Characterization of Real-World Data Sources to Support Observational Studies of Biologics, Including Biosimilars in Oncology

Ryan M Seals, ScD[1*], Catherine M Lockhart, PharmD, PhD[2], Nancy Lin, ScD[3], John D Seeger, PharmD, PhD[1]

1 Optum Epidemiology, Boston, MA; 2 Biologics and Biosimilars Collective Intelligence Consortium, Alexandria, VA, 3 Health Catalyst, Salt Lake City, Utah

***Corresponding author: ryan.m.seals@optum.com.**

Background

Use of large US health insurance claims databases for postmarketing assessment of medical products is well-established. However, certain oncology-relevant measures are often difficult to assess in claims datasets. The Biologics and Biosimilars Collective Intelligence Consortium (BBCIC) is a multi-stakeholder consortium established as a neutral convener to support transparent, methodologically rigorous research to generate real world evidence on the use, safety, and effectiveness of biologics, including biosimilars. This includes the BBCIC Distributed Research Network (DRN), a network of national and regional health insurers and integrated delivery networks that encompass administrative health care claims data on nearly 95 million patient lives. However, as administrative claims data alone may be insufficient to answer particular research questions, this project was undertaken to identify potential RWD sources to enrich the BBCIC's existing capabilities.

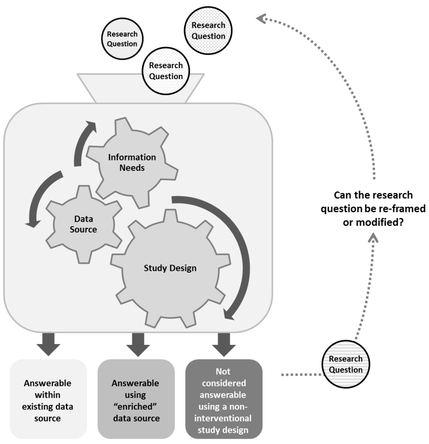
Objectives

- Convene a multi-stakeholder Workgroup of BBCIC industry sponsors, researchers, and practicing clinicians to develop recommendations for non-interventional CSR/CES studies of oncology biosimilars and reference biologic products, and;
- Identify and assess the utility of potential real world data (RWD) sources for the comparative safety and effectiveness of oncology biosimilar and reference products.

Conceptual Framework

To structure the assessment of potential RWD sources, a conceptual framework based on prior work was adopted.[1] Within this framework, generation of robust real-world evidence relies on the assessment of several components during study planning (**Figure**).

The component of interest for the current study is the evaluation of the fitness of potential data sources and/or determining if supplemental data collection approaches will be required to meet the needs of the specific research questions of interest. While limitations of the data may put limits on the research questions that can be asked, a range of research questions can usefully contribute to filling gaps in information. The process of refining the research question(s) in light of potential data source limitations can determine whether or not particular RWD sources are a suitable fit, or whether different sources of data will be required. This approach can identify areas of uncertainty regarding the appropriateness of the data source for the scientific question(s) under consideration, and inform planning around additional data characterization or feasibility assessments.



[1] Lin ND, Bosco JL, Holmes CH, Lockhart CM, Seeger JD. Methodologic Considerations for Data Source Selection and Study Design of Non-Interventional Studies Comparing the Safety and Effectiveness of Biosimilars and Reference Biologics: Insulin Glargine Products as a Case Example. 35th International Conference on Pharmacoepidemiology and Therapeutic Risk Management. Philadelphia, PA, USA: WILEY 111 RIVER ST, HOBOKEN 07030-5774, NJ USA, 2019;52-53.

Methods

BBCIC convened a Workgroup (WG) of researchers and clinicians from payers, industry, and academia. The WG developed recommendations to identify and assess the utility of potential real-world data (RWD) sources for conducting comparative safety and effectiveness studies of cancer care. We organized discussions around a hypothetical comparative study of trastuzumab originator and biosimilars in breast cancer patients. WG members developed and disseminated a questionnaire to potential RWD sources, soliciting:

1) characteristics of the RWD source,
2) collaboration experience,
3) availability of select data elements considered necessary for study conduct,
4) ability to link to external data sources, and,
5) counts of patients treated with trastuzumab products.

