



Generating Real-World Tumor Burden Endpoints from Electronic Health Record Data: Comparison of RECIST, Radiology-Anchored, and Clinician-Anchored Approaches for Abstracting Real-World Progression in Non-Small Cell Lung Cancer

Sandra D. Griffith · Melisa Tucker · Bryan Bowser · Geoffrey Calkins · Che-hsu (Joe) Chang · Ellie Guardino · Sean Khozin · Josh Kraut · Paul You · Deb Schrag · Rebecca A. Miksad

Received: March 18, 2019 / Published online: May 28, 2019
© The Author(s) 2019

ABSTRACT

Introduction: Real-world evidence derived from electronic health records (EHRs) is increasingly recognized as a supplement to evidence generated from traditional clinical trials. In oncology, tumor-based Response Evaluation Criteria in Solid Tumors (RECIST) endpoints are standard clinical trial metrics.

Enhanced Digital Features To view enhanced digital features for this article go to <https://doi.org/10.6084/m9.figshare.8026466>.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s12325-019-00970-1>) contains supplementary material, which is available to authorized users.

S. D. Griffith · M. Tucker · B. Bowser · G. Calkins · Che-hsu (Joe) Chang · J. Kraut · P. You · R. A. Miksad (✉)
Flatiron Health Inc., New York, NY, USA
e-mail: rmiksad@flatiron.com

E. Guardino
Genentech, Inc., San Francisco, CA, USA

S. Khozin
United States Food and Drug Administration, Silver Spring, MD, USA

D. Schrag
Division of Population Sciences, Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA

D. Schrag
Harvard Medical School, Boston, MA, USA

The best approach for collecting similar endpoints from EHRs remains unknown. We evaluated the feasibility of a RECIST-based methodology to assess EHR-derived real-world progression (rwP) and explored non-RECIST-based approaches.

Methods: In this retrospective study, cohorts were randomly selected from Flatiron Health's database of de-identified patient-level EHR data in advanced non-small cell lung cancer. A RECIST-based approach tested for feasibility ($N = 26$). Three non-RECIST approaches were tested for feasibility, reliability, and validity ($N = 200$): (1) radiology-anchored, (2) clinician-anchored, and (3) combined. Qualitative and quantitative methods were used.

Results: A RECIST-based approach was not feasible: cancer progression could be ascertained for 23% (6/26 patients). Radiology- and clinician-anchored approaches identified at least one rwP event for 87% (173/200 patients). rwP dates matched 90% of the time. In 72% of patients (124/173), the first clinician-anchored rwP event was accompanied by a downstream event (e.g., treatment change); the association was slightly lower for the radiology-anchored approach (67%; 121/180). Median overall survival (OS) was 17 months [95% confidence interval (CI) 14, 19]. Median real-world progression-free survival (rwPFS) was 5.5 months (95% CI 4.6, 6.3) and 4.9 months (95% CI 4.2, 5.6) for clinician-anchored and radiology-anchored approaches, respectively. Correlations

between rwPFS and OS were similar across approaches (Spearman's ρ 0.65–0.66). Abstractors preferred the clinician-anchored approach as it provided more comprehensive context.

Conclusions: RECIST cannot adequately assess cancer progression in EHR-derived data because of missing data and lack of clarity in radiology reports. We found a clinician-anchored approach supported by radiology report data to be the optimal, and most practical, method for characterizing tumor-based endpoints from EHR-sourced data.

Funding: Flatiron Health Inc., which is an independent subsidiary of the Roche group.

Keywords: Carcinoma, non-small cell lung; Endpoints; Immunotherapy; PD-1; PD-L1; Real-world evidence

INTRODUCTION

As oncology clinical trials have become increasingly complex, their efficacy results may not directly translate into real-world effectiveness [1]. Real-world evidence (RWE) sources, including data derived from electronic health records (EHRs), can augment insights from traditional clinical trials by better understanding the realities of real-world patients [2, 3]. RWE can be leveraged to improve therapy development programs, increasing the external validity of evidence available to support treatment decisions, and the FDA recently issued the Framework for its Real-World Evidence Program [4–7]. The key to fully unlocking the value of the EHR for these purposes is reliable and obtainable outcome metrics for patients in real-world settings. However, whether traditional cancer endpoints other than mortality can be gleaned from EHR data remains unknown. We need to determine whether we can apply traditionally defined clinical trial tumor endpoints to EHR data, or whether new definitions are needed.

Endpoints based on tumor size changes are often treatment efficacy metrics in solid tumor clinical trials [8, 9], where radiographic images are evaluated using the Response Evaluation

Criteria in Solid Tumors (RECIST) and reviewed by an independent central committee to improve assessment objectivity [10, 11]. RECIST assesses size changes of designated tumors over time (“target lesions”) combined with the presence or absence of new tumors to categorize patients’ disease status: “response,” “stable disease,” or “progression.”

Routine clinical practice follows a similar assessment paradigm. Treatment effectiveness is often determined by periodically evaluating a variety of clinical parameters (e.g., imaging, physical exam, biomarkers, pathology specimen, patient-reported concerns). Depending on the clinical context, EHR documentation of changes in one or more of these parameters may reflect the outcome of an intervention and tumor burden dynamics may be more globally summarized as “improved” (i.e., “tumor response”), “no change” (i.e., “stable disease”), or “worse” (i.e., “tumor progression,” an event for which the patient may need a new treatment). In addition, outcome appraisals recorded in the EHR may also incorporate quality of life metrics, and other quantitative and qualitative assessments and test results.

EHR data provide significant opportunities to develop real-world data (RWD)-specific metrics but may also add challenges to the interpretation of real-world outcomes. For example, the relevant information may reside within structured sources (i.e., highly organized data, such as white blood cell count) or unstructured (e.g., free-text clinical notes or radiology or biomarker reports). In other words, the EHR holds information, but differently from a clinical trial data set. Therefore, we may need to process the available EHR data in a different way, culling it through manual curation by trained abstractor experts following precise, pre-defined policies and procedures, through computer-based algorithms mimicking the manual approach or through a combination of both.

