Considerations to Ensure Transportability When Using Open Claims Data to Conduct Natural History and Event Rate Estimation Studies

Abstract #35

Nazha M. Diwan,[1,2] Paulina Colombo,[2,3] Lili Zhou,[2] Shicheng Weng,[2] Soraya Azmi,[2] Jamie Colasurdo,[2] Tzuyung Douglas Kou[2]

[1]Emory University, Atlanta, GA, USA; [2]BeiGene USA, Inc, San Mateo, CA, USA; [3]University of Arizona, Tucson, AZ, USA

Disclosures: NMD, PC, LZ, SW, SA, JC, TDK: Employment: BeiGene.

BACKGROUND

- With real-world data (RWD), such as insurance claims, electronic health records, and disease registries, pharmacoepidemiology studies are now able to examine the effectiveness, safety, and tolerability of a drug across large patient populations over an extended follow-up period
- Two types of insurance claims are commonly used in pharmacovigilance settings: closed-payer claims and open claims
- The increased availability of open claims data with linkage across multiple different data sources through tokenization provides pharmacoepidemiologists with a unique opportunity to generate contemporaneous insights with very little lag time in larger patient cohorts
- However, the construct of this type of data asset has inherent gaps in observations as it is subject to missing data and duplication issues that are thought to limit the internal and external validity of cancer outcomes research studies, which have typically relied on closed-payer claims to generate real-world evidence
- One such open claims database is the Symphony Health Solutions database, which contains deidentified and tokenized information that allows linkage of patient-level data from varied sources such as hospital claims, physician offices, as well as prescription (RX) data (dispensed and specialty Rx), with record dates as recent as one month prior
- The objectives of this study were:
  - To assess the impact of multiple cohort selection criteria that mimic different patterns of healthcare-seeking behaviors on cohort sample size of patients with chronic lymphocytic leukemia (CLL) in the Symphony database, an open claims database
  - To compare the demographic and clinical characteristics of patients with CLL in the Symphony database to those of patients in Surveillance, Epidemiology, and End Results (SEER)–Medicare data, a closed claims database that reflects a patient's completed encounters

METHODS

Design

- A descriptive analysis using the Symphony Health Solutions Database from 2013 to 2022

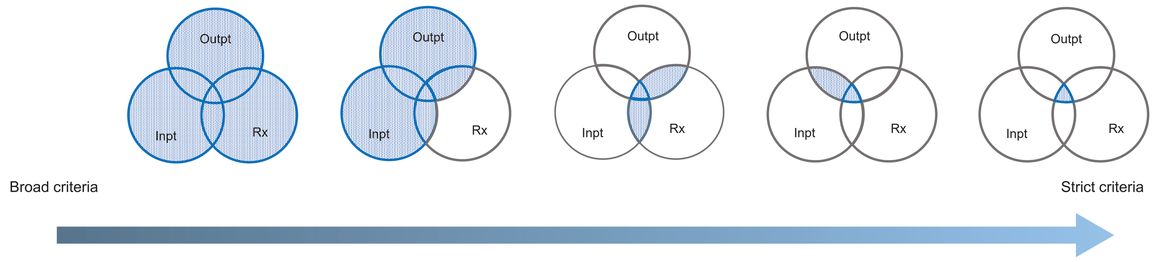
Cohort

- Patients diagnosed with CLL from 2013 to 2022 in the Symphony database, with available records as recent as 1 month prior, were selected using distinct criteria based on the presence or absence of any inpatient, outpatient, and/or Rx claims over a 1-year period before diagnosis (**Figure 1**)
- Presence of CLL was defined as ≥2 medical encounters with CLL (ICD-9-CM: 204.1x, 200.8x; ICD-10-CM: C91.1x, C83.0x) on different days

