Integration of claims and clinical data to enable real-world evidence in oncology

Chi Nguyen[1], Hanke Zheng[2], Michael Grabner[1], Shiva K Vojjala[1], John Barron[1], Saurabh Ray[2], Brian Sweet[2], Nathan Hill[2]

[1]Carelon Research, Inc., Wilmington, DE, USA; [2]Bristol Myers Squibb, Princeton, New Jersey, USA

Background

- The Generating Evidence Excellence (GEx) research environment integrates multiple real-world data sources, including administrative claims, electronic medical records (EMR), a payer-driven cancer care quality program, mortality, and social determinants of health (SDoH) data, to create a unique and clinically rich platform for oncology research.
- The GEx research environment analyzed in this study comprises two distinct data containers:
  - Healthcare Integrated Research Database (HIRD®) + Cancer Care Quality Program (CCQP): contains closed administrative claims data from the HIRD® of commercially-insured and Medicare Advantage members and clinical data from a CCQP where oncology practices submit information as part of incentive and utilization management.
  - HIRD® + Blue Health Intelligence (BHI) + IntrinsiQ Specialty Solutions (IQSS): contains closed administrative claims data from HIRD® and BHI, along with EMR data from IQSS.
- As use of the GEx environment for research is growing, it is important to gain an understanding of the data sources, patient characteristics, and representativeness of the GEx data containers.

Objective

To describe GEx patient characteristics, as well as overall survival, and contrast them to US national data from the American Community Survey (ACS) and the US Cancer Statistics Public Use Database (USCS). This analysis focused on non-small cell lung cancer (NSCLC).

Methods

Study Design

This was a retrospective observational study of patients with newly diagnosed NSCLC.

National databases

- **USCS**: combines cancer incidence data from the National Program of Cancer Registries (NPCR) and Surveillance, Epidemiology, and End Results (SEER) in all 50 states and territories.
- **ACS**: collects information from approximately two million households per year, including social determinants of health data at the Census block group level. This study used 2019 5-year estimates.

Inclusion and exclusion criteria

GEx Research Environment

- Patients with newly diagnosed lung cancer as documented in claims and with confirmation of NSCLC pathology documented in the clinical data from 01/01/2015 to 12/31/2019 were identified from each of the two GEx data containers.
- The first observed diagnosis claim for NSCLC or the first treatment (if first diagnosis was unavailable) was set as the index date.
- ≥18 years old as of index date.
- Continuous health plan enrollment (including both medical and pharmacy benefits) for ≥6 months prior to the index date.
- Patients with diagnosis claims for other primary cancers prior to index date were excluded.

US Cancer Statistics Public Use Database (USCS)

- Patients with newly diagnosed lung and bronchus cancer and non-small cell pathology were identified using ICD-O-3, a standard coding system for tumors, for the period between 01/01/2015 and 12/31/2019.
- ≥18 years old at the time of diagnosis.

Outcomes

- Patient characteristics, including race, cancer stage at diagnosis, and Census block level socio-economic status (SES) were described for each of the GEx data containers and the USCS (as available).
- Mortality data for the GEx research environment were obtained from a combination of EMR, claims, and third-party obituary data.
- Mortality for the USCS was obtained from the National Center for Health Statistics' National Vital Statistics System.

Statistical analysis

- Patient demographic and clinical characteristics were assessed via descriptive statistics.
- For each GEx data container, the all-cause mortality rate was calculated as the number of deaths during the entire follow-up divided by the total time at risk and reported as number of deaths per 100 person-years. Kaplan-Meier curves and log-rank tests were used to examine the survival time of patients with NSCLC by cancer stage at diagnosis.
- We calculated the overall survival rate at 12, 24, and 36 months and compared it to the survival of patients identified from the SEER database, where the same measure was provided.

Results

Patient population

- Within the GEx research environment, we identified a total of 6,233 eligible patients with NSCLC in the HIRD+CCQP container and 1,176 in the HIRD+BHI+IQSS container (Table 1).
- The mean and median duration of follow-up was approximately 13 and 9 months, respectively. Approximately 64.7% of the HIRD+CCQP study population and 61.7% of the HIRD+BHI+IQSS study population had more than 6 months of follow-up.
- A total of 827,882 patients with NSCLC were identified from the USCS, of which 132,998 patients were from the SEER registries.