Which EHR evidence source is best to anchor cancer outcomes against? Radiology reports may not include declarative RECIST assessments but typically have an “Impression” section written by the reviewing radiologist. Clinician assessments, findings, and interpretation of relevant results and reports are often

summarized in EHR notes by the clinician caring for the patient, especially when they inform treatment decisions. Therefore, clinician notes may serve as an alternative or supplement to radiology reports to curate cancer outcomes from EHR data.

Our objective was to identify a practical and efficient large-scale data abstraction method for estimation of cancer outcomes from the EHR. As a result of the variety of EHR formats and locations where documentation of outcome assessments may exist, such an abstraction approach must (1) be applicable across multiple EHR systems; (2) be amenable to manual and/or electronic abstraction from unstructured documents; (3) accommodate clinical judgement nuances; (4) be reliable despite possibly missing data points; and (5) be efficient to support scaling for large cohort assessments.

We anticipated that traditional clinical trial approaches to collecting endpoints, such as cancer progression, may need to be modified for the unique features of EHR data. We tested this hypothesis in a preliminary experiment in a small cohort of patients with advanced non-small cell lung cancer (aNSCLC). We then compared several alternative approaches for abstraction of cancer progression events from the EHR in a larger cohort of aNSCLC patients.

METHODS

We conducted experiments to answer the following questions: (1) can RECIST be feasibly applied to EHR cancer progression data, and (2) how do alternate abstraction approaches perform?

Data Sources

The overall cohort was selected from Flatiron Health's longitudinal EHR-derived database, which, at the time of this study, included over 210 cancer clinics representing more than 1.2 million active patients across the USA. We identified more than 120,000 and 25,000 patients diagnosed with lung cancer and aNSCLC, respectively. The majority of patients were treated in community oncology clinics.

Demographic, clinical, and outcomes data were extracted out of the source EHR, including structured data and unstructured documents. To create the database, we aggregated, normalized, de-identified, and harmonized patient-level data. Data were processed centrally and stored in a secure format and environment. Structured data (e.g., treatments, labs, diagnosis codes) were mapped to standard ontologies. Dates of death were obtained from a composite mortality variable comprising the EHR structured data linked to commercial mortality data and the Social Security Death Index [12]. Unstructured data (e.g., clinician notes, radiology reports, death notices) were extracted from EHR-based digital documents via "technology-enabled" chart abstraction [13]. Every data point sourced from unstructured documents was manually reviewed by trained chart abstractors (clinical oncology nurses and tumor registrars, with oversight from medical oncologists). Quality control included duplicate chart abstraction of a sample of abstracted variables as well as logic checks based on clinical and data considerations.

Study Design and Data Collection

This retrospective observational study investigated methods to assess cancer progression from EHR data through two objectives (Fig. 1): (1) to evaluate the feasibility of a RECIST approach (experiment 1) and (2) if a RECIST approach is not feasible, to evaluate three alternative non-RECIST abstraction approaches (experiment 2).

Inclusion criteria for the study cohort from which patients were randomly selected for both experiments were (1) NSCLC patients diagnosed with advanced disease between January 1, 2011 and April 1, 2016; (2) at least two clinical visits on or after January 1, 2011 documented in the EHR; and (3) documentation of initiation of at least two lines of systemic therapy after advanced diagnosis. Advanced NSCLC was defined as a diagnosis of stage IIIB or metastatic stage IV, or recurrent early disease. At the time of this study, staging of patients diagnosed with aNSCLC followed the criteria from the

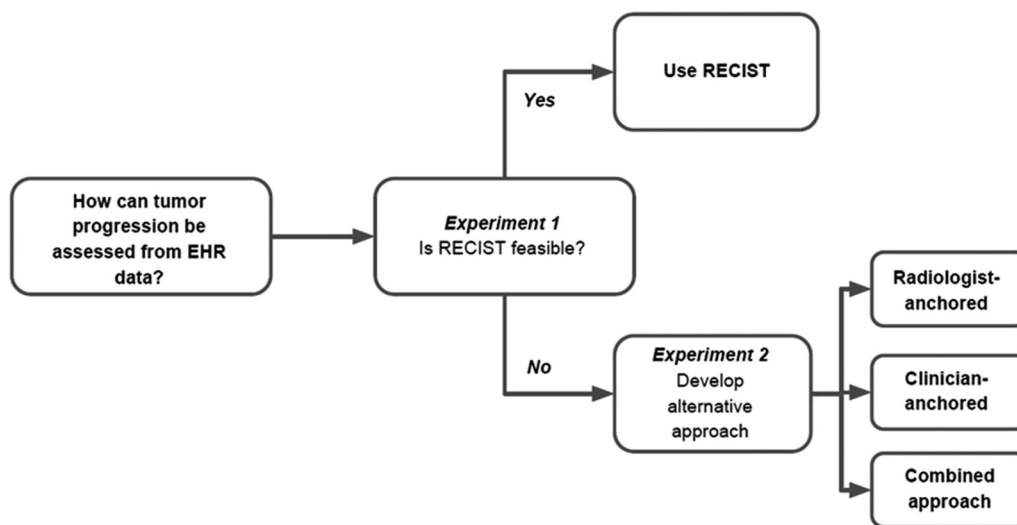


Fig. 1 Using the EHR to generate a cancer progression endpoint

American Joint Commission of Cancer/Union for International Cancer Control/International Association for the Study of Lung Cancer (AJCC/UICC/IASLC) 7th edition manual [14].

To assess the feasibility of the RECIST approach (experiment 1), a total of 26 patients were randomly selected. The sample size was chosen to achieve expected theme saturation consistent with an initial feasibility study to assess usability [15, 16]. To study the non-RECIST abstraction approaches (experiment 2), 200 patients were randomly selected. The sample size was chosen to balance feasibility and the need for a sufficient number of progression and death events to inform descriptive analyses.

Institutional review board (IRB) approval of the study protocol (IRB # RWE-001, “The Flatiron Health Real-World Evidence Parent Protocol”, Tracking # FLI1-18-044) by the Copernicus Group IRB, with waiver of informed consent, was obtained prior to study conduct, and covers the data from all sites represented.