Statistical analysis

- Five patient cohorts were developed, and resulting demographics including age at cancer diagnosis, sex, follow-up post diagnosis, and Charlson comorbidity index were compared across the 5 cohorts in Symphony
  - End of follow-up definition 1: Last medical record OR 2: last medical or Rx record
  - Charlson comorbidities were identified based on encounters with medical diagnoses within 1 year prior to index date; presence of each comorbidity was defined as having ≥1 inpatient OR ≥2 outpatient claims at least 30 days apart
- Further, the demographics and prevalence of comorbidities, such as chronic pulmonary disease, congestive heart failure, and diabetes, were compared to a published study by Diamond et al. (2023)[1] that focused on a similar target population of CLL patients diagnosed between 2007-2015 but used SEER-Medicare linked data, a closed-payer claims database

Figure 1. Definitions of Cohort Selection Criteria of Patients With CLL Using Inpatient, Outpatient, and/or Rx Claims in Symphony Database (2013-2022)



Selection criteria	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Cohort 5
Inpatient claims	✓	✓	✓	✓	✓
and/or	or	or	or	and	and
Outpatient claims	✓	✓	✓	✓	✓
and/or	or		and		and
Pharmacy claims	✓		✓		✓

CLL, chronic lymphocytic leukemia; inpt, inpatient; outpt, outpatient; Rx, prescription.

CONCLUSIONS

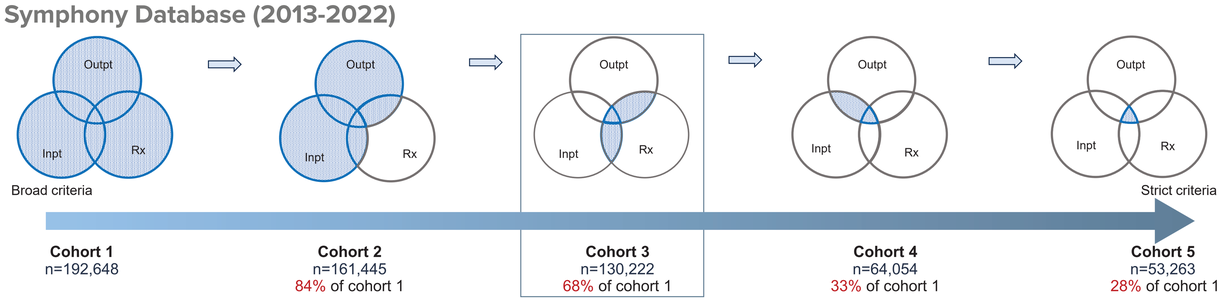
- While recognizing the constraints associated with open claims data, it is important to highlight the significant benefits they offer, including the recency of the data and wide coverage. These advantages are particularly valuable in fields characterized by a limited patient population, such as rare types of cancers
- This RWD analysis attempted to overcome some limitations, in particular the gaps in the dataset, by using varying cohort criteria to define a population of CLL patients. This approach allowed the development of a cohort that mimics a more standardized cohort, namely SEER-Medicare
- In a subset of patients with presence of all 3 types of claims in the Symphony database, our findings suggest that this method can be used to achieve comparable patient characteristics to those found in a closed claims database
- The use of tailored patient selection criteria could potentially help overcome the known limitations of inherent observation gaps in open claims data and generate meaningful insights for pharmacoepidemiology studies, such as natural history and event rate estimation studies

RESULTS

Symphony Cohorts

- Sample size varied significantly among the 5 Symphony cohorts, ranging from 192,648 to 53,263, with a substantial decrease in patient count with each addition of claims data criteria (ie, inpatient, outpatient, and Rx)
- Requiring the presence of Rx claims reduced the sample size by 32%, while restricting analysis to patients with presence of each claim type resulted in a 72% reduction (**Figure 2**)

Figure 2. Variations in Sample Size Based on Requirement of Inpatient, Outpatient, and/or Rx Claims in Symphony Database (2013-2022)



inpt, inpatient; outpt, outpatient; Rx, prescription.