Patient characteristics

- Both GEx data containers showed similar patient demographics, with mean age at first diagnosis of 63 years, and 71% of patients being White.
- The majority of the GEx study population resided in the South (32.6% for HIRD+CCQP and 61.6% for HIRD+BHI+IQSS) and Midwest (41.0% for HIRD+CCQP and 24.6% for HIRD+BHI+IQSS).
- As seen in Table 1, patients in the GEx research environment overrepresented the working-age population less than 65 years old (64.0% for both GEx data containers vs. 29.9% for the USCS), had better socio-economic status, and appeared to have more advanced disease (stage IV) compared to the national databases.
- Patients with stage IV NSCLC at diagnosis accounted for 59.5% and 53.4% of the study population in HIRD+CCQP and HIRD+BHI+IQSS, respectively, compared to 46.1% in the USCS.

Table 1. Baseline characteristics

	HIRD+CCQP (N=6,233)	HIRD+BHI+IQSS (N=1,176)	U.S Cancer Statistics Public Use Databases (N=827,882)
Mean age (SD)	62.8(8.8)	63.1(8.3)	Not available
Age categories			
18-44	2.1%	1.0%	0.9%
45-54	12.8%	10.5%	5.9%
55-64	49.2%	53.1%	23.1%
65-74	24.0%	23.1%	36.0%
75+	11.9%	12.2%	34.0%
Female	48.0%	51.7%	47.8%
Race			
N with available data	5,221	877	823,905
White	71.3%	71.3%	84.6%
Black or African American	7.4%	3.6%	11.4%
Asian	1.9%	1.5%	3.3%
Other	19.4%	23.6%	0.6%
Geographic region			
West	11.0%	5.6%	16.6%
South	32.6%	61.6%	40.0%
Northeast	15.4%	8.2%	19.5%
Midwest	41.0%	24.6%	23.9%
Medicare Advantage	27.6%	15.6%	N/A
Lung cancer stage at diagnosis			
N with available data	6,223	959	797,002
Stage I/II (localized)	11.4%	18.8%	29.6%
Stage III (regional)	29.1%	27.8%	24.3%
Stage IV (distant)	59.5%	53.4%	46.1%
Socioeconomic status (SES) quintile distribution*			
N with available data	5,802	494	N/A
1-Worst SES status	13.3%	9.1%	
2	21.2%	19.6%	
3	24.4%	28.7%	
4	23.1%	25.9%	
5-Best SES status	18.0%	16.6%	
Urban/rural			
N with available data	8,526	496	N/A
Rural	32.7%	25.2%	
Urban	67.3%	74.8%	

*Quintile thresholds were based on the full US population as represented in the 2019 5-year ACS estimates. For example, in the HIRD+CCQP container, 13.3% of patients were in the lowest quintile (which represents the lowest 20% of the US).

Mortality

- The all-cause mortality rate was 38.8 (HIRD+CCQP) and 27.3 (HIRD+BHI+IQSS) per 100 person-years.
- The Kaplan-Meier curves indicated that patients with stage IV disease had statistically significantly lower survival than those with stages I or II (log-rank test p-value <0.01)
- The median survival times of NSCLC patients identified from the HIRD+CCQP data container were 30.6 months (stage III) and 13.7 months (stage IV), while those identified from the HIRD+BHI+IQSS had median survival times of 30.3 months (stage III) and 19.0 months (stage IV).