Experiment 1: Feasibility of RECIST Criteria

To determine if the elements required for RECIST version 1.1 [11] evaluation could be abstracted from patient charts (i.e., usability), we evaluated radiology reports in the EHR for

baseline and first treatment response imaging assessment after first line (1L) systemic therapy initiation. Radiology reports were evaluated for explicit descriptions of “target lesions” as required by RECIST. The subgroup of patients for whom an explicit description of target lesions was not found were also evaluated with a more lenient approach: abstractors identified target lesions on the basis of available radiology reports. Lesions of at least 1 cm were classified as measured if size was numerically documented in the radiology report. The RECIST approach was determined to be potentially feasible for a patient if (1) their baseline scan was conducted within 2 months prior to the referenced 1L therapy start date; (2) their initial follow-up scan was performed at least 28 days after therapy start date; (3) documentation indicated that the follow-up scan was compared to the baseline scan; (4) documentation existed for lesion measurements on both the baseline and follow-up scans; and (5) all measured and non-measured lesions were specifically described in each imaging report. Although RECIST can be applied to multiple imaging modalities, CT is preferred for evaluation of lung cancer [11]. Therefore, we focused on chest CT and PET/CT radiology reports.

Experiment 2: Comparison of Alternative Abstraction Methods for Assessing Real-World Cancer Progression

Three approaches for determining real-world cancer progression were used to abstract outcomes data from individual charts: (1) a radiology-anchored approach; (2) a clinician-anchored approach; and (3) a combination of both. The radiology-anchored approach, conceptually similar to RECIST, focused on identifying new lesions and quantifying tumor burden changes observed by imaging, but was optimized for the type of radiological sources routinely found in EHR data (e.g., “Impression” section within radiology reports). Reports from any imaging modality were considered. The clinician-anchored approach was conceptually focused on the clinician as synthesizer of signals from the entire patient’s chart. The combined approach considers both radiology reports and the clinician’s synthesis: the date of progression is determined by the earliest source document for the event. Definitions of each approach, evidence sources evaluated, and the logic for calculating progression dates are described in Table 1.

Analysis

For experiment 1, descriptive statistics (frequencies and percentages) were calculated for patients meeting RECIST feasibility criteria. Results were summarized for the percentage of cases that met all criteria, under both the strictest definition (explicit target lesion mention) and the more lenient definition. The saturation sampling completeness threshold, a qualitative research method, was 75% (i.e., at least 75% of patients had sufficient data to allow radiology reports for the RECIST approach to be usable).

For experiment 2, demographics, clinical and tumor characteristics, and treatment types were summarized using quantitative descriptive statistics (medians and interquartile ranges or frequencies and percentages, as appropriate). Progression events within the first 14 days of 1L therapy start were excluded from all

calculations as they occurred too early to reflect treatment effectiveness.

To compare the pre-defined abstraction approaches (Fig. 1), the proportions of patients with a progression event with accompanying 95% confidence intervals (CIs) were calculated. The concordance of progression events and associated dates were also assessed for each approach. Discordant cases were reviewed to determine the discrepancy sources. Among patients with at least one progression event, we assessed frequency (95% CIs) of near-term (within 15 days prior to treatment initiation and within 60 days post progression date), clinically relevant downstream events: death, start of new therapy line (second or subsequent lines), or therapy stop.

RWD-based time-to-event analyses for PFS (rwPFS) and time to progression (rwTTP) as well as overall survival (OS) were performed using standard Kaplan–Meier methods [14]. The index date was defined as the date of 1L therapy start, and the event date was defined as the date of the first progression event (rwTTP); the first progression event or date of death (rwPFS); or the date of death due to any cause (OS). In cases with inter-abstractor disagreement on the event date, the date from the abstractor who documented the event first was used in the Kaplan–Meier analysis. Patients lost to follow-up or without the relevant event by the end of the study period were censored at the date of last confirmed activity (last clinical visit or drug administration date). Medians, 95% CIs, and Kaplan–Meier curves were calculated for each outcome.

To assess the correlation between rwPFS and OS, the cohort was restricted to patients with a documented death event ($N = 123$). For the correlation between rwTTP and OS, the cohort was further restricted to patients with documentation of both progression and death events ($N = 112$ for the clinician anchored approach; 113 for the radiology anchored approach and for the combined approach). Spearman’s rank correlation coefficient (Spearman’s rho) was used for all correlation analyses.

The reliability of progression abstraction was assessed by calculating inter-rater agreement for the first progression event in a random subset of

Table 1 Non-RECIST-based approaches to determining cancer progression using EHR data (experiment 2)

	Radiology-anchored approach^a	Clinician-anchored approach^a	Combined approach^a
Definition	Documented in the radiology report as progression based on the radiologist’s interpretation of the imaging	Documented in a clinician’s note as cancer progression based on the clinician’s interpretation of the entire patient chart, including diagnostic procedures and tests	Documented in either the radiology report or clinician note as cancer progression
Primary and corroborating evidence sources	Primary: radiology reports	<i>Primary:</i> clinician notes <i>Corroborating:</i> radiology reports	<i>Primary:</i> radiology reports and clinician notes
Cancer progression date	Date of the first radiology report that indicated a progression event	Date of the first radiology report referenced by the clinician assessment when available, or date of clinician note when no corresponding radiology was conducted or documented	The earliest report date of the available sources of documentation of a progression event

Three different abstraction approaches for determining real-world cancer progression: (1) radiology-anchored approach, (2) clinician-anchored approach, and (3) combined approach. For each abstraction approach, the approach definitions, source evidence evaluated, and progression date assignment rules are described

^a Each approach also considered pathology reports as potential evidence for progression. However, there were no instances of pathology reports preceding radiology reports in the cohort analyzed for this study. In addition, there were no instances of conflicting information between radiology and pathology reports. For simplicity of presentation, pathology reports were excluded from this description as a potential evidence source

