Real-world Overall Survival Using Oncology Electronic Health Record Data: Friends of Cancer Research Pilot

Laura Lasiter[1], Olga Tymejczyk[2], Elizabeth Garrett-Mayer[3], Shrujal Baxi[2], Andrew J. Belli[4], Marley Boyd[5], Jennifer B. Christian[6], Aaron B. Cohen[2,7], Janet L. Espirito[5], Eric Hansen[4], Connor Sweetnam[8], Nicholas J. Robert[5], Mackenzie Small[6], Mark D. Stewart[1], Monika A. Izano[8], Joseph Wagner[6], Yanina Natanzon[9], Donna R. Rivera[10] and Jeff Allen[1,*]

In prior work, Friends of Cancer Research convened multiple data partners to establish standardized definitions for oncology real-world end points derived from electronic health records (EHRs) and claims data. Here, we assessed the performance of real-world overall survival (rwOS) from data sets sourced from EHRs by evaluating the ability of the end point to reflect expected differences from a previous randomized controlled trial across five data sources, after applying inclusion/exclusion criteria. The KEYNOTE-189 clinical trial protocol of platinum doublet chemotherapy (chemotherapy) vs. programmed cell death protein 1 (PD-1) in combination with platinum doublet chemotherapy (PD-1 combination) in first-line nonsquamous metastatic non-small cell lung cancer guided retrospective cohort selection. The Kaplan-Meier product limit estimator was used to calculate 12-month rwOS with 95% confidence intervals (CIs) in each data source. Cox proportional hazards models estimated hazard ratios (HRs) and associated 95% CIs, controlled for prognostic factors. Once the inclusion/exclusion criteria were applied, the five resulting data sets included 155 to 1,501 patients in the chemotherapy cohort and 36 to 405 patients in the PD-1 combination cohort. Twelve-month rwOS ranged from 45% to 58% in the chemotherapy cohort and 44% to 68% in the PD-1 combination cohort. The adjusted HR for death ranged from 0.80 (95% CI: 0.69, 0.93) to 1.15 (95% CI: 0.71, 1.85), controlling for age, gender, performance status, and smoking status. This study yielded insights regarding data capture, including ability of real-world data to precisely identify patient populations and the impact of criteria on end points. Sensitivity analyses could elucidate data set–specific factors that drive results.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

☑ Real-world data (RWD) have the potential to complement clinical trial data and fill gaps in knowledge about the performance of approved treatments used in routine care settings, including patient populations excluded from clinical trials, or where limited clinical trial data exist. There is interest in using real-world evidence to support regulatory decisions in rare cancer patient populations, new indications, alternative doses, and schedules.

WHAT QUESTION DID THIS STUDY ADDRESS?

☑ This study sought to evaluate the performance of rwOS and the considerations necessary to assess directionality of treatment associations in a real-world population across five US oncology electronic health record RWD providers with different sources of patient data by aligning the patient population with key inclusion/exclusion criteria from the KEYNOTE-189 study.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

☑ Insights were yielded regarding data capture, including ability of RWD to precisely identify patient populations and the impact of criteria on end points.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?

☑ Sensitivity analyses could elucidate data set–specific factors that drive results.

[1]Friends of Cancer Research, Washington, District of Columbia, USA; [2]Flatiron Health, New York, New York, USA; [3]American Society of Clinical Oncology, Alexandria, Virginia, USA; [4]COTA, Inc., Boston, Massachusetts, USA; [5]Ontada, Irving, Texas, USA; [6]IQVIA, Durham, North Carolina, USA; [7]New York University School of Medicine, New York, New York, USA; [8]Syapse, San Francisco, California, USA; [9]ConcertAI, Boston, Massachusetts, USA; [10]National Cancer Institute, Rockville, Maryland, USA. *Correspondence: Jeff Allen (jallen@focr.org)

