2018

Real World Evidence - what it is and what it can tell us according to the International Society for Pharmacoepidemiology (ISPE) Comparative Effectiveness Research (CER) Special Interest Group (SIG)

Hongbo Yuan

M. Sanni Ali

Emily S. Brouwer

Cynthis J. Girman

Jeff J. Guo

Citation/Publisher Attribution

Yuan, H., Ali, M.S., Brouwer, E.S., Girman, C.J., Guo, J.J., Lund, J.L., & Bennett, D.(2018), Real-World Evidence: What It Is and What It Can Tell Us According to the International Society for Pharmacoepidemiology (ISPE) Comparative Effectiveness Research (CER) Special Interest Group (SIG). *Clin. Pharmacol. Ther., 104*(2), 239-241. https://doi.org/10.1002/cpt.1086 Available at: https://doi.org/10.1002/cpt.1086

Title page

Manuscript title:

Real World Evidence – what it is and what it can tell us according to the International Society for Pharmacoepidemiology (ISPE) Comparative Effectiveness Research (CER) Special Interest Group (SIG)

Authors list and affiliations:

Hongbo Yuan, MD, MSc, PhD[1], M Sanni Ali, DVM, MSc, PhD[2,3], Emily S. Brouwer, MPH, PharmD, PhD[4,5], Cynthia J. Girman, DrPH, FISPE[6], Jeff J. Guo, PhD[7], Jennifer L. Lund, PhD[8], Elisabetta Patorno, MD, DrPH[9], Jonathan L. Slaughter, MD, MPH[10], Xuerong Wen, PhD[11], Dimitri Bennett, MD, MPH, FISPE, FACE[12], on behalf of the ISPE Comparative Effectiveness Research Special Interest Group.

[1]Canadian Agency for Drugs and Technologies in Health, Ottawa, Ontario, Canada
[2]Faculty of Epidemiology and Population Health, Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK
[3]Center for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences (NDORMS), University of Oxford, Oxford, UK.
[4]Shire Pharmaceuticals, Lexington MA, USA
[5]Department of Pharmacy Practice and Science, University of Kentucky College of Pharmacy, Lexington, KY, USA
[6]CERobs Consulting, LLC, Chapel Hill, NC, USA
[7]Pharmacy Practice & Administrative Sciences, University of Cincinnati College of Pharmacy, Cincinnati, OH, USA
[8]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA;
[9]Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
[10]Center for Perinatal Research, Nationwide Children's Hospital and Department of Pediatrics, The Ohio State University, Columbus, OH, USA
[11]Health Outcomes, Department of Pharmacy Practice, College of Pharmacy, University of Rhode Island, Kingston, RI, USA
[12]Department of Pharmacoepidemiology, Takeda Pharmaceuticals International Co., Cambridge, MA, USA

Contact details of corresponding author:
Name: Hongbo Yuan, MD, MSc, PhD
Mailing address: 865 Carling Ave., Suite 600; K1S 5S8 Ottawa, Ontario Canada
Fax number: not available
Telephone number: 613-226-4969
e-mail address: HongboY@cadth.ca

Conflict of Interest statement:

Dr. Yuan has nothing to disclose. Dr. M Sanni Ali has nothing to disclose. Dr. Brouwer reports employment at Shire Pharmaceuticals, formerly employed at Boehringer Ingelheim Pharmaceuticals. Submitted work completed as a part of extracurricular involvement as chair of ISPE Comparative Effectiveness Research Special Interest Group. Dr. Girman reports grants from pharmaceutical companies including CSL Behring, Roivant, Adgero, Regeneron, Prolong, Ritter, CytoSorbents, Boehringer-Ingelheim, outside the submitted work; and serves on the Methodology Committee of the Patient-Centered Outcomes Research Institute, and ex-officio on their Clinical Trials Advisory Panel. Dr. Guo did not receive any financial support or material gain that may involve the subject matter of the article. Dr. Guo was a principal or co-investigator on research studies sponsored by Novartis, AstraZeneca, Bristol-Myers Squibb, Janssen Ortho-McNeil, Roche-Genentech, and Eli Lilly. None of them involved in this manuscript development. Dr. Lund reports that her spouse is a full-time, paid employee of GlaxoSmithKline. Elisabetta Patorno was supported by a career development grant K08AG055670 from the National Institute on Aging and reports research funding from GSK and Boehringer-Ingelheim, outside the submitted work. Dr. Slaughter has nothing to disclose. Dr. Wen has nothing to disclose. Dr. Bennett is an employee of Takeda Pharmaceuticals International Co.