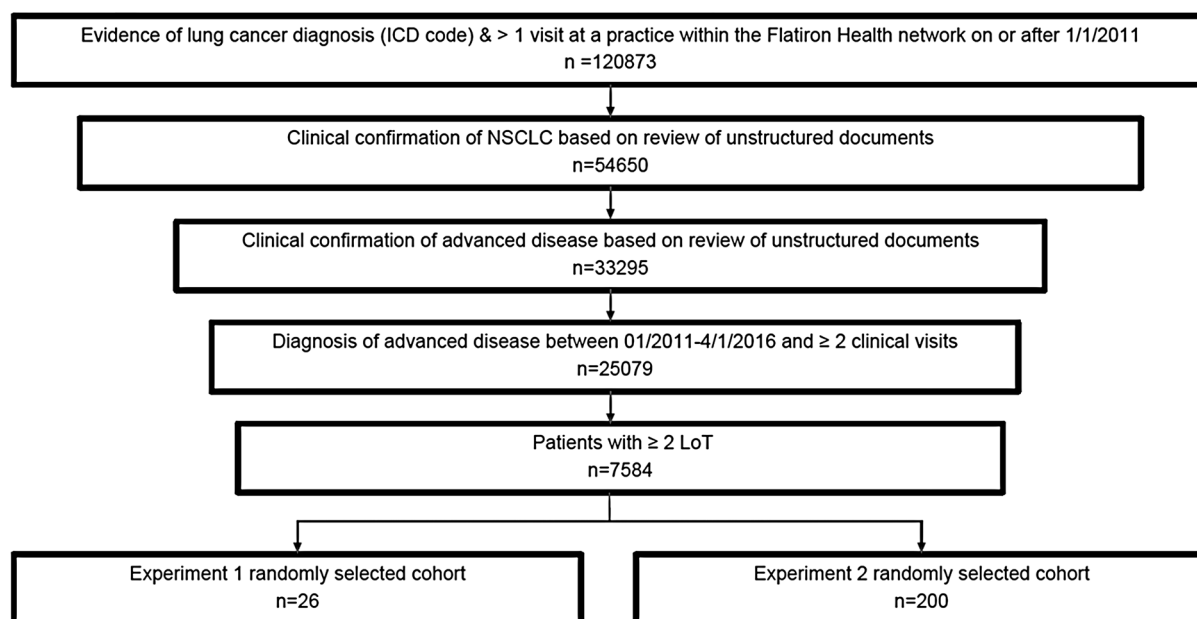
patients ($N = 55$) abstracted in duplicate (two independent abstractors per chart). Patients may have had any number of progression events within or across multiple lines of therapy, but the inter-rater agreement analysis included only up to the first event. Agreement was calculated at patient level and considered from multiple perspectives: (1) agreement on the presence or absence of at least one progression event (event agreement), regardless of date; (2) in cases where both abstractors found at least one progression event (cases with event disagreement or no events were excluded), agreement on when the progression occurred (date agreement); and, (3) a combined approach where both absence or presence of a progression event and the event date, if one was found, contribute toward agreement (overall agreement). Date and overall agreement were calculated for exact date, as well as for 15- and 30-day windows between dates.

Open-ended qualitative feedback was collected from abstractors regarding the usability, feasibility, and consistency of each approach.

RESULTS

Experiment 1: Feasibility of RECIST Criteria

Experiment 1 evaluated a cohort of 26 aNSCLC patients (Fig. 2). With a strict RECIST definition (radiologist-defined target lesions required in imaging report), no patient chart (0%) yielded data suitable for assessing cancer progression. With more lenient criteria (no explicit mention of target lesions required), only 15 charts (58% of experiment 1 cohort) had radiology reports with descriptions appropriate for RECIST assessment. Of these, only 8 (31% of experiment 1 cohort) directly compared all measured



Appropriate scans were defined as a baseline scan (PET/CT or CT chest) within 2 months prior to therapy start AND a scan (PET/CT or CT chest) to follow-up response at least 28 days after therapy start.
To be "directly followed", radiology reports needed to sequentially describe all documented lesions of >1 cm to track changes in size.

Fig. 2 Assessing applicability of RECIST for defining cancer progression in real-world EHR data in experiment 1. Twenty-six patient charts were randomly selected from the overall cohort of 7584 patients with at least 2 clinical

visits and 2 lines of therapy (LoT). RECIST criteria were applied and the numbers of patients meeting the various criteria were recorded

key lesions between two time points, as required by RECIST. Even fewer (6, 23% of experiment 1 cohort) had evidence that all non-measured key lesions were followed between the two time points, also a RECIST component. Completeness did not reach the saturation sampling threshold of 75% completeness: 0–23%, depending on requirement for explicit documentation of key lesions. As a result of infeasibility (lack of usability), output from the RECIST abstraction approach was not further analyzed or compared.

Experiment 2: Comparison of Different Abstraction Methods for Assessing Real-World Cancer Progression

In experiment 2 ($N = 200$), the median age at advanced diagnosis was 66 years, half were women, 63% were stage IV at diagnosis, most

had non-squamous cell histology (73%), and most had a smoking history (85%) (Table 2).

For most patients ($n = 173$, 87% of experiment 2 cohort) both radiology- and clinician-anchored approaches identified at least one progression event. In addition, 7 patients (4% of experiment 2 cohort) had radiology-anchored progression events without confirmation in a clinician note, for a total 180 patients with a radiology-anchored progression event (90% of experiment 2 cohort). Of those 7 cases without clinician confirmation, chart review determined that the discrepancy could be attributed to (1) the radiology report was the last available document in the chart and the clinician assessment was inexistent or unavailable in the chart (5/7), signifying possible death, pending clinical visit, hospice referral, or loss to follow-up, or (2) there was disagreement between the radiology report and the assessment in the clinician note (2/7),

Table 2 Demographic and clinical characteristics of experiment 2 cohort

Variable	N = 200
Demographics	
Median age at advanced diagnosis, years [IQR]	65.5 [57.0; 72.0]
Age at advanced diagnosis, <i>n</i> (%)	
< 55 years	37 (18.0)
55–64 years	58 (29.0)
65+	105 (52.5)
Gender, <i>n</i> (%)	
Female	100 (50.0)
Male	100 (50.0)
Race/ethnicity, <i>n</i> (%)	
White	137 (68.5)
Black or African American	15 (7.5)
Asian	6 (3.0)
Other race	15 (7.5)
Unknown/missing	27 (13.5)
Region, <i>n</i> (%)	
Northeast	62 (31.0)
Midwest	36 (18.0)
South	67 (33.5)
West	35 (17.5)
Clinical characteristics	
Stage at diagnosis, <i>n</i> (%)	
Stage I	13 (6.5)
Stage II	8 (4.0)
Stage III	44 (22.0)
Stage IV	125 (62.5)
Not reported	10 (5.0)
Histology, <i>n</i> (%)	
Non-squamous cell carcinoma	145 (72.5)
Squamous cell carcinoma	46 (23.0)
NSCLC histology NOS	9 (4.5)