Received July 14, 2021; accepted September 28, 2021. doi:10.1002/cpt.2443

Randomized controlled trials (RCTs) evaluate the safety and efficacy of medical products in specific patient populations under rigorously controlled conditions. While adherence to structured protocols, use of restrictive eligibility criteria, and patient randomization maximize the internal validity of RCT results, eligibility and protocol-directed care may reduce the relevancy of study results for broader patient populations receiving approved drugs subsequently in routine clinical practice. Limitations regarding the generalizability and transportability of trial findings create a value basis for the evaluation of patient outcomes in real-world settings.[1] Real-world data (RWD) have the potential to complement clinical trial data and fill gaps in knowledge about the performance of approved treatments used in routine care settings, including patient populations excluded from clinical trials, or where limited clinical trial data exist. As such, there is interest in using real-world evidence (RWE) to support regulatory decisions in rare cancer patient populations, new indications, alternative doses, and schedules.[2] Real-world end points measurable across data sources, including administrative claims and electronic health records (EHRs), that are consistently implemented across studies are needed to optimize data collection and accurate interpretation of real-world study findings.[3]

Replication of study findings across multiple data sources using common harmonized data elements and analytical framework is essential to evaluate the potential applications of RWE. In prior work, we established standardized definitions for oncology real-world end points derived from both EHR and claims data.[4] In this study, we assessed the performance of real-world overall survival (rwOS) by evaluating the ability of the end point to reflect expected differences from a previous RCT across multiple data sources, after applying inclusion/exclusion criteria.[5] Although not designed to replicate the clinical trial, this study used KEYNOTE-189, an RCT published in 2018, as a relative benchmark to explore the performance of rwOS across data sources. KEYNOTE-189 demonstrated improved outcomes of pembrolizumab in combination with platinum therapy (cisplatin or carboplatin) and pemetrexed, compared with chemotherapy alone for frontline treatment of patients with nonsquamous metastatic non-small cell lung cancer (metastatic non-small cell lung cancer). By estimating treatment effects in each data set, we sought to distinguish the effects of underlying patient characteristics (e.g., age, biomarker status, and health performance status) from data set–specific considerations (e.g., completeness of key variables). Furthermore, this work generated insights into methodological transparency, data quality, and reporting standards for real-world outcome measures that can inform the interpretation of the results of real-world studies in oncology.

METHODS

Data sources

Five organizations supplying EHR data aligned on a common set of definitions and protocols (**Table 1**, **Table 2**, **and Table S1**). Each data partner conducted data extraction and statistical analyses using deidentified patient data from their respective real-world population and reported aggregated data only. EHR data was sourced through structured (programmatic database extractions) and/or unstructured (chart review) methods conducted in accordance with abstraction rules and quality processes established within each organization.

Study population

Similar to KEYNOTE-189, the real-world population selected for this study included patients with metastatic non-small cell lung cancer (mNSCLC) who initiated frontline treatment in the metastatic setting with combination platinum therapy (chemotherapy) or pembrolizumab plus combination platinum therapy (programmed cell death protein 1 [PD-1] combination), where combination is defined as cisplatin or carboplatin plus pemetrexed (**Figure S1**). Eligible patients had a documented encounter (defined as a physician visit, drug administration, or vitals documentation) in each database on two or more separate occasions on or after January 1, 2011 through March 31, 2018. Frontline treatment was defined as the first regimen subsequent to the date of metastatic diagnosis and included all agents received within 30 days following the day of first administration or noncanceled order after metastatic diagnosis. The index date was the date frontline therapy was initiated. Patients with an index date on or after January 1, 2015 but no later than March 31, 2018 were included in order to improve temporal proximity of treatment groups. Patients with a gap of greater than 120 days from the date of metastatic diagnosis to the first clinical encounter (structured and/or unstructured) were excluded as having possibly incomplete early therapy data. Data cutoff was March 31, 2019, to allow a minimum of 12 months' potential follow-up.