Figure 1A. Kaplan-Meier curves examining survival time stratified by cancer stage at diagnosis (HIRE +CCQP)

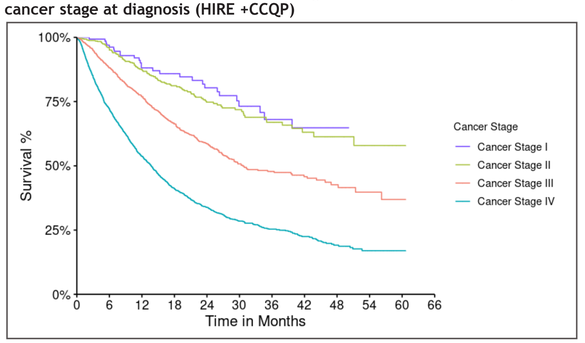
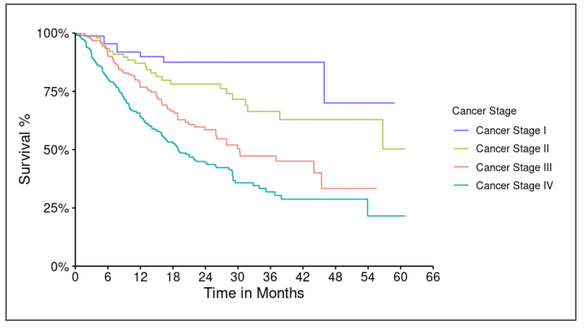


Figure 1B. Kaplan-Meier curves examining survival time stratified by cancer stage at diagnosis (HIRE +BHI+ION)



- The unadjusted overall cumulative survival for non-metastatic NSCLC in the GEx data containers, which included stages I, II, and III, was found to be similar to that reported in the SEER registries.
- The unadjusted overall cumulative survival at 36 months for stage IV NSCLC was 25.4% (HIRD+CCQP) and 31.9% (HIRD+BHI+IQSS), which was higher compared to the SEER registries (14.8%).

Table 2. Overall cumulative survival

	HIRD+CCQP (N=6,233*)	HIRD+BHI+IQSS (N=1,176*)	U.S Cancer Statistics Public Use Databases (SEER members N=132,998*)
Stage I/II (localized), N	710	180	30,383
12-month cumulative survival, %	87.5%	88.2%	86.4%
24-month cumulative survival, %	76.0%	81.9%	75.7%
36-month cumulative survival, %	67.0%	73.4%	67.3%
Stage III (regional), N	1,809	303	29,331
12-month cumulative survival, %	76.9%	76.8%	69.4%
24-month cumulative survival, %	58.6%	58.6%	53.6%
36-month cumulative survival, %	47.8%	47.3%	44.2%
Stage IV (distant), N	3,701	517	69,965
12-month cumulative survival, %	53.7%	64.0%	35.2%
24-month cumulative survival, %	33.8%	44.3%	21.3%
36-month cumulative survival, %	25.4%	31.9%	14.8%

* Patients with unknown or missing cancer stage were not included.

Discussion and limitations

- The study provides insights into the demographics, clinical characteristics, and survival of patients contained in the GEx research environment and compares them to US national databases.
- The GEx research environment also contains data on healthcare costs, enabling the conduct of studies on the total costs of care; this is a topic for future research.
- It is important to take into account the study population, data collection methods, and data completeness when considering the differences between the databases:
  - The HIRD includes commercially insured and Medicare Advantage members, who were typically employed, younger, appeared to have more advanced disease than the USCS/SEER sample, and generally had higher SES status than the overall US population.
  - The USCS gathers information on patient and tumor characteristics, including cancer stage, at the time of diagnosis. On the other hand, the GEx data environment collects information from oncology practices, either at the time of diagnosis or treatment initiation.
  - Mortality data may be underreported in the GEx research environment and efforts are currently underway to quantify the extent of this issue.
- The GEx research environment is striving to enhance the quality and completeness of data and broaden its scope to other therapeutic areas, including immunology.

Conclusions

The linked GEx research environment provides a sophisticated resource to enhance understanding of real-world oncology patient outcomes. The commercially insured/Medicare Advantage population is younger and more economically advantaged than the national NSCLC population and most relevant when used for understanding working-age patients.

Disclosure

- This study was funded by Bristol Myers Squibb.
- CN, MG, SV, JB are employees of Carelon Research (a wholly owned subsidiary of Elevance Health), which received funding from BMS for the conduct of this study. MG is a shareholder of Elevance Health. HZ, SR, BS, NH are employees and shareholders of BMS.

Presented at ISPOR 2023; May 7-10, 2023; Boston, MA

Email: Hanke.Zheng@bms.com