Table 2 continued

Variable	N = 200
Smoking status, <i>n</i> (%)	
History of smoking	169 (84.5)
No history of smoking	25 (12.5)
Unknown/not documented	6 (3.0)
First-line therapy class, <i>n</i> (%)	
Platinum-based chemotherapy combinations	103 (51.5)
Anti-VEGF-based therapies	48 (24.0)
Single agent chemotherapies	28 (14.0)
EGFR TKIs	18 (9.0)
Non-platinum-based chemotherapy combinations	1 (0.5)
PD-1/PD-L1-based therapies	1 (0.5)
Clinical study drug-based therapies	1 (0.5)
Treatment setting, <i>n</i> (%)	
Community	194 (97.0)
Academic	6 (3.0)
Median follow-up time from advanced diagnosis, months [IQR]	13 [9.0; 21.0]

(e.g., radiology report recorded progression but the clinician determined stable disease).

For the 173 patients with a progression event identified by both radiology- and clinician-anchored approaches, progression dates matched between the two approaches for 156 patients (90%). Among the 17 patients (10%) with differing progression event dates, the radiology-anchored approach identified an earlier progression date in almost all cases (*n* = 16, 93%). In all 16 cases, the clinician recorded stable disease immediately following the radiology-anchored progression event but subsequently documented a progression event. In the one final discordant case, the clinician-anchored progression event did not have a corresponding scan documented in the chart.

Table 3 Likelihood of predicting downstream events in experiment 2

	Abstraction approach		
	Radiology-anchored	Clinician-anchored	Combined
Number of patients with at least one progression event, <i>n</i> (% of experiment 2 cohort)	180 (90.0%)	173 (86.5%)	180 (90.0%)
Number of patients with a downstream event ^a	121	124	121
Proportion of patients with an associated downstream event, % (95% CI)	67.2 (60–74)	71.7 (65–79)	67.2 (60–74)

^a Clinically relevant downstream events defined as death, start of new therapy line (second or subsequent lines), or therapy stop. Downstream events occurred 15 days prior and up to 60 days after the progression date

Table 4 Correlation between rwPFS or rwTTP and OS in experiment 2

	Abstraction approach		
	Radiology-anchored	Clinician-anchored	Combined
rwPFS			
Median, months (95% CI)	4.9 (4.2–5.6)	5.5 (4.6–6.3)	4.9 (4.2–5.6)
Correlation with OS, % (95% CI) ^{a,b}	65 (53–74)	66 (55–75)	65 (53–74)
rwTTP			
Median, months (95% CI)	5.0 (4.2–6.1)	5.6 (4.8–6.5)	5.0 (4.2–6.1)
Correlation with OS, % (95% CI) ^{a,c}	70 (59–78)	70 (59–78)	70 (59–78)

^a Spearman's rho

^b Includes only patients with an observed death (*n* = 123)

^c Includes only patients with an observed death and a cancer progression event preceding death (*n* = 112 for the clinician-anchored approach; *n* = 113 for the radiology-anchored and the combined approach)

A near-term, downstream event (treatment stop, treatment change, or death, Table 3) was present for 124/173 (71.7%) and 121/180 (67.2%) of patients with cancer progression identified by the clinician-anchored approach and the radiology-anchored approach, respectively. The combined abstraction approach using the first clinician- or radiology-anchored event showed identical results to the radiology-anchored approach.

Clinician-anchored median rwPFS (Table 4) was higher than radiology-anchored (5.5 [95% CI 4.6–6.3] vs. 4.9 months [95% CI 4.2–5.6]). A similar pattern was observed for rwTTP (Fig. 3). Median OS for all patients was 17 months (95%

CI 14–19). Correlations between rwPFS or rwTTP and OS (Table 4) were very similar across abstraction approaches and slightly higher for rwTTP (Spearman's rho 0.70; 95% CI 0.59–0.78) than for rwPFS (Spearman's rho [95% CI] ranged from 0.65 [0.53–0.74] to 0.66 [0.55–0.75], depending on approach).

Inter-rater agreement on presence or absence of a first progression event was 96–98%, depending on approach (Table 5). Overall agreement was also similar for all approaches, 71–73%, when 30-day windows between progression event dates were allowed. When considering only cases where both abstractors agreed that at least one progression event

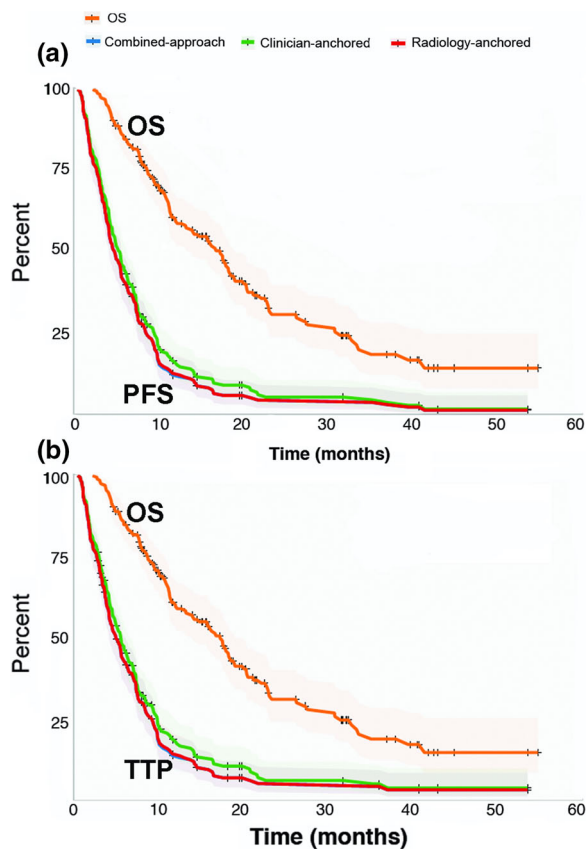


Fig. 3 rwPFS, rwTTP, and OS in experiment 2. Kaplan–Meier estimate curves for overall survival and **a** progression-free survival (PFS) or **b** time to progression (TTP), for all three non-RECIST abstraction approaches

occurred ($n = 48, 49$), the progression event dates were within 30 days of each other in 69–71% of patients.

