this paper is to discuss the potential utility of social media as a unique source of PGHD to capture the patient perspective and patient-experienced outcomes in the real world.

Harnessing social media for real-world data

Social media platforms such as Facebook, Twitter and patient networks have created abundant opportunities for patients and their carers to create and exchange health-related information. Previous work has found that patients tend to use social media platforms to increase knowledge, for social support, to exchange advice and to improve self-care and doctor–patient communication [11–13]. This has in turn generated a potentially rich but analytically 'messy' source of RWD; the ability to harness these data for medical research has been assisted in recent years by the application of advanced analytics. Approaches such as natural language processing coupled with machine learning are now able to effectively deal with the many complexities of the data extracted from social media, including multiplicity of terms, duplicate posts, misspellings and abbreviations (among others) [14]. Furthermore, in place of manual coding, machine learning algorithms can be developed which accurately and automatically identify features of posted content, such as adverse events (AEs), enabling the analysis of hundreds of thousands of text-based posts [15, 16]. Data can also be easily extracted from publicly available sites, bypassing many logistical hurdles associated with traditional approaches and allowing for accelerated, real-time and cost-effective data collection.

Pharmacovigilance in particular has been an area of early development in the utilisation of social media data. This is because, outside of clinical trials, more than 95% of treatment-related AEs are estimated to remain undocumented by healthcare professionals [17]. Because social media is adopted by patients to seek advice and share experiences, it is thought these data may enable greater capture of AEs, augment real-time reporting and in turn enable expedited signal detection. Indeed, approximately 12–62% of all posts on patient forums have been found to include information related to an AE [18]. Initial work has explored the extent to which these data correspond with existing pharmacovigilance sources, and a recent systematic review found good concordance (between 57% and 99%) for AEs reported in social media [19]. Although concordance is generally good [20], where differences have been observed it has been found that social media data tend to include a higher frequency of AEs relating to milder, unpleasant or quality of life events, with severe events requiring clinical diagnosis

being underrepresented [17]. However, it is important to consider that rather than being a limitation with respect to the validity of the data, these differences may instead reflect nuances in data capture. Indeed, other work has shown that patient and clinical agreement tends to be higher for observable symptoms but poorer for subjectively experienced symptoms such as fatigue [21]. By integrating the patient perspective, PGHD from social media may offer additional dimensionality to the routine monitoring of drug safety, as well as more broadly capture symptoms or experiences relevant to patients that may otherwise remain under-recorded. Reflecting the potential importance of social media data for pharmacovigilance, the US FDA signed an agreement in 2015 with PatientsLikeMe (a patient network) to determine how patient-reported data from the platform could help to generate insight into drug safety [22].

Beyond pharmacovigilance, other studies have shown that social media data can be used meaningfully to understand patient experience with their disease or treatment more broadly. For example, a recent study extracted >10,000 data points from a variety of social media platforms and developed a machine learning algorithm to automatically identify mentions of treatment switching among patients with multiple sclerosis. The most common reasons for switching were then mapped and found to be comparable to those obtained from published data [23]. Sentiment analysis is another promising area for this type of social media analytics [24]. This approach involves assessing the ratio of positive to negative words contained in a post to ascribe positive, negative or neutral sentiment to opinion-based text. This approach has previously been applied to understand experience with systemic treatment options among patients with multiple sclerosis [25], attitudes towards vaccinations [26, 27], and to monitor mood among cancer patients online [28]. More traditional qualitative content analysis can also be applied to extracted text from social media, albeit on a smaller scale owing to the manual nature of these techniques. This approach has also been successfully applied, for example, to understand patient perception of care quality [29].

Potential limitations and challenges

Despite a number of potential applications, using social media to capture the patient perspective is not without challenges. Exploration of topics can often be limited; Twitter, for example, only allows individuals to write 280 characters. Many discussions also take place in private patient forums, largely inaccessible to researchers. Although analytical techniques to deal with complexities inherent in social media data continue to advance, it may often be the case that there is too much noise to generate meaningful insight.

McDonald *et al. BMC Medicine*   (2019) 17:11

Page 3 of 5

Beyond technical issues surrounding data capture, issues concerning the representativeness of the patient population are also important. Indeed, the demographics of individuals posting on social media are rarely known. Where it is possible to garner this information, data show active users tend to be younger, women, more highly educated and less acutely ill or functionally impaired [30, 31], presenting issues of external validity. Indeed, the 'digital divide' in Internet usage has been well documented; although recent reports suggest Internet usage in adults aged over 65 years has doubled in recent decades, older adults (>75 years) and those with functional impairment remain less likely to engage in health-related Internet usage [31]. It is also possible that, for older adults, younger carers or relatives may be engaging online on the patient's behalf. It is essential to quantify demographic disparities in order to apply analytical strategies that help mitigate biases in patient representativeness (e.g. stratified sampling). Identifying proxies for demographic information is one potential solution for this; recent work has used machine learning techniques to show that features extracted from patients' user names can be used to accurately infer patient demographics [32].

