REAL WORLD DATA: ASSESSING ELECTRONIC HEALTH RECORDS AND MEDICAL CLAIMS DATA TO SUPPORT REGULATORY DECISION-MAKING FOR DRUG AND BIOLOGICAL PRODUCTS

A Commentary from ISPE

Cynthia J Girman, DrPH, FISPE,[1] Mary E Ritchey, PhD, FISPE,[1-3] Vincent Lo Re, III, MD, MSCE, FISPE[4]

[1]Real World Evidence and Patient Outcomes, CERobs Consulting, LLC, Wrightsville Beach, NC, USA

[2]Med Tech Epi, LLC, Philadelphia, PA USA

[3]Center for Pharmacoepidemiology & Treatment Science, Rutgers University, New Brunswick, NJ USA

[4]Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA USA

For submission to*: Pharmacoepidemiology & Drug Safety as a Commentary*

This Commentary is endorsed by the International Society for Pharmacoepidemiology (ISPE).

Corresponding author:

Cynthia J Girman, DrPH, FISPE

CERobs Consulting, LLC

2612 Lumina Ave North

Wrightsville Beach, NC 28480

Email: cindy.girman@cerobs.com

Mobile: +1 919 606-8723

Word Count: not counting tables and text boxes: 1497 (max 1500); references: 15 (max 15)

1. Introduction

On September 29, 2021, the U.S. Food and Drug Administration (FDA) released a draft guidance for industry on assessing electronic health records (EHR) and medical claims to support regulatory decisions for drug and biological products, culminating years of discussion, stakeholder feedback, and workshops in professional society forums.[2] This draft guidance focuses on the adequacy and relevance of real world data (RWD) sources for specific research questions and whether follow-up (i.e., person-time) is sufficient to ascertain outcomes involving regulatory decisions on effectiveness or safety. In addition, the guidance addresses whether study elements of a research question can be characterized with appropriate data quality and accuracy, and if missing data limit interpretation of study results.[2,4-6] Notably, the recommendations are not limited to new product indications or labeling changes. The draft guidance does not review choice of study design or analytic techniques, including handling of confounding, but these will be addressed in forthcoming documents. While this draft guidance moves closer to use of RWD for regulatory decisions, it incorporates inefficiencies from randomized clinical trials (RCTs) into studies using RWD and appears to be broadly applicable to studies in EHR and claims without considering important differences between the types of data sources.

2. Adequacy and Relevance of the Data Source

> **Box 1. Conceptual and Operational Definitions**
>
> **Conceptual definition** of study elements (PICO) is the high-level clinical criteria to define a condition, exposure, or outcome. This corresponds to how the study element might be worded in a well-articulated study question.
>
> **Operational definition** based on the conceptual definition is the criteria for defining a study variable within the data source and extracting the data, along with any plan for validation. This corresponds to how the data elements might be detailed in a protocol in the Exposures, Outcomes, or Variables sections.

As with any scientific research study, pre-specification of objectives, hypotheses, study design, and statistical analysis plans (SAP) lends more confidence to the findings, especially if protocols are registered.[7] Pre-specified documentation should include conceptual and operational definitions (**Box 1**) of all elements of the study question, using the PICO framework (-Population with inclusion and exclusion criteria, Intervention (treatment), Comparator, Outcomes, and important covariates) as previously suggested.[5-6] Data source adequacy is contextual, with context given by the research question and regulatory decision (e.g., approval, labeling change, new indication).[3-5,11] Sensitivity analyses to assess the robustness of findings should also be pre-specified,[8] and quantitative bias analysis (QBA) should be considered based on performance of operational definitions in validation studies and bias parameters.[9-10]

The draft RWD guidance appropriately highlights critical information not in EHR or claims data that pharmacoepidemiologists routinely consider when assessing data sources for specific research questions, such as relevant formulary restrictions, tiering or stepped therapy, and prior authorizations that limit medication use.[2] Variation in medical practices, diagnostic criteria, and treatment patterns across health care systems are also important considerations. Because U.S. residents often switch medical insurance with change in employer, continuity of care is critical to allow sufficient longitudinal follow-up (person-time) and capture of study outcomes. The time periods for PICO elements (pre-treatment period for baseline information, washout for therapy initiation,[12] and follow-up) should be clearly delineated in the protocol, avoiding immortal time bias[13] and accounting for temporal changes during the study that could impact findings.

