Poster Code: MSR79

Evaluation of real-world response rate in clinical trial-aligned cohorts of patients with lung, colorectal and breast cancer using machine learning

Qianyi Zhang, MS[1]; Konstantin Krismer, PhD[1]; Xiaoyan Wang, MS[1]; Christina Fullerton, PharmD, BCOP[1]; Aaron Dolor, PhD[1]; Niquelle Brown Wadé, PhD[1]; Prakrit Baruah, BS[1]; Erica Yim, BA[1]; Tori Williams, BA[1]; Qianyu Yuan, PhD[1]; Samuel Wilkinson, BA[1]; Sheila Nemeth, PhD[1]; Nisha Singh, MS[1]; Mengru Wang, MS[1]; Madeline Richey, MPH[1]; Aaron B Cohen, MD, MSCE,[1]; Erin Fidyk, MSN, MBA[1]; Auriane Blarre, MS[1]; Kelly Magee, MS[1]

[1] Flatiron Health, New York, NY

Scan for Abstract, Digital Poster, and Supplemental Materials

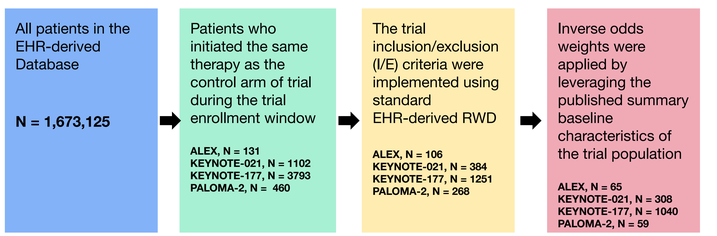
Background

- Response to treatment and related endpoints are essential to oncology clinical research. Rapid evidence generation in real-world cohorts can inform clinical trial study design and drug development.
- We developed a real-world response (rwR) approach ("Scaled rwR") by leveraging natural language processing (NLP)-based deep learning models trained on expert human abstracted data to generate rapid insights across large cohorts of patients.
- We evaluated real-world response rate (rwRR) using Scaled rwR in clinical trial-aligned cohorts of patients with lung, colorectal, and breast cancer across therapy classes to better understand relationships between response-based endpoints in clinical trials and real-world data (RWD)

Methods

- Data sources:
  - The US nationwide Flatiron Health electronic health record(EHR)-derived deidentified database, comprising patient-level structured and unstructured data,[1,2] originating from ~280 cancer clinics (~800 sites of care). The majority of patients in the database originate from community oncology settings.
  - The published results from the control arms of ALEX, KEYNOTE-021, KEYNOTE-177, and PALOMA-2 trials
- Inclusion criteria: Real-world cohorts were generated to align with the patient population in the control arm of each trial (Figure 1)
- Variable: Scaled rwR was generated by leveraging NLP-based deep learning models trained on expert human abstracted data (training set N = ~12k patients) to extract clinician's documentation of change in disease burden (i.e., complete response, partial response, stable disease, progressive disease, unknown) at each instance of disease assessment imaging.
- Statistical methods: rwRR (proportion of patients with at least one rw partial response (rwPR) or rw complete response (rwCR) assessment determination during the course of treatment) was calculated with and without real-world confirmation (a subsequent rwPR, rwCR or rw stable disease (rwSD)) including and excluding patients with no response assessments as non-responders.

Figure 1. Cohort Selection



References
1. Ma X et al. MedRxiv. 2023. doi:10.1101/2020.03.16.20037143
2. Birnbaum B et al. arXiv. 2020. doi:10.48550/arxiv.2001.09765.
3. Ma X, Bellomo L, Magee K, et al. Characterization of a Real-World Response Variable and Comparison with RECIST-Based Response Rates from Clinical Trials in Advanced NSCLC. Adv Ther 2021;38:1843-1859

Acknowledgments: Madeline Morenberg (Flatiron Health, Inc.) provided design support.
Disclosures: This study was sponsored by Flatiron Health, Inc.—an independent member of the Roche Group. During the study period, QZ, KK, XW, CF, AD, NBW, PB, EY, TW, QY, SW, SN, NS, MW, MR, ABC, EF, AB, KM reported employment with Flatiron Health, Inc. and stock ownership in Roche.
Author Contact Info: Qianyi Zhang, qianyi.zhang@flatiron.com

