Leveraging Machine Learning to Assess the Association of Rash and Survival in Patients With Advanced NSCLC

MSR24

Scan for abstract

Qianyu Yuan[1], Aaron Dolor[1], Yunzhi Qian[1], Doug Donnelly[1], Melissa Estevez[1], Yulia Kuznetsova[1], Nisha Singh[1], Prakirthi Yerram[1]

[1]Flatiron Health, New York, NY

Background

- The association between rash and survival is well-documented for first and second-generation epidermal growth factor receptor-tyrosine kinase inhibitors (EGFR-TKIs), but less for third-generation
- This study leveraged machine learning (ML)-extracted real-world adverse events (rwAEs) to evaluate incidence and association between rash incidence and survival outcomes in patients with non-small cell lung cancer (NSCLC) treated with EGFR-TKIs

Results

- The analysis included 5606 patients. Baseline characteristics of the overall patient population and those stratified by first-, second-generation, and third-generation TKIs are shown in **Table 1**
- Compared with first- and second-generation TKIs, third-generation TKIs showed higher incidences of anemia, and QT prolongation and lower rash **(Table 2)**, aligning with clinical trials

Table 1. Basic Characteristics of Study Population

Characteristic	All patients N = 5606	1st generation n = 1988	2nd generation n = 669	3rd generation n = 2949
Age at diagnosis (years), n (%)				
<50	309 (5.5)	108 (5.4)	44 (6.6)	157 (5.3)
50-64	1619 (29.0)	601 (30.0)	194 (29.0)	824 (28.0)
65-74	1782 (32.0)	612 (31.0)	202 (30.0)	968 (33.0)
75+	1896 (34.0)	667 (34.0)	229 (34.0)	1000 (34.0)
Gender, n (%)				
Female	3804 (68.0)	1360 (68.0)	454 (68.0)	1990 (67.0)
Male	1801 (32.0)	628 (32.0)	214 (32.0)	959 (33.0)
Race[a], n (%)				
White	3192 (66.0)	1185 (66.0)	388 (65.0)	1619 (65.0)
Asian	741 (15.0)	244 (14.0)	80 (13.0)	417 (17.0)
Black or African American	415 (8.5)	142 (8.0)	58 (9.7)	215 (8.7)
Other Race	512 (10.0)	213 (12.0)	69 (11.0)	230 (9.3)
Stage at diagnosis[a], n (%)				
I	473 (8.4)	164 (8.2)	58 (8.7)	251 (8.5)
II	221 (3.9)	85 (4.3)	28 (4.2)	108 (3.7)
III	516 (9.2)	203 (10.0)	71 (11.0)	242 (8.2)
IV	4281 (76.0)	1488 (75.0)	493 (74.0)	2300 (78.0)
Unknown/Other	115 (2.1)	48 (2.4)	19 (2.8)	48 (1.6)
Histology[a], n (%)				
Non-squamous cell carcinoma	5424 (97.0)	1923 (97.0)	647 (97.0)	2854 (97.0)
NSCLC histology NOS	92 (1.6)	44 (2.2)	11 (1.6)	37 (1.3)
Squamous cell carcinoma	90 (1.6)	21 (1.1)	11 (1.6)	58 (2.0)
ECOG[b], n (%)				
0	1463 (26.0)	400 (20.0)	185 (28.0)	878 (30.0)
1	1874 (33.0)	598 (30.0)	222 (33.0)	1054 (36.0)
2 - 4	840 (15.0)	276 (14.0)	97 (14.0)	467 (16.0)
Unknown	1429 (25.0)	714 (36.0)	165 (25.0)	550 (19.0)

[a] Variables with *P* value <.05; [b] Variables with *P* value <.001.
Abbreviations: ECOG, Eastern Cooperative Oncology Group; NOS, not otherwise specified; NSCLC, non-small cell lung cancer.

Table 2. Incidence of AEs Across TKI Generations

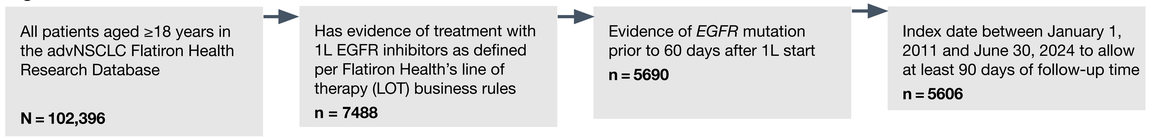
Incidence (%)	All patients N = 5606	1st generation n = 1988	2nd generation n = 669	3rd generation n = 2949
Alopecia	10.0	14.0	8.2	8.3
Amylase increase	<0.1	<0.1	0.0	<0.1
Anemia	20.0	15.0	18.0	23.0
Conjunctivitis	1.8	2.9	1.8	1.0
Cough	31.0	33.0	26.0	32.0
Diarrhea	56.0	57.0	72.0	52.0
Diplopia	2.0	1.9	1.5	2.2
Dry eye	0.9	0.9	0.6	1.1
Dyspnea	29.0	30.0	22.0	31.0
Hand foot syndrome	0.4	0.4	1.0	0.2
Increased bilirubin	1.9	4.1	0.6	0.7
Increased lacrimation	0.8	1.2	0.1	0.7
Keratitis	0.4	0.4	0.6	0.3
Leukopenia	4.0	2.0	1.2	5.9
Lipase increase	0.4	0.4	0.6	0.4
Liver failure	0.1	0.2	0.1	0.1
Loss of appetite	38.0	38.0	34.0	38.0
Nausea/vomiting	35.0	36.0	35.0	34.0
Photophobia	0.9	0.5	0.4	1.2
Photopsia	0.4	0.2	0.7	0.4
Pneumonitis	5.1	3.1	4.3	6.6
Pruritus	22.0	28.0	23.0	18.0
QT prolongation	1.0	<0.1	0.1	1.8
Rash	51.0	68.0	60.0	37.0
Stomatitis	10.0	7.5	21.0	9.2
Uveitis	0.1	0.1	0.1	0.1

Abbreviations: AEs, adverse events.

