Characterizing the Feasibility and Performance of Real-World Tumor Progression End Points and Their Association With Overall Survival in a Large Advanced Non–Small-Cell Lung Cancer Data Set

Sandra D. Griffith, PhD[1]; Rebecca A. Miksad, MD, MPH[1]; Geoff Calkins[1]; Paul You, MPH[1]; Nicole G. Lipitz[1]; Ariel B. Bourla, MD, PhD[1]; Erin Williams[1]; Daniel J. George, MD[2]; Deborah Schrag, MD[3,4]; Sean Khozin, MD, MPH[5]; William B. Capra, PhD[6]; Michael D. Taylor, PhD, PharmD[6]; and Amy P. Abernethy, MD, PhD[1]

abstract

PURPOSE Large, generalizable real-world data can enhance traditional clinical trial results. The current study evaluates reliability, clinical relevance, and large-scale feasibility for a previously documented method with which to characterize cancer progression outcomes in advanced non–small-cell lung cancer from electronic health record (EHR) data.

METHODS Patients who were diagnosed with advanced non–small-cell lung cancer between January 1, 2011, and February 28, 2018, with two or more EHR-documented visits and one or more systemic therapy line initiated were identified in Flatiron Health's longitudinal EHR-derived database. After institutional review board approval, we retrospectively characterized real-world progression (rwP) dates, with a random duplicate sample to ascertain interabstractor agreement. We calculated real-world progression-free survival, real-world time to progression, real-world time to next treatment, and overall survival (OS) using the Kaplan-Meier method (index date was the date of first-line therapy initiation), and correlations between OS and other end points were assessed at the patient level (Spearman's ρ).

RESULTS Of 30,276 eligible patients, 16,606 (55%) had one or more rwP event. Of these patients, 11,366 (68%) had subsequent death, treatment discontinuation, or new treatment initiation. Correlation of real-world progression-free survival with OS was moderate to high (Spearman's ρ, 0.76; 95% CI, 0.75 to 0.77; evaluable patients, n = 20,020), and for real-world time to progression correlation with OS was lower (Spearman's ρ, 0.69; 95% CI, 0.68 to 0.70; evaluable patients, n = 11,902). Interabstractor agreement on rwP occurrence was 0.94 (duplicate sample, n = 1,065) and on rwP date 0.85 (95% CI, 0.81 to 0.89; evaluable patients n = 358 [patients with two independent event captures within 30 days]). Median rwP abstraction time from individual EHRs was 18.0 minutes (interquartile range, 9.7 to 34.4 minutes).

CONCLUSION We demonstrated that rwP-based end points correlate with OS, and that rwP curation from a large, contemporary EHR data set can be reliable, clinically relevant, and feasible on a large scale.

INTRODUCTION

Although overall survival (OS) is a benchmark of treatment efficacy in oncology, it may not be reliably determined during limited study periods. Intermediate tumor burden–based end points, such as progression-free survival (PFS), may evaluate clinically meaningful treatment effects more efficiently and have supported the regulatory approval of anticancer therapies.[1-3] Randomized controlled trials (RCTs) use the Response Evaluation Criteria in Solid Tumors (RECIST)[4] on imaging studies to determine tumor burden changes—for example,

progression and/or response—which, in turn, affect clinical decisions and treatment benefit predictions.

Whereas RCTs are a gold standard in clinical evidence, boundaries imposed by predefined eligibility criteria and sample size create a divide between the scope of an RCT and the disease population at large. Real-world data (RWD) that are gathered during routine care, such as in electronic health records (EHRs), can complement RCT results.[5] Collected retrospectively and prospectively from contemporary, generalizable, and often large cohorts as RWD, the resulting real-world

CONTEXT

Key Objective

The current study evaluated the large-scale feasibility of a previously described approach for the curation of a tumor progression variable from unstructured electronic health record (EHR)–based real-word data as well as the meaningfulness and reliability of the corresponding real-world progression-based end points.

Knowledge Generated

Our approach to real-world progression curation from EHRs is feasible and scalable to a database of more than 30,000 patients. Preliminary evaluations indicated that the associated progression end points are reliable and seem to be valid in the context of their relationship with clinical events documented in the EHR database from which they were generated and of published clinical trial results.

Relevance

We present a feasible and replicable approach to yield an EHR-generated progression variable that is clinically meaningful and that could be incorporated into large, contemporary real-world studies.

evidence (RWE) can assess effectiveness, externally validating trial efficacy.[6]

Translating RCT end point standards into RWD may be challenging. In EHRs, tumor burden information may need to be abstracted from unstructured documents—for example, radiology reports and/or clinician notes—or inferred from structured data, such as laboratory findings and/or treatment orders. We previously found that applying RECIST retrospectively to EHRs—generally lacking raw images—is unfeasible as a result of incomplete and/or unclear real-world sources.[4] The question is how to unlock real-world clinical data as valid study end points.

To generate clinically meaningful conclusions, EHR-based characterization of relevant outcomes must accommodate RWD strengths and mitigate its limitations—for example, heterogeneous practice patterns and variable documentation. The approach needs to be transparent, replicable, and portable across data collection and clinical settings.

Our prior work compared multiple methods for curating real-world progression (rwP) from unstructured EHR documents.[7] Here, we evaluate one of those methods in a large, EHR-based cohort of patients with advanced non–small-cell lung cancer (aNSCLC), the leading cause of cancer mortality in the United States.[8] On the basis of a proposed development framework,[9] we examined three questions: Is capturing rwP retrospectively and with sufficient completeness feasible in large cohorts? Is rwP data sufficiently meaningful and reliable? and How do rwP-based end points—real-world PFS (rwPFS) and real-world time to progression (rwTTP)—compare with treatment-based end points, such as real-world time to next treatment (rwTTNT), including correlation with OS? Our objective was understanding the performance characteristics of rwP to enhance our ability to draw clinically meaningful RWD-based conclusions.

METHODS

For topics in this section, additional information can be found in the Appendix.