Scaled real-world response enhances understanding of relationships between response-based endpoints in clinical trials and RWD

Table 2. Objective Response Rate (ORRs) vs rwRRs

Trial Results		Real-world cohort after weighting			
N	ORR (%)	Include vs Exclude patients with no response assessments as non-responders	N	rwRR (%)	rwRR with confirmation (%)
ALEX -- *First line crizotinib in ALK positive aNSCLC*					
151	75.5 (67.8, 82.1) **w/o confirmation	Include	64.6	37.9 (27.1, 50.1)	29.9 (20.1, 42.0)
		Exclude	56.8	43.2 (31.1, 56.1)	34.1 (23.1, 47.0)
KEYNOTE-021 -- *First line carboplatin plus pemetrexed in aNSCLC*					
63	28.6 (18, 41) **w/ confirmation	Include	308.3	43.9 (38.5, 49.5)	31.3 (26.4, 36.7)
		Exclude	263.6	51.4 (45.4, 57.4)	36.6 (31.1, 42.6)
KEYNOTE-177 -- First line FOLFOX/FOLFIRI +/- bevacizumab or cetuximab in dMMR/MSI-H mCRC					
154	33.1 (25.8, 41.1) **w/ confirmation	Include	1040.0	51.8 (48.8, 54.9)	37.0 (34.1, 39.9)
		Exclude	870.2	62.0 (58.7, 65.1)	44.2 (40.9, 47.5)
PALOMA-2 -- First line letrozole in ER+/HER2- mBC					
222	34.7 (28.4, 41.3) **w/ confirmation	Include	59.0	35.9 (24.9, 48.7)	19.3 (11.2, 31.1)
		Exclude	45.2	46.9 (33.1, 61.1)	25.2 (14.8, 39.4)

Results

- Participants: Cohorts totals are displayed in Table 1
- Results of rwRRs vs ORRs for each trial were shown in Table 2
- The absolute differences between weighted rwRRs including patients with no response assessments as non-responders, with confirmation when required per trial protocol vs ORRs were -37.6%, 2.3%, 3.9%, -15.4%

Conclusion

- ORRs and rwRRs aligned closely in the control arms of KEYNOTE-021/-177, with chemotherapy as the regimens of interest.
- We observed differences in ALEX and PALOMA-2, where oral targeted and hormonal therapies were the regimens of interest. This discrepancy between rwRR and ORR likely reflects inherent differences between clinical trials and RWD.
- These results highlight the considerations in identifying I/E criteria that may be more impactful depending on the patient population, also generate signals that alignment of real world and trial cohorts and outcomes may vary by therapy class.

Discussion

- The real-world cohorts were contemporaneous with trial enrollment periods and optimized through weighting to enhance their relevance.
- Applying multiple approaches to calculate rwRR improves the robustness of results and offers more context to interpret comparisons between real-world and trial outcomes.
- Although inherent differences exist between rwR and RECIST-based response, several hypotheses may explain the observed variation between rwRR and ORR by cohort (e.g., infeasibility of replicating some trial I/E criteria, discrepancies in biomarker status determination, differences in oral therapy adherence, and differences in imaging cadence).
- Previous studies (e.g., comparing rwRR vs ORR in ALEX and KEYNOTE-021[3], both using detailed rwR) have provided similar results which further affirmed the study's findings and hypotheses.