Initial trial-related inclusion and exclusion criteria ("baseline cohort")

The real-world cohorts were identified following key inclusion and exclusion criteria reported by KEYNOTE-189 (**Table S2**). Patients were included in the real-world cohorts if they had evidence of pathologically confirmed mNSCLC (patients metastatic at diagnosis or patients with earlier-stage disease who progressed to metastatic disease), received no previous systemic antineoplastic therapy at any point in the metastatic setting or in the advanced setting within 12 months prior to metastatic diagnosis date, and received at least one dose of either a PD-1 combination or chemotherapy frontline regimen. Patients were included regardless of programmed death-ligand 1 (PD-L1) testing or status. The cohort meeting these criteria is referred to as the "baseline cohort" and included many patients who would likely not have qualified for a typical clinical trial in mNSCLC.

Additional trial-related exclusion criteria ("fully restricted cohort")

To explore potential associations between survival estimates and prognostic factors typically excluded from clinical trials, the following exclusion criteria were applied to the baseline cohort: squamous cell carcinoma or non-small cell lung cancer (NSCLC) not otherwise specified, evidence of inadequate kidney or liver organ function, Eastern Cooperative Oncology Group (ECOG) performance status (PS) ≥2, and evidence of EGFR/ALK (epidermal growth factor receptor / anaplastic lymphoma kinase) sensitizing mutations, all at index date. Patients without organ function, ECOG, or EGFR/ALK data were included. Exact definitions of the criteria, as implemented by each group, are reported in **Table 2**. The cohort meeting all these additional exclusion criteria is referred to as the "fully restricted cohort" and is intended to represent the strict eligibility requirements for patient populations typically enrolled in clinical trials.

End-point definitions

Real-world OS was defined as the length of time from the index date to the date of death. If there was no evidence of death, patients were censored at the last recorded clinical activity prior to data cutoff. The implementation of rwOS by each group is reported in **Table 1**.

Table S1 describes implementation of key descriptive variables and model covariates: disease stage (0, I, II, III, IV, unknown), smoking status (known history of smoking, no known history of smoking), and PD-L1

Table 1 Variations across data sources and implementation of end-point definition

Data set	Description of data source	Population	Derivation of date of death	Censor date
A	Structured and unstructured EHR data and commercial obituary data.	Academic and community practice patients in the United States; outpatient; initiated frontline therapy between 2015 and 2018.	An algorithm is used. If dates agree across the three data sources, the date is selected. If discrepancy <7 days exists, EHR data is preferentially captured. If discrepancy >7 days exists, EHR data with accompanying source documentation (e.g., death certificate) is prioritized, otherwise commercial obituary data is captured.	Date of last clinical activity prior to data cutoff, defined as in-person visit event with healthcare provider such as treatment administration or collected test.
B	Structured and unstructured EHR data, commercial obituary data, and Social Security Death Index (SSDI) data.	Academic and community practice patients in the United States; outpatient; initiated frontline therapy between 2015 and 2018.	An algorithm is used. If all dates agree across the three data sources, the date is selected. If any two dates agree, that date is selected. If all three dates disagree, the following hierarchy is applied: SSDI, obituary, EHR data. If a day level DoD is available in abstracted EHR data, that date is selected over the consensus structured date. Exact date was used, where available. If only month-level date was available, it was generalized to the end of month. If only year-level date was available, if was generalized to the end of year.	Date of last structured activity, defined as the most recent visit prior to data cutoff.
C	Structured and unstructured EHR data; hospital-based, enterprise-wide, and national cancer registries; commercial obituary data; and digitized obituaries.	Community practice patients in the United States; inpatient and outpatient; initiated frontline therapy between 2015 and 2018.	An algorithm is used. Tumor registry (hospital, enterprise-wide, national) dates were preferentially selected, followed by structured EHR, followed by commercial obituary data.	Date of last contact (physical encounter, medication order, or medication administration, from structured or unstructured EHR data) prior to the data cutoff.
D	Structured EHR data.	Community practice patients in the United States; outpatient; initiated frontline therapy between 2015 and 2018.	Actual date of death documented from EHR or DMF.	Date of last structured clinical activity prior to data cutoff.
E	Structured EHR data, structured claims data.	Academic and community practice patients in the United States; primarily outpatient; initiated frontline therapy between 2017 and 2018.	Mortality algorithm incorporating EHR and Claims.	Date of last clinical encounter with healthcare provider prior to data cutoff.