To supplement EHR or claims data when key data elements are missing or poorly measured, the FDA guidance recommends linkage to other data sources. This is an important consideration, though may not always be an option due to the lack of linking variables, privacy issues, and legal considerations for sharing data. Even with linkage, the resulting cohort can be limited due to loss of adequate patient sample. The impact of the missing data on interpretation of findings might be addressed by sensitivity analyses, E-values and QBA.[9-10] The FDA guidance acknowledges that information for some data elements may be contained in unstructured fields within EHR data (e.g., clinical notes), and recommends thoroughly describing the process for extracting such data, whether manually or using automated technology.

3. Characterization and Validation of Exposures, Outcomes, and Important Covariates

The guidance indicates that the medication exposure definition should include dose, formulation, strength, route of administration, timing, frequency, duration studied (if relevant), and potentially manufacturer, some of which may not be available in EHR and claims databases. The degree that each is crucial to the *specific* research question should be considered.

Algorithms to identify exposures, outcomes, and important covariates should be operationally defined by data such as medical diagnoses, procedural codes, or dispensed medications. As with clinical assessments in RCTs, algorithms may not always be accurate, which could lead to misclassification. However, misclassification may or may not lead to biased estimates of treatment effect, and sensitivity analyses and QBA can inform on the potential impact of such biases.[8-10]

Throughout the guidance, FDA refers to complete (100%) verification of study variables (PICO and covariates) as the "most rigorous approach" to using RWD but acknowledges that it may not be feasible. The use of the term 'verification' may be conflated with 'validation.' Data *verification* typically involves confirming the correct data point for each patient (or subset), which is impossible for *all* variables in any RWD study. Such an approach is not expected for

RCTs (**Table**) nor is it performed by the FDA in the Sentinel System to evaluate safety signals associated with medical products. In contrast, *validation* assesses the accuracy of an operational definition compared to a reference standard to yield test characteristics such as positive and negative predictive value, sensitivity, and specificity, contributing to understanding of the degree of potential misclassification. If the reference standard is the medical record, the guidance appropriately recommends blinding data abstractors and adjudicators to medication exposure to reduce bias and using standardized and reproducible abstraction and adjudication processes to minimize intra- and inter-rater error. However, the guidance should clarify the appropriate validation approaches to consider when EHR data are used for the study and the same EHR *is* the medical record reference standard.

The guidance fails to distinguish between or prioritize critical variables in recommending validation. In RCTs, data are typically collected at each clinical site through direct clinical assessment of a patient or from medical records. While procedures are available to ensure the accuracy of these data, risk-based quality assurance procedures drive focus on more critical variables,[15] and 100% verification is rare. Achieving 100% accuracy (and no misclassification) takes extraordinary effort and expense and may never be achievable. Selecting only patients whose outcome can be 'validated' could result in a biased sample of patients with more complete records, decreasing generalizability. Validating each covariate in a study is inefficient, costly, and will yield little that might change the overall interpretation of the study. Instead, Sponsors should *justify* the adequacy of critical study elements (exposure, primary outcomes, and key confounders) and perform analyses to assess the robustness of results for varying degrees of potential misclassification (**Box 2**).

> **Box 2. Adequacy and Robustness of Critical Data Elements**
>
> The degree of validation for any PICO element or covariate should depend on whether there is a plausible biological explanation that if the variable were measured with greater accuracy, it would change the conclusions. Sponsors should be expected to justify the adequacy of critical study variables and perform sensitivity and quantitative bias analyses to demonstrate robustness of results to varying degrees of misclassification.