Abstractors reported preferring the clinician-anchored approach because it afforded a more comprehensive context. One abstractor reported that the clinician-anchored approach “provide[s] a true comparison of the scan results and guide to overall patient treatment.” Abstractors also reported the clinician-anchored approach to be faster and more straightforward compared to the radiology-anchored approach.

DISCUSSION

EHR-based real-world research requires validated approaches to abstract outcomes in order

to glean meaningful insights and ensure consistent descriptions across different data sets and studies. What we really want to know about an intervention is whether patients are doing “better,” “worse,” or “unchanged.” Many proxies have been proposed to answer this question. In prospective clinical trials, RECIST tumor-size measurements on imaging are a common metric. However, it was unknown how clinical trial endpoint language translates to assessments of outcomes documented during routine clinical care. Additionally, EHR outcome abstraction needs to facilitate, rather than hinder, research on large, contemporary real-world patient cohorts. Therefore, we evaluated one RECIST-based and three non-RECIST-based approaches to identify a feasible method suitable for large-scale abstraction of cancer progression from EHRs.

We found that it is not feasible to use RECIST (standard or lenient) to abstract cancer progression from EHRs as outcomes are missing at least 75% of the time. This finding is unsurprising given that RECIST application in clinical trial settings often requires enormous resources. While quantitative methods were used exclusively for experiment 2, we determined infeasibility (lack of usability) for experiment 1 using the saturation sampling method, a common qualitative approach [16, 17] that enables early stopping if additional cases are unlikely to yield a positive outcome. Feasibility requires adequate completeness of data; however, at 20 patients the RECIST approach failed to reach a very lenient 75% completeness threshold. Further enrichment to 26 patients showed persistently poor data completeness.

We then assessed three non-RECIST approaches to define cancer progression from the EHR using technology-enabled human curation (Table 1); all had more complete endpoint information than the RECIST-based approach, yielded similar analytic results, and identified progression events at a frequency consistent with expected clinical benchmarks in this disease based on clinical studies. Progression events were identified in 86–90% of cases, regardless of approach, and predicted near-term downstream clinical events, such as treatment change and death, more than two-thirds of the

Table 5 Inter-rater agreement reliability in experiment 2

Approach	Agreement level	N	Agreement, % (95% CI)		
			Exact	15-day window	30-day window
Radiology-anchored	Event	55	98 (94–100)	–	–
	Date	49	61 (47–75)	67 (54–80)	69 (56–82)
	Overall	55	64 (51–77)	69 (57–81)	71 (59–83)
Clinician-anchored	Event	55	96 (91–100)	–	–
	Date	48	60 (46–74)	67 (54–80)	71 (58–84)
	Overall	55	62 (49–75)	67 (55–79)	71 (59–83)
Combined	Event	55	98 (94–100)	–	–
	Date	49	61 (47–75)	69 (56–82)	71 (58–84)
	Overall	55	64 (51–77)	71 (59–83)	73 (61–85)

Patient charts were abstracted in duplicate by different abstractors and agreement (95% CIs) is reported. Event agreement is based on the presence or absence of at least one cancer progression event. Date agreement is based on when the progression occurred, and only calculated in cases where both abstractors recorded a cancer progression. Overall agreement is based on a combined approach where both the absence or presence of a progression event and the date of the event, if one was found, contribute toward agreement

time. For progression events without a near-term, downstream clinical event, chart review confirmed that these were due to clinical care realities (e.g., few subsequent treatment options or a treatment “holiday”). Regardless of abstraction approach, rwP performed as expected in time-to-event analyses; a median rwPFS of approximately 5 months is similar to that observed in published aNSCLC clinical trials [18, 19], and correlations between rwPFS/rwTTP and OS were consistent with, or higher than, results from clinical trials [18, 20, 21]. Given our reliance on RECIST in clinical trials, one might anticipate greater reliability from the radiology-anchored abstraction approach; however, the clinician-anchored approach was similarly reliable and abstractor feedback indicated that it may be more scalable. The similar reliability may be due to the overlap between clinician and radiology assessments of disease progression in real-world settings.

Modest rwPFS differences for the clinician- and radiology-anchored approaches likely reflect differences in the underlying conceptual structure. The radiology-anchored approach identified more progression events, but the clinician-anchored approach was more likely to

be associated with a treatment change or death. This finding suggests that the clinician-anchored approach may benefit from clinician notes that synthesize information and adjudicate potentially conflicting findings. In the radiology-anchored approach abstractors are directed to consider radiology reports as the primary source of truth; however, radiologists may lack access to EHRs to support interpretation of imaging findings. Median rwPFS was approximately 2 weeks shorter for the radiology-anchored approach when compared with the clinician-anchored approach, likely because imaging often precedes a clinician visit and/or assessment in real-world settings. Nonetheless, modest differences and overlapping CIs preclude strong conclusions. Importantly, the association between rwPFS and OS in this cohort showed no meaningful differences between approaches.

Despite these similar quantitative results, abstractor feedback differed. The clinician-anchored approach was favored, as a clinician synthesis resonated with the broader context of the patient’s clinical journey. In addition, abstractors reported shorter abstraction times for the clinician-anchored approach. These

features support the feasibility and scalability of abstracting progression from the EHR with a clinician-anchored approach. In contrast, the radiology-anchored approach is potentially less scalable, at least for community oncology medical records (most patients in this cohort). Further, it is plausible that the radiology-anchored approach may not perform as well when determination of progression is based on non-radiologic findings (e.g., cutaneous progression, symptom history, biomarker evidence), and in resource- or access-limited settings.

Overall, the clinician-anchored cancer progression abstraction method is the most practical of those studied for abstraction of tumor-based endpoints in EHR data. A scalable approach can support research on a large patient population with a rapid turnaround time, generating contemporary data to answer questions of relevance to current and future patients.