Data Sources

Flatiron Health's longitudinal EHR-derived database contains aggregated, normalized, and harmonized deidentified patient-level data curated from structured and unstructured data via technology-enabled chart abstraction.[10] Normalization and harmonization involved mapping to a common terminology across different source systems. At the time of this analysis—April 2018—the database included more than 265 US cancer clinics using a variety of EHR systems and representing more than 2 million patients. Dates of death were sourced from a composite mortality variable comprised of EHR data linked to commercial mortality data and the Social Security Administration's Death Master File.[11]

Institutional review board approval of the study protocol (#RWE-001; "The Flatiron Health Real-World Evidence Parent Protocol"; tracking: #FLI118044), with a waiver of informed consent, was obtained before study conduct, covering the data from all sites represented.

Cohort Selection

The cohort included patients who were diagnosed with aNSCLC—stage IIIB or metastatic stage IV, or recurrent advanced disease—between January 1, 2011, and February 28, 2018, with at least two EHR-documented clinical visits on or after January 1, 2011, and documentation of the initiation of at least one line of systemic therapy after advanced diagnosis (Fig 1). Patients who died 14 days or fewer from first-line systemic therapy initiation were excluded.

rwP Definition

EHR data were curated to retrospectively determine rwP events using a previously proposed clinician-anchored abstraction approach (Appendix).[7]

The date of cancer progression was defined as the date of the first source evidence for progression referenced by the clinician—for example, radiology report, pathology, or
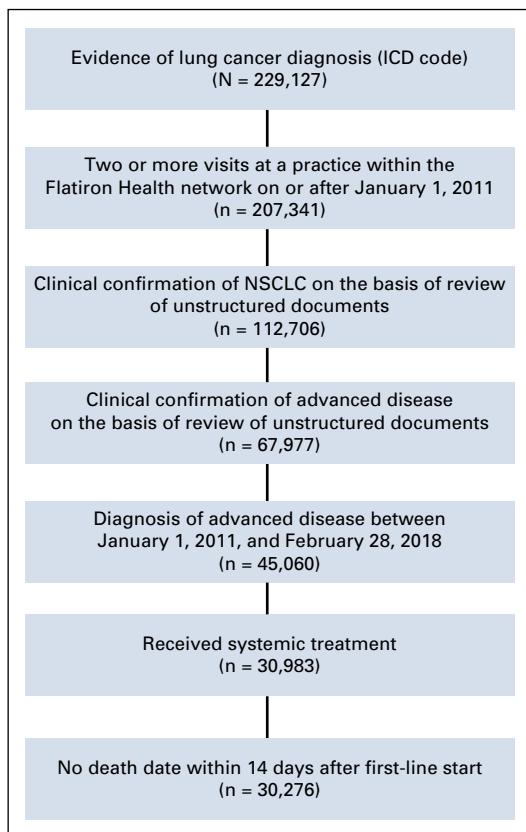
FIG 1. Patients meeting all inclusion and exclusion criteria. The cohort was identified through a combination of structured data—for example, International Classification of Diseases (ICD) codes—and abstraction of unstructured documents. This resulted in 30,276 patients in the final cohort available for analysis. NSCLC, non–small-cell lung cancer.

exam—or the clinician note date when no evidence sources were documented.

The entire patient chart was abstracted, including multiple rwP events when present. This analysis focused on the first rwP event after systemic therapy initiation for the treatment of aNSCLC, with most thus occurring during first-line treatment. A continuous monthly audit to assess only interabstractor

agreement was performed by duplicate abstracting a random subsample of 10% of newly accruing patients—that is, each new patient had an independent 10% chance of duplicate auditing—up to 100 patients per month.

Statistical Analyses

Calculation of real-world end points. Real-world end points—rwPFS, rwTTP, rwTTNT, and OS—in patients with at least one relevant event (Table 1) were calculated using the Kaplan-Meier method,[12] indexed to the start of first-line systemic therapy. Kaplan-Meier curves and median time-to-event estimates were reported.

Feasibility. We measured the completeness of data capture from imaging sources to assess the feasibility of abstracting rwP at scale from a large, generalizable EHR database, with a three-factor framework: Assessment, Documentation, and Abstraction (Fig 2).

The frequency of relevant imaging was estimated to determine whether patients were evaluated for cancer progression at expected intervals (Fig 2; Assessment). For an rwP event to be abstracted from the EHR, the clinician note referenced and documented disease status results (Fig 2; Documentation). Patients were censored at the time of death.

As a proxy for completeness of rwP capture by an abstractor (Fig 2; Abstraction), we measured inter-rater event agreement in the duplicate subgroup for the presence or absence of at least one progression event.

Abstraction time per patient chart was computed by the abstractor software interface. Average times were analyzed with descriptive statistics.

Meaningfulness and reliability. Meaningfulness of the rwP data was evaluated using three criteria:

Face validity—the extent to which the method seems to measure the variable[13]: Descriptive survival analyses for rwPFS, rwTTP, and OS were determined and compared with similar statistics in the literature.

Association with near-term, clinically relevant downstream events, defined as death, start of a new therapy line, and stop or discontinuation of active systemic treatment:

TABLE 1. End Point Definitions and Censoring Approach

End Point	Abbreviation	Index Date	Event Date	Censor Date*
Real-world progression-free survival	rwPFS	First-line systemic therapy start	First progression event or death	Last clinician note
Real-world time to progression	rwTTP	First-line systemic therapy start	First progression event	Last clinician note
Overall survival	OS	First-line systemic therapy start	Death as a result of any cause	Last confirmed activity (last clinical visit or drug administration)
Real-world time to next treatment	rwTTNT	First-line systemic therapy start	Second-line systemic therapy start	Last confirmed activity (last clinical visit or drug administration)

Abbreviations: OS, overall survival; rwPFS, real-world progression-free survival; rwTTP, real-world time to progression; rwTTNT, real-world time to next treatment.

