RWD198

Evaluating Real World Data Fitness-for-use in Early-stage NSCLC within a U.S. Based Community Health System Network

Yu-Han Kao[1], Xiaohan Hu[1], Hozefa A. Divan[1], Vladmir Turzhitsky[1], Kaushal Desai[1], Thomas D. Brown[2]

1. Merck & Co., Inc., Rahway, NJ, USA,
2. Syapse Holdings, Inc., 1442 Pottstown Pike, Unit #3008, West Chester, PA 19380

Background

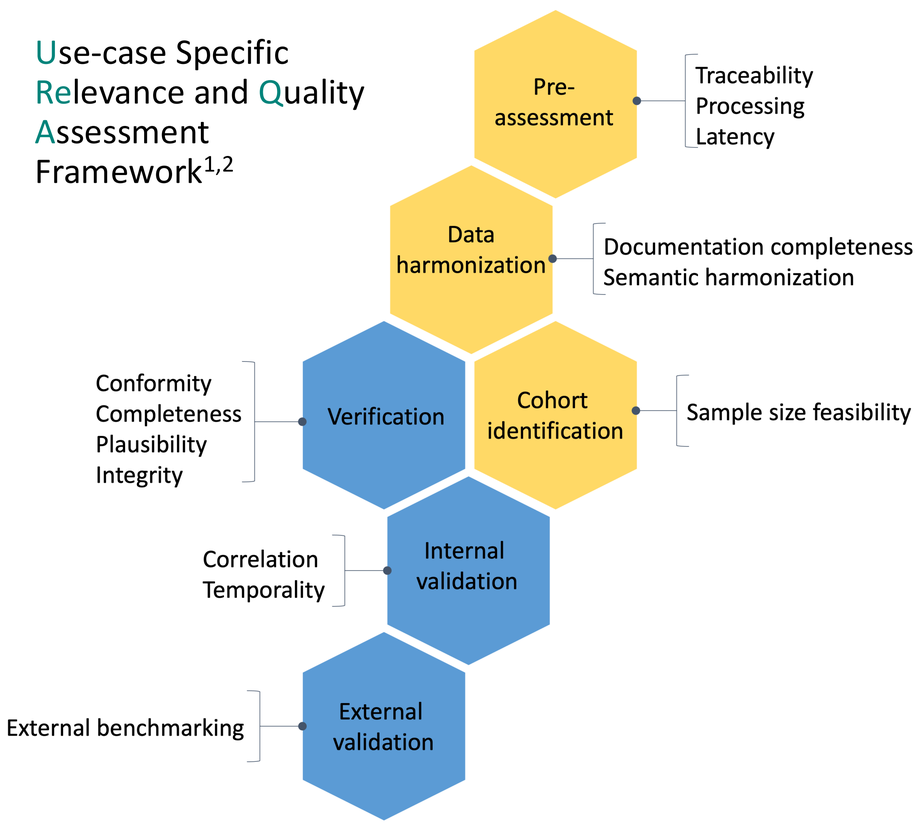
- The early-stage cancer treatment landscape is evolving rapidly as precision oncology extends its applications beyond advanced disease. These advancements necessitate a thorough understanding of patient populations, treatment patterns, and patient outcomes in real-world settings to optimize care and enhance clinical decision-making.
- Real-world data (RWD) has become increasingly valuable for capturing the complexity of cancer care. However, compiling robust RWD for early-stage cancers presents significant challenges. This is due to the necessity of acquiring comprehensive information on a wide range of interventions (e.g., surgical resection, radiation therapy, systemic therapies) and reporting intermediate endpoints from various data sources.
- The Syapse Learning Health Network (LHN) integrates data from community health systems across 25 states in the U.S., encompassing 457 hospitals and 1300 oncologists. By combining data from several sources, including electronic health records (EHRs), laboratory and radiology/imaging systems, and hospital-based cancer registries, the LHN presents a unique opportunity to understand disease management of early-stage cancers in real-world settings.