There are several limitations to consider in this study. There are several sources of potential subjectivity (e.g., radiologists, clinicians, and abstractors). However, rwPFS correlates with OS in ways similar to RECIST-based PFS in clinical trial studies, and the correlation between the radiology- and clinician-anchored approaches was similar. We implemented clear instructions and abstractor training to mitigate subjectivity during the abstraction process. Although similar inter-rater reliability was observed across approaches, further training and experience could improve abstractor reliability. Another limitation of this analysis is the grouping of patients who presented with aNSCLC (only treated with palliative intent) with those who initially presented with an earlier stage of NSCLC and progressed to aNSCLC after curative intent treatment. Correlations between OS and intermediate endpoints may vary by these subgroups; future studies could stratify by stage at initial presentation. In addition, future research with patients diagnosed in 2018 or later (when the AJCC 8th edition staging system came into effect, and after our study enrollment ended) could evaluate the potential impact of this staging update [22, 23].

More broadly, there may be limitations when applying this work to other diseases and settings. Although real-world studies in other high-prevalence tumor types, such as breast

cancer, have analyzed intermediate endpoints also based on clinician assessments [24–26], cross-tumor validation studies are lacking. Tumor, treatment, and assessment differences (e.g., tumor kinetics, tumor response pattern, assessment cadence and modality, role of biomarker data, and the availability and type of subsequent treatment options) vary across tumor types and may affect the performance of these endpoint variables. Development of a set of rules to reconcile conflicting progression data and integrate additional data sources may help address these differences. Our work predominantly included EHR data generated in the community oncology setting. Since management and documentation patterns may vary across treatment settings, additional exploration focused on academic centers may help ascertain generalizability. Whether this clinician-anchored approach will hold true for other real-world endpoints such as tumor response requires further examination. Lastly, any approach utilizing time-dependent endpoints is susceptible to bias due to data missingness or the frequency of assessments, particularly if unbalanced across treatment groups. Further study is required to characterize the extent of potential bias.

CONCLUSIONS

This study lays the foundations for developing and validating surrogate endpoints of cancer progression suitable for generating high-quality RWD based on information routinely found in the EHR. Identification of a practical clinician-anchored abstraction method for characterizing the response of real-world cancer patients to treatment enhances the relevance and interpretability of RWE in oncology. On the basis of the study findings, we foresee two important next steps. In order to fully harness the potential value of EHR data, outcome measures tailored to the unique features of the data source are needed. Once those outcome measures are developed, such as the methodology for abstraction of cancer progression described here, a comprehensive validation framework must be established to enable robust and

standardized characterization of such approaches. Second, we need to tailor and expand this outcomes development and validation approach to different contexts and cancer types. It is also important to explore the broader opportunities to assess other outcome metrics suitable for the real-world setting such as quality of life.

ACKNOWLEDGEMENTS

The study team wishes to thank the patients whose data contributed to this study.

Funding. This study was sponsored by Flatiron Health Inc., which is an independent subsidiary of the Roche group. All authors had full access to all of the data in this study and take complete responsibility for the integrity of the data and accuracy of the data analysis. Article processing charges and open access fee have been funded by Flatiron Health Inc.

Assistance. The authors would like to thank Rana Elkhali, PhD, Nicole Lipitz and Julia Saiz, PhD for editorial assistance and Rachael Sorg, MPH, for assistance with statistical analyses. We also thank Amy P. Abernethy, MD, PhD, for her contributions while Chief Medical Officer and Chief Scientific Officer of Flatiron Health. Dr. Abernethy participated in this work prior to joining the FDA. This work and related conclusions reflect the independent work of study authors and do not necessarily represent the views of the FDA or USA.

Authorship. All named authors meet the International Committee of Medical Journal Editors (ICMJE) criteria for authorship for this article, take responsibility for the integrity of the work as a whole, and have given their approval for this version to be published.

Authorship Contributions. RAM, SDG, BB and MT led the overall experimental design, execution and interpretation. JK and PY conducted the analyses. All authors contributed to the interpretation of results and preparation of

the manuscript, and provided final approval of the manuscript version to be published.

Prior Presentation. An earlier version of this work was presented at the Friends of Cancer Research Blueprint for Breakthrough Forum on June 16, 2016 and summarized in an article describing the meeting (Eastman, P. *Oncology Times*, 38(14): 1,10–11. <https://doi.org/10.1097/01.cot.0000490048.59723.41>). A version of this manuscript containing results from a preliminary analysis is available as non-peer reviewed preprint at bioRxiv, <https://doi.org/10.1101/504878>.

Disclosures. Sandra D Griffith reports employment at Flatiron Health Inc., which is an independent subsidiary of the Roche Group, also reports equity ownership in Flatiron Health Inc., and stock ownership in Roche. Melisa Tucker reports employment at Flatiron Health Inc., which is an independent subsidiary of the Roche Group, also reports equity ownership in Flatiron Health Inc., and stock ownership in Roche. Bryan Bowser reports employment at Flatiron Health Inc., which is an independent subsidiary of the Roche Group, also reports equity ownership in Flatiron Health Inc., and stock ownership in Roche. Geoffrey Calkins reports employment at Flatiron Health Inc., which is an independent subsidiary of the Roche Group, also reports equity ownership in Flatiron Health Inc., and stock ownership in Roche. Che-hsu (Joe) Chang reports employment at Flatiron Health Inc., which is an independent subsidiary of the Roche Group, also reports equity ownership in Flatiron Health Inc. Josh Kraut reports employment at Flatiron Health Inc., which is an independent subsidiary of the Roche Group, also reports equity ownership in Flatiron Health Inc., and stock ownership in Roche. Paul You reports employment at Flatiron Health Inc., which is an independent subsidiary of the Roche Group, also reports equity ownership in Flatiron Health Inc., and stock ownership in Roche. Rebecca A. Miksad, reports employment at Flatiron Health Inc., which is an independent subsidiary of the Roche Group, also reports equity ownership in Flatiron Health Inc., and stock ownership in Roche. She also reports being an advisor for the

De Luca Foundation and a grant review committee member for the American Association for Cancer Research. Ellie Guardino reports employment at Genentech. Deb Schrag reports work on related projects to augment capacity for using “real world” data in partnership with AACR’s project GENIE and research funding for GENIE has been awarded to Dana Farber Cancer Institute, she is a compensated Associate Editor of the journal *JAMA*. Sean Khozin declares that he has no conflicts to disclose.