*Patients are censored only when no end point event was identified within the study period.
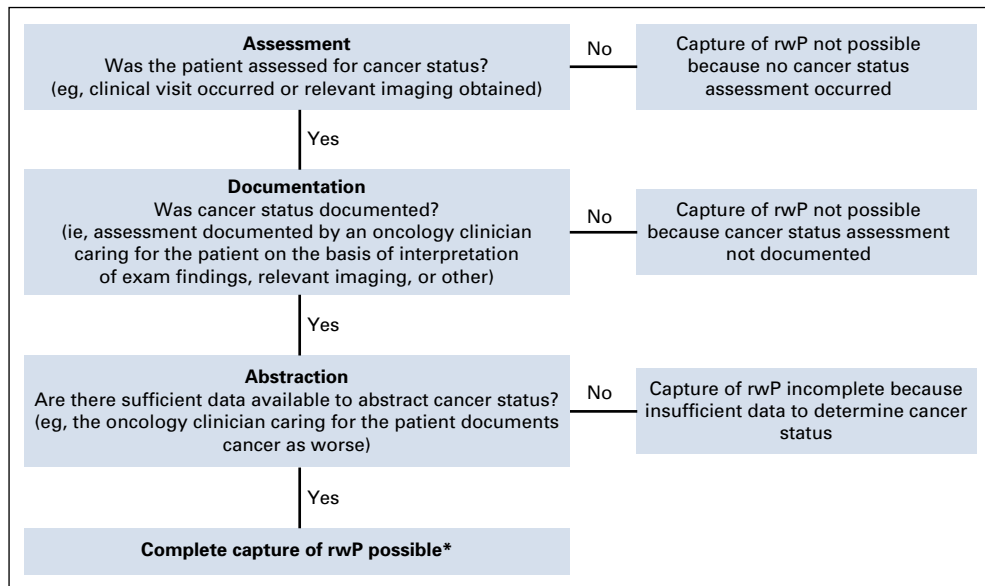
FIG 2. Framework for measuring the completeness of electronic health record (EHR) end point capture, such as the cancer status of progression. The ability to completely capture end points requires three dependent factors: Assessment, Documentation, and Abstraction. If any of the components of this framework—Assessment, Abstraction, or Evidence—are missing, capture of end points is incomplete. Our clinician-anchored real-world progression (rwP) curation approach focuses on the oncology clinician as the synthesizer of signals from the entire patient chart; imaging, when described in the oncology clinician note, is considered a piece of source evidence. (*) Other sources of missingness include abstractor error, missing documents in the EHR, etc.

Near term was defined as 15 or fewer days before—allowing for short delays in documentation—and 60 or fewer days after—clinically appropriate for aNSCLC—the rwP date. Treatment discontinuation was defined as 60 days of active follow-up without structured documentation of additional systemic therapy.

- Correlation of rwP-based end points—rwPFS and rwTTP—with OS, assessed at the patient level (Spearman's ρ): For correlation calculations, the cohort was restricted to patients with a documented death event (n = 20,020 for rwPFS) or with both rwP and death events documented (n = 11,902 for rwTTP).

We assessed reliability by calculating agreement rates for the occurrence of an rwP event (event agreement), defined as the proportion of patients with an rwP event captured by two independent abstractors, and similarly for the documented event date (date agreement), in those with event agreement. Agreement for four date windows was also evaluated: 3, 7, 15, and 30 days.

Comparison with treatment-based end points. rwP-based end points rwPFS and rwTTP were compared with rwTTNT, a treatment-based end point derived from structured data (Table 1). Correlation between rwTTNT and OS was calculated (Spearman's ρ) among patients with

both a documented death event and a second-line systemic treatment start (n = 9,269).

Sensitivity analyses. We conducted sensitivity analyses to assess the robustness of the correlation between real-world end points—rwPFS, rwTTP, and rwTTNT—and OS. Three correlation calculations were reperformed and three additional sensitivity analyses were conducted (Appendix Table A1).

RESULTS

Table 2 lists the demographic and clinical characteristics of the 30,276 eligible patients. Median follow-up after advanced diagnosis was 9.3 months (interquartile range, 4.3 to 19.4 months).

Feasibility

Feasibility analyses were based on the completeness of the information collected in the eligible overall population (N = 30,276). The probability that a patient's first, second, and third imaging set occurred within a 3-month window of first-line therapy start was 85%, 47%, and 18%, respectively. This increased to 94%, 89%, and 83%, respectively, for a 12-month window. At 3 months after aNSCLC diagnosis, 96% of patients had at least one clinician note. These frequencies of imaging and clinical assessments seemed to be consistent with realistic expectations, which suggested

TABLE 2. Demographic and Clinical Characteristics (N = 30,276)

Characteristic	Value
Age at advanced diagnosis category, years	
< 35	63 (0.21)
35-44	380 (1.26)
45-54	2,635 (8.70)
55-64	7,667 (25.3)
≥ 65	19,531 (64.5)
Median age at advanced diagnosis, years (IQR)	69.0 (61.0-75.0)
Practice type	
Academic	2,249 (7.43)
Community	28,027 (92.6)
Sex	
Female	14,339 (47.4)
Male	15,937 (52.6)
Region	
Northeast	6,479 (21.4)
Midwest	5,575 (18.4)
South	11,120 (36.7)
West	4,424 (14.6)
Other/missing	2,678 (8.9)
Race	
White	20,765 (68.6)
Black or African American	2,481 (8.2)
Asian	874 (2.9)
Other	2,723 (9.0)
Not reported	3,418 (11.3)
Stage at diagnosis	
I	2,129 (7.0)
II	1,327 (4.4)
III	6,169 (20.4)
IV	19,668 (65.0)
Group stage is not reported	981 (3.2)
Histology	
Nonsquamous-cell carcinoma	21,394 (70.7)
Squamous-cell carcinoma	7,330 (24.2)
NSCLC histology NOS	1,552 (5.1)
Smoking status	
History of smoking	25,656 (84.7)
No history of smoking	4,169 (13.8)
Unknown/not documented	451 (1.5)
ALK status	
Rearrangement present	653 (2.2)
Rearrangement not present	16,214 (53.6)
Unsuccessful/indeterminate test	990 (3.3)

