SA4

Characteristics, Treatment Patterns, and Outcomes of Patients With Non–Small-Cell Lung Cancer Across Clinico-Genomic Databases and Electronic Medical Records

D. Lin[1]; C. Shao[1]; X. Liu[1]; H.M. Sineshaw[1]

[1]Merck & Co., Inc., Rahway, NJ, USA

Background

- Despite treatment advances, 5-year survival in non–small-cell lung cancer (NSCLC) remains poor,[1] highlighting the need for personalized therapies
- Clinico-genomic databases (CGDBs) combine clinical records with next-generation sequencing (NGS)-derived genomic data to support evidence generation to inform precision oncology clinical development and decision making[2]
- Assessing whether CGDBs reflect the real-world NSCLC population by comparing their patient profiles with those in traditional electronic medical record (EMR) datasets is critical because[3]
  - CGDB cohorts may differ from NSCLC without genomic data, introducing selection bias
  - Uneven NGS access and uptake across demographic groups can lead to underrepresentation and constrain outcome interpretation

Objectives

- Describe and compare patients with NSCLC by stage across Flatiron Health–Foundation Medicine (FH-FMI) CGDB, Flatiron Health Research Databases (FHRD), and National Cancer Institute Surveillance, Epidemiology, and End Results (SEER) 21 registry in terms of demographic and clinical characteristics, treatment patterns, and NGS testing utilization

Methods

Study design

- This was a retrospective cohort study of patients ≥18 years of age with a diagnosis of NSCLC in the following 3 data sources

1) FH-FMI CGDB
  - Advanced NSCLC cohort: Diagnosed with advanced NSCLC between January 1, 2018, and January 30, 2024

2) FHRD
  - Advanced NSCLC cohort: Diagnosed with advanced NSCLC on or after January 1, 2018, or diagnosed with early-stage NSCLC and subsequently developed recurrent or progressive disease on or after January 1, 2018, and up to 6 months before August 31, 2025

3) SEER 21 registry
  - Diagnosed with NSCLC between January 1, 2018, and December 31, 2021
  - Cutoff date of December 31, 2022

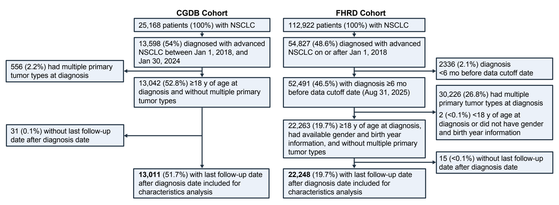
Assessments

- Outcomes included
  - Patient demographic and clinical characteristics in the study data sources
  - Treatment patterns in CGDB and FHRD data sources by treatment setting
  - NGS utilization in CGDB and FHRD
  - Overall survival (OS) in patients with NSCLC from CGDB and FHRD
- Descriptive statistics were used to summarize patient characteristics, treatment patterns, and NGS testing utilization
- OS was estimated, overall and by cancer stage, using the Kaplan-Meier method (including medians and 95% CIs)

Results

Advanced NSCLC

Figure 1. Attrition of patients in the FH-FMI CGDB and FHRD cohorts



- Demographics were generally similar across the CGDB and FHRD cohorts; the CGDB cohort was more community-based, while the FHRD cohort had more academic representation
- FH cohorts have relatively lower proportions of non-Hispanic Black and Hispanic patients, compared to SEER

Table 1. Demographic characteristics of patients with advanced NSCLC

Characteristic	CGDB (N = 13,011)	FHRD (N = 22,248)	FHRD with NGS (N = 16,278)	SEER 21 (N = 82,255)
Sex				
Female	6555 (50.4)	10,939 (49.2)	8190 (50.3)	38,718 (47.1)
Male	6456 (49.6)	11,309 (50.8)	8088 (49.7)	43,537 (52.9)
Age at initial diagnosis, median (IQR), y	70.0 (15.0)	69.0 (14.0)	67.8 (14.0)	-
Year of diagnosis				
2018–2019	4481 (34.4)	5694 (25.6)	3471 (21.3)	43,005 (52.3)
2020–2021	4557 (35.0)	6975 (31.4)	5063 (31.1)	39,250 (47.7)
2022–2023	3821 (29.4)	6447 (29.0)	5171 (31.8)	-
2024–2025	152 (1.2)	3132 (14.1)	2573 (15.8)	-
Race and ethnicity				
Non-Hispanic White	7107 (54.6)	11,912 (53.5)	8707 (53.5)	56,461 (68.7)
Non-Hispanic Black or African American	904 (6.9)	1880 (8.5)	1236 (7.6)	10,261 (12.5)
Non-Hispanic Asian	301 (2.3)	594 (2.7)	462 (2.8)	8030 (9.8)
Non-Hispanic other/unknown race	1028 (7.9)	642 (2.9)	471 (2.9)	667 (0.8)
Hispanic	400 (3.1)	870 (3.9)	627 (3.9)	6836 (8.3)
Ethnicity unknown	3271 (25.1)	6350 (28.5)	4775 (29.3)	-
Practice type				
Academic	1496 (11.5)	6069 (27.3)	3580 (22.0)	-
Community	10,846 (83.4)	15,574 (70.0)	12,202 (75.0)	-
Community/academic	669 (5.1)	605 (2.7)	496 (3.0)	-