Compliance with Ethics Guidelines. IRB approval of the study protocol (IRB # RWE-001, “The Flatiron Health Real-World Evidence Parent Protocol”, Tracking # FLI1-18-044) by the Copernicus Group IRB, with waiver of informed consent, was obtained prior to study conduct, covering the data from all sites represented.

Data Availability. The data sets generated during and/or analyzed during the current study are not publicly available. The data supporting the findings of this study are available upon request from Flatiron Health Inc. Restrictions apply to the availability of these data which are subject to the de-identification requirements of the Health Insurance Portability and Accountability Act of 1996 (HIPAA) and implementing regulations, as amended.

Open Access. This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

REFERENCES

- Sox HE, Greenfield S. Comparative effectiveness research: a report from the Institute of Medicine. *Ann Intern Med*. 2009;151(3):203–5.
- Berger ML, Curtis MD, Smith G, Harnett J, Abernethy AP. Opportunities and challenges in leveraging electronic health record data in oncology. *Future Oncol*. 2016;12(10):1261–74.
- Khozin S, Blumenthal GM, Pazdur R. Real-world data for clinical evidence generation in oncology. *J Natl Cancer Inst*. 2017;109(11):dix187.
- Martell RE, Sermer D, Getz K, Kaitin KI. Oncology drug development and approval of systemic anti-cancer therapy by the U.S. Food and Drug Administration. *Oncologist*. 2013;18(1):104–11.
- Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-world evidence—what is it and what can it tell us? *N Engl J Med*. 2016;375(23):2293–7.
- U.S Food & Drug Administration. Framework for FDA’s real-world evidence program. Released on December 6, 2018. <https://www.fda.gov/downloads/ScienceResearch/SpecialTopics/RealWorldEvidence/UCM627769.pdf>. Accessed 4 Jan 2019.
- Miksad RA, Abernethy AP. Harnessing the power of real-world evidence (RWE): a checklist to ensure regulatory-grade data quality. *Clin Pharmacol Ther*. 2018;103(2):202–5.
- Johnson JR, Williams G, Pazdur R. End points and United States Food and Drug Administration approval of oncology drugs. *J Clin Oncol*. 2003;21(7):1404–11.
- Pazdur R. Endpoints for assessing drug activity in clinical trials. *Oncologist*. 2008;13(Suppl 2):19–21.
- Therasse P, Arbuck SG, Eisenhauer EA, et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J Natl Cancer Inst*. 2000;92(3):205–16.
- Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer*. 2009;45(2):228–47.
- Curtis MD, Griffith SD, Tucker M, et al. Development and validation of a high-quality composite real-world mortality endpoint. *Health Serv Res*. 2018;53(6):4460–76.
- Abernethy AP, Gippetti J, Parulkar R, Revol C. Use of electronic health record data for quality reporting. *J Oncol Pract*. 2017;13(8):530–4.
- Edge SB, Byrd DR, Compton CC, editors. *AJCC cancer staging manual*. 7th ed. New York: Springer; 2010.

15. Virzi RA. Refining the test phase of usability evaluation: how many subjects is enough? *Hum Factors*. 1992;34(4):457–68.
16. Nielsen J, Landauer, TK. A mathematical model of the finding of usability problems. In: CHI '93 Proceedings of the INTERACT '93 and CHI '93 conference on human factors in computing systems. 1993 May 1, pp 206–213.
17. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc*. 1958;53(282):457–81.
18. Johnson KR, Ringland C, Stokes BJ, et al. Response rate or time to progression as predictors of survival in trials of metastatic colorectal cancer or non-small-cell lung cancer: a meta-analysis. *Lancet Oncol*. 2006;7(9):741–6.
19. Laporte S, Squifflet P, Baroux N, et al. Prediction of survival benefits from progression-free survival benefits in advanced non-small-cell lung cancer: evidence from a meta-analysis of 2334 patients from 5 randomised trials. *BMJ Open*. 2013;3(3):e001802.
20. Yoshino R, Imai H, Mori K, et al. Surrogate endpoints for overall survival in advanced non-small-cell lung cancer patients with mutations of the epidermal growth factor receptor gene. *Mol Clin Oncol*. 2014;2(5):731–6.
21. Blumenthal GM, Karuri SW, Zhang H, et al. Overall response rate, progression-free survival, and overall survival with targeted and standard therapies in advanced non-small-cell lung cancer: US Food and Drug Administration trial-level and patient-level analyses. *J Clin Oncol*. 2015;33(9):1008–14.
22. Detterbeck FC, Boffa DJ, Kim AW, Tanoue LT. The eighth edition lung cancer stage classification. *Chest*. 2017;151(1):193–203.
23. Rami-Porta R, Asamura H, Travis WD, Rusch VW. Lung cancer—major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. *CA Cancer J Clin*. 2017;67(2):138–55.
24. Taylor-Stokes G, Mitra D, Waller J, Gibson K, Milligan G, Iyer S. Treatment patterns and clinical outcomes among patients receiving palbociclib in combination with an aromatase inhibitor or fulvestrant for HR+/HER2-negative advanced/metastatic breast cancer in real-world settings in the US: results from the IRIS study. *Breast*. 2019;43:22–7.
25. Bartlett CH, Mardekian J, Cotter M, et al. Abstract P3-17-03: Concordance of real world progression free survival (PFS) on endocrine therapy as first line treatment for metastatic breast cancer using electronic health record with proper quality control versus conventional PFS from a phase 3 trial. *Cancer Res* 2018;78(4 Suppl):Abstract nr P3-17-03. In: Proceedings of the 2017 San Antonio Breast Cancer Symposium; 2017 Dec 5-9; San Antonio, TX. Philadelphia (PA): AACR.
26. Kurosky SK, Mitra D, Zanotti G, Kaye JA. Treatment patterns and outcomes of patients with metastatic ER+/HER2- breast cancer: a multicountry retrospective medical record review. *Clin Breast Cancer*. 2018;18(4):e529–38.