(Continued in next column)

TABLE 2. Demographic and Clinical Characteristics (N = 30,276) (Continued)

Characteristic	Value
Unknown/results pending	541 (1.8)
Not tested	11,878 (39.2)
EGFR status	
Mutation positive	3,003 (9.9)
Mutation negative	15,706 (51.9)
Unsuccessful/indeterminate test	656 (2.2)
Unknown/results pending	385 (1.3)
Not tested	10,526 (34.8)
ROS-1 status	
Rearrangement present	111 (0.4)
Rearrangement not present	8,699 (28.7)
Unsuccessful/indeterminate test	572 (1.9)
Unknown/results pending	274 (0.9)
Not tested	20,616 (68.1)
First-line therapy class	
ALK inhibitors	561 (1.9)
Anti–VEGF-based therapies	5,405 (17.9)
Clinical study drug–based therapies	292 (1.0)
EGFR TKIs	3,054 (10.1)
EGFR antibody-based therapies	121 (0.4)
No first-line therapy	0 (0.0)
Nonplatinum-based chemotherapy combinations	120 (0.4)
Other therapies	49 (0.2)
PD-1/PD-L1–based therapies	3,014 (10.0)
Platinum-based chemotherapy combinations	15,274 (50.4)
Single-agent chemotherapies	2,386 (7.9)
Median follow-up time from advanced diagnosis, months (IQR)	9.30 (4.3-19.4)

NOTE. Data are given as No. (%), unless otherwise noted.

Abbreviations: ALK, anaplastic lymphoma kinase; EGFR, epidermal growth factor receptor; IQR, interquartile range; NOS, not otherwise specified; NSCLC, non–small-cell lung cancer; PD-1, programmed death 1; PD-L1, programmed death ligand 1; ROS-1, human ros homolog; TKI, tyrosine kinase inhibitor.

that there was no meaningful impact of missing evaluations on data completeness. Among duplicate abstracted patients (n = 1,065), agreement on the presence or absence of a first progression event was 94%, irrespective of event date.

Using the clinician-anchored approach, median time for an oncology-trained abstractor to abstract all EHR-documented rwP events for a patient was 18.0 minutes (interquartile range, 9.7 to 34.4 minutes).

Meaningfulness and Reliability

We conducted validity and reliability analyses in population subsets in which relevant events had been documented (Appendix Fig A1).

Face validity. We identified at least one rwP event after the initiation of systemic therapy for aNSCLC in 16,606 patients (55%). The proportion of patients with an rwP event increased among those who started second-line systemic treatment during the observation period (11,937 [87%] of 13,653) and among patients with at least 6 months of active follow-up from the start of the observation period (14 days after the initiation of first-line therapy; n = 16,025 [77%]). Figure 3 shows the Kaplan-Meier curves for rwPFS and rwTTP accompanied by OS.

Association with near-term, clinically relevant, downstream events. Among the evaluable population of 16,606 patients—those with at least one rwP event—11,366 patients (68%) had a documented near-term, downstream, clinically relevant event. In sensitivity analysis, which excluded follow-up losses within 60 days of an rwP event, the proportion with near-term, downstream events increased to 74%.

Correlation of rwP-based end points (rwPFS and rwTTP) with OS. Table 3 lists the correlation between real-world end points and OS. Among patients who died during follow-up (n = 20,020), correlation between rwPFS and OS was moderately high (Spearman's ρ, 0.76; 95% CI, 0.75 to 0.77). Among patients with both a progression event and

death (n = 11,902), correlation between rwTTP and OS was lower (Spearman's ρ, 0.69; 95% CI, 0.68 to 0.70).

Reliability. As noted above, agreement on the occurrence of an rwP event in the duplicate abstraction patient subgroup (n = 1,065) without comparing any of the dates identified was 0.94 (Appendix Fig A2). Among those with at least one rwP captured by two independent abstractors (n = 358), agreement on the rwP date was 0.73 (95% CI, 0.68 to 0.78) when an exact date match was required, and 0.85 (95% CI, 0.81 to 0.89) within a 30-day date window. Reasons for date disagreement included ambiguous language (eg, "Progression on imaging last month") or referencing different source evidence from different time points (eg, computed tomography chest v magnetic resonance imaging brain). Date agreement gains seemed to plateau at 7 to 15 days.

Comparison With Treatment-Based End Points

The rwTTNT Kaplan-Meier curve contour was visually similar to those of the progression-based end points (rwPFS and rwTTP; Fig 3). Median rwTTNT was 8.7 months (95% CI, 8.5 to 8.8 months), longer than that of median rwPFS (5.2 months; 95% CI, 5.1 to 5.3 months) and rwTTP (7.1 months; 95% CI, 7.0 to 7.3 months). Association between rwTTNT and OS was lower than that of rwPFS and rwTTP (Spearman's ρ, 0.61; 95% CI, 0.60 to 0.62).

