Published in final edited form as:

Curr Epidemiol Rep. 2017 December ; 4(4): 262–265. doi:10.1007/s40471-017-0123-y.

Pharmacoepidemiology in the era of real-world evidence

Sengwee Toh, ScD

Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA, USA

Keywords

Pharmacoepidemiology; Real-world data; Real-world evidence; Distributed data networks

Long before the terms real-world data (RWD) and real-world evidence (RWE) were coined, researchers had been using data collected as part of routine healthcare delivery to generate evidence about the utilization, benefits, and risks of medical products [1–4]. There are several variations to the definitions of RWD and RWE, but most are similar to the ones used by the U.S. Food and Drug Administration (FDA), which defines RWD as “data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources” and RWE as “the clinical evidence regarding the usage, and potential benefits or risks, of a medical product derived from analysis of RWD” [5].

It is not uncommon to re-label an existing construct with a more contemporary descriptor. The term RWD provides a more unified framework to broadly capture the various types of data collected outside of traditional randomized controlled trials. Examples of RWD include electronic health record data, administrative claims data, product and disease registry data, and data collected from patients through smart devices or social media. But more important, the introduction of RWD and RWE signifies a shift in the paradigm of evidence generation that could have major impacts on research, regulatory science, and public health. The 21st Century Cures Act, passed by the U.S. Congress in 2016, directed the FDA to evaluate how RWE may be used to support approval of new indications for approved drugs or to support or satisfy post-approval study requirements [6]. The European Medicines Agency is also considering the inclusion of RWE as part of adaptive licensing [7]. In other words, RWE generated from observational studies or pragmatic trials that leverage routinely collected electronic health data may soon be used to supplement findings from traditional randomized controlled trials in the approval of medical products [8, 9]. Safety findings from post-marketing RWD-based studies, which have been more formally integrated into regulatory review in the U.S. recently [10], may gain additional regulatory importance.

Corresponding Author: Sengwee Toh, Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, 401 Park Drive, Suite 401 East, Boston, MA 02215, USA, darren_toh@harvardpilgrim.org, Phone: 617-867-4818, Fax: 617-867-4276.

Compliance with Ethical Standards

Conflict of Interest: The author does not have any conflict of interest to disclose.

Human and Animal Rights and Informed Consent: This article does not contain any studies with human or animal subjects performed by the author.

It will be interesting to see how regulatory agencies and the research community react and position themselves in the coming years. Pharmacoepidemiology is a research field that involves analysis of routinely collected electronic health data. The field has developed good practices on the conduct and report of RWD-based studies [11–15]. Building upon its solid foundation, pharmacoepidemiology can move forward in several domains in the era of RWD and RWE.

Better

Pharmacoepidemiologists will continue to find ways to reduce confounding and other biases inherent in observational RWD-based studies. No one particular type of data source is necessarily better than the other. There are also substantial variations in the size, patient characteristics, and quality of databases within each type of data source. In general, administrative claims databases have better defined observation periods during which medically attended events can be captured, but they do not contain certain clinical data important for some studies [16]. Electronic health record databases are enriched with more clinical information, but they generally do not have complete capture of events that occur across multiple health systems [17]. Linkages among complementary data sources improve data completeness and help reduce some biases, but the linked databases have their own limitations, e.g., complete data is only available in a subset of patients who may be quite different from other patients. Data linkage will become more technically feasible but will continue to be challenging due to privacy and other concerns. Regardless of data source, careful curation to make it fit-for-purpose is paramount [18, 19]. Continued development of valid and practical methods that analyze the increasingly complex RWD, especially if multiple data sources are linked, will also be needed.

Bigger

Pharmacoepidemiologists will continue to identify appropriate data sources with sufficient sample sizes to answer clinical questions that are expected to be increasingly targeted. Despite what is implied in its name, individualized or precision medicine research [20] often cannot be done in individual patients and requires a relatively large number of patients with similar characteristics to draw valid inference. Even if the interest is to generate RWE in the broader population, the assessment of rare treatments (e.g., orphan drugs) or rare outcomes (e.g., sudden cardiac death) generally requires large data sources as well. Although some studies can be done with one database, many studies now require data from multiple databases to achieve adequate statistical power or more generalizable findings. Distributed data networks have been proven to be a viable and preferred approach to analyzing multiple databases [21, 22]. In the North America alone, there are distributed networks created to support medical product safety surveillance [10, 23], comparative effectiveness research [24], pragmatic trials embedded within health systems [25], and public health surveillance [26]. The creation, maintenance, and expansion of distributed data networks require stable funding, proper governance, and analytic methods that produce valid results while protecting patient privacy [22, 27].