DMF, death master file; DoD, date of death; EHR, electronic health record.

expression status (<1%, 1–49%, ≥50%, unknown) and evidence of brain metastases (yes, no) at index date.

Statistical analysis

The Kaplan-Meier product limit estimator was used to calculate 12-month and median rwOS with 95% confidence intervals (CIs) for the baseline and fully restricted cohorts in each data source. Cox proportional hazards models were used to estimate hazard ratios (HRs) and associated 95% CIs for the associations between treatment groups and rwOS in the fully restricted cohort. Unadjusted HRs for the association between treatment groups and rwOS were calculated in the baseline and the fully restricted cohorts. Additionally, unadjusted models for the fully restricted cohort were stratified by age, gender, PD-L1 expression status (<1%, ≥1%, 1–49%, ≥50%), evidence of brain metastases (yes, no), and platinum drug agent (carboplatin, cisplatin). Lastly, we adjusted for age (<65, ≥65 years), gender (male, female), smoking status (known history of smoking, no known history of smoking), and ECOG PS (0, 1, unknown) to account for confounding by prognostic factors not homogenized in inclusion/exclusion criteria. Forest plots were used to visualize the range of estimates across data sets.

Post hoc analyses

To explore potential associations between estimates of survival and individual prognostic factors, and evaluate the impact of methodological approaches to deriving them from RWD, exclusion criteria were applied sequentially to the baseline cohort in a stepwise manner and unadjusted HRs for the associations between treatment groups and rwOS were calculated after each restriction step:

Step 0: Baseline cohort (as defined above)
Exclude patients with:
Step 1: Squamous cell carcinoma or NSCLC not otherwise specified
Step 2: Inadequate kidney or liver organ function at index date
Step 3: Eastern Cooperative Oncology Group (ECOG) performance status at index date ≥2
Step 4: Evidence of EGFR/ALK sensitizing mutations at index date

Table 2 Definitions of inclusion and exclusion criteria for alignment with Keynote-189

Data set	Advanced diagnosis	Metastatic status	Histology	Organ function	ECOG PS	EGFR/ALK sensitizing mutations
Overall	Diagnosis with advanced disease defined as American Joint Committee on Cancer stage: Stage IIIB, IIIC or IV NSCLC at initial diagnosis or Early stage (stages I, II, and IIIA) NSCLC with a recurrence or progression to metastatic status.	Each group determined metastatic status according to internally consistent protocols.	Patients were placed into one of the following categories: Non-squamous cell carcinoma Squamous cell carcinoma NSCLC histology not otherwise specified (NOS)	Patient is excluded if there is evidence of inadequate kidney/liver function based on laboratory values in the 90 days prior to and including the index date based upon structured lab data. Inadequate renal function is defined as creatinine clearance of <50 ml/min (creatinine clearance calculated using the Cockroft-Gault equation). Inadequate liver function is calculated as a serum total bilirubin of ≥1.5× ULN (unless direct bilirubin was measured on the same date and <ULN) and AST or ALT ≥5× ULN NOTE: If there are multiple lab values of interest in the time window, the value closest to index date should be used. If there are multiple values for a given lab on the same day, lower creatinine clearance values/ higher bilirubin, AST, and ALT values should be used (i.e. prioritize values that exclude patients)	Patient's ECOG PS at the time of the index date: 0 1 2+ Unknown NOTE: ECOG PS may have been recorded up to 30 days prior to the index date, *OR* up to 7 days after the index date, whichever is closest to the index date. If there are multiple ECOG PS values at the same absolute distance from the index date, priority is given to the ECOG PS value that precedes the index date. For patients with multiple ECOG PS values recorded on the same day, the highest value will be selected.	Approach 1 to the ALK/EGFR exclusion criterion: Test result showing ALK/EGFR rearrangement/mutation at any point before or up to 30 days after the index date. Approach 2 to the ALK/EGFR exclusion criterion: (alone or in combination with approach 1 above) Patients who are prescribed EGFR or ALK targeting therapies in the first 6 months of the index date will be excluded Erlotinib, Afatinib, Osimertinib, Gefitinib, Crizotinib, Ceritinib, Alectinib
A	Data source: Abstracted EHR data. Definition: No deviation.	Data source: Abstracted EHR data Definition: Defined as at least one of the following: stage IV, TNM M1, and/or metastatic progression	Data source: Abstracted EHR data. Definition: No deviation.	Data source: Structured EHR data. Definition: No deviation.	Data source: Structured EHR data. Definition: No deviation.	Data source: Abstracted EHR data Definition: Used a combination of approach 1 and 2. No deviation from definitions.
B	Data source: Abstracted EHR data. Definition: No deviation.	Data source: Abstracted EHR data. Definition: Stage IV only.	Data source: Abstracted EHR data. Definition: No deviation.	Data source: Structured EHR data. Definition: No deviation.	Data source: Structured EHR data. Definition: No deviation.	Data source: Abstracted EHR data. Definition: Approach 1 without deviation. Test date was identified as the most recent date across the "specimen collected" date, "specimen received" date, and "result date" variables. If multiple tests were recorded, the one closest to the index date was used. If multiple tests were recorded on the same date, the result showing rearrangement / mutation was used.