Objectives

- The objective of this study was to evaluate the characteristics of the early-stage non-small cell lung cancer (eNSCLC) cohort within the Syapse analytic dataset (ADS) to assess its fitness-for-use in generating real world evidence (RWE) in various scenarios.

Methods

- We applied a previously developed data assessment framework focused on early-stage cancers to evaluate the eNSCLC cohort with initial diagnosis of stage I-III since 2016 from Syapse ADS released on Oct 15th, 2024, and assessed its fitness for three use cases:
  - Characterizing the patient cohort,
  - Understanding biomarker testing trends over time
  - Analyzing primary treatment patterns.



Results

Cohort Identification

Attrition steps	N	Percent included from prior step	Needed variable name (% missing)
Syapse enriched ADS cohort	32396	-	
Patients diagnosed with NSCLC	26664	82.30%	Histology (0%)
Diagnosed since 2016	22797	85.50%	Diagnosis date (0%)
Diagnosis at age >= 18	22797	100%	Diagnosis date (0%), Diagnosis age (0%)
Prioritized stage group **I-III**[a]	11112	48.70%	Prioritized stage group (3.9%)
Exclude other primary cancer prior to diagnosis	**8228**	74%	Other cancer diagnosis (1.2%), Other cancer date (21.5%)

[a.] Summarized stage based on clinical stage and pathological stage

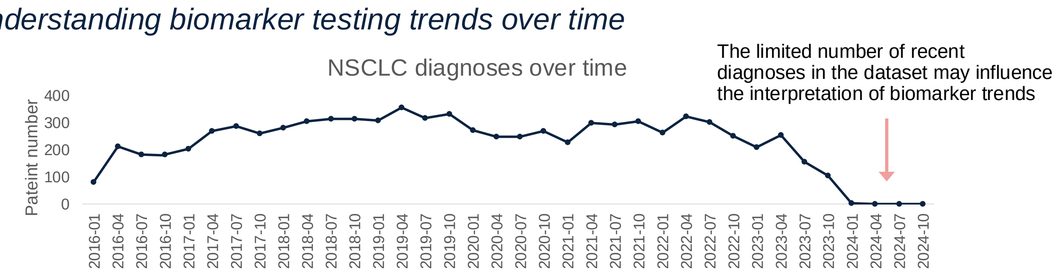
Use Case:

Characterizing patient cohort

Needed variable name (% missing)	
Demographic characteristics	**Clinical characteristics**
• Sex (0.1%) • Race (1.1%) • Ethnicity (3.6%) • Region[a] (0%) • Social Vulnerability Index (9.6%) • Insurance status (4%) • Primary Care Health Professional Shortage Areas (9.4%) • Median household income (4.7%)	• Histology subgroup[b] (5.4%) • Smoking status (1.9%) • Charlson Comorbidity Index (26.2%) • Clinical stage group (11.3%) • Pathological stage group (15.4%) • ECOG performance status (59.9%)

[a.] Region – Midwest: 71.8%, Northeast: 8.2%, South: 20%, West 0.1%
[b.] Histology subgroup – Adenocarcinoma: 58.5%, Squamous cell carcinoma: 36.1%, Missing/Unknown: 5.4%

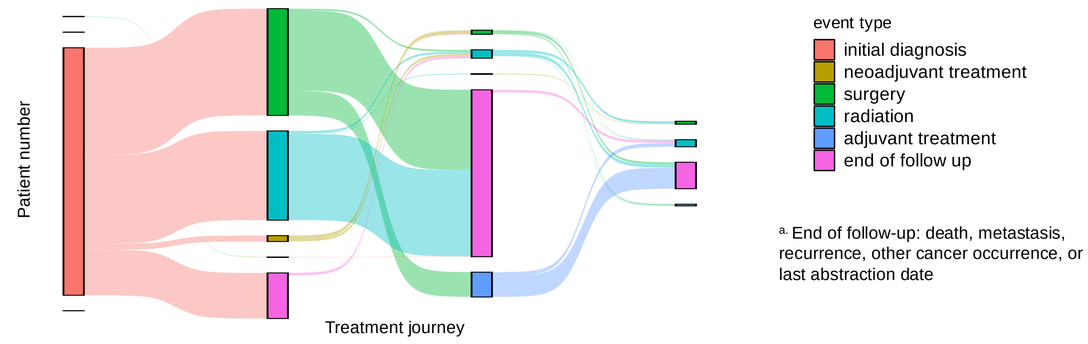
Understanding biomarker testing trends over time



