

# Pharmacovigilance for a Revolving World: Prospects of Patient-Generated Data on the Internet

G. Niklas Norén

Published online: 6 August 2014  
© Springer International Publishing Switzerland 2014

*“The whole world turns upside down in 10 years, but you turn upside down with it”*

Spider Robinson, 1977<sup>1</sup>

The past 10 years have seen the advent of smartphones and a massive increase in the use and influence of the internet and social media in our daily lives. The means by which we interact with one another and with the world’s information have fundamentally changed. Navigation is one example: when I drive from Uppsala to Stockholm, not only does my phone now direct me where to go, but its instructions adapt in real-time based on the current flow of traffic, as crowd-sourced from the phones of fellow drivers. In health, we have seen the emerging movement of the quantified self where individuals track exercise and sleep patterns, along with information on diet, and aspects of their physical and mental state, with the intent of understanding what promotes their own good health and well-being (a daunting causal inference task!). So, too, have we seen the growth of internet communities where patients store and interact with their own health data, and share experiences, with the ultimate aim of learning from one another for better health outcomes.

We have yet to see the full impact of these developments in pharmacovigilance, but movement has begun. The important role of patients in identifying, describing, and

ultimately avoiding, harm from medicines is increasingly recognized [1, 2], and modern technologies are being explored to facilitate patient engagement [3]. Information on direct patient reports in pharmacovigilance differs in certain respects from information provided by health professionals, but has been found to be of complementary value [4]. Interestingly, in one study, direct patient reports in Denmark and Norway (collected electronically) were found to be better documented, on average, than reports from health professionals [5]. Mobile applications for spontaneous reporting are being piloted, and we have seen initiatives to use social media sites to stimulate submission of individual case reports [6]. This may add to our understanding of adverse drug reactions and provide opportunities to directly engage with and support patients [7], but also comes with challenges such as how to effectively protect sensitive information, account for selection biases, and minimise the risk of intended misuse [7, 8].

In parallel, patient-generated data on the internet is beginning to be explored as a primary basis for analysis of possible adverse drug reactions. In a paper recently published in *Drug Safety*, Freifeld et al. [9] described an analysis of Twitter microblog posts for references to drugs and adverse events, with comparison to reporting patterns in the US FDA Adverse Event Reporting System (FAERS). Other researchers have explored online discussion fora and internet search patterns for similar purposes [10–12]. Together, these studies begin to accumulate evidence for the technical feasibility of identifying references to possible adverse drug reactions in patient-generated data on the internet, and of analysing these patterns to characterise and understand patients’ experiences and concerns. Such

---

G. N. Norén (✉)  
Uppsala Monitoring Centre, WHO Collaborating Centre for  
International Drug Monitoring, Box 1051, 751 40 Uppsala,  
Sweden  
e-mail: niklas.noren@who-umc.org

G. N. Norén  
Department of Mathematics, Stockholm University, Stockholm,  
Sweden

<sup>1</sup> Via *Flash mob gone wrong*, Tom Scott 2010, <http://www.youtube.com/watch?v=RyMdOT8YJgY>.

descriptive analyses may provide valuable insights into factors that affect compliance as well as treatment outcomes. They may also help us understand which adverse effects have the greatest impact on patients. In contrast, it remains an open question to what extent and in what way each of these data sources could support traditional signal detection. In this context, the different sources of internet-based data may need to be considered separately as they are different in nature and carry their own separate challenges. They vary on one hand in the richness of the provided information and on the other hand (usually inversely so) in their scope and coverage.

At one end of the spectrum we have the more detailed case descriptions, and the longitudinal medical information captured on dedicated social network sites. The former resemble the free-text narratives of individual case reports, should be encoded in standard E2B format<sup>2</sup>, and captured by the regular pharmacovigilance system, presuming that they fulfil reporting requirements. This will enable their inclusion in regular pharmacovigilance signal detection and analysis. If the descriptions do not fulfil reporting requirements, they might perhaps still be captured in E2B format to facilitate analysis but be kept separate; their appropriate use requires careful consideration [7–9]. In turn, the longitudinal data shared by patients on dedicated social network sites represent a form of patient-maintained electronic medical record that could be subjected to analytics developed for other sources of longitudinal observational data [13]. Beyond the general lack of medical validation common to many sources of patient-generated data, the biggest challenge in using these data sources for signal detection may be selection bias and lack of power. The largest internet-based patient communities today gather hundreds of thousands of members, and in that are dwarfed by databases of administrative claims and electronic medical records currently explored for the same purpose. It is encouraging to see that many patients are prepared to share their personal information with others, often altruistically [14, 15]. Overall, they represent a minority of all patients but, for particular conditions and treatments, the coverage will be greater, and with the current growth in popularity of these communities, the richness of their information makes them an interesting potential resource for the future [3].