Table 1. Baseline Characteristics Table*

Characteristics	Real-world cohort before weighting	Real-world cohort after weighting	Trial
ALEX			
Total N	106	64.6	151
Median Age	59.85 (13.2)	55.57 (14.3)	54.0 (25, 88)
Gender			
Female	59 (55.7)	37.5 (58.0)	87 (58.0)
Male	47 (44.3)	27.1 (42.0)	64 (42.0)
ECOG performance status			
0 or 1	55 (50.9)	60.1 (93.0)	141 (93.0)
2	11 (10.4)	0.8 (1.3)	10 (7.0)
Missing	40 (37.7)	3.7 (5.7)	--
Smoking status			
Active smoker	46 (43.4)	22.6 (35.0)	5 (3.0)
Former smoker			48 (32.0)
No history of smoking	60 (56.6)	42.0 (65.0)	98 (65.0)
KEYNOTE-021			
Total N	384	308.3	63
Median Age	66.40 (9.5)	63.65 (9.5)	63.2 (58, 70)
Gender			
Female	174 (45.3)	181.9 (59.0)	37 (59.0)
Male	210 (54.7)	126.4 (41.0)	26 (41.0)
Race			
White	283 (73.7)	283.7 (92.0)	58 (92.0)
Black or African American	38 (9.9)	11.8 (3.8)	0 (0.0)
Other	63 (16.4)	12.8 (4.2)	5 (8.0)
ECOG performance status			
0	98 (25.5)	69.1 (22.4)	29 (46.0)
1	183 (47.7)	166.5 (54.0)	34 (54.0)
Missing	103 (26.8)	72.7 (23.6)	0 (0.0)
Smoking status			
History of smoking	355 (92.4)	265.2 (86.0)	54 (86.0)
No history of smoking	27 (7.0)	40.3 (13.1)	9 (14.0)
Unknown/Not documented	2 (0.5)	2.8 (0.9)	0 (0.0)
KEYNOTE-177			
Total N	1251	1040	154
Median age	61.00 (11.4)	61.60 (11.4)	71 (46.0)
Gender = Male	730 (58.4)	551.2 (53.0)	82 (53.0)
ECOG performance status = 0	488 (39.0)	572.0 (55.0)	84 (55.0)
Treatment			
FOLFIRI	78 (6.2)	114.4 (11.0)	16 (19.0)
FOLFIRI,Bevacizumab	293 (23.4)	260.0 (25.0)	36 (23.0)
FOLFIRI,Cetuximab	37 (3.0)	83.2 (8.0)	11 (7.0)
FOLFOX	286 (22.9)	83.2 (8.0)	11 (7.0)
FOLFOX,Bevacizumab	546 (43.6)	468.0 (45.0)	64 (42.0)
FOLFOX,Cetuximab	11 (0.9)	31.2 (3.0)	5 (3.0)
PALOMA-2			
Total N	268	59	222
Median age	66.51 (10.9)	62.26 (12.5)	61 (28, 88)
Race			
Asian	7 (2.6)	8.3 (14.0)	30 (13.5)
White	187 (69.8)	46.0 (78.0)	172 (77.5)
Other	74 (27.6)	4.7 (8.0)	20 (9.0)
ECOG performance status			
0	56 (20.9)	27.1 (46.0)	102 (45.9)
1	43 (16.0)	31.3 (53.0)	117 (52.7)
2	17 (6.3)	0.0 (0.0)	3 (1.4)
Missing	152 (56.7)	0.6 (1.0)	--
Disease stage at initial diagnosis			
I	37 (13.8)	8.3 (14.0)	30 (13.5)
II	65 (24.3)	18.3 (31.0)	68 (30.6)
III	34 (12.7)	10.6 (18.0)	39 (17.6)
IV	96 (35.8)	16.1 (27.3)	72 (32.4)
Unknown/Not documented	36 (13.4)	5.7 (9.7)	13 (5.9)
No. of disease sites			
1	153 (61.0)	17.7 (33.1)	66 (29.7)
2	64 (25.5)	13.6 (25.4)	52 (23.4)
3	27 (10.8)	16.5 (30.9)	61 (27.5)
>= 4	7 (2.8)	5.7 (10.6)	43 (19.4)
Disease site			
Visceral	78 (31.1)	29.5 (55.2)	110 (49.5)
Non visceral	173 (68.9)	24.0 (44.8)	112 (50.5)
Bone only	115 (45.8)	13.0 (24.3)	48 (21.6)

*Categorical variables are N (%) and continuous variables are median (IQR) with the exception of age in ALEX, KEYNOTE-021 and PALOMA-2 which appear to be median (range); the categories of variables were reported to match the original trial and may exhibit variability across trials