Brisker

Pharmacoepidemiologists will continue to develop methods that generate RWE faster without sacrificing the scientific rigor of the analysis. The key is to differentiate between study decisions that can only be made by thoughtful deliberation and other mechanical steps that can largely be done with minimal human input. It is possible to conduct semi-automated pharmacoepidemiologic studies using pre-tested and customizable analytic tools and produce results comparable to those from traditional protocol-based studies [28, 29]. Future work will continue to develop new analytic tools to accommodate more complex analyses. It is critical to ensure the transparency of the development to facilitate independent validation of the tools.

Broader

Pharmacoepidemiologists will continue to broaden the spectrum of clinical questions that can be addressed by RWD. The availability of new data elements enables assessments of questions that had been difficult to study in the past. For example, better integration of patient-reported data with other routinely collected data (e.g., electronic health record data) offers new opportunities to examine the effects of medical treatments from patients' perspectives [30]. The emergence of new clinical scenarios also give rise to new research questions. For example, the increasing availability of biosimilars raises questions about the comparative effectiveness and safety of these products compared to their innovator products [31]. RWE includes findings from pragmatic trials that leverage data collected as part of routine clinical care [8, 9]. It is critical to build upon the success of completed and ongoing pragmatic trials embedded within health systems [25, 32, 33] to conduct additional studies that address other pressing clinical questions.

Bolder

Pharmacoepidemiologists will continue to think outside the box. Pharmacoepidemiology has benefited from its ability to synergize multiple disciplines, including epidemiology, pharmacology, medicine, biostatistics, and social science. The era of RWD and RWE calls for closer collaborations with experts from computer science, data science, informatics, genomic research, and other disciplines. Data-adaptive techniques (such as machine learning) combined with thoughtful human input are increasingly being used to mine electronic health record databases [34, 35] and improve analytic methods commonly used in pharmacoepidemiology [36, 37]. The ability to collect more data from mobile devices enables exploration of new issues, e.g., the relation between weather and joint pain in patients with rheumatoid arthritis [38]. Pushing the boundaries and leveraging progress made in other research areas will continue to help advance the field of pharmacoepidemiology.

Conclusion

Pharmacoepidemiologists have been using electronic health data to generate evidence on the risks and benefits of medical products for years. They are well-positioned to embrace the new challenges and opportunities in the (new) era of RWD and RWE.

Acknowledgments

Dr. Toh is partially supported by the National Institute of Biomedical Imaging and Bioengineering (U01EB023683).

References

1. Dunlop DM. Drug control and the British Health Service. *Ann Intern Med.* 1969; 71(2):237–44. [PubMed: 4979214]
2. Stolley PD, Lasagna L. Prescribing patterns of physicians. *J Chronic Dis.* 1969; 22(6):395–405. [PubMed: 5370512]
3. Federspiel CF, Ray WA, Schaffner W. Medicaid records as a valid data source: the Tennessee experience. *Med Care.* 1976; 14(2):166–72. [PubMed: 768652]
4. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol.* 2005; 58(4):323–37. DOI: 10.1016/j.jclinepi.2004.10.012 [PubMed: 15862718]
5. U.S. Food and Drug Administration. [Accessed October 1, 2017] Use of real-world evidence to support regulatory decision-making for medical devices. Guidance for industry and Food and Drug Administration staff. 2017. Available from: <https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm513027.pdf>
6. Public Law 114–255; 114th Congress; 2016. Available from: <https://www.congress.gov/114/plaws/publ255/PLAW-114publ255.pdf>
7. European Medicines Agency. [Accessed October 1, 2017] Guidance for companies considering the adaptive pathways approach. 2016. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2015/11/WC500196726.pdf
8. Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, et al. Real-World Evidence - What Is It and What Can It Tell Us? *N Engl J Med.* 2016; 375(23):2293–7. DOI: 10.1056/NEJMsbl609216 [PubMed: 27959688]
9. Jarow JP, LaVange L, Woodcock J. Multidimensional Evidence Generation and FDA Regulatory Decision Making: Defining and Using “Real-World” Data. *JAMA.* 2017; 318(8):703–4. DOI: 10.1001/jama.2017.9991 [PubMed: 28715550]
10. Ball R, Robb M, Anderson SA, Dal Pan G. The FDA’s sentinel initiative--A comprehensive approach to medical product surveillance. *Clin Pharmacol Ther.* 2016; 99(3):265–8. DOI: 10.1002/cpt.320 [PubMed: 26667601]
11. U.S. Food and Drug Administration. [Accessed October 1, 2017] Guidance for industry and FDA staff: Best practices for conducting and reporting pharmacoepidemiologic safety studies using electronic healthcare data. 2013. Available from: <https://www.fda.gov/downloads/drugs/guidances/ucm243537.pdf>
12. The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). [Accessed October 1, 2017] Guide on Methodological Standards in Pharmacoepidemiology (Revision 5). 2016. Available from: http://www.encepp.eu/standards_and_guidances/documents/ENCePPGuideofMethStandardsinPE_Rev5.pdf
13. Public Policy Committee of the International Society for Pharmacoepidemiology. Guidelines for good pharmacoepidemiology practice (GPP). *Pharmacoepidemiol Drug Saf.* 2016; 25(1):2–10. DOI: 10.1002/pds.3891 [PubMed: 26537534]
14. Benchimol EI, Smeeth L, Guttman A, Harron K, Moher D, Petersen I, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med.* 2015; 12(10):e1001885.doi: 10.1371/journal.pmed.1001885 [PubMed: 26440803]