There may also be nuances in the data captured within social media. For studies that have attempted to validate data from social media with data obtained from traditional sources, further investigation is needed into the extent to which observed differences reflect issues in data quality (e.g. as a result of the limited representation of certain groups) as opposed to more general intricacies in the type of information patients may be more likely to share in online communities (e.g. quality of life events). Importantly, the current notion is not that social media should replace existing patient-reported data, but rather that the observed benefits of these data (rapid, cost-effective and large-scale access to real-world PGHD) should be harnessed to complement existing data sources. However, as the world continues to get more and more connected this needs to be continually assessed.

Privacy concerns also remain a fundamental challenge and a barrier to effectively harnessing social media data for public health. Even though text extraction takes place on content that is posted 'publicly', it can be contested whether or not it is correct to presume consent for the use of these data. Other publications have provided more detailed discussions regarding the ethical considerations associated with using these data [33, 34]. It should nonetheless be noted that privacy concerns are not unique to social media and are seen in other areas in which patient data are used for public health research or surveillance [35]. In these domains, effective communication and patient engagement are known to be key [36]. Indeed, studies have shown that the more patients know about how their data are used, the more accepting they are of data sharing [37, 38]. The same considerations will likely apply when targeting patient acceptability for social media. Encouragingly, early data show good acceptability, with 71% of patients recruited at emergency departments in the US identified as being willing to share their social media data for public health research [30].

Future perspective

As the science for extracting and analysing data from social media continues to advance, there are a number of interesting future applications which may further extend the utility of these data. For example, some initial work using machine learning algorithms has shown that it is possible to predict a diagnosis of depression recorded in a patient's medical record up to 6 months prior using only the language content of their Facebook posts [39]. Other studies have shown similar feasibility for detecting depression using only data from Twitter [40]. The implication is that these data could be used in the future to facilitate a scalable screening tool for detecting mental illness. Of course, there are ethical and regulatory logistical challenges that would need to be addressed in order to effectively implement such a programme. However, that data posted on social media could identify patients who may benefit from targeted interventions who remain undetected because they fail to present to or discuss symptoms with their clinician is an exciting area for development. Equally, as patients continue to use social media to seek information related to their health, data from social media could be used to develop patient-focused strategies or interventions aimed at better supporting patients' needs by providing targeted information and support.

Linking social media profiles to electronic medical records may also offer an opportunity to further extend the utility of these data in the future [13]. From an epidemiological perspective, this would allow background demographic and health information to be captured for these digital cohorts. In turn, it would allow analyses to be extended to include, for example, comparative effectiveness research. From a care perspective, these data could assist with improved patient-centred management. For example, AEs reported on social media could be communicated back to healthcare professionals. In doing do, these data could provide a means by which to encourage more open and sustained patient–clinician communication, a concept fundamental to patient-centred care [41]. There are of course challenges associated with data extraction and linkage that would need to be overcome before this next-generation of health-enabled social media can be realised, but this reflects an exciting area of future research.

Conclusion

As patients increasingly turn to social media as a means of seeking information or sharing experiences, these data offer a unique opportunity to capture patient-generated data in the real world. The feasibility of harnessing social media data has been assisted in recent years by the evolution of advanced analytics. The ability to generate insights from these data relevant to public health has already been demonstrated with some success. Although there are a number of exciting potential future applications of these data, privacy and governance considerations remain a fundamental concern for advancement of the field.

Abbreviations
AE: Adverse events; FDA: Food and Drug Administration; PGHD: Patient-generated health data; PRO: Patient-reported outcome; RWD: Real-world data

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Centre for Observational Research and Data Sciences, Bristol-Myers Squibb, Uxbridge, UK. [2]Bristol-Myers Squibb, Uxbridge, UK. [3]Evidera, London, UK.