Sensitivity Analyses

For progression-based end points (rwPFS and rwTTP), sensitivity analyses demonstrated slight variability in OS
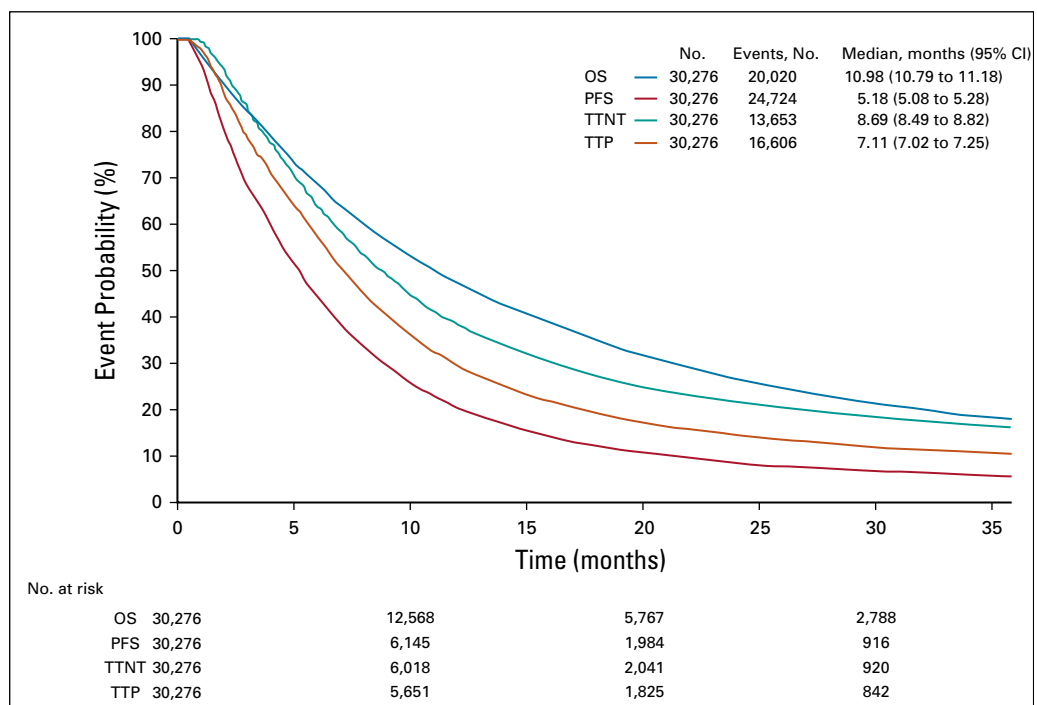
FIG 3. Kaplan-Meier survival curves for real-world progression-free survival (PFS), real-world time to progression (TTP), time to next treatment (TTNT), and overall survival (OS).

TABLE 3. Association Between OS and Real-World End Points (rwP and treatment based): Baseline and Sensitivity Analyses

Cohort	rwPFS		rwTTP		rwTTNT	
	No.	Correlation With OS	No.	Correlation With OS	No.	Correlation With OS
Baseline	20,020	0.76 (0.75 to 0.77)	11,902	0.69 (0.68 to 0.70)	9,269	0.61 (0.60 to 0.62)
Excluding patients diagnosed pre–stage IV	13,381	0.76 (0.75 to 0.77)	7,918	0.69 (0.68 to 0.70)	6,134	0.62 (0.60 to 0.64)
Including censored patients	30,276	0.78 (0.78 to 0.78)	16,606	0.67 (0.66 to 0.68)	13,653	0.58 (0.57 to 0.59)
Excluding patients with potentially missing* treatment data	18,506	0.76 (0.75 to 0.77)	10,977	0.69 (0.68 to 0.70)	8,610	0.61 (0.60 to 0.62)
Excluding events in which the clinician mentions possible pseudoprogression	20,020	0.76 (0.75 to 0.77)	11,873	0.69 (0.68 to 0.70)	—	—†
Restricted to patients who have never received immunotherapy	15,355	0.78 (0.77 to 0.79)	8,210	0.70 (0.69 to 0.71)	5,706	0.61 (0.59 to 0.63)
Restricted to rwP (rwPFS, rwTTP) and second-line start (rwTTNT) occurring before January 1, 2015	14,067	0.79 (0.78 to 0.80)	5,949	0.65 (0.64 to 0.66)	4,493	0.54 (0.52 to 0.56)

NOTE. Correlations are expressed using the Spearman's ρ coefficient (95% CI).

Abbreviations: OS, overall survival; rwPFS, real-world progression-free survival; rwTTP, real-world time to progression; rwTTNT, real-world time to next treatment.

*Defined as gaps of more than 90 days between advanced diagnosis date and first treatment administration or visit date.

†rwTTNT was not calculated within this subgroup as this end point is less relevant for patients whose documentation explicitly mentioned pseudoprogression. These patients are not likely to change treatment immediately.

correlations, but consistency with the primary analysis in magnitude and direction. The correlation between rwTTNT and OS was lower for some subgroups, dropping to 0.54 (95% CI, 0.52 to 0.56) when restricted to patients who started second-line therapy before 2015.

DISCUSSION

In oncology, tumor burden worsening—that is, progression—is a fundamental treatment metric. If tumor-based and time-to-event end points are to be characterized on the basis of real-world experience at the point of routine clinical care, a transparent, well-developed approach is needed to process the complexity—and volume—of EHR data. The current study extends a previous pilot study that demonstrated the feasibility of a clinician-anchored approach to abstract clinically relevant end points from EHR data.[7] We confirm reasonably high completeness, meaningfulness, and reliability of retrospectively EHR-abstracted rwP events for a cohort of more than 30,000 patients with aNSCLC.

Scalability addresses challenges unique to real-world studies, in particular, rapidly producing a large, reliable data set that is suitable for analysis. Our clinician-anchored rwP abstraction approach required less than 20 minutes per chart for most patients. Such factors as abstractor background and training and the patient's clinical course—for example, number of scans—may affect abstraction time. With this approach, rwP was characterized for the original cohort of approximately 24,000 patients within 10 months (in the course of the study, the cohort subsequently grew to more than 30,000 patients). As long as imaging reports are available, this approach is EHR agnostic and therefore portable across data collection settings. Continued EHR and

procedural advancements may further facilitate scaling and efficiency.

Of importance, there is no currently accepted framework for developing EHR-abstracted real-world end points for RWE studies, nor is there consensus on which criteria an end point would need to meet to clear a potential validation hurdle. We focused on three questions for our rwP variable: feasibility of the approach, reliability and meaningfulness, as well as comparison with other end points.

To address feasibility, we explored source data completeness. Evaluation metrics indicated that patients in our database seemed to be imaged and assessed by clinicians regularly, with an overall frequency that was consistent with our own practical experience. This finding suggests no meaningful impact of missing evaluations on data completeness.

