Ethical and Legal Issues concerning the Use of Social Media to Get to the "Real World"

Andrew Cox, PhD, Research Scientist, Evidera, London, UK; Seye Abogunrin, Research Associate, Evidera, London, UK; H. Keri Yang, PhD, MPH, MS, Director, Center for Observational and Real World Evidence, Merck & Co., Inc., West Point, PA, USA; and Ruth Suter, BS, MBA, Executive Director - Market Access and Patient Services, BioMarin Pharmaceuticals, San Rafael, CA, USA

KEY POINTS . . .

Currently, there are more than 2 billion patient posts from over 50 million healthcare users available to view online.

There is a small but growing industry and increasing interest in utilising these posts as an information and data source.

There is currently no clarity of ethical and legal issues or any guidelines around the use of patient posts.

A recent, large US study estimated that looking for health information is the third most popular online activity, with 80% of internet users looking online for health information. Because 25% of the US population does not go online [1], however, the study more accurately suggests that 59% of the entire US adult population seek health information online (80% of 75% that go online). The rise of social media (SM) and health-related SM has been phenomenal. Health-related conditions are discussed on major SM sites such as Twitter, Facebook, and Google, in addition to many other generic SM sites. Furthermore, an increasing number of patients are discussing their conditions in health-specific forums. For example, the breast cancer forum on the MacMillan Cancer website [2] includes over 48,000 posts spanning 7 years. One start-up company, which regularly monitors data from health-related social media, reports accessing over 2.5 billion patient posts from 50 million health care users [3]. Patients commonly discuss treatments and the associated side effects in great detail. In the case of rare diseases, SM is a vital tool in connecting patients with others diagnosed with the same condition those who often know the most about the condition. It is not surprising that publications, research projects, and commercial ventures focusing on health-related SM are increasing. A recent review of publications found 284 publications connected with SM since 2002 [4].

There has been a great deal of interest in patient postings for surveillance of adverse drug reactions, summarising the patient voice (e.g., patient opinions about specific treatments), and discovering unmet needs in the treatment of diseases (Fig. 1). The technology to "scrape" tens of thousands of patient posts from the internet is widely available and can be accessed in a matter of minutes. As of yet, however, the technology to process text data is still in its relatively early evolution. Computer programs still have difficulties distilling meaningful evidence from textual data, especially the somewhat messier format of communication that appears in social media. Yet natural language processing (NLP) technology is progressing rapidly, with many active research groups and publications in the field.

Already new start-up companies, such as Treato and dMetrics [3], are collecting and indexing billions of patient posts, compiling them into a large database, and commercially marketing proprietary software and services based on the ability to interrogate the data source. Web crawlers of such companies constantly revisit and update information from SM sites. All of this is done based on the tenet that the information is in the public domain (e.g., free for all) with no permissions, ethics, or terms and conditions. Though this may or may not be the case, the purpose of this article is to highlight that there has been no formal discussion or consideration of the ethical and legal dimensions.

Figure 1. A Fictional Patient Posting and How it Could be Informative.
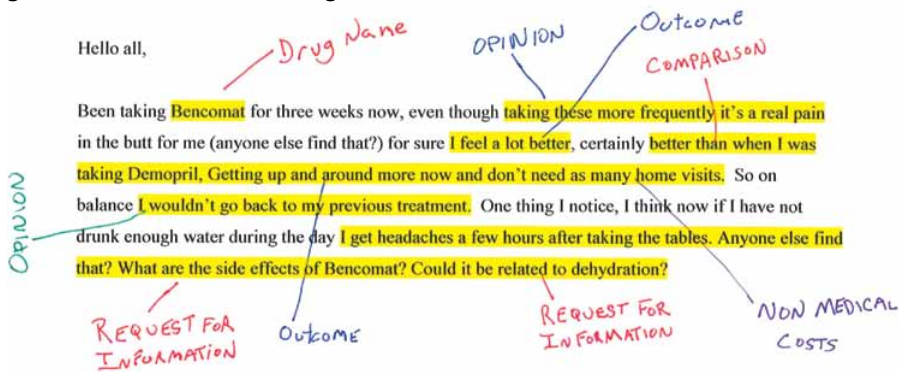
There is undoubtedly an ethical dimension to research using health-related social media. At one end of the scale, a researcher can read hundreds of posts from an open access site and make notes, tabulate counts of comments on specific treatments or side effects, and then report or publish the findings. This activity may be regarded as fairly benign. Moving a little further along the ethical scale, a company may scrape and analyse hundreds of thousands of patient posts, searching and tabulating adverse drug reactions for the purpose of safety monitoring. At the other end of the scale, however, a researcher could scrape hundreds of thousands of patient posts, even accessing posts on sites that require memberships and are password protected. The researcher could publish the resulting research, showing individual word-for-word examples of posted items. All of this can take place without any regard to permissions, ethical behaviour, or terms and conditions of website proprietors. This is clearly not quite so benign—any display of patient posts, in full or in part, allows anyone to search and locate the source almost instantly, opening the door to identification of individuals and personal information.