Results

- The questionnaire was distributed to 18 potential RWD sources identified from literature review and suggestions from individual WG members, chosen to reflect a mix of data source types. All candidate sources were contacted in July 2020, with up to 2 follow-up emails in August 2020.
- A total of 7 (39%) sources responded.
  - 2 primarily EHR-based
  - 1 primarily claims-based
  - 4 combination of EHR and claims

Table 1. Select data source characteristics of RFI respondents.

	Data Source						
	1	2	3	4	5	6	7
General data source questions							
Underlying data	EHR	Claims	EHR + claims	EHR + claims	EHR + claims (Medicare/ Medicaid)	EHR	EHR + claims
Approx. N	~66k (breast cancer)	~1m current (total)	~800k (EHR; breast cancer)	~1.2m (EHR; total)	--	~280k (breast cancer)	~3.2m (EHR; 2018)
Access/collaboration	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Average follow-up	1.8y	2-2.5y	5y (breast cancer)	2y	--	--	2.6-3y
Collaboration Experience and Requirements							
CDM	Yes	Yes	Yes	Yes	Yes	No	Yes
Licensed extracts	Yes	No	Yes	Yes	No	Yes	Yes
Ability to Link with Other Sources of Data							
Claims/EHR linkage	No	No	Yes	Yes	Yes	No	Yes
Cancer registry linkage	No	Yes	Yes	Possible	Yes	Partial	No
Death linkage	SS DMF, obituaries	State death registry	SS DMF, obituaries	No	SS DMF, tumor registry	3rd party	SS DMF

Results

RWD respondents indicated the availability of the following key data elements:

- Biomarkers (e.g., estrogen receptor +/- status)
  - n = 4 (57%)
- Ability to distinguish between reference biologics and biosimilars
  - n = 5 (71%)
- Ability to link to external data sources (e.g., cancer registries, death registries)
  - n = 3 (43%)

Table 2. Select data elements reported by RFI respondents.

	Data Source						
	1	2	3	4	5	6	7
Baseline Covariates							
Race/BMI	Yes	Partial	Partial	Yes	Yes	Partial	Yes
SES	No	No	No	Yes (via linkage)	Partial	No	Yes
Prior/concurrent dx/rx/px	Yes	Yes	Yes	Partial (via linked claims)	Yes	Partial	Yes
Exposure Assessment/Cohort Identification							
Biomarkers (e.g., HER2)	Yes	Partial	Partial	Yes	Yes	Partial	Yes
Grade/stage	Yes	Partial	Partial	Yes	Yes	Partial	Yes
IV/administered therapies	Yes	Yes; partial inpatient	Partial	Yes	Yes	Yes	Yes
Biosimilar differentiation	Yes	Yes	Partial	Yes	Yes	No; possible	Yes
Outcomes							
Patient-reported outcomes (e.g., QoL)	Yes	No	No	No	No; potential for patient surveys	No	No
Survival	Yes	Yes (death registry linkage)	Partial	No	Yes	Partial	Yes
Progression	Yes	Yes, from linked registry, medical records	Partial	Yes	Yes	Partial	Yes

- 3 respondents provided feasibility counts of a population of interest for a hypothetical study of trastuzumab
  - Population of breast cancer patients ranged from 66k – 128k
    - All 3 respondents also reported the size of the population with human epidermal growth factor receptor 2 (HER2) and with Stage IV or metastatic disease
- Population of patients receiving trastuzumab ranged from 3k – 12k
  - 1 respondent was able to differentiate and provide counts of trastuzumab reference biologic vs biosimilars

Discussion

- Non-interventional studies of biosimilars and their reference biologics can generate valuable real-world evidence to support clinical and policy decisions, but such studies must carefully consider the fitness of available data sources for particular research questions.
  - For studies requiring detailed clinical oncology data, the feasibility of potential RWD sources should be directly assessed via a questionnaire similar to that employed in this study
- Some oncology research questions may not be answerable with claims-only data sources, given the following key limitations:
  - Lack of biomarker information
  - Lack of information on cancer progression
  - Incomplete mortality data
- Oncology research involving biologics and biosimilars will likely need to be completed with enriched RWD sources
  - Claims + linkage to external registry data
    - E.g., cancer registries, mortality registries
  - Claims + linkage to EHR data
    - EHR data can capture rich clinical data (e.g., stage, grade, radiographic and laboratory findings) not captured through administrative claims alone

Presented at the 2021 AMCP Virtual Conference