15. Wang SV, Schneeweiss S, Berger ML, Brown J, de Vries F, Douglas I, et al. Reporting to Improve Reproducibility and Facilitate Validity Assessment for Healthcare Database Studies V1.0. *Pharmacoepidemiol Drug Saf.* 2017; 26(9):1018–32. DOI: 10.1002/pds.4295 [PubMed: 28913963]
16. Suissa S, Garbe E. Primer: administrative health databases in observational studies of drug effects--advantages and disadvantages. *Nat Clin Pract Rheumatol.* 2007; 3(12):725–32. DOI: 10.1038/ncprheum0652 [PubMed: 18037932]
17. Hersh WR, Weiner MG, Embi PJ, Logan JR, Payne PR, Bernstam EV, et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. *Med Care.* 2013; 51(8 Suppl 3):S30–7. DOI: 10.1097/MLR.0b013e31829b1dbd [PubMed: 23774517]
18. Brown JS, Kahn M, Toh S. Data quality assessment for comparative effectiveness research in distributed data networks. *Med Care.* 2013; 51(8 Suppl 3):S22–9. DOI: 10.1097/MLR.0b013e31829b1e2c [PubMed: 23793049]
19. Kahn MG, Raebel MA, Glanz JM, Riedlinger K, Steiner JF. A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research. *Med Care.* 2012; 50(Suppl):S21–9. DOI: 10.1097/MLR.0b013e318257dd67 [PubMed: 22692254]
20. Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med.* 2015; 372(9):793–5. DOI: 10.1056/NEJMp1500523 [PubMed: 25635347]
21. Brown JS, Holmes JH, Shah K, Hall K, Lazarus R, Platt R. Distributed health data networks: a practical and preferred approach to multi-institutional evaluations of comparative effectiveness, safety, and quality of care. *Med Care.* 2010; 48(6 Suppl):S45–51. DOI: 10.1097/MLR.0b013e3181d9919f [PubMed: 20473204]
22. Curtis LH, Brown J, Platt R. Four health data networks illustrate the potential for a shared national multipurpose big-data network. *Health Aff (Millwood).* 2014; 33(7):1178–86. DOI: 10.1377/hlthaff.2014.0121 [PubMed: 25006144]
23. Suissa S, Henry D, Caetano P, Dormuth CR, Ernst P, Hemmelgarn B, et al. CNODES: the Canadian Network for Observational Drug Effect Studies. *Open Med.* 2012; 6(4):e134–40. [PubMed: 23687528]
24. Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet, a national patient-centered clinical research network. *J Am Med Inform Assoc.* 2014; 21(4):578–82. DOI: 10.1136/amiajnl-2014-002747 [PubMed: 24821743]
25. Richesson RL, Hammond WE, Nahm M, Wixted D, Simon GE, Robinson JG, et al. Electronic health records based phenotyping in next-generation clinical trials: a perspective from the NIH Health Care Systems Collaboratory. *J Am Med Inform Assoc.* 2013; 20(e2):e226–31. DOI: 10.1136/amiajnl-2013-001926 [PubMed: 23956018]
26. Vogel J, Brown JS, Land T, Platt R, Klompas M. MDPHnet: secure, distributed sharing of electronic health record data for public health surveillance, evaluation, and planning. *Am J Public Health.* 2014; 104(12):2265–70. DOI: 10.2105/AJPH.2014.302103 [PubMed: 25322301]
27. Toh S, Gagne JJ, Rassen JA, Fireman BH, Kulldorff M, Brown JS. Confounding adjustment in comparative effectiveness research conducted within distributed research networks. *Med Care.* 2013; 51(8 Suppl 3):S4–10. DOI: 10.1097/MLR.0b013e31829b1bb1
28. Gagne JJ, Han X, Hennessy S, Leonard CE, Chrischilles EA, Carnahan RM, et al. Successful comparison of US Food and Drug Administration Sentinel analysis tools to traditional approaches in quantifying a known drug-adverse event association. *Clin Pharmacol Ther.* 2016; doi: 10.1002/cpt.429
29. Zhou M, Wang SV, Leonard CE, Gagne JJ, Fuller C, Hampp C, et al. Sentinel Modular Program for Propensity Score-Matched Cohort Analyses: Application to Glyburide, Glipizide, and Serious Hypoglycemia. *Epidemiology.* 2017; 28(6):838–46. DOI: 10.1097/EDE.0000000000000709 [PubMed: 28682851]
30. Lavalley DC, Chenok KE, Love RM, Petersen C, Holve E, Segal CD, et al. Incorporating Patient-Reported Outcomes Into Health Care To Engage Patients And Enhance Care. *Health Aff (Millwood).* 2016; 35(4):575–82. DOI: 10.1377/hlthaff.2015.1362 [PubMed: 27044954]
31. Baldziki M, Brown J, Chan H, Cheetham TC, Conn T, et al. AMCP Task Force on Biosimilar Collective Intelligence Systems. Utilizing data consortia to monitor safety and effectiveness of