The FDA guidance suggests that the performance of algorithms or operational definitions for key variables should be demonstrated using sufficiently large samples, appropriate sampling techniques, and reasonable reference standards, but gives little guidance on how to define a 'sufficiently large' sample or the appropriate sampling methods. Appropriate reference standards are also not recommended.

Studies to validate health outcomes of interest can be performed, or a prior study or publication can be used if conducted in a "similar population or data source." The accuracy of case-identifying algorithms can vary across populations, healthcare settings, coding systems, or calendar time periods, and each of these factors can affect an algorithm's performance and transportability to other health databases. Thus, when considering applying algorithms in a

different setting or database, researchers should consider if the prevalence of the outcome might differ between the validation population and the new population, and whether restrictions on access to medications are comparable.

Validation may be performed in cases when only false-positives are of concern, and in cases and non-cases when false-negatives are also of concern. However, validating non-cases in very large studies of rare outcomes may be impractical; a random sample of non-cases may be more practical.[14]

Table: FDA expectations when using rea-world data to evaluate medical product effectiveness or safety compared to using randomized clinical trials for product approval.

FDA Recommendation/ Requirement	Recommendation in FDA Guidance for Effectiveness/Safety Studies in EHR/Claims	Requirements for RCTs for Product Approval
Pre-specified protocol and SAP	Yes	Yes
Pre-specified sensitivity and subgroup analyses	Yes	Yes
Definitions of outcomes	Yes	Yes
Verification of outcomes	"Most rigorous approach"	Dependent on study outcome, often not required
Validation of outcomes	Yes	For outcomes trials; adjudication may be used
Validation of variables to define study population	Yes	No
Validation of treatment definitions	Yes	Data collection and pill counts; crossover assessed
Validation of covariates	Yes	No
QA/QC at time of data collection	Yes, but may be impractical for Sponsors (and Data Providers) to implement	Yes
QA/QC at data checking/cleaning	Yes, but impractical for Sponsors to implement; documentation from data provider may not be obtainable	Yes, procedures documented
QA/QC at transformation to analytic file	Yes	Yes
Traceability/Auditable	Yes, but detailed documentation from data provider may not be obtainable	Yes

Abbreviations: FDA=US Food and Drug Administration; QA/QC=quality assurance/quality control; RCT=randomized controlled trial; SAP=statistical analysis plan

4. Missing Data

FDA appropriately expresses concerns about missing data for prescriptions (e.g., samples, out-of-pocket purchase) and outcomes (e.g., out-of-network care), noting that linkage to data sources to enhance capture of these elements may be useful. Additional sensitivity analyses, akin to those in RCTs for missing outcome or exposure data, should also be considered.

Demonstrating that the results are robust and yield the same conclusion provides reassurance that the missing data are not influential.

The guidance outlines the need to have documentation on all data quality assurance and quality control (QA/QC) procedures during data accrual, curation, and transformation to the final analytic dataset. While data provenance is important, it may be challenging for Sponsors to verify data against original source records at the point of care if they are using commercially licensed databases from database providers . Sponsors do not have the ability to either access point-of-care data or to change processes for commercially licensed claims or EHR databases. In general, the FDA draft guidance describes QA/QC for data accrual and curation similar requirements in primary data collection for RCTs conducted by Sponsors, procedures which significantly contribute to the inefficiency and expense of RCTs. Guidance for industry should focus on QA/QC that Sponsors can readily influence and document, such as those from the time of the receipt of the data. These should be driven by how potential misclassification and errors would affect the conclusions of the study, using the philosophy behind risk-based monitoring.[15]

Throughout the draft guidance, the FDA recommends that Sponsors discuss specific issues with the relevant review division. However, defining the process and timing whereby such discussion could take place is essential, since none currently exists.