We explored interabstractor consistency and correlation with downstream events, including mortality, as reliability and validity points. Consistency was high. Abstractors agreed on progression dates—within a reasonable window—85% of the time. Future studies may define reasonable windows or high agreement rates according to their specific clinical questions, disease settings, or desired level of certainty. Additional research may help understand the impact on estimates when agreement fluctuates across specific subgroups. Our results are clinically plausible in lung cancer. Approximately one third of patients did not receive a new treatment or experience another event after progression, and estimates of rwPFS, rwTTP, rwTTNT, and OS and Kaplan-Meier contours aligned with expectations for this population. Correlations between rwPFS and rwTTP and OS were similar or higher than clinical trial results.[14,15] The determination of

whether the value of an rwP end point lies in estimating clinical benefit intrinsically or in correlating with traditional clinical trial benefit measures, such as PFS, will continue to unfold as technology enhancements expedite research in increasingly large and comprehensive data sets. Studies that replicate clinical trial findings using rwP and, where parallel, retrospective or prospective single data set collection of rwP and RECIST-based progression, which enables direct comparison, will be critical to determine when rwP is good enough.

For our third question, we compared our curated rwP-based end points with treatment-based end points, such as rwTTNT, that are readily obtainable from structured EHR information. rwTTNT captures treatment patterns, but unlike rwP it may conflate toxicity and progression as treatment change reasons, which may be valuable in some instances but suboptimal when the focus is strictly on tumor burden changes. Furthermore, rwTTNT calculation is negated for patients without subsequent treatment after progression. Of note, rwTTNT has a lower correlation with OS than rwPFS, which suggests that it is a weaker intermediate end point.

In the current study, rwP events were not associated with near-term clinically relevant downstream events for approximately one third of patients. A subset chart review confirmed no treatment changes despite evidence of worsening disease in a substantial portion of patients. In light of recent studies that have suggested a benefit from continuing immunotherapy postprogression, understanding potential differences between progression- and treatment-based end points is critical.[15-17] Ideally, a set of end points that measure multiple outcomes—for example, OS, PFS, and/or quality of life—will improve our understanding of patients' cancer course.

Despite conceptual similarities, there are differences between rwP and RECIST-based progression, including the anchoring evidence source and assessment timing. Unlike the narrow RCT populations, the rwP variable performed within clinical expectations for a broad cohort of patients treated with a range of systemic therapies. Median rwPFS of approximately 5 months falls within the range observed in aNSCLC clinical trials for cytotoxic therapies, which the majority of this cohort received,[14,15,18] but is lower than that observed in recent immunotherapy and targeted therapy trials.[15,19-21] This trend may be more related to variations between populations—selected in clinical trials versus unselected in the real world—than between end points.

When data are not collected intentionally for research purposes, missing data, treatment compliance, measurement error, or unmeasured confounders create limitations and biases. In our case, additional work is needed to understand the potential impact of abstractor subjectivity or simplification in rwP assessment, as well as that of unbalanced evaluation frequency across groups. Because the study was restricted to patients with aNSCLC who were mostly from community settings, results may not be generalizable to other tumor types and patient populations. Our study did not include subgroup analyses that could require the abstraction of additional clinical variables from unstructured data (eg, extent of metastatic disease) and/or the incorporation of detailed structured data (eg, dose and timing of individual agents). As treatment options expand in aNSCLC, the association between rwPFS and OS may be affected by outcomes from subsequent therapy lines, by outlier exceptional responders to targeted therapy, or by the evolution of the progression concept in the immunotherapy era. Duplicate abstraction for agreement assessment ultimately represented a small portion of the final cohort size—approximately 1,000 patients of more than 30,000, of which only one third were evaluable for date agreement.

As survival rates have improved and more than one therapy line has become typical, OS has become an inadequate metric for ascertaining the benefit of some cancer treatment interventions. Intermediate end points are essential for evaluating treatment benefits in real-world contexts. This study presents a scalable, feasible, and replicable approach to yield an EHR-generated progression variable ready for incorporation into large, contemporary real-world analyses. More work is needed to document how rwP under different treatments relates to treatment effects observed in clinical trials; understand the impact of different imaging and assessment cadences—for example, informative censoring; and clarify the relationship between rwP and other end points, such as real-world tumor response. Similar work is also needed in other tumor types. Systematic assessment of real-world end points will improve confidence in RWD and resulting RWE.

AFFILIATIONS
[1]Flatiron Health, New York, NY
[2]Duke University, Durham, NC
[3]Dana-Farber Cancer Institute, Boston, MA
[4]Harvard Medical School, Boston, MA
[5]US Food and Drug Administration, Silver Spring, MD
[6]Genentech, South San Francisco, CA

CORRESPONDING AUTHOR
Rebecca A. Miksad, MD, MPH, Flatiron Health, 233 Spring St, 5th Floor, New York, NY 10013, e-mail: rmiksad@flatiron.com.

REFERENCES

1. Kim C, Prasad V: Cancer drugs approved on the basis of a surrogate end point and subsequent overall survival: An analysis of 5 years of US Food and Drug Administration approvals. JAMA Intern Med 175:1992-1994, 2015
2. Kemp R, Prasad V: Surrogate endpoints in oncology: When are they acceptable for regulatory and clinical decisions, and are they currently overused? BMC Med 15:134, 2017
3. Wilson MK, Karakasis K, Oza AM: Outcomes and endpoints in trials of cancer treatment: The past, present, and future. Lancet Oncol 16:e32-e42, 2015

4. Eisenhauer EA, Therasse P, Bogaerts J, et al: New Response Evaluation Criteria in Solid Tumours: Revised RECIST guideline (version 1.1). Eur J Cancer 45:228-247, 2009

5. Khozin S, Abernethy AP, Nussbaum NC, et al: Characteristics of real-world metastatic non-small cell lung cancer patients treated with nivolumab and pembrolizumab during the year following approval. Oncologist 23:328-336, 2018

6. 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 375:2293-2297, 2016

7. Griffith SD, Tucker M, Bowser B, et al: 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. Adv Ther doi: 10.1007/s12325-019-00970-1 [epub ahead of print on May 28, 2019]

8. Cronin KA, Lake AJ, Scott S, et al: Annual report to the nation on the status of cancer, part I: National cancer statistics. Cancer 124:2785-2800, 2018