Funding:

The authors declare that there was no financial support provided for this manuscript.

Keywords:

Real world evidence, confounding, decision making, observational

Disclaimer:

The views and recommendations expressed in this article are those of the authors and the Special Interest Group on Comparative Effectiveness Research and do not necessarily reflect those of the International Society for Pharmacoepidemiology (ISPE) and this article is not an ISPE document.

Introduction

On December 8, 2016, the *NEJM* published a sounding board on Real World Evidence (RWE) [1] by the US FDA leadership. While the value of RWE based on non-randomized observational studies was appreciated, such as for hypothesis generating, safety and measuring quality in healthcare delivery, the authors expressed concerns on the quality of data sources and the ability of methodologies to control for confounding. In response, we offer a few considerations regarding these concerns.

What are the challenges and the limitations of non-randomized observational studies? Can the advancement in the current methodology overcome these obstacles?

Currently, our evidence-based decision making largely focuses on randomized controlled trial (RCT) based 'efficacy' studies rather than 'effectiveness' studies. Non-randomized observational studies make up the bulk of evidence in 'real world'. Although, different types of RWE studies exist, the issues that we tend to address in this commentary are about observational studies, primarily the use of healthcare longitudinal databases with health insurance claims or electronic medical records (EMRs).[2] These two types of databases are the most commonly used data sources to generate real world evidence when evaluating the use of medications and their effectiveness and safety in clinical practice or studying unmet need/disease burden.

Confounding, selection bias and information bias constitute a major threat to the validity of non-randomized observational studies. However, with appropriate newer study designs and advanced statistical analysis, such as appropriate use of propensity score methods, confounding or biases can often be eliminated or largely reduced. Incomplete data on key outcome measures, pre-existing medical conditions/complications, or laboratory values from existing data sources can lead to skepticism on the quality and value of evidence based on those databases. Yet, validation

efforts, quality assurance programs, and sensitivity analysis can help improve or ensure the robustness of findings.

The past decade witnessed some breakthroughs in methodological advancement for non-randomized observational studies. In some cases, well-done analysis with complex methodology of real world evidence could sometimes substitute or replicate the findings from RCT[2]. Uptake of the new user and active comparator study designs[3] have also gone a long way in mitigating confounding, as well as other forms of bias common in observational studies. The use of multiple imputation with chained equations in combination with other analytic techniques have improved the way we deal with incomplete or missing data on important confounding variables.[4,5] With the application of propensity score methodology and analytic techniques specifically targeting big data, including machine learning, we are becoming much closer to the quality of evidence expected in the context of RCTs. Propensity score methodology can address measured confounding factors well, even in the setting of rare outcomes, whereas conventional RCTs may be low of power.[6] In the context of large datasets, predictive models with machine learning techniques can guide in the appropriate selection of variables for the estimation of propensity scores. In certain settings, high-dimensional propensity score methodology can generate valid effect estimates when benchmarked against results expected from randomized trials.[7] Sensitivity analyses can help researchers understand the robustness of results, just as such analyses play a role in RCTs. The choice of the study design and the selection of the data source should be directed by the scientific question of interest, and should consider whether a RCT is feasible and/or appropriate. In the context of a non-interventional observational study, researchers should consider a feasibility assessment to determine if a selected data source

contains sufficient and reliable information on outcomes and important confounding variables to answer the specific research question of interest.

