HHS Public Access

Author manuscript

Clin Pharmacol Ther. Author manuscript; available in PMC 2023 January 01.

Published in final edited form as:

Clin Pharmacol Ther. 2022 January ; 111(1): 145–149. doi:10.1002/cpt.2398.

Assessing and Interpreting Real-World Evidence Studies: Introductory Points for New Reviewers

Shirley V. Wang[1,*], **Sebastian Schneeweiss**[1]

[1]Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA

Abstract

Nonrandomized real-world evidence (RWE) studies are conducted using healthcare data collected as part of clinical practice. As RWE studies are increasingly considered for regulatory, coverage, and other clinical decision making, nonspecialists will find themselves in the position of assessing the validity of RWE studies, a field that may be less familiar to them. This introductory guide provides conceptual guidance for reviewing RWE studies and is particularly directed at professionals for whom this is new or whose prior experience has primarily been in reviewing randomized controlled trial evidence. We focus on RWE studies that make causal inference, evaluating whether one treatment option is better, worse, or neutral compared to another. Although we provide citations to direct the reader to resources with more details on complex issues, this guide cannot substitute for years of training and expertise in the field.

Where to begin?

Imagine that you are given a real-world evidence (RWE) study and told to evaluate whether the quality of the evidence is high enough that it should be used to inform a clinical decision or policy for your organization. You need to make a determination of whether the evidence is fit-for-purpose. What should you think about?

The literature is full of articles and books on causal inference in medical research, many of which focus on randomized experiments, some concentrate on noninterventional research, but few highlight the similarities and differences that would help a reviewer who is comfortable with randomized clinical trials (RCTs) adapt to confidently assess the validity of RWE. Concato and colleagues from the US Food and Drug Administration (FDA) discussed the false dichotomy between trials and RWE, highlighting variations in design features within trials, including randomization, primary vs. secondary data collection mechanisms, and use of external control groups.[1] Hernan and Robins reminded us in a few noteworthy papers that designing and analyzing RWE studies, like a hypothetical randomized trial that could have been done, the target trial, would provide clarity on the exact study question being asked, and naturally lead to improved design and analytic choices.[2,3] This framework includes consideration of the eligibility criteria, treatment strategy, assignment procedures, follow-up window, outcome, causal contrast of interest,

*Correspondence: Shirley V. Wang (Swang1@bwh.harvard.edu).

and analysis plan for the target trial to address the question of interest. This target trial framework can be equally useful for those who review RWE studies as for those who conduct RWE studies.

Consider the triad of question, design, and data

There are three major components of a RWE study that determine whether the findings are decision-relevant. First, the research question must align with the question that you or your organization is trying to address. Second, the study design must use methods that are appropriate for validly addressing the question using that data source. Third, the data must be suitable to address the question. Only if all three align, can the evidence be considered fit-for-purpose.

QUESTION

One of the first things to consider is the precise research question being evaluated in the RWE study. Breaking down the question according to the Population, Intervention, Comparator, Outcome, and Timing (PICOT) framework[4] can help you assess in a first pass how relevant the question addressed in the research study is to the question that you or your organization are trying to address.

For example, your organization needs to make a policy decision based on the relative benefit of drug X compared to drug Y. Upon breaking down the PICOT components of the research question, you recognize that the research study was focused on patients aged 40–65 years, whereas the patient population relevant for your policy decision is 65+ years (Table 1). You realize that the research study is comparing drug X to drug Z rather than drug Y of the same class, and that whereas you are interested in understanding the effects of the drugs during treatment, the RWE study only provided results after following patients for a fixed window, regardless of whether they discontinued therapy, as frequently observed in clinical practice (i.e., an as-started analysis similar to an intention-to-treat analysis in a trial with random treatment assignment). Depending on the context, you may consider drug Z close enough to drug Y for the study to be informative. You may or may not be willing to extrapolate the effects of the drug across age groups. Because you are interested in effects of drug X while patients are on treatment rather than understanding effectiveness in the context of real-world nonadherence, you may consider the fixed follow-up window used in the RWE study relevant for informing your decision if the fixed window used is close to the average duration of treatment, but not if the window extends far beyond the point when most patients have discontinued therapy.

Similar to assessment of trial evidence, breaking down the research question into component pieces and relating these to the specific clinical or policy question you need to resolve can help you quickly determine whether a given RWE study has the potential to be fit for your purpose or not. Any given RWE study may be fit for some purposes and not for others.