Methods

- **Data source:** The US-based, longitudinal Flatiron Health Research Database—an electronic health record-derived, deidentified database, with patient-level data originated from ~280 US cancer clinics (~800 sites of care; primarily community oncology settings) and curated via technology-enabled abstraction[1]
- **Setting:** The study included adults aged ≥18 years with advanced *EGFR*-mutated NSCLC, treated with first-line (1L) EGFR-TKI monotherapy between January 1, 2011, and June 30, 2024 **(Figure 1)**
- **Main outcome measures:** A natural language processing model was used to extract rwAEs
- **Statistical analysis:** Descriptive statistics were used to compare the incidence of 37 rwAEs overall and by TKI generation. Kaplan-Meier and Cox models evaluated the association between rash incidence and real-world overall survival (rwOS) and progression-free survival (rwPFS). This study also evaluated International Classification of Diseases (ICD) codes and ML extraction, alone and combined, for identifying rash and its relationship with survival outcomes

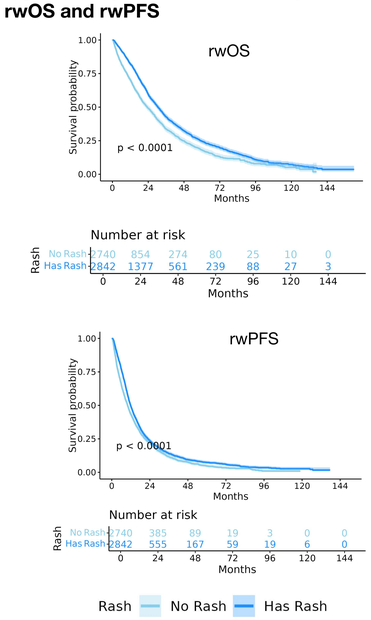
Figure 1. Cohort Selection



Abbreviations: 1L, first-line; advNSCLC, advanced non-small cell lung cancer; EGFR, epidermal growth factor receptor.

Results (Continued)

Figure 2. Impact of Rash Development on rwOS and rwPFS



- Rash development was associated with improved rwOS and rwPFS for patients who received EGFR TKIs **(Figure 2)**
- Specifically, patients with rash showed an improved rwOS and rwPFS in overall population. Notably, this association was even stronger with third-generation TKIs **(Table 3 and Table 4)**
- Using ICD codes alone showed lower rash incidence (11%) than combining ML extraction with ICD codes (52%). Rash development utilizing either method, was associated with improved rwOS

Table 3. Adjusted Hazard Ratios for rwOS Comparing Rash vs No Rash

Rash	All study population		1st generation TKI		2nd generation TKI		3rd generation TKI	
	Adjusted HR (95% CI)	*P* value	Adjusted HR (95% CI)	*P* value	Adjusted HR (95% CI)	*P* value	Adjusted HR (95% CI)	*P* value
No	ref		ref		ref		ref	
Yes	0.75 (0.7-0.81)	<.01	0.73 (0.65-0.82)	<.01	0.73 (0.59-0.89)	<.01	0.64 (0.57-0.72)	<.01

Table 4. Adjusted Hazard Ratios for rwPFS Comparing Rash vs No Rash

Rash	All study population		1st generation TKI		2nd generation TKI		3rd generation TKI	
	Adjusted HR (95% CI)	*P* value	Adjusted HR (95% CI)	*P* value	Adjusted HR (95% CI)	*P* value	Adjusted HR (95% CI)	*P* value
No	ref		ref		ref		ref	
Yes	0.85 (0.8-0.9)	<.01	0.81 (0.72-0.89)	<.01	0.66 (0.55-0.79)	<.01	0.73 (0.67-0.81)	<.01

Abbreviation: HR, hazard ratio.

Conclusions

- Using ML models successfully scaled multiple AEs across a large patient cohort, aligning the observed AE rates with clinical expectations[2]
- The study confirmed the correlation between the incidence of rash and improved survival outcomes in all three TKI generations
- The use of ICD codes for rash showed limited completeness, suggesting that relying solely on this method may under represent the actual incidence of AEs

Disclosures: This study was sponsored by Flatiron Health, Inc.—an independent member of the Roche Group. During the study period, QY, AD, YQ, DD, ME, YK, NS, and PY reported employment with Flatiron Health, Inc. and stock ownership in Roche. Data first presented at ISPOR 2025 in Montreal, QC, Canada on May 14, 2025

Author contact information: Qianyu Yuan, qianyu.yuan@flatiron.com

Reference

1. Flatiron Health. Database Characterization Guide. Flatiron.com. Published March 18, 2025. Accessed April 18, 2025. https://flatiron.com/database-characterization
2. Ramalingam SS, et al. *N Engl J Med*. 2019;382(1):41-50. doi:10.1056/nejmoa1913662