9. Eastman P: Blueprint proposes using real-world data to speed drug approvals. Oncol Times 38:1, 10-11, 2016

10. Abernethy AP, Gippetti J, Parulkar R, et al: Use of electronic health record data for quality reporting. J Oncol Pract 13:530-534, 2017

11. Curtis MD, Griffith SD, Tucker M, et al: Development and validation of a high-quality composite real-world mortality endpoint. Health Serv Res 53:4460-4476, 2018

12. Kaplan EL, Meier P: Nonparametric estimation from incomplete cbservations. J Am Stat Assoc 53:457-481, 1958

13. Gravetter FJ, Forzano LAB: Research Methods for the Behavioral Sciences (ed 4). Belmont, CA, Wadsworth, 2012, pp 78

14. 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 7:741-746, 2006

15. 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 33:1008-1014, 2015

16. Blumenthal GM, Kluetz PG, Schneider J, et al: Oncology drug approvals: Evaluating endpoints and evidence in an era of breakthrough therapies. Oncologist 22:762-767, 2017

17. Cappuzzo F, Delmonte A, Capici S, et al: Treatment beyond progression in patients with advanced squamous NSCLC participating in the Expanded Access Programme (EAP). J Thorac Oncol 12:S667-S668, 2017 (suppl; abstr P1.06-006)

18. 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 3:e001802, 2013

19. Gandhi L, Rodríguez-Abreu D, Gadgeel S, et al: Pembrolizumab plus chemotherapy in metastatic non-small-cell lung cancer. N Engl J Med 378:2078-2092, 2018

20. Hellmann MD, Ciuleanu TE, Pluzanski A, et al: Nivolumab plus ipilimumab in lung cancer with a high tumor mutational burden. N Engl J Med 378:2093-2104, 2018

21. 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 2:731-736, 2014

■■■

APPENDIX

Methods

Data Collection

The Flatiron electronic health record (EHR)–derived database includes structured data (eg, laboratory test results, information on prescribed drugs) and unstructured data (eg, radiology reports, pathology reports, medical care notes, biomarker tests). In order to prepare for analysis, the information is extracted from digital documents via technology-enabled chart abstraction. Every data point sourced from unstructured documents was abstracted manually from patient charts following precise prespecified policies and procedures that have been developed iteratively and tested for feasibility. Trained chart abstractors are experts (clinical oncology nurses and tumor registrars, with oversight from oncologists) tested through standard procedures, with both a conceptual test (eg, what constitutes real-world progression [rwP]) and a standardized testing module using the abstraction technology interface. Abstractor quality over time is monitored through dedicated performance procedures.

Variables in the Analyses

Mortality. Information about mortality from the study database was supplemented with mortality data sourced from Flatiron's composite mortality dataset (version 2). Structured plus unstructured EHR data from the Flatiron Health database were linked to a commercial death data source and US Social Security Death Index (Curtis MD, et al: Health Services Res 53:4460-4476, 2018 doi: https://doi.org/10.1111/1475-6773.12872). The composite dataset for non–small-cell lung cancer has sensitivity of 91% as benchmarked to the National Death Index. Measurement of overall survival in a sample cohort of advanced non–small-cell lung cancer (aNSCLC) patients revealed median survival very similar to that calculated using National Death Index data, establishing reliability of survival estimates generated from these data (Curtis MD, et al: Health Services Res 53:4460-4476, 2018 doi: https://doi.org/10.1111/1475-6773.12872).

Disease stage. The TNM group stage reported by the clinician was recorded. If group stage was not reported, it was calculated on the basis of reported T, N, and M stages. When applicable, group stage IV was inferred by the presence of metastatic disease. American Joint Committee on Cancer version 7 guidelines (Edge SB, et al [eds]. New York, NY, Springer, 2010) were used for initial diagnoses prior to July 1, 2018, and American Joint Committee on Cancer version 8 guidelines (Amin MB, et al [eds]. New York, NY, Springer, 2017) were used thereafter.

rwP Definition

For patients who met the study inclusion criteria, EHR data were curated to retrospectively determine rwP events. A "clinician-anchored" abstraction approach previously proposed and tested for cancer progression end points was used. This approach anchors on clinician-documented cancer progression based on the clinician's interpretation of the entire patient chart, including results of diagnostic procedures and tests (eg, radiology and pathology reports) (Griffith SD, et al: bioRxiv 504878, 2019 [preprint], doi: https://doi.org/10.1101/504878).

The date of cancer progression was defined as the date of the first source evidence for progression referenced by the clinician (eg, radiology report, pathology, exam) or the date of clinician note when no other corresponding evidence sources were documented. In prior experiments, this approach was more feasible and, potentially, more scalable than other methods to identify if (and when) a progression event occurred (eg, first searching for radiology reports /"radiology-anchored"), and resulted in similar data quality (Griffith SD, et al: bioRxiv 504878, 2019 [preprint], doi: https://doi.org/10.1101/504878). In cases where the clinician note indicated a mixed response, the event was recorded as rwP only if there was an associated treatment change; in case where an initial radiological study showed findings as "possible" (or other non-diagnostic term) progression, a rwP event was recorded only if progression is subsequently confirmed within a 30-day window.

The rwP variable was curated for all patients with aNSCLC in the cohort. Curation procedures were specific to aNSCLC. The entire patient chart was abstracted; including multiple rwP events when present. This analysis focused on the first rwP event after initiation of systemic therapy for advanced disease. Thus, most rwP events occurred for first-line treatment. Chart reviews indicated that imaging results obtained shortly after treatment initiation did not reflect treatment effectiveness, thus progression events occurring ≤ 14 days of therapy start were not analyzed.

Statistical Analysis

Feasibility analyses. The metrics used to evaluate completeness of the data were adjusted as follows: disease status assessments utilizing multiple imaging modalities (eg computed tomography chest and magnetic resonance imaging brain) within a 10-day window period were grouped together; clinician notes were grouped using a 3-day window. The median time to imaging and clinician assessment was calculated. The probability that at least the first (or second or third) imaging and clinician assessments were performed by 3 (and 12) months after the start of first-line systemic therapy was determined. Patients were censored at the time of death.