In a second pass, while studying the methods section of the RWE study more closely, one may go further and attempt to reconstruct the (hypothetical) randomized trial that the investigators emulate with their RWE study. Such a reverse-engineered RCT may appear

unusual and give rise to specific concerns in the RWE study that can now be more precisely articulated. This activity often allows the reviewer to more precisely pinpoint what question the RWE study answers. Because RWE studies are by definition based on treatment choices and data that are collected from clinical practice, one may quickly realize that the reverse-engineered hypothetical RCT is what is often referred to as a pragmatic RCT.[5] Highlighting the differences between a trial conducted in a highly controlled research environment with features like run-in phase, blinding of treatment and outcome assessment, adherence reminders, etc. vs. a trial with more pragmatic attributes will sharpen the reviewer's eye on the precise study question and possibly design-related biases.

DESIGN

There are many types of study designs for trials, for example, crossover trials, parallel group trials, and cluster trials.[6] Similarly, there are a variety of study designs that can be used to conduct RWE studies.[7,8] Each has properties that make them better suited for addressing certain types of research questions.[9] An overarching goal of study design for RWE studies is to minimize potential bias from nonrandomized treatment assignment and ascertainment of outcomes based on documentation from clinical practice rather than a protocol-based standard assessment. Randomization and protocol-based outcome measurement are two key features of RCTs that have made them the gold standard for evidence on treatment effects.[10] However, there are well known limitations of trials, namely the highly selected participant populations, tightly controlled conditions under which they are conducted, high cost and ethical considerations, as well as lack of power to detect rare but serious adverse effects. Because no single source of evidence is perfect, a fuller picture of the effects of drugs can be obtained with the appropriate use of evidence from RWE studies to complement RCT findings.

Appropriate interpretation and evaluation of RWE study design will include assessment of three inter-related sources of systematic bias[8]: selection bias, information bias, and confounding (Table 2).

Contemplating the re-engineered target trial helps not only with clarifying the studied research question (above) but major sources of bias in nonrandomized RWE studies can be addressed by designing them to mimic target trials.[2,11] This framework can be used to clarify thinking around many alternative nonrandomized study designs, for example, parallel arm trials emulated by cohort or cohort sampling studies, such as nested case-control designs, or crossover trials emulated by self-controlled studies.[12] The framework is also quite illuminating to uncover major design-related biases, like selection bias, information bias, confounding, and other specific time-related sources of bias, like immortal time, reverse causation, inadequate capture of latency, or misclassification of the exposure effect window, depletion of the susceptible, or immeasurable time.[13–15] In short, using the target trial framework can help with planning and design of RWE studies; conversely, re-engineering the hypothetical trial based on the design choices of an RWE study can help identify problematic decisions that result in biases.

Evaluation of an RWE study within the target trial framework will center on time 0, which parallels the point of randomization in an RCT. The timing of measurement of inclusion-exclusion criteria, exposure, outcome, follow-up, and covariates can all be indexed against this temporal anchor. Many biases related to inappropriate handling of person-time in RWE studies can be mitigated with thoughtful definition of time 0 (e.g., new user vs. nonuser or prevalent user designs).[3,16,17] To summarize and assess the appropriateness of the temporality in the study design, if a design diagram is not provided by the authors, making a diagram can help the reader more effectively assess and interpret the study.[18]

Similar to the direct comparison of PICOT policy question and RWE study question in Table 1, the components of the target trial can be laid out for side-by-side comparison with RWE study parameters. Mapping out how key study parameters are measured and when they are measured relative to time 0 can help the reader identify misalignment in scientific choices compared to the target trial. For example, a side-by-side comparison of a target trial's intended parameter vs. the actual RWD implementation could help highlight that while the population of interest is patients with type 2 diabetes, the RWE study inclusion criterion used a broad set of diabetes-related codes that included gestational diabetes and type 1 diabetes. Alternatively, the contrast could highlight that while in the target trial, follow-up for exposed outcomes would begin after initiation of therapy, the RWE study under review started follow-up after discharge from a hospital, with exposure status assigned based on future dispensation of drug (causing immortal time bias).