Data are n (%) unless otherwise noted.

- The FHRD with NGS cohort had a higher percentage of nonsquamous NSCLC than the other 3 cohorts
- The FHRD and SEER 21 cohorts had higher percentages of stage IV disease at initial diagnosis than the CGDB cohort
- The CGDB cohort was healthier and had better Eastern Cooperative Oncology Group (ECOG) performance status capture than the FHRD cohort

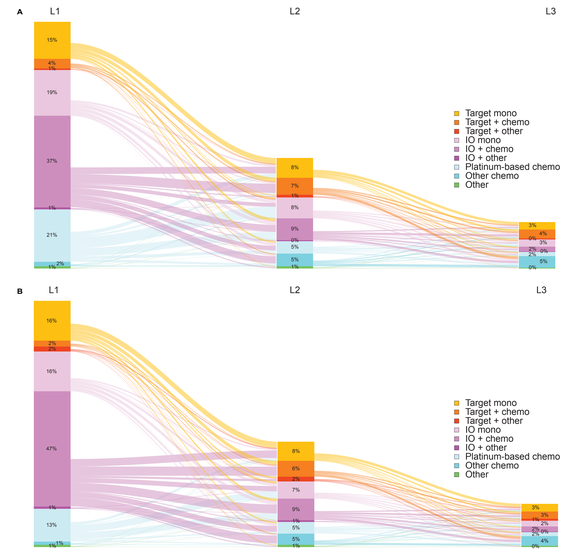
Table 2. Clinical characteristics of patients with advanced NSCLC

Characteristic	CGDB (N = 13,011)	FHRD (N = 22,248)	FHRD with NGS (N = 16,278)	SEER 21 (N = 82,255)
Histology				
Squamous	2874 (22.1)	4717 (21.2)	2879 (17.7)	20,047 (24.4)
Nonsquamous	9594 (73.7)	16,506 (74.2)	12,672 (77.8)	58,076 (70.6)
Not otherwise specified	543 (3.2)	1025 (4.6)	727 (4.5)	4132 (5.0)
Tumor stage at initial diagnosis				
I–II	1903 (14.6)	2362 (10.6)	1665 (10.2)	-
III	123 (0.9)	-	-	-
IIIA	1191 (9.2)	1115 (5.0)	805 (4.9)	-
IIIB–IIIC	1452 (11.2)	2507 (11.3)	1574 (9.7)	12,405 (15.1)
IV	8068 (62.0)	15,709 (70.6)	11,886 (73.0)	69,850 (84.9)
Unknown	284 (2.1)	555 (2.5)	348 (2.1)	-
ECOG performance status at diagnosis				
0 or 1	8057 (61.9)	9014 (40.5)	7084 (43.5)	-
≥2	1932 (14.8)	4239 (19.1)	3061 (18.8)	-
Unknown	3022 (23.2)	8995 (40.4)	6133 (37.7)	-
CCI score				
Subjects with data, n	5844	11,028	7852	-
Median (IQR)	2.0 (2.0)	2.0 (4.0)	2.0 (3.0)	-
Follow-up time from diagnosis				
Subjects with data, n	13,011	22,248	16,278	64,925
Median (IQR), mo	13.8 (23.4)	12.2 (22.2)	13.2 (22.6)	8.0 (18.0)

Data are n (%) unless otherwise noted.
CCI, Charlson comorbidity index.

- Among total NSCLC patients, the distribution of treatment categories was comparable between the CGDB and FHRD cohorts across lines of therapy (LoTs)

Figure 2. Treatment category for patients with advanced NSCLC by LoT in the (A) CGDB and (B) FHRD cohorts



IO, immunotherapy.