Table 2 (Continued)

Data set	Advanced diagnosis	Metastatic status	Histology	Organ function	ECOG PS	EGFR/ALK sensitizing mutations
C	Data source: Unstructured EHR data and structured pathology reports. Definition: Recurrence or progression to metastatic status is the date of medical oncologist stated metastasis or distant recurrence collected from unstructured EHR.	Data source: Unstructured EHR. Definition: Metastatic status is based on a medical oncologist statement.	Data source: Unstructured EHR and structured pathology reports. Definition: Main sources of histology are not codified in the source data but are applied post data collection. Agreed upon vocabulary based on WHO lung cancer histology hierarchy management is utilized to derive NSCLC grouping.	Data source: Structured EHR. Definition: Organ function was not evaluated if value, unit, or range were missing.	Data source: Unstructured and structured EHR. Definition: If Karnofsky performance status (KPS) was available instead of ECOG PS, KPS was converted to ECOG PS.	Data source: Unstructured and structured EHR; commercial laboratory electronic reports. Definition: Used a combination of approach 1 and 2. Definition of ALK rearrangement is based on laboratory report or physician statement. EGFR mutations are defined as any reported mutation on exon 18 through exon 21 regardless of classification. Approach 2 did not deviate.
D	Data source: Structured EHR data. Definition: No deviation.	Data source: EHR data. Definition: Metastatic status identified by having at least one of the following: Stage IV or M1 disease; documented organ site of metastasis; or disease status as metastatic disease	Data source: EHR data. Definition: No deviation.	Data source: EHR data. Definition: No deviation.	Data source: EHR data. Definition: No deviation.	Data source: EHR data. Definition: No deviation.
E	Data source: Structured EHR. Definition: To identify staging with ICD9/10 used to classify patients with early stage who progressed.	Data source: Structured EHR. Definition: Diagnosis, staging and M values.	Data source: Structured EHR histology codes and descriptions. Definition: For a subset of patients, NSCLC was indicated, but not Squamous vs. Non-Squamous. These patients were kept since they could not be definitively classified as either.	Data source: Structured EHR. Definition: No deviation.	Data source: Structured EHR. Definition: No deviation.	Data source: Structured and unstructured EHR. Definition: No deviation.

ALT, alanine aminotransferase; AST, aspartate aminotransferase; ECOG PS, Eastern Cooperative Oncology Group performance status; EGFR/ALK, epidermal growth factor receptor/anaplastic lymphoma kinase; EHR, electronic health record; ICD 9/10, International Classification of Diseases, Ninth and Tenth Revisions; NSCLC, non-small cell lung cancer; TNM M1, tumor, node, and metastasis metastasis 1; ULN, upper limit of normal; WHO, World Health Organization.