5. Conclusion

The draft guidance outlining specific issues of concern to FDA is a welcome leap forward to carve a path for use of RWD to support regulatory decision-making. However, the guidance recommends practices that would significantly limit the ability to use RWD to generate reliable and robust results of effectiveness and safety for regulatory purposes. The guidance could be enhanced by implementing a risk-based approach to prioritize operational definition validation and QA/QC as well as advising on types of sensitivity analyses and QBA that would give FDA confidence in the robustness of results. These additions would better address the congressional charge to FDA to identify ways that RWD can be used to *more efficiently and more rapidly generate evidence on product effectiveness and safety*.

References

1. Energy and Commerce Committee. 21st Century Cures Act. Retrieved 01Oct2018. https://energycommerce.house.gov/cures

2. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER) and Center for Biologics Evaluation and Research (CBER) and Oncology Center for Excellence (OCE). Real-World Data: Assessing electronic health records and medical claims to support regulatory decision-making for drug and biological products: Guidance for Industry, Draft Guidance. Sept 2021. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/real-world-data-assessing-electronic-health-records-and-medical-claims-data-support-regulatory

3. https://www.sentinelinitiative.org. Accessed Oct 3, 2021.

4. Daniel G, Silcox C, Bryan J, McClellan M, Romine M, Frank K. Characterizing RWD quality and relevancy for regulatory purposes. Duke-Margolis Center for Health Policy, retrieved 01Oct2018. https://healthpolicy.duke.edu/sites/default/files/atoms/files/characterizing_rwd.pdf

5. Girman CJ, Ritchey ME, Zhou W, Dreyer NA. Considerations in characterizing real-world relevance and quality for regulatory purposes: A commentary. *Pharmacoepidemiol Drug Saf* 2018; 1-4. https://doi.org/10.1002/pdf.4697.

6. Ritchey ME, Girman CJ. Evaluating the feasibility of electronic health records and claims data sources for specific research purposes. *Ther Innov Reg Sci* 2020. https://doi.org/10.1007/s43441-020-00139-x.

7. Wang SV, Schneeweiss S, Berger ML et al. Reporting to improve reproducibility and facilitate validity assessment for healthcare database studies. v1.0. *Pharmacoepidemiol Drug Saf* 2017: 26(9): 1018-32.

8. Girman CJ, Faries D, Ryan P et al. Pre-study feasibility and identifying sensitivity analyses for protocol pre-specification in comparative effectiveness research. J *Compar Effect Res* 2014; 3(3): 259-270.

9. Lash, TL, Fox MP, Fink AK. Applying Quantitative Bias Analysis to Epidemiologic Data, New York (NY): Springer, 2009.

10. Jonsson-Funk M, Landi SN. Misclassification in administrative claims data: quantifying the impact on treatment effects. *Curr Epidemiol Rep* 2014; 1: 175-85.

11. Girman CJ. Ritchey ME, McNeill AM, Andersson Sundell K, Meyer RJ. Demonstrating Real-World Evidence is fit for purpose to support labeling: Parallels in pursuing labeling claims for patient reported outcomes. *Ther Innov Reg Sci* 2021 https://10.1007/s43441-020-00252-x

12. Wang SV, Pinheiro S, Hua W, Arlett P, Uyama Y, Berlin JA, Bartels DB, Kahler KH, Bessette LG, Schneeweiss S. STaRT-RWE: structured template for planning and reporting on the implementation of real world evidence studies. *BMJ* 2021; 372 doi: https://doi.org/10.1136/bmj.m4856

13. Suissa S. Effectiveness of inhaled corticosteroids in chronic obstructive pulmonary disease: immortal time bias in observational studies. *Am J Respir Crit Care Med* 2003; 168(1): 49-53.

14. Thurin NH, Boxco-Levy P, Blin P et al. Intra-database validation of case-identifying algorithms using reconstituted electronic health records from healthcare claims data. BMC Med Res Method 2021; 21: 95. https://doi.org/10.1186/s12874-021-01285-y

15. Hurley C, Shiely F, Power J, Clarke M, Eustace JA, Flanagan E, Kearney PM. Risk based monitoring (RBM) for clinical trials: A systematic review. *Contemporary Clinical Trials* 2016; 51: 15-27