As our data collection systems evolve, the completeness and data quality of large longitudinal health insurance claims and EMRs databases are being enhanced on an ongoing basis. With rapidly evolving technology supporting linkage across multiple sources of information, e.g., claims, registries, EMRs and biobanks, ability to control for residual confounding are rapidly improving. Moreover, other available methods such as instrumental variable analysis may help in the context of unmeasured confounding if used appropriately.[8]

Can non-randomized observational studies generate high-level quality evidence for decision making? Does evidence quality rely on randomized control trials alone?

Well-designed and controlled RCTs are recognized as the 'gold standard' to assess medication efficacy, whereas non-randomized observational studies are often proposed as a substitute when RCTs are not available or not feasible. Blinding is an important methodologic feature of RCTs to minimize bias and maximize validity of results, and can have an important impact on outcome ascertainment and post-randomization decisions. Typically, observational studies are not blinded but certain aspects (outcome assessment or adjudication) can be blinded to reduce bias. However, the bias from lack of blinding is less pronounced for objective outcomes. The concerns regarding the quality of evidence from non-randomized observational studies are largely attributable to the 'non-randomized' design; yet methodological advances can often address confounding. More attention <rather than on randomization and blinding> is warranted on limiting potential selection bias and on assessing whether the population and outcomes can be reliably defined for the specific scientific question in the selected data source.

Many systematic reviews have compared the evidence obtained from RCTs versus non-randomized studies, and shown little evidence for meaningful differences in effect estimates between non-randomized studies and RCTs when the same research question was being addressed in the same population, regardless of specific observational study design, heterogeneity, or inclusion of studies of pharmacological interventions.[9] Evaluation of several key questions could serve to inform when and how the findings from non-randomized observational studies could substitute for that of RCTs.[2]

Evidence from non-randomized observational studies and RCTs are complementary and additive to the overall body of evidence. Both types of studies should be critically appraised and used for decision making. The availability of RCTs should not make non-randomized observational studies unnecessary, and vice versa. If both types of studies are available and answer the same question, they should all contribute to the body of evidence. Existing guidelines on the evaluation of the quality of evidence based on non-randomized observational studies can assist in this regard. While RCTs will continue to generate high-quality evidence that will determine and guide our treatments, large non-randomized observational studies are also powerful resources, especially in the area of comparative effectiveness research and can help answer appropriate scientific questions.[10]

The following key guidelines can be used to evaluate the quality of published non-randomized observational studies where health insurance claims and EMR databases are used:

- A checklist For Retrospective Database Studies – Report of The ISPOR Task Force On Retrospective Databases (can be found here: https://www.ispor.org/workpaper/healthscience/FinalReportRetroR.pdf)

- Good ReseArch for Comparative Effectiveness (GRACE) Checklist and Principles (can be found here: https://www.graceprinciples.org/grace.html)
- International Society for Pharmacoepidemiology (ISPE) Guidelines for Good Pharmacoepidemiology Practices (GPP) (can be found here: https://www.pharmacoepi.org/resources/policies/guidelines-08027/)
- STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) (can be found here: https://www.strobe-statement.org/index.php?id=strobe-home)
- The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement (can be found here: http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001885)
- International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Good Practices for Outcomes Research Reports (can be found here: https://www.ispor.org/research_initiatives/hs_initiatives.asp)
- Methodological Standards for Patient-Centered Outcomes Research (PCOR) by the Patient Centered Outcomes Research Institute (PCORI) (can be found here: https://www.pcori.org/research-results/about-our-research/research-methodology/pcori-methodology-standards)
- Guide on Methodological Standards in Pharmacoepidemiology by the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) (can be found here: http://www.encepp.eu/standards_and_guidances/methodologicalGuide.shtml)

Most recently, the ISPOR and the ISPE created the joint Task Force to make recommendations on non-randomized observational studies regarding 1) transparency in the reporting of study methodology, 2) reproducibility and replicability of findings. These documents highlight the

importance of replicability across different data sources and different methods, and the importance of providing sufficient detail to allow such replication. Transparency and replicability would substantially increase our confidence on the validity of observational studies and can offer a trustworthy base for the expanded use of RWE in healthcare decision making.