DATA

The third cornerstone of assessing whether an RWE study is able to answer the research question is consideration of the data source(s) being used. In many settings. this may be the most difficult task. as much of the information on how the data were exactly generated and recorded in clinical practice remains hidden from the reviewer.[19] Two key areas of assessing the appropriateness of electronic healthcare research databases are reliability and relevance.[20,21]

Reliability

Evaluating data source reliability targets the question "does the data adequately capture the intended concepts?"[22] At heart, it is about the completeness and accuracy of measurement of clinical concepts relevant for research studies. Evaluation of database reliability is broader than validation studies to evaluate the performance of algorithms used to measure specific study parameters. It can include consideration of many aspects of data preparation, such as how the data were collected, data cleaning, and quality control processes used to create the research database. Database reliability can be evaluated by considering whether data elements match expectation (plausibility, e.g., observed age-sex distribution matches expected age-sex distribution in the covered population); the completeness of data capture and reasons for missingness (e.g., X proportion missing because laboratory results are available only for tests processed by one national vendor); as well as assessing the logical consistency of data transformations when moving from raw to more processed data fields (e.g., body mass index is a derived variable based on a function of height and weight).

Ideally, information about database reliability will be made available by the researchers or organizations creating and maintaining the research database.

Relevance

Evaluating the relevance of a data source involves assessing whether the data elements available in the research database(s) are sufficient to address the study question.[20] Relevant data sources include data on the population of interest over the relevant time frame, have sufficient persons and follow-up time, and have information that captures key study parameters, such as inclusion criteria, exposures, outcomes, and baseline characteristics.

For example, if the research question is about the effectiveness of a drug dispensed in the inpatient setting, a research database comprised solely of insurance claims might not be able to capture details of exposure as medication use during a hospitalization are not individually billed. A research database comprised of inpatient electronic health records (EHRs) would be able to capture inpatient exposure but would have incomplete capture of pre-hospitalization covariates and post-hospitalization outcomes. Depending on context, a linked EHR-claims data source or data from an integrated healthcare delivery system might be better able to capture key study parameters for this type of question.[23]

It may be worth noting that most secondary data sources will not perfectly measure all parameters of interest because the investigator is not deciding what to measure, nor how and when to measure it. However, reasonably close proxy measures can result in findings similar to studies with primary data collection.[24,25]

DISCUSSION

In this paper, we aimed to provide a digestible, high-level introductory points on assessing and interpreting RWE studies, designed for people who are comfortable with reviewing RCTs but are new to evaluating RWE studies (Table 3). The focus on the triad of question, design, and data was deliberate because most review issues that will compromise the utility of an RWE study will be found among them. The analytic strategy of an RWE study is largely similar to that of RCTs except for the statistical approach to deal with the lack of baseline randomization. This, in turn, is largely a problem of data reliability and capture of relevant preexposure patient characteristics.

Currently, the clarity of reporting on the triad of question, design, and data is variable for RWE studies,[26–28] however, there are recent and ongoing efforts to improve documentation and reporting on the conduct of RWE studies,[29–33] which will facilitate interpretation by the reviewer. We hope that this guide provides a useful introduction of things to look for when evaluating RWE studies for regulatory, coverage, clinical, or other decision making.

FUNDING

The authors were supported by funding from the National Institutes of Health (NIH): NHLBI R01HL141505 and NIA R01AG053302. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

CONFLICTS OF INTEREST

S.V.W. has no conflicts of interest to disclose, S.S. is principal investigator of the FDA Sentinel Innovation Center funded by the FDA, co-principal investigator of an investigator-initiated grant to the Brigham and Women's Hospital from Boehringer Ingelheim unrelated to the topic of this study, He is a consultant to Aetion Inc., a software manufacturer of which he owns equity, His interests were declared, reviewed, and approved by the Brigham and Women's Hospital and Partners HealthCare System in accordance with their institutional compliance policies.

