A Proposed Framework for Evaluating Continuity of Data Coverage in Electronic Health Record and Administrative Claims Data in Real-World Evidence Studies

E. Susan Amirian, PhD; Joseph T. Dye, PhD; Thomas Wilson, PhD; Janet Espirito, PharmD; Nicholas Robert, MD; **John Bian, PhD**
Ontada, Irving, TX, USA

Background

Regulators have demonstrated increasing interest in leveraging RWE for decision making. The FDA's 2020 RWE guidance documents emphasize the importance of assessing data coverage continuity (e.g., enrollment/ disenrollment in payer data; evidence of continuous care through patient encounters in EHR data) in candidate real-world data (RWD) sources. Although there are considerations specific to the research question and nuances of each data source, development of customizable evaluation frameworks can facilitate the conduct of clear, cogent analyses that demonstrate whether the candidate RWD source has the requisite availability and comprehensiveness over time to be utilized for regulatory studies. Here, we provide a high-level framework for examining data coverage continuity in EHR and administrative claims databases and describe opportunities to mitigate challenges from lack of data continuity.

Data Continuity Appraisal

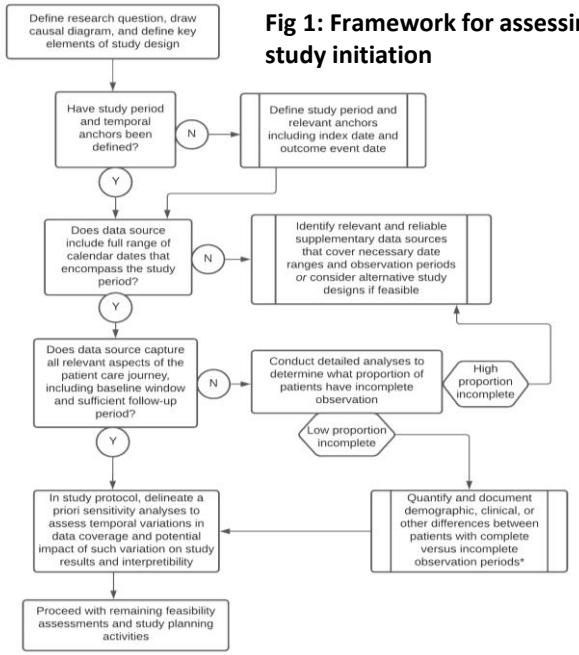

Fig 1: Framework for assessing data coverage at study initiation

Examples of Challenges	Examples of Solutions

EHR and administrative claims data are *encounter-based/transactional* and have variable encounter density per patient.	• Use of supplementary data sources, such as an all-payers claims dataset, may provide additional patient data *prior to the first visit* in the EHR and *past the censoring date* (last visit in EHR) for those lost to follow-up.
	• Development and maintenance of a registry-style database can allow for continuous data curation, allowing identification of data gaps.
	  • Regularly calculated metrics can include participant-level percentage of days observable in the dataset.
	  • Multi-source data from EHR, claims, billing, etc. can be aggregated into a "continuous observation file" to maximize the encounter density represented in the database.

Baseline & Identification Period: Participants tend to have variable time present in study database prior to index; therefore, baseline data and diagnostic details may be missing for some patients.	
Follow-Up Period: Loss to follow-up (censoring) may be differential by exposure and outcome.	• Inverse probability of censor weighting. • Assessing the range of possible effects on results (e.g., quantitative bias analysis), esp. when lost to follow-up dates differs between data sources or is assigned per different study definitions.

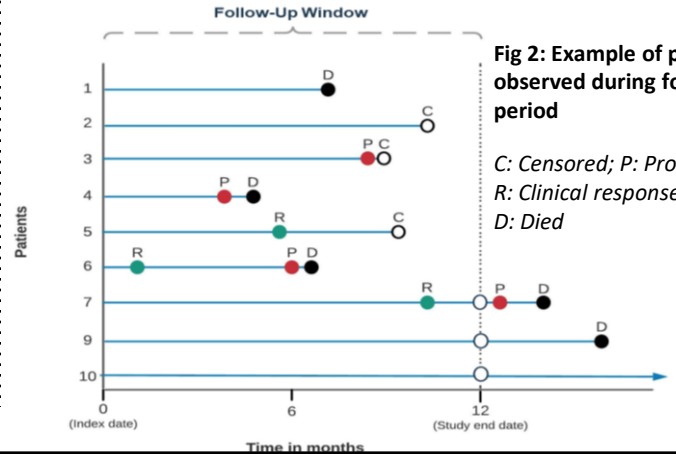

Fig 2: Example of patterns observed during follow-up period

C: Censored; P: Progressed; R: Clinical response achieved; D: Died

Discussion

- Careful consideration must be given to the different types of challenges associated with data coverage in each study window (Table).
- Solutions specific to the research question and study design/time window should be developed with incomplete data coverage in mind.
- External data sources may be needed for data coverage appraisal or supplementation. Each source will have its own strengths and limitations and should be evaluated accordingly.
  - For example, scope of data is associated with types of health care services available and received in the provider network.
    - Practice-specific EHR is a deep view of the patient journey at a given practice (e.g., diagnosis, treatment, lab tests, prescriptions).
    - Claims data can be either specific to one practice or can capture multiple practices in the patient journey but does not include deep clinical information.
  - Encounter density and scope of data should both be considered when assessing data coverage.
- Straightforward tools and frameworks can be developed (e.g., Fig 1) to facilitate processes related to data coverage evaluation and decisions regarding supplementation with additional data sources.

References

1. US Food and Drug Administration. Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products. September 2021. Accessed April 20, 2022.

2. Duke University. Margolis Center for Health Policy. Determining Real-World Data's Fitness for Use and the Role of Reliability. September 26, 2019. Accessed April 20, 2022.

3. Gokhale M, et al. Real-world Evidence: The Devil Is In the Detail. Diabetologia 2020;63(9):1694-1705

4. Schneeweis S, et al. Graphical Depiction of Longitudinal Study Designs in Health Care Databases. Ann Intern Med 2019;170(6):398-406

5. Gatto NM, et al. The Structured Process to Identify Fit-For-Purpose Data: A Data Feasibility Assessment Framework. Clin Pharmacol Ther. 2022;111(1):122-134

6. Jaksa A, et al. Organized structure of real-world evidence best practices: moving from fragmented recommendations to comprehensive guidance. 2021;10(9):711-731