- Among patients with NSCLC and actionable genomic alterations (AGAs), the distribution of treatment categories was comparable between the CGDB and FHRD cohorts across LoTs
- Among patients with non-AGA NSCLC, a lower percentage of patients in the CGDB cohort received immunotherapy as a first LoT compared with the FHRD cohort

Table 3. Treatment category by LoT

Treatment category	AGA		Non-AGA	
	CGDB[a]	FHRD[b]	CGDB	FHRD
1L	**N = 2410**	**N = 5286**	**N = 8088**	**N = 12,734**
Targeted therapy	1448 (60.1)	3109 (58.8)	595 (7.4)	602 (4.7)
Immunotherapy	628 (26.1)	1654 (31.3)	5314 (65.6)	9853 (77.4)
Chemotherapy	308 (12.7)	465 (8.8)	2116 (26.2)	2167 (17.0)
Other	26 (1.1)	58 (1.1)	63 (0.8)	112 (0.9)
2L	**N = 1211**	**N = 2618**	**N = 3492**	**N = 5088**
Targeted therapy	770 (63.6)	1628 (62.2)	909 (26.0)	1274 (25.0)
Immunotherapy	281 (23.2)	636 (24.3)	1583 (45.4)	2336 (45.9)
Chemotherapy	130 (10.7)	283 (10.8)	939 (26.9)	1393 (27.4)
Other	30 (2.5)	75 (2.7)	61 (1.7)	81 (1.7)
3L	**N = 545**	**N = 1187**	**N = 1428**	**N = 1980**
Targeted therapy	322 (58.5)	690 (58.1)	421 (29.5)	550 (27.8)
Immunotherapy	115 (21.0)	234 (19.7)	418 (29.3)	603 (30.5)
Chemotherapy	101 (18.7)	235 (19.8)	563 (39.4)	769 (38.8)
Other	10 (1.8)	28 (2.4)	26 (1.8)	58 (2.9)

All data are n (%).
1L, first line; 2L, second line; 3L, third line.
[a]In CGDB, AGAs were defined as patients with any of the following: EGFR mutation, ALK fusion, ROS1 fusion, NTRK fusion, BRAF mutation, BRAF V600E mutation, MET exon 14 skipping mutation, MET amplification, or RET mutation.
[b]In FHRD, AGAs were defined as either EGFR, ALK, ROS1, NTRK, BRAF, MET, or RET pathological positive, due to the lack of mutation-level detail.

- Among patients with NSCLC, OS from index date of LoT initiation was generally comparable between the CGDB and FHRD cohorts
- OS of patients assessed from different index dates of LoT initiation was comparable between the CGDB and FHRD cohorts among patients with AGA and non-AGA NSCLC

Table 4. OS in total and by AGA status in patients with advanced NSCLC

OS	CGDB	CGDB (left truncation)	FHRD
Total population			
Index from date of 1L initiation			
Patients, N	10,460	10,388	17,636
Median OS (95% CI), mo	17.3 (16.7–18.0)	14.5 (14.1–15.11)	17.4 (16.9–17.9)
Index from date of 2L initiation			
Patients, N	4687	4674	7546
Median OS (95% CI), mo	12.4 (11.8–13.0)	11.2 (10.7–11.8)	12.0 (11.6–12.5)
Index from date of 3L initiation			
Patients, N	1975	1972	3115
Median OS (95% CI), mo	9.5 (8.8–10.1)	8.7 (8.2–9.4)	9.4 (8.9–10.0)
Population with AGAs			
Index from date of 1L initiation			
Patients, N	2404	2385	5245
Median OS (95% CI), mo	26.6 (25.1–29.1)	23.4 (21.9–24.9)	27.3 (25.9–28.6)
Index from date of 2L initiation			
Patients, N	1209	1206	2575
Median OS (95% CI), mo	16.5 (15.9–18.6)	16.0 (14.3–16.8)	17.1 (16.1–18.2)
Index from date of 3L initiation			
Patients, N	549	548	1166
Median OS (95% CI), mo	11.2 (9.6–12.5)	10.2 (9.0–11.9)	11.8 (10.4–12.7)
Population with non-AGAs			
Index from date of 1L initiation			
Patients, N	8056	8003	12,391
Median OS (95% CI), mo	15.1 (14.4–15.7)	12.6 (12.0–13.2)	13.9 (13.5–14.5)
Index from date of 2L initiation			
Patients, N	3478	3468	4971
Median OS (95% CI), mo	11.3 (10.7–11.9)	10.1 (9.5–10.7)	10.2 (9.8–10.8)
Index from date of 3L initiation			
Patients, N	1426	1424	1949
Median OS (95% CI), mo	8.9 (8.2–9.7)	8.2 (7.6–8.9)	8.6 (7.9–9.1)

Across the CGDB and the FHRD with NGS, most patient samples were collected before 1L initiation. In both datasets, the majority of patients were sequenced before 2L initiation; however, a larger proportion of patients in the FHRD with NGS were sequenced before 1L initiation than the CGDB, and the interval between sample collection and NGS testing was shorter in the FHRD.