References
1. Garrison LP, Neumann PJ, Erickson P, Marshall D, Mullins CD. Using real-world data for coverage and payment decisions: the ISPOR Real-World Data Task Force report. Value Health. 2007;10(5):326–35 https://doi.org/10.1111/j.1524-4733.2007.00186.x.
2. Kanapuru B, Singh H, Kim J, Kluetz PG. Patient-reported outcomes (PRO) in cancer trials submitted to the FDA from 2012-2015. J Clin Oncol. 2017; 35(Suppl 15):e14024 https://doi.org/10.1200/JCO.2017.35.15_suppl.e14024.
3. Pakhomov SV, Jacobsen SJ, Chute CG, Roger VL. Agreement between patient-reported symptoms and their documentation in the medical record. Am J Manag Care. 2008;14(8):530–9.
4. Valikodath NG, Newman-Casey PA, Lee PP, Musch DC, Niziol LM, Woodward MA. Agreement of ocular symptom reporting between patient-reported outcomes and medical records. JAMA Ophthalmol. 2017;135(3):225 https://doi.org/10.1001/jamaophthalmol.2016.5551.
5. Basch E, Deal A, Dueck A, Scher HI, Kris MG, Hudis C, et al. Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. JAMA. 2017;318(2):197–8 https://doi.org/10.1001/jama.2017.7156.
6. Stewart MA. Effective physician-patient communication and health outcomes: a review. CMAJ. 1995;152(9):1423–33.
7. Little P, Everitt H, Williamson I, et al. Observational study of effect of patient centredness and consultations. BMJ. 2001;323:908–11.
8. Hutchings A, Grosse Frie K, Neuburger J, van der Meulen J, Black N. Late response to patient-reported outcome questionnaires after surgery was associated with worse outcome. J Clin Epidemiol. 2013;66(2):218–25 https://doi.org/10.1016/j.jclinepi.2012.09.001.
9. Wood WA, Bennett AV, Basch E. Emerging uses of patient generated health data in clinical research. Mol Oncol. 2015;9(5):1018–24 https://doi.org/10.1016/j.molonc.2014.08.006.
10. United States Food and Drug Administration. Patient-focused drug development: collecting comprehensive and representative input guidance for industry, food and drug administration staff, and other stakeholders. 2018. https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm610442.pdf. Accessed 29 Oct 2018.
11. Antheunis ML, Tates K, Nieboer TE. Patients' and health professionals' use of social media in health care: motives, barriers and expectations. Patient Educ Couns. 2013;92(3):426–31 https://doi.org/10.1016/J.PEC.2013.06.020.
12. Greene JA, Choudhry NK, Kilabuk E, Shrank WH. Online social networking by patients with diabetes: a qualitative evaluation of communication with Facebook. J Gen Intern Med. 2011;26(3):287–92 https://doi.org/10.1007/s11606-010-1526-3.
13. De Simoni A, Shanks A, Balasooriya-Smeekens C, Mant J. Stroke survivors and their families receive information and support on an individual basis from an online forum: descriptive analysis of a population of 2348 patients and qualitative analysis of a sample of participants. BMJ Open. 2016;6(4): e010501 https://doi.org/10.1136/bmjopen-2015-010501.
14. Segura-Bedmar I, Martínez P. Pharmacovigilance through the development of text mining and natural language processing techniques. J Biomed Inform. 2015;58:288–91 https://doi.org/10.1016/J.JBI.2015.11.001.
15. Liu J, Zhao S, Zhang X. An ensemble method for extracting adverse drug events from social media. Artif Intell Med. 2016;70:62–76 https://doi.org/10.1016/J.ARTMED.2016.05.004.
16. 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 https://doi.org/10.1016/J.JBI.2015.10.011.
17. Hazell L, Shakir SAW. Under-reporting of adverse drug reactions: a systematic review. Drug Saf. 2006;29(5):385–96.
18. 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 https://doi.org/10.1111/bcp.12746.
19. Yang CC, Yang H, Jiang L. Postmarketing drug safety surveillance using publicly available health-consumer-contributed content in social media. ACM Trans Manag Inf Syst. 2014;5(1):1–21 https://doi.org/10.1145/2576233.
20. Topaz M, Lai K, Dhopeshwarkar N, et al. Clinicians' reports in electronic health records versus patients' concerns in social media: a pilot study of adverse drug reactions of aspirin and atorvastatin. Drug Saf. 2016; 39(3):241–50 https://doi.org/10.1007/s40264-015-0381-x.
21. Basch E, Iasonos A, McDonough T, et al. Patient versus clinician symptom reporting using the National Cancer Institute Common Terminology Criteria for Adverse Events: results of a questionnaire-based study. Lancet Oncol. 2006;7(11):903–9 https://doi.org/10.1016/S1470-2045(06)70910-X.
22. PatientsLikeMe and the FDA sign research collaboration agreement. https://blog.patientslikeme.com/openness/patientslikeme-and-the-fda-sign-research-collaboration-agreement/. Accessed 29 Oct 2018.
23. Risson V, Saini D, Bonzani I, Huisman A, Olson M. Patterns of treatment switching in multiple sclerosis therapies in US patients active on social media: application of social media content analysis to health outcomes research. J Med Internet Res. 2016;18(3):e62 https://doi.org/10.2196/jmir.5409.
24. Pang B, Lee L. Opinion mining and sentiment analysis. Found Trends in Inf Retr. 2008;2(1–2):1–135 http://www.cs.cornell.edu/home/llee/omsa/omsa.pdf. Accessed 29 Oct 2018.