References

1. Concato J, Stein P, Dal Pan GJ, Ball R, & Corrigan-Curay J Randomized, observational, interventional, and real-world-What's in a name? Pharmacoepidemiol. Drug Saf 29, 1514–1517 (2020). [PubMed: 32940401]
2. Hernán MA & Robins JM Using big data to emulate a target trial when a randomized trial is not available. Am. J. Epidemiol 183, 758–764 (2016). [PubMed: 26994063]
3. Hernán MA, Sauer BC, Hernández-Díaz S, Platt R & Shrier I Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. J. Clin. Epidemiol 79, 70–75 (2016). [PubMed: 27237061]
4. Straus SE Evidence-based medicine: how to practice and teach EBM 3rd edn, Elsevier/Churchill Livingstone (2005).
5. Thorpe KE et al. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers, J. Clin. Epidemiol 62, 464–475 (2009). [PubMed: 19348971]
6. Friedman LM, Furberg C & DeMets DL Fundamentals of Clinical Trials. Springer Mosby, Inc., New York, NY 1996).
7. Gagne JN et al. Taxonomy for monitoring methods within a medical product safety surveillance system: year two report of the Mini-Sentinel Taxonomy Project Workgroup, Accessed August 28, 2018, <https://www.sentinelinitiative.org/sites/default/files/Methods/Mini-Sentinel_Methods_Taxonomy-Year-2-Report.pdf>.
8. Schneeweiss SHK Pharmacoepidemiology, In Modern Pharmacoepidemiology, 4th edn (eds, Rothman KLT, Greenland S, and Vanderweele T) (Wolters Kluwer, New York, NY 2021).
9. Schneeweiss S & Patorno E Conducting real-world evidence studies on the clinical outcomes of diabetes treatments. Endocr. Rev 10.1210/endrev/bnab007, [e-pub ahead of print].
10. Franklin JM & Schneeweiss S When and how can real world data analyses substitute for randomized controlled trials? Clin. Pharmacol. Ther 102, 924–933 (2017). [PubMed: 28836267]
11. Hernán MA et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease, Epidemiology 19, 766–779 (2008). [PubMed: 18854702]
12. Schneeweiss S A basic study design for expedited safety signal evaluation based on electronic healthcare data. Pharmacoepidemiol. Drug Saf 19, 858–868 (2010). [PubMed: 20681003]
13. Suissa S Immortal time bias in pharmaco-epidemiology. Am. J. Epidemiol 167, 492–499 (2008). [PubMed: 18056625]
14. Suissa S Immeasurable time bias in observational studies of drug effects on mortality. Am. J. Epidemiol 168, 329–335 (2008). [PubMed: 18515793]
15. Suissa S & Dell'Aniello S Time-related biases in pharmacoepidemiology. Pharmacoepidemiol. Drug Saf 29, 1101–1110 (2020). [PubMed: 32783283]
16. Ray WA Evaluating medication effects outside of clinical trials: new-user designs. Am. J. Epidemiol 158, 915–920 (2003). [PubMed: 14585769]
17. Renoux C, Azoulay L & Suissa S Biases in evaluating the safety and effectiveness of drugs for the treatment of COVID-19: designing real-world evidence studies. Am. J. Epidemiol 190, 1452–1456 (2021). [PubMed: 33564823]
18. Schneeweiss S et al. Graphical depiction of longitudinal study designs in health care databases. Ann. Intern. Med 170, 398–406 (2019). [PubMed: 30856654]
19. Schneeweiss S & Avorn J A review of uses of health care utilization databases for epidemiologic research on therapeutics. J. Clin. Epidemiol 58(4), 323–337 (2005). [PubMed: 15862718]

20. Characterizing RWD quality and relevancy for regulatory purposes. October 1, 2018. <https://healthpolicy.duke.edu/sites/default/files/2020-03/characterizing_rwd.pdf>.
21. U.S. Food & Drug Administration. Framework for FDA's real world evidence program <https://www.fda.gov/downloads/ScienceResearch/SpecialTopics/RealWorldEvidence/UCM627769.pdf> (2018). Accessed January 31, 2019.
22. Duke Margolis Health Policy. Determining real-world data's fitness for use and the role of reliability <https://healthpolicy.duke.edu/sites/default/files/2019-11/rwd_reliability.pdf>. Accessed March 30, 2021.
23. Lin KJ & Schneeweiss S Considerations for the analysis of longitudinal electronic health records linked to claims data to study the effectiveness and safety of drugs. Clin. Pharmacol. Ther 100, 147–159 (2016). [PubMed: 26916672]
24. Patorno E, Schneeweiss S, Gopalakrishnan C, Martin D & Franklin JM Using real-world data to predict findings of an ongoing phase IV cardiovascular outcome trial – cardiovascular safety of linagliptin vs. Glimepiride. Diabetes Care 42, 2204–2210 (2019). [PubMed: 31239281]
25. Franklin JM, Patorno E, Desai RJ, Glynn RJ, Martin D, Quinto K et al. Emulating randomized clinical trials with nonrandomized real-world evidence studies. Circulation 143, 1002–1013 (2021). [PubMed: 33327727]
26. Reproducible Evidence: Practices to Enhance and Achieve Transparency (REPEAT) initiative. Accessed 5/21/2018, <www.repeatinitiative.org/>.
27. Wang S, Verpillat P, Rassen J, Patrick A, Garry E & Bartels D Transparency and reproducibility of observational cohort studies using large healthcare databases. Clin. Pharmacol. Ther 99, 325–332 (2016). [PubMed: 26690726]
28. Benchimol EI, Moher D, Ehrenstein V & Langan SM Retraction of COVID-19 pharmacoepidemiology research could have been avoided by effective use of reporting guidelines. Clin. Epidemiol 12, 1403–1420 (2020). [PubMed: 33376409]
29. Orsini LS et al. Improving transparency to build trust in real-world secondary data studies for hypothesis testing—why, what, and how: recommendations and a road map from the real-world evidence transparency initiative. Value Health 23, 1128–1136 (2020). [PubMed: 32940229]
30. Wang SV et al. STaRT-RWE: structured template for planning and reporting on the implementation of real world evidence studies. BMJ 372, m4856 (2021). [PubMed: 33436424]
31. Wang SV et al. Reporting to improve reproducibility and facilitate validity assessment for healthcare database studies V1.0. Pharmacoepidemiol. Drug Saf 26, 1018–1032 (2017). [PubMed: 28913963]
32. Langan SM et al. The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE). BMJ 363, k3532 (2018). [PubMed: 30429167]
33. Berger ML et al. Good practices for real-world data studies of treatment and/or comparative effectiveness: recommendations from the Joint ISPOR-ISPE Special Task Force on real-world evidence in health care decision making. Value Health 20, 1003–1008 (2017). [PubMed: 28964430]