Sensitivity analyses. A series of sensitivity analyses were conducted to assess the robustness of the correlation between real-world end points (real-world progression-free survival, real-world time to progression, real-world time to next treatment) and overall survival. Three correlation calculations were re-performed, and three additional sensitivity analyses were conducted (Table A1): (1) for the subgroup of patients originally diagnosed with stage IV disease in order to remove patients with a different tumor biology and/or prior curative intent therapy; (2) including patients who had been censored in the primary analysis (censoring date serves as the event date); and (3) excluding patients with potential documentation gaps at the start of treatment that may represent variations in documentation.

Three additional sensitivity analyses were conducted to determine the potential impact of immunotherapy-related pseudoprogression (Wolchok JD, et al: Clin Cancer Res 15:7412-7420, 2009 doi: 10.1158/1078-0432.CCR-09-1624) on the correlations between rwP-based end points and overall survival: (1) exclude events for which the clinician mentioned suspected pseudoprogression in the note; (2) exclude patients with EHR-documented immunotherapy treatment because clinicians may not explicitly mention pseudoprogression; (3) restrict rwP events were restricted to those occurring before 1/1/2015 (ie, before approval of immunotherapy for aNSCLC) because treatments may not have been fully captured in the EHR data.
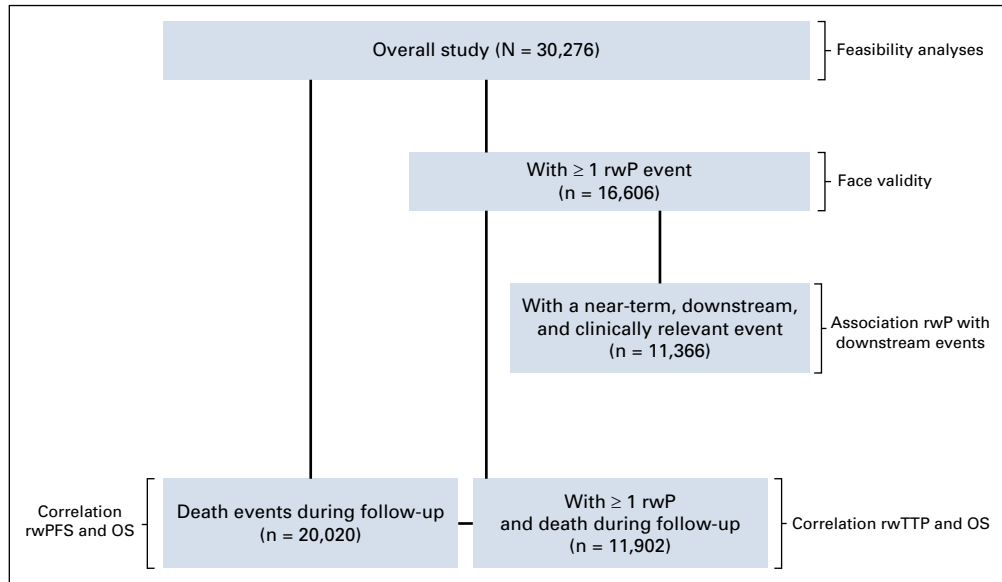
FIG A1. Patient flow and evaluable samples within the study population. OS, overall survival; rwP, real-world progression; rwPFS, real-world progression-free survival; rwTTP, real-world time to progression.
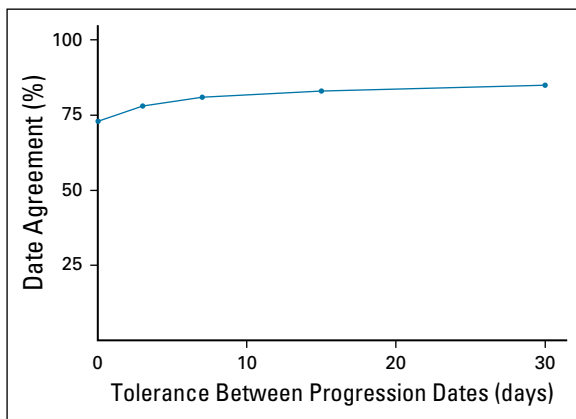
FIG A2. Inter-rater date agreement among duplicate abstracted patients. Event agreement, which indicates whether abstractors agreed on the presence or absence of at least one progression event, was 0.94 (95% CI, 0.93 to 0.95). Date agreement for those patients with at least one real-world progression event captured by two independent abstractors (n = 358) is displayed as a function of the date window (in days) applied.

TABLE A1. Sensitivity Analyses

Evaluation	Sensitivity Analysis	Rationale
Impact of clinical factors on the correlation of rwPFS, rwTTP, and rwTTNT with OS	Limiting cohort to those with stage IV disease at diagnosis	Excluding patients who may have a different tumor biology and/or prior treatment with curative intent
	Expanding cohort to include patients who were censored in the primary analysis	Censoring date serves as the event date
	Excluding patients with potential documentation gaps at the start of treatment	Documenting gaps may represent variations in documentation of disease status assessments
Impact of potential pseudoprogression from immunotherapy on the correlation of rwP-based end points (rwPFS and rwTTP) with OS[5]	Excluding events for which the clinician mentioned suspected pseudoprogression in the note	Reflecting that pseudoprogression may have different characteristics than rwP
	Excluding patients with EHR-documented immunotherapy treatment	Reflecting that pseudoprogression may not always be documented in the chart
	Restricting rwP events to those occurring before January 1, 2015 (ie, before approval of immunotherapy for aNSCLC)	Reflecting that treatments may not have been fully captured in EHR data

Abbreviations: aNSCLC, advanced non–small-cell lung cancer; EHR, electronic health record; OS, overall survival; rwP, real-world progression; rwPFS, real-world progression-free survival; rwTTNT, real-world time to next treatment; rwTTP, real-world time to progression.