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

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

Affiliations: *[1]BeiGene USA, Inc., San Mateo, CA, USA*

ABSTRACT

Background: The increased availability of open claims data with linkage across multiple and different data sources (medical records, insurance or pharmacy claims records, etc) through tokenization provides pharmacoepidemiologists with a unique opportunity to generate contemporaneous insight with very little lag time. However, the construct of this type of data asset has inherent gaps in observations and are subject to missing data issues that are thought to limit the external validity of observational studies utilizing open claims data.

Objectives: Using Symphony Health Data, this study assessed the impact of multiple cohort selection criteria that mimic different patterns of health seeking behaviors on sample size and compared the demographic characteristics of chronic lymphocytic leukemia (CLL) patients to those in the SEER-Medicare data, a closed claim database that is reflective of a patient's completed encounters.

Methods: Patients diagnosed with CLL from 2013-2022 in Symphony database, with available records as recent as one month prior, were selected using distinct criteria based on the presence or absence of any inpatient (inpt), outpatient (outpt), and/or prescription (rx) claims over 1-year and 2-year periods before diagnosis. In total, 10 patient cohorts were evaluated and resulting demographic and preliminary clinical characteristics were compared to CLL patients in SEER-Medicare linked database. Further, of those treated with ibrutinib, prevalence rates of chronic diseases were compared to the published study focusing on a similar target population (Diamond et al 2023).

Results: There were significant sample size variations among the 10 Symphony cohorts. In cohorts 1-5, the sample size ranged from 192,648 to 53,263, with a substantial decrease in patient count with each addition of claims data criteria (e.g., inpt, outpt, 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 72% reduction. Comparing the various cohorts in Symphony to the patients identified in SEER-Medicare, we found similar age and post-index follow-up between the two databases (median follow-up 3.1-3.5 years in Symphony vs 3.8 years in SEER-Medicare). Prevalence rates of common chronic diseases, such as diabetes, for those treated with ibrutinib were comparable to the aforementioned study. In comparing SEER-Medicare data to Symphony, ibrutinib treated patients demonstrated prevalence proportions of 18.1% and 15.8% for diabetes without complications, respectively.

Conclusion: In a subset of patients with presence of all three types of claims in Symphony database, study findings suggest comparable patient characteristics to studies using closed claims database. The use of tailored patient selection criteria could potentially help overcome the known limitations of potential observation gaps in open claims data and generate meaningful insights for pharmacoepidemiology studies.