Table 1

Breaking down the PICOT question can highlight mismatches between the question of interest and question being asked

	Your clinical or policy question	RWE study question
Population	Patients 65+ with type 2 diabetes	Patients 40–65 with diagnosis of type 2 diabetes measured in the 180 days prior to and including the date of treatment initiation
Intervention	Drug X	Drug X
Comparator	Drug Y	Drug Z
Outcome	Hospitalization for heart failure	Hospitalization with heart failure diagnosis as primary discharge diagnosis
Timing	On-treatment follow-up Starts: Day after treatment initiation Stops: End of treatment, switching therapies, death or disenrollment	Intention-to-treat follow-up Starts: Day after treatment initiation Stops: Death, disenrollment, 730 days after initiation

PICOT, Population, Intervention, Comparator, Outcome, and Timing; RWE, real-world evidence.

Table 2

Three major sources of systematic bias in RWE studies

Bias	Description	Example
Selection bias	Bias in an effect estimate arises when there are differences in how patients enter or exit the study population in ways that are related to both the exposure and outcome.	1. The study selects initiators of drug X and prevalent users of drug Y as comparison groups. Prevalent users are patients who have "survived" without developing the outcome. This imbalance in susceptibility to develop the outcome will be reflected in the effect estimate. 2. About 40% of patients exposed to drug X and drug Y are lost to follow-up. Twice as many outcomes are missed due to censoring in drug X exposed patients than drug Y.
Information (measurement) bias	Bias in an effect estimate arises when there is incorrect classification of a key study parameter such as exposure or outcome.	The code algorithm used to measure an outcome of myocardial infarction has a PPV of 90%. This means that 10% of patients classified as having the outcome did not actually have the outcome. The PPV is the same across compared exposure groups, thus the estimated effect will be biased toward the null.
Confounding	Bias in an effect estimate arises when risk factors for the outcome are imbalanced between compared groups.	Patients who initiate treatment with drug X are more likely to be over 65, women, and of low socioeconomic status than patients who initiate treatment with drug Y. These 3 factors are risk factors for the outcome, thus the apparent risks of drug X will be inflated unless the imbalance in these baseline confounding factors are addressed in design or analysis.

PPV, positive predictive value; RWE, real-world evidence.

Table 3

Summary of introductory points on assessing and interpreting RWE studies

Question	Compare the components of the PICOT question that you want to address to the research question that is actually addressed by the study. Re-engineering the hypothetical trial that was emulated by the RWE study is a helpful thought experiment to identify the precise study question.
Design	Map out a hypothetical target trial that would address your PICOT question of interest and compare to the RWE study parameters implemented to help think about potential sources of design-related biases. Consider creating a design diagram to be clearly summarize time 0 in the RWE study (parallel to randomization date for a trial) and other temporal anchors.
Data	
Reliability	Consider how closely the elements of the database reflect the clinical constructs they are supposed to.
Relevance	Consider whether and how well key study parameters are measurable in the data.

PICOT, Population, Intervention, Comparator, Outcome, and Timing; RWE, real-world evidence.

Clin Pharmacol Ther. Author manuscript; available in PMC 2023 January 01.