At the other end of the spectrum, there are the less detailed case descriptions, as shared on social network sites and in microblog posts, along with the correlations that can be identified between drugs and adverse events entered

together or in sequence in internet searches. The search patterns, in particular, provide lower barriers for contribution that may increase their value for early detection; many who would hesitate to share their adverse experiences in public might still use internet search engines to seek relevant information, and agree to share this data. Discussion posts in online patient communities that may reflect possible adverse drug reactions have been correlated with regulatory activity [10], and internet search logs have been explored for correlations with known adverse drug reactions [11] and interactions [12]. These publications include some examples of emerging adverse drug reactions or interactions, such as the association in internet searches between pravastatin, paroxetine, and hyperglycaemia [12]. However, they are primarily based on evaluation against established adverse drug reactions, which must be interpreted with caution in light of the differences in empirical patterns between emerging and established adverse drug reactions [16]. A separate challenge is the limited information to support causality assessment and the risk of intentional manipulation [8, 17]. This may make such patterns difficult to act upon, even if proven predictive for emerging adverse drug reactions, on average. For reference, consider the rofecoxib withdrawal of 2003, which illustrates many of the challenges surrounding the complex decisions that need to be made in pharmacovigilance [18]. Would we be prepared to make such decisions based on associations derived from internet searches or brief comments on microblogs? Most likely not! Instead, the real promise of more terse patient-generated data on the internet may lie in its combination with traditional pharmacovigilance data such as individual case reports and observational studies. An often-cited precedent for the successful use of internet information to serve public health is the early detection of influenza outbreaks via search patterns [19]. This has been placed in new light following recent analyses indicating that the predictive value of internet search patterns alone was previously over-estimated; however, their combination with traditional surveillance methods has been found to outperform each method when used separately [17]. In a similar vein, the pooled analyses of internet search patterns and spontaneous reports was found to outperform separate analyses of each in recalling established adverse drug reactions [11]. In the context of initial data preparation, Freifeld et al. [9] reported that information extracted from product labels improved accuracy, which is another way that joint consideration of distinct information sources may bring benefit.

There may be a more prominent role to be played by patient-generated data on the internet in the detection of specific types of risk related to medicine use. Causality assessment for harm related to substandard or counterfeit medicines, or to medicine (mis-)use, does not rely on

<sup>2</sup> The standard adopted for electronic transmission of Individual Case Safety Reports by the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH).

epidemiology but on root cause analysis for the individual case. Such risks may lend themselves more naturally to evaluation without access to external data, on condition that the original reporter may be reached, technically and ethically—the delicate balance between patient safety and privacy must be respected, and this is true for all of the above-mentioned sources of data.

Our analytics must improve and adapt to evaluate the above opportunities—let alone incorporate these information sources in routine pharmacovigilance work further down the line. Freifeld et al. [9] described several of the key challenges, most notably the analysis of unstructured free text. The detection of references to drugs and medical events is a first step but not straightforward at that, given the multitude of ways in which we may refer to the same adverse event or drug. Freifeld et al. [9] described some of the challenges in mapping internet vernacular to Medical Dictionary for Regulatory Activities (MedDRA®) codes for analysis. Similarly, drugs may have hundreds of commercial names worldwide and require a global reference such as the WHO Drug Dictionary Enhanced™ for sensitivity [20]. At the same time, many common English words coincide with a drug product name in some part of the world, which calls for disambiguation methods for reliable recognition. Ideally, we should employ semantic analysis that goes beyond the bag-of-words perspective and looks for meaning in free text. We will also need effective methods for de-duplication and identification of complementary descriptions of the same case, especially if we seek to use separate information sources, synergistically. Here, probabilistic methods for record matching successfully applied to spontaneous reports should provide a good starting point [21]. New analytical strategies for de-identification will be an important component in ensuring patient confidentiality while not compromising the accuracy of our analyses. Finally, we may need to adapt and implement methods from areas such as fraud detection as a basis to detect suspected misuse or manipulation. Several of these topics will be addressed by the European Innovative Medicines Initiative project, WEB-Recognizing Adverse Drug Reactions (RADR), scheduled to start in the autumn of 2014.

If the world turns upside down every 10 years, will another decade take us back to 2004? Certainly not! Effective pharmacovigilance will remain reliant on reliable and relevant data, and research over coming years will shed light on the appropriate place and optimal analysis of patient-generated data on the internet. We shall learn how it may best be combined with other data sources to bring value to pharmacovigilance. Along the way, the necessary technological advances should benefit old and new pharmacovigilance information sources alike. All the while, we will need to broaden our perspective to include the whole

of the world, accounting for regional and demographic differences and addressing language barriers, with the aim of supporting patients and health professionals, worldwide.