Table 3 Characteristics of fully restricted cohorts

	KEYNOTE-189	A	B	C	D	E
Characteristic	**Chemotherapy/PD-1 combination treated patient characteristics**					
Total number of patients/ Treatment, no.	410/206	346/54	1,501/405	232/36	748/132	155/125
Age						
Median, yrs (IQR)	65, 34, 84/64, 34, 84	68, 60, 74/65, 60, 72	67, 59, 74/65, 59, 72	66, 59, 73/64, 58, 71	67, 60, 74/64, 58, 72	68, 59, 74/65, 60, 73
<65 yr, %	48.0/55.8%	38.7/50.0%	42.4/47.2%	46.1/50.5%	41.7/52.3%	37.4/43.2%
Gender, Male, %	62.0/52.9%	45.7/55.6%	45.7/55.6%	52.6/63.9%	46.9/58.3%	47.7/52.8%
ECOG, %						
0	45.4/38.8%	25.4/24.1%	21.0/31.4%	14.7/22.2%	14.7/30.3%	19.4/24.8%
1	53.9/60.7%	46.2/44.4%	33.7/37.5%	29.7/38.9%	63.2/43.2%	35.5/31.2%
Unknown	0.5/0.5%	28.3/31.5%	45.3/31.1%	55.6/38.9%	22.1/26.5%	45.2/44.0%
Smoking status, %						
Evidence of smoking	88.3/87.9%	88.2/87.0%	88.3/89.6%	13.8/36.1%	84.9/84.8%	25.2/25.6%
No evidence	11.7/12.1%	11.8/13.0%	11.7/10.4%	86.2/63.9%	15.1/15.2%	74.8/74.4%
Histology, %						
Non-squamous cell carcinoma	96.1/96.1%	100/100%	100/100%	100/100%	100/100%	100/100%
NOS	2.4/1.9%	Restricted	Restricted	Restricted	Restricted	Restricted
Brain metastases, %						
Evidence of	17.8/17.0%	31.8/29.6%	18.5/14.1%	13.8/19.5%	13.2/9.1%	16.1/20.0%
No evidence	82.2/83.0%	68.2/70.4%	81.5/85.9%	86.2/80.6%	86.8/90.9%	83.9/80.0%
PD-L1 expression status, %						
<1%	31.0/30.6%	11.9/12.8%	23.7/16.3%	50.9/20.8%	50.5/32.3%	NA[a]
>1%	63.4/62.1%	19.3/42.6%	51.9/65.4%	49.1/79.2%	43.5/64.6%	NA
1–49%	31.2/28.2%	14.8/34.0%	39.0/34.6%	26.4/41.7%	39.6/38.5%	NA
>50%	32.2/34.0%	5.9/6.4%	12.9/30.9%	22.6/37.5%	3.8/26.2%	NA
Unknown	5.6/7.3%	67.4/46.8%	24.4/18.3%		6.0/3.1%	NA
Renal function, %						
No Evidence of Inadequate Function		70.5%/64.8%	89.5/93.1%	100/100%	81.0/77.3%	13.5/16.8%
Unknown%		29.5%/35.2%	10.5/6.9%	0.0/0.0%	19.0/22.7%	86.5/83.2%
Hepatic function, %						
No evidence of inadequate function		70.5%/59.3%	83.3/87.4%	100/100%	79.9/76.5%	49.7/71.2%
Unknown		29.5%/40.7%	16.7/12.6%	0.0/0.0%	20.1/23.5%	50.3/28.8%
Median time from advanced diagnosis to frontline therapy initiation, months (IQR)		1.00, 0.57, 1.63/0.97, 0.62, 1.58	1.2,0.8,1.7/ 1.1,0.7,1.5	1.25, 0.90, 1.80/1.17, 0.70, 1.95	0.83, 0.37, 1.47/0.72, 0.37, 1.38	1.03, 0.53, 1.77/0.93, 0.50, 1.77
Median structured follow-up time from frontline therapy initiation, months (IQR)		10.98, 5.14, 21.28/11.85, 5.60, 14.84	7.2,2.8,17.6/ 10.1,3.5,15.6	12.27, 4.93, 25.04/10.37, 5.36, 17.12	10.08, 3.67, 19.73/12.98, 3.55, 16.10	13.53, 5.93, 14.06/14.06, 6.10, 21.13
Status at initial diagnosis, %						
Advanced at diagnosis		87.6/88.9%	100/100%	88.8/83.3%	95.6/97.0%	96.1/92.0%
Progressed after initial diagnosis		12.4/11.1%		11.2/16.7%	4.4/3.0%	3.9/8.0%
Stage, %						
0		0.0/0.0%		0.0/0.0%	0.0/0.0%	1.3/0.8%