	PD-L1 test[a]	EGFR test[b]
Overall, N = 8228	N (%)	N (%)
Patient ever tested since diagnosis	4345 (52.8%)	4053 (49.3%)
Patients with test date[c]	4219 (97.1%)	4025 (99.3%)
Patients with test results[d]	TPS: 3113 (71.6%) State: 4311 (99.2%)	4050 (99.8%)

[a.] Only tests using IHC assay are included here. [b.] Only tests using NGS, PCR, and DNA sequencing are included here.
[c.] Date granularity: month or day. [d.] PD-L1 results include tumor proportion score (TPS): 0-100%; state: positive, negative, equivocal, etc.; EGFR results include state: present, absent, etc.

Analyzing primary treatment patterns



[a.] End of follow-up: death, metastasis, recurrence, other cancer occurrence, or last abstraction date

	Systemic therapy	Surgery	Radiation
Overall, N = 8228	N (%)	N (%)	N (%)
Patients with any record of the treatment[a]	3570 (43.4%)	3897 (47.4%)	3510 (42.7%)
Start date is available for the treatment[b]	3469 (97.2%)	3888 (99.8%)	3490 (99.4%)

[a.] Any treatment setting throughout patient journey since diagnosis
[b.] Any date granularity (i.e., year, month or day)

Conclusions

- The results demonstrated high completeness of crucial variables enabling the selection of a substantial eNSCLC cohort of over 8,000 patients. The baseline demographic characteristics exhibited low missingness and comprehensive information for PD-L1 and EGFR testing. Additionally, more than 90% of patients had detailed treatment information.
- Approximately 70% of patients were from the Midwest U.S., and about 70% had Charlson Comorbidity Index at diagnosis. Only 40% had documented ECOG performance status, likely due to less frequent documentation in early-stage settings. The dataset also showed a limited number of patients from the past 12 months, likely because of data latency from cancer registries.
- External benchmarking was performed by comparing our data to SEER data (not shown), demonstrating a comparable distribution of patient characteristics and initial treatment patterns between the two datasets. We also noted an increase in PDL1 and EGFR testing in recent years, aligning with our expectations for evolving clinical standards.
- These findings suggest that the community health system-based eNSCLC cohort demonstrated fitness-for-use for characterizing patient cohorts, understanding biomarker testing trends over time, and analyzing primary treatment patterns. The dataset has the potential to provide valuable insights into broader clinical practices beyond clinical trials or academic settings, supporting better understanding standard of care in eNSCLC patients.
- Consideration should be given to aspects such as data recency and geographic representation in assessing fitness-for-use. Researchers should take these factors into account, particularly for studies requiring the most recent data or those focusing on specific populations. Acknowledging these considerations will ensure the responsible and accurate application of RWE in understanding and improving cancer care.

References

1. Desai, K., Chandwani, S., Ru, B., Reynolds, M., Christian, J. B., Estiri, H. (2021). PCN37 An Oncology Real-World Data Assessment Framework for Outcomes Research. Value in Health, 24(Supplement 1), S25. https://doi.org/10.1016/j.jval.2021.04.129
2. Kao, Y. H., Markt, S., Ru, B., Sillah, A., Desai, K., Meng, W. (2023). RWD87 Real-World Data Quality Assessment for Out comes Research in Early-Stage Cancers. Value in Health, 26(6), S376-S377. https://doi.org/10.1016/j.jval.2023.03.2116

Contact information

Yu-Han Kao, yu-han.kao@merck.com

Presented at ISPOR EU, Barcelona, November 20, 2024