Approximately 80% would share information with drug companies to help make safer products or learn more about their disease.

Social media is a new frontier, with important implications for patient safety and understanding of the patient perspective. Currently, there is little thought given to the ethical and legal issues surrounding use of what is in reality, highly personal and potentially sensitive information. It may seem strange to suggest that data posted on the world-wide-web (WWW) is personal. But if you have spent any time working with or reading these patient posts, you cannot fail to acknowledge that there are truly personal, harrowing, and moving accounts of people's experiences with diseases. A number of important issues should be considered and debated to set standards of professional conduct involving these types of SM websites and posts.

What are the wishes of patients that post information on these types of SM websites? Are they agreeable that their conversations and experiences may be used for research purposes?

It is fair to assume that patients themselves are most likely to oppose the use of their SM posts for research use. Surprisingly, results of a survey of over 3,000 SM users showed that approximately 70% of patients with a medical condition believed that information they posted could be used to discriminate against them. Conversely, in spite of this strong reservation, more than 90% were willing to share their health data to help improve care or help research [5]. Approximately 80% would share information with drug companies to help make safer products or learn more about their disease.

Are permissions from website proprietors needed or required? Website proprietors clearly have an important stake in their site content being used for research purposes. Web crawlers and scraping software used by researchers to obtain posted information can cause a site to crash if they are not professionally and responsibly programmed. Many website proprietors also own the

copyright of the material within the website, so there are multiple stakeholders. In our experience, when asked for permission to use posts, website proprietors most often responded by neither refusing nor granting research requests, seeming to sit on the fence. However, it is likely that the responses are actually from webmasters who do not understand the nature or issues around requests from researchers.

Do these types of research projects violate website terms and conditions?

All websites and SM sources have terms and conditions, the complexity and coverage vary widely, typically such terms and conditions (T&Cs) contain a restriction of use for personal and non-commercial use of the site and contents. Most contain some statements of copyright ownership and that permission should be sought before use. Occasionally T&Cs will prohibit access or scraping the website by crawlers or scraping software. Sometimes terms imply permission to use the information, if used in a synthesised form. Overwhelmingly, it is unclear how these terms and conditions apply explicitly to the research use.

How does copyright law apply to these kinds of studies?

Copyright law applies to material available on the WWW; wholesale copying and reuse of material is a breach of copyright. For research purposes, however, this can be avoided easily by not reproducing text verbatim (directly copying) and presenting only a summary level synthesis of the material.

What about identification of individuals?

Undoubtedly, scattered over multiple postings SM users give away a staggering amount of personal information that could be used toward identification of the individual—a real possibility. The concern is that identification could lead to discrimination, such as denial of health benefits or job opportunities or for marketing of products or services. In fact, outside of the health field, SM has been used to identify and criminally prosecute individuals [6,7].

There are a number of different perspectives

Different perspectives are available for framing ethical considerations of the use of SM. Patients with a rare disease may be more likely to risk (inadvertent) identification. In fact, the community of patients with rare diseases often have a high use of SM, which offers vital communication and networking, especially when cases are spread across diverse geographical locations. Where more traditional sources of clinical information and other background knowledge are sparse or non-existent, SM serves as a key source of information and experience exchange for patients. The point that dialogues are available for the world to view is important in allowing new patients to be able to connect with the patient network. To argue that because these dialogues have been posted publically therefore, it should be public domain seems disingenuous.

The case may be slightly different regarding chronic disease. There are likely large volumes of posts from multiple sites and, consequently, risk of identification of individuals may be less of a concern. Some conditions, however, may carry social stigma.

Patients with HIV or with mental health conditions may be much more averse to the potential risk of being identified. On the other hand, some health-related SM users may not suffer from a condition at all. A case in point is the large volume of SM discussions available regarding the issues of whether to vaccinate and the safety of vaccines. Patients, researchers, the biomedical industry, academia, legal specialities, third party commercial companies, and data analysts are all likely to have differing views on the ethical and legal issues involved.

The Way Forward

At present, there are no clear guidelines or legislation addressing the use of SM in research. Many who hold the view this type of posted text is an immense resource, free for the taking. Two issues, however, emerge: one must take the trouble to become familiar with the content of health-related posts, and this is highly personal and sensitive information that should be treated in an ethical and professional manner. To that end, there is a strong need for industry guidelines, a concise guide to professional best practices. Creation of guidelines involves input from a range of interested parties. Potential best practice starting points may be:

■ Report synthesised or summary level data from forums, posts, or parts of posts. Do not repeat or show directly copied text. This minimises the possibility of identifying individuals and guards against infringement of copyright.

■ Do not access or use material present on sites that require registration and passwords to access; deal only with material that is in the public domain, unless explicit permission has been granted to use the protected material.

■ Collect data from websites in a responsible manner, build safeguards into crawling and scraping programs (used to obtain the data) to avoid overloading or crashing websites.

■ When publishing research results, include a statement describing the ethical and legal policy safeguards used. Journals should require demonstration of ethical conduct from authors.

■ The intended research should demonstrate intent that it is being conducted for the benefit of the patient.

Glossary

• **Scraping / Web Scraping** - A computer software technique for extracting information from websites.
• **Natural Language Processing (NLP)** - A computer software technique used to understand and process human speech as it is spoken and human text as it is written.
• **Web Crawler** - A program or automated script which browses the World Wide Web in a methodical, automated manner.

References

[1] Pew Research Internet Project: Health Topics. Available at: http://pewinternet.org/Reports/2011/HealthTopics.aspx. [Accessed March 15, 2015]. [2] We are MacMillan Cancer Support. Available at: http://community.macmillan.org.uk/cancer_types/breast-cancer/discussions.aspx. [Accessed March 15, 2015]. [3] dmetrics Website. Available at: https://dmetrics.com/. [Accessed March 15, 2015]. [4] Treato Website. Available at: http://treato.com. [Accessed March 15, 2015]. [4] Hamm MP, Chishold A, Shulhan J, et al. Social media use among patients and caregivers: a scoping review. BMJ Open 2013;3:e002819.

[5] Grajales F, Clifford D, Loupos P, et al. Social Networking Sites and the Continuously Learning Health System: A Survey. Discussion Paper. (2014). Available at: http://www.iom.edu/~/media/Files/Perspectives-Files/2014/Discussion-Papers/VSRT-PatientDataSharing.pdf. [Accessed on November 6, 2014]. [6] Michigan Office of Highway Safety Planning 2012 Annual Evaluation Report (pp. 21)". 2012. Retrieved May 20, 2014. [7] "Police overwhelmed by social media evidence". CBC News. 2011-06-20. Retrieved June 20, 2011. ■

Additional information:
The preceding article was based on the workshop of the same name at the ISPOR 17th Annual European Congress, 8-12, 2014, Amsterdam, The Netherlands. To view the presentation, go to: http://www.ispor.org/Event/Released Presentations/2014Amsterdam#workshoppresentations.

For more information on the uses of social media, you may attend the following sessions at the ISPOR 18th Annual European Congress in Milan, Italy:
W12: Optimizing Patient Involvement in Payer Health Care Decisions to Access New Therapies, Tuesday 10 November 2015, 8:45AM, and IP5: Blog it, Tweet It, Like It, or Bin It The Role of Social Media Data in PRO Research, Monday 9 November 2015, 11:15AM.