Table 3 (Continued)

Characteristic	KEYNOTE-189	A	B	C	D	E
	Chemotherapy/PD-1 combination treated patient characteristics					
I		6.6/5.6%		6.9/5.6%	0.9/0.0%	0.6/0.0%
II		2.6/0.0%		0.9/2.8%	0.0/0.0%	0.0/0.0%
III		3.2/5.6%		3.4/8.3%	2.9/0.0%	0.0/0.8%
IV		87.6/88.9%	100/100%	87.9/83.3%	95.6/97.0%	95.5/89.6%
Unknown		0.0/0.0%	0.0/0.0%	0.9/0.0%	0.0/0.0%	2.6/8.8%
Index year, %						
2015		35.5/1.9%	36.2/0.0%	34.5/0.0%	29.7/0.0%	0.0/0.0%
2016		36.1/0.0%	36.8/0.2%	35.3/0.0%	33.7/0.0%	0.0/0.0%
2017		23.7/63.0%	22.7/71.6%	23.7/75.0%	30.5/79.5%	83.9/69.6%
2018		4.6/35.2%	4.2/28.1%	6.5/25.0%	6.1/17.4%	16.1/30.4%

ECOG, Eastern Cooperative Oncology Group; NOS, not otherwise specified; PD-1, programmed cell death protein 1; PD-L1, programmed death-ligand 1.
[a]Data were not provided for this analysis.

To evaluate the potential impact of crossover on results, the proportion of crossover in the chemotherapy group, defined as initiation of an immunotherapy (pembrolizumab, nivolumab, or atezolizumab)–containing second line treatment prior to the date of data cutoff was calculated.

RESULTS

Characteristics of patients in each group's fully restricted and baseline cohorts, as well as of patients included in KEYNOTE-189, are described in **Table 3 and Table S3**, respectively. The fully restricted cohorts included 155 to 1,501 chemotherapy and 36 to 542 PD-1 combination patients. Median age in chemotherapy and PD-1 combination groups ranged between 66–68 and 64–65, respectively. The proportion of patients <65 years of age was higher in the PD-1 combination group than in the chemotherapy group by a margin ranging from 4 to 11 percentage points across data sets. The proportion of males in the PD-1 combination groups was 53% to 64% while the chemotherapy group had a slightly lower proportion (46% to 53%). A substantial proportion of patients had unknown ECOG (22–56% across cohorts and data sets), unknown liver function (0–50%), or unknown kidney function (0–87%). History of smoking was identified in >84% of patients in three data sets, and in <40% of patients in the remaining two data sets, due to data missingness. PD-L1 expression status was provided in four data sets, with a proportion of unknown values ranging from 0 to 67% across treatment groups and data sets. A higher proportion of chemotherapy patients had evidence of a PD-L1 expression status <1% compared with PD-1 combination patients, (where percent unknown was low). Patients whose cancer progressed to metastatic after initial diagnosis comprised up to 17% of chemotherapy or PD-1 combination cohorts in four data sets, but were excluded from one data set (i.e., only stage IV patients were included). PD-1 combination patients initiated treatment in 2017–2018, while patients in the chemotherapy group initiated treatment from 2015 to 2018