Table 5. NGS utilization by cohort

NGS utilization	CGDB (N = 13,011)	FHRD with NGS (N = 16,278)
Specimen site		
Metastatic	8353 (64.2)	2891 (17.8)
Primary	4658 (35.8)	6131 (37.7)
Unknown	-	7256 (44.6)
Specimen type		
Blood	3952 (30.4)	6894 (42.3)
Tissue	8594 (66.1)	9064 (55.7)
Unknown/other	465 (3.6)	320 (2.0)
Timing of specimen sample collection relevant to disease course		
Before 1L initiation	8598 (66.1)	12,499 (89.7)
After 1L initiation and before 2L initiation	1367 (10.5)	718 (5.2)
After 2L initiation and before 3L initiation	350 (2.7)	139 (1.0)
After 3L initiation	180 (1.4)	585 (4.2)
Missing	2516 (19.3)	-
Timing of NGS testing relevant to disease course		
Before 1L initiation	5540 (42.6)	10,153 (72.8)
After 1L initiation and before 2L initiation	4099 (31.5)	1743 (12.5)
After 2L initiation and before 3L initiation	596 (4.6)	178 (1.3)
After 3L initiation	263 (2.0)	1870 (13.4)
Missing	2513 (19.3)	-
Timing of NGS testing relevant to sample collection		
<2 wk	3661 (28.1)	9035 (55.5)
>2 to ≤4 wk	3654 (28.1)	3639 (22.4)
>4 to ≤12 wk	3960 (30.4)	2834 (17.4)
>12 wk	1730 (13.3)	770 (4.7)
Missing	6 (0.0)	-
FMI FoundationOne test platform version		
FoundationOne (T4b/T5a/T7)	500 (3.8)	NA
FoundationOne CDx (DX1/DX2)	8536 (65.6)	NA
FoundationOne Heme (D2/T6b)	23 (0.2)	NA
FoundationOne Liquid (CF2/CF3)	959 (7.4)	NA
FoundationOne Liquid CDx (AB1)	2993 (23.0)	NA
Lab name		
Guardant	NA	5012 (30.8)
Foundation Medicine	13,011 (100)	2981 (18.3)
In-house pathology lab	NA	1872 (11.5)
Caris Life Sciences	NA	1422 (8.7)
NeoGenomics Laboratories	NA	1284 (7.9)
Other	NA	3707 (22.8)

All data are n (%).
NA, not applicable.

Conclusions

- Among patients with advanced NSCLC, the demographic distribution was generally comparable between the FH-FMI CGDB and FHRD cohorts
- For patients with advanced NSCLC, treatment distribution and OS from different index dates of LoT initiation across LoTs were comparable between the CGDB and FHRD cohorts
- The findings suggest that the FH-FMI CGDB and FHRD cohorts provide complementary and comparable real-world evidence for NSCLC research; however, potential overlap in patient populations between the 2 cohorts cannot be ruled out

References

1. American Cancer Society. Lung Cancer Survival Rates. https://www.cancer.org/cancer/types/lung-cancer/detection-diagnosis-staging/survival-rates.html.
2. Flatiron. Clinico-Genomic Data is a Game Changer for Precision Oncology. https://resources.flatiron.com/real-world-evidence/clinico-genomic-data-is-a-game-changer-for-precision-oncology#:~:text=Real%2Dworld%20clinico%2Dgenomic%20data,of%20patients%20living%20with%20cancer.
3. Dienstmann R, et al. *ESMO Real World Data Digit Oncol.* 2025;7:100117.

Acknowledgments

This study was sponsored by Merck Sharp & Dohme LLC, a subsidiary of Merck & Co., Inc., Rahway, NJ, USA (MSD). The authors thank Lihai Song and Fansen Kong from the Merck BARDS programming team. Medical writing assistance was provided by John Teiber, PhD, of ICON plc (Blue Bell, PA, USA). This assistance was funded by MSD.

Contact information

Contact Dr Dan Lin at dan.lin@merck.com for questions or comments

Poster presented

https://bit.ly/4283lx1

Presented at the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Conference; Philadelphia, PA; May 17–20, 2026.