What is our recommendation regarding the use of RWE for decision making?

The current paradigm of evidence hierarchy in evidence-based decision making needs to be reviewed. We believe the validity of evidence may not necessarily solely depend on whether it is generated by 'randomized' or 'non-randomized' studies. Well-designed, well-executed and well-reported non-randomized observational studies are a critical component of the overall body of evidence. A framework from the FDA with scientific considerations for study design, population identification, outcome ascertainment, reducing potential biases, and analysis could be helpful in encouraging researchers to apply best practices. The use of RWE based on such studies should be considered more broadly in various decision making settings, including clinical practice, Health Technology Assessment (HTA), regulatory and post-marketing lifecycle management. In an era of pay-for-value and adaptive pathway in formulary and reimbursement decision making, we recommend a paradigm shift from largely relying on idealized RCT evidence, which may fail to reflect the true value of the drug or medical device to more realistic RWE supported by high-quality data and robust methodology where the strengths and limitations of the generated data can be acknowledged. Another important future step would be for academic institutions to develop, and funding agencies to support development of high quality healthcare databases, and rigorous educational training and fellowship programs to foster next-generation researchers and leaders. Educational efforts should also be extended to other healthcare stakeholders, including

clinicians, policy decisions makers and industry, to guide the appropriate understanding and use of RWE.

We are encouraged by the FDA's effort to enhance the use of RWE in regulatory decision making, e.g. the recent draft guidance on RWE for medical devices. As RWE can be critical in the evaluation of both safety and effectiveness of medications, we appreciate that FDA is taking further steps moving in this direction by developing a framework and draft guidance (no later than the end of 2021) on how RWE can contribute to the assessment of safety and effectiveness in regulatory assessment, such as in the approval of new indications.

Acknowledgments:

We would like to thank Dr. Sebastian Schneeweiss for reviewing the article and for his constructive comments.

Author contributions

HY and DB wrote the first manuscript draft. All authors critically reviewed, revised and approved the final manuscript.

References:

1. Sherman RE, Anderson SA, Dal Pan GJ et al. Real-World Evidence — What Is It and What Can It Tell Us? N Engl J Med. 2016 Dec 8;375(23):2293-2297.
2. Franklin JM, Schneeweiss S. When and how can real world data analyses substitute for randomized controlled trials? Clin Pharmacol Ther. 2017 Dec;102(6):924-933.
3. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. Am J Epidemiol. 2003 Nov 1;158(9):915-20.
4. Rubin DB. Nested multiple imputation of NMES via partially incompatible MCMC. Statistica Neerlandica 2003; 57(1): 3-18.
5. van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. Stat Methods Med Res. 2007 Jun;16(3):219-42.
6. Franklin JM, Eddings W, Austin PC, Stuart EA, and Schneeweiss S. Comparing the performance of propensity score methods in healthcare database studies with rare outcomes. Stat Med. 2017 May 30;36(12):1946-1963.
7. Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, and Brookhart MA. Epidemiology. 2009 July ; 20(4): 512–522.
8. Myers JA, Rassen JA, Gagne JJ et al. Effects of adjusting for instrumental variables on bias and precision of effect estimates. American journal of epidemiology 2011; 174(11):1213–1222.,.
9. Anglemyer A, Horvath HT, Bero L. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. Cochrane Database Syst Rev. 2014 Apr 29;(4):MR000034.

10. Morton JB, McConeghy R, Heinrich K, Gatto NM, Caffrey AR. Consensus of recommendations guiding comparative effectiveness research methods. Pharmacoepidemiology and Drug Safety 2016; DOI: 10.1002/pds/4051.