25. Ramagopalan S, Wasiak R, Cox AP. Using Twitter to investigate opinions about multiple sclerosis treatments: a descriptive, exploratory study. F1000Res. 2014;216:1–9 https://doi.org/10.12688/f1000research.5263.1.

26. Salathé M, Khandelwal S. Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. PLoS Comput Biol. 2011;7(10):e1002199 https://doi.org/10.1371/journal.pcbi.1002199.

27. Dunn AG, Leask J, Zhou X, Mandl KD, Coiera E. Associations Between exposure to and expression of negative opinions about human papillomavirus vaccines on social media: an observational study. J Med Internet Res. 2015;17(6):e144 https://doi.org/10.2196/jmir.4343.

28. Rodrigues RG, das Dores RM, Camilo-Junior CG, Rosa TC. SentiHealth-Cancer: a sentiment analysis tool to help detecting mood of patients in online social networks. Int J Med Inform. 2016;85(1):80–95 https://doi.org/10.1016/J.IJMEDINF.2015.09.007.

29. López A, Detz A, Ratanawongsa N, Sarkar U. What patients say about their doctors online: a qualitative content analysis. J Gen Intern Med. 2012;27(6):685–92 https://doi.org/10.1007/s11606-011-1958-4.

30. Padrez KA, Ungar L, Schwartz HA, et al. Linking social media and medical record data: a study of adults presenting to an academic, urban emergency department. BMJ Qual Saf. 2015;25:414–23 https://doi.org/10.1136/bmjqs-2015-004489.

31. Greysen SR, Garcia CC, Sudore RL, et al. Functional impairment and internet use among older adults: implications for meaningful use of patient portals. JAMA. 2014;174(7):1188–90 https://doi.org/10.1001/jamainternmed.2014.1864.

32. Cesare N, Grant C, Hawkins JB, Brownstein JS, Nsoesie EO. Demographics in social media data for public health research: does it matter? 2017. https://arxiv.org/ftp/arxiv/papers/1710/1710.11048.pdf. Accessed 12 Feb 2018.

33. Rivers CM, Lewis BL. Ethical research standards in a world of big data. F1000Res. 2014;3:38 https://doi.org/10.12688/f1000research.3-38.v1.

34. Vayena E, Salathé M, Madoff LC, Brownstein JS. Ethical challenges of big data in public health. PLOS Comput Biol. 2015;11(2):e1003904 https://doi.org/10.1371/journal.pcbi.1003904.

35. Godlee F. What can we salvage from care.data? BMJ. 2016;354:i3907 https://doi.org/10.1136/bmj.i3907.

36. Barrett G, Cassell JA, Peacock JL, Coleman MP. National survey of British public's views on use of identifiable medical data by the National Cancer Registry. BMJ. 2006;332(7549):1068–72 https://doi.org/10.1136/bmj.38805.473738.7C.

37. Kaye J, Whitley EA, Lund D, Morrison M, Teare H, Melham K. Dynamic consent: a patient interface for twenty-first century research networks. Eur J Hum Genet. 2014;23(10):141–6 https://doi.org/10.1038/ejhg.2014.71.

38. Grande D, Mitra N, Shah A, Wan F, Asch DA, Health P. The importance of purpose: moving beyond consent in the societal use of personal health information. Ann Intern Med. 2015;161(12):855–62 https://doi.org/10.7326/M14-1118.

39. Eichstaedt JC, Smith RJ, Merchant RM, Ungar LH, Crutchley P, Preoţiuc-Pietro D, et al. Facebook language predicts depression in medical records. Proc Natl Acad Sci U S A. 2018;115:11203–8 https://doi.org/10.1073/pnas.1802331115.

40. Reece AG, Reagan AJ, Lix KLM, Dodds PS, Danforth CM, Langer EJ. Forecasting the onset and course of mental illness with Twitter data. Sci Rep. 2017;7(1):13006 https://doi.org/10.1038/s41598-017-12961-9.

41. Yeoman G, Furlong P, Seres M, Binder H, Chung H, Garzya V, et al. Defining patient centricity with patients for patients and caregivers: a collaborative endeavour. BMJ Innov. 2017;3(2):76–83 https://doi.org/10.1136/bmjinnov-2016-000157.