- The demographics of the 5 Symphony cohorts were similar, including median age at diagnosis (70-71 years), sex distribution (55.2%-57.3% male), and median follow-up years post index (3.1-3.4 years based on end follow-up defined as last medical record; 3.4-3.7 years based on end follow-up defined as last medical or Rx record) (**Table 1**)
- However, the prevalence of comorbidities increased with the requirement for more types of claims, doubling in cohort 5 compared with cohort 1
  - The prevalence of chronic pulmonary disease was 10.8% in cohort 1 vs 20.5% in cohort 5, congestive heart failure was 6.2% in cohort 1 vs 12.4% in cohort 5, and diabetes with complication was 5.3% in cohort 1 vs 9.9% in cohort 5 (**Table 1**)
- Comparing the various Symphony cohorts with the patients identified in SEER-Medicare, there was similar age distribution, post-index follow-up, and comorbidities between cohort 3 in Symphony, with inpatient OR outpatient AND Rx records, and SEER-Medicare (**Table 1**)
  - The age at diagnosis was 70 years in Symphony vs 72 years in SEER-Medicare
  - The median follow-up after cancer diagnosis was 3.4 to 3.7 years in Symphony vs 3.8 years in SEER-Medicare
  - The prevalence of common chronic diseases in Symphony patients was comparable to that in patients with CLL identified in the SEER-Medicare data for diabetes (with complications, 6.4% vs 6.6%), chronic pulmonary disease (12.9% vs 12.6%), and congestive heart failure (7.4% vs 8.7%) (**Table 1**)

Table 1. Demographic and Clinical Characteristics of 5 Patient Cohorts in Symphony vs SEER-Medicare

	Symphony					SEER-Medicare
Patient selection criteria	**Cohort 1**	**Cohort 2**	**Cohort 3**	**Cohort 4**	**Cohort 5**	
Cohort Sample Size	192,648	161,445	130,222	64,054	53,263	98,529
Enrollment/coverage file	N/A	N/A	N/A	N/A	N/A	Medicare coverage file
Inpatient claims	✓	✓	✓	✓	✓	Part A
and/or	or	or	or	and	and	and
Outpatient claims	✓	✓	✓	✓	✓	Part B
and/or	or		and		and	and
Pharmacy claims	✓		✓		✓	Part D
Demographics						
Median age at diagnosis, years	70	70	70	71	71	72
Male, %	57.3	56.7	56.1	56.0	55.2	42.6
Median follow-up period post diagnosis, years						
Definition 1: Medical claims	3.3	3.3	3.4	3.1	3.2	3.8
Definition 2: Medical claims **OR** Rx record	3.6	3.6	3.7	3.4	3.5	3.8
Selected Comorbidities,[a] %						
Chronic pulmonary disease	10.8	12.5	12.9	20.1	20.5	12.6[b]
Congestive heart failure	6.2	7.1	7.4	12.1	12.4	8.7[b]
Diabetes with complications	5.3	6.2	6.4	9.8	9.9	6.6[b]
Diabetes without complications	14.8	17.1	17.7	24.6	24.9	21.5[b]
Peripheral vascular disease	5.8	6.7	6.9	11.3	11.3	11.2[b]

[a] Comorbidities were identified based on encounters with medical diagnoses using the medical claims within 1 year prior to index date. Presence of each comorbidity was defined as having ≥1 inpatient OR ≥2 outpatient claims at least 30 days apart. [b] From Diamond et al[1] (n=4958).

REFERENCES

1. Diamond A, et al. *JACC Cardio Oncol*. 2023;5(2):233-243.

ACKNOWLEDGMENTS

This study was sponsored by BeiGene, Ltd. Editorial assistance was provided by Nucleus Global, an Inizio Company, and was supported by BeiGene.

CORRESPONDENCE: Nazha M. Diwan, nazha.diwan@beigene.com

Presented at the ISPE 2024 Mid-Year Meeting; April 13-16, 2024; Orlando, FL