- biosimilars and their innovator products. *J Manag Care Spec Pharm*. 2015; 21(1):23–34. DOI: 10.18553/jmcp.2015.21.1.23 [PubMed: 25562770]
32. Huang SS, Septimus E, Kleinman K, Moody J, Hickok J, Avery TR, et al. Targeted versus universal decolonization to prevent ICU infection. *N Engl J Med*. 2013; 368(24):2255–65. DOI: 10.1056/NEJMoa1207290 [PubMed: 23718152]
33. Hernandez AF, Fleurence RL, Rothman RL. The ADAPTABLE Trial and PCORnet: Shining Light on a New Research Paradigm. *Ann Intern Med*. 2015; 163(8):635–6. DOI: 10.7326/M15-1460 [PubMed: 26301537]
34. Walker AM, Zhou X, Ananthakrishnan AN, Weiss LS, Shen R, Sobel RE, et al. Computer-assisted expert case definition in electronic health records. *Int J Med Inform*. 2016; 86:62–70. DOI: 10.1016/j.ijmedinf.2015.10.005 [PubMed: 26725697]
35. Schuemie MJ, Sen E, t Jong GW, van Soest EM, Sturkenboom MC, Kors JA. Automating classification of free-text electronic health records for epidemiological studies. *Pharmacoepidemiol Drug Saf*. 2012; 21(6):651–8. DOI: 10.1002/pds.3205 [PubMed: 22271492]
36. Lee BK, Lessler J, Stuart EA. Improving propensity score weighting using machine learning. *Stat Med*. 2010; 29(3):337–46. DOI: 10.1002/sim.3782 [PubMed: 19960510]
37. Westreich D, Lessler J, Funk MJ. Propensity score estimation: neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression. *J Clin Epidemiol*. 2010; 63(8):826–33. DOI: 10.1016/j.jclinepi.2009.11.020 [PubMed: 20630332]
38. Reade S, Spencer K, Sergeant JC, Sperrin M, Schultz DM, Ainsworth J, et al. Cloudy with a Chance of Pain: Engagement and Subsequent Attrition of Daily Data Entry in a Smartphone Pilot Study Tracking Weather, Disease Severity, and Physical Activity in Patients With Rheumatoid Arthritis. *JMIR Mhealth Uhealth*. 2017; 5(3):e37.doi: 10.2196/mhealth.6496 [PubMed: 28341616]