**Acknowledgments** G. Niklas Norén is a member of the Innovative Medicines Initiative project (IMI) WEB-RADR consortium, a 3-year public-private partnership to explore reporting by healthcare professionals and patients by means of mobile devices and the strengthening of safety signals from social media platforms. IMI WEB-RADR is funded by the European Commission and the European Federation of Pharmaceutical Industries and Associations through the IMI 9th Call and scheduled to start in the autumn of 2014. The views expressed here are those of the author and should not be interpreted as the position of the IMI WEB-RADR Consortium or any of its participating organisations. No sources of funding were used to assist in the preparation of this editorial. G. Niklas Norén has no conflicts of interest that are directly relevant to the content of this editorial.

## References

1. Basch E. The missing voice of patients in drug-safety reporting. *N Engl J Med*. 2010;362(10):865–9.
2. Banerjee A, Okun S, Edwards IR, et al. Patient-reported outcome measures in safety event reporting: PROSPER consortium guidance. *Drug Saf*. 2013;36(12):1129–49.
3. Wicks P, Vaughan TE, Massagli MP, Heywood J. Accelerated clinical discovery using self-reported patient data collected online and a patient-matching algorithm. *Nat Biotech*. 2011;29(5):411–4.
4. de Langen J, van Hunsel F, Passier A, den Berg L-V, van Groothoest K. Adverse drug reaction reporting by patients in The Netherlands: three years of experience. *Drug Saf*. 2008;31(6):515–24.
5. Bergvall T, Norén GN, Lindquist M. vigiGrade: a tool to identify well-documented individual case reports and highlight systematic data quality issues. *Drug Saf*. 2014;37(1):65–77.
6. Knezevic MZ, Bivolarevic IC, Peric TS, Jankovic SM. Using Facebook to increase spontaneous reporting of adverse drug reactions. *Drug Saf*. 2011;34(4):351–2.
7. Edwards IR, Lindquist M. Social media and networks in pharmacovigilance: boon or bane? *Drug Saf*. 2011;34(4):267–71.
8. Franzen W. Can social media benefit drug safety? *Drug Saf*. 2011;34(9):793.
9. Freifeld CC, Brownstein JS, Menone CM, et al. Digital drug safety surveillance: monitoring pharmaceutical products in Twitter. *Drug Saf*. 2014;37(5):343–50.
10. Chee BW, Berlin R, Schatz B. Predicting adverse drug events from personal health messages. *AMIA Annu Symp Proc*. 2011;2011:217–26.
11. White RW, Harpaz R, Shah NH, DuMouchel W, Horvitz E. Toward enhanced pharmacovigilance using patient-generated data on the internet. *Clin Pharmacol Ther*. 2014;96(2):239–46. doi:10.1038/clpt.2014.77.
12. White RW, Tatonetti NP, Shah NH, Altman RB, Horvitz E. Web-scale pharmacovigilance: listening to signals from the crowd. *J Am Med Inform Assoc*. 2013;20(3):404–8.
13. Norén GN, Hopstadius J, Bate A, Star K, Edwards IR. Temporal pattern discovery in longitudinal electronic patient records. *Data Min Knowl Disc*. 2010;20(3):361–87.
14. van Hunsel F, van der Welle C, Passier A, van Puijenbroek E, van Groothoest K. Motives for reporting adverse drug reactions by patient-reporters in The Netherlands. *Eur J Clin Pharmacol*. 2010;66(11):1143–50.

15. Grajales F, Clifford D, Loupos P, et al. Social networking sites and the continuously learning health system: a survey: Institute of Medicine; 23 Jan 2014.
16. Norén GN, Caster O, Juhlin K, Lindquist M. Zoo or Savannah? Choice of training ground for evidence-based pharmacovigilance. *Drug Saf. Epub* 9 Jul 2014. doi:[10.1007/s40264-014-0198-z](https://doi.org/10.1007/s40264-014-0198-z).
17. Lazer D, Kennedy R, King G, Vespignani A. The parable of Google Flu: traps in big data analysis. *Science*. 2014;343(6176):1203–5.
18. Edwards IR. What are the real lessons from Vioxx? *Drug Saf.* 2005;28(8):651–8.
19. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature*. 2009;457(7232):1012–4.
20. Lindquist M. VigiBase, the WHO Global ICSR Database System: basic facts. *Drug Inf J.* 2008;42(5):409–19.
21. Norén GN, Orre R, Bate A, Edwards IR. Duplicate detection in adverse drug reaction surveillance. *Data Min Knowl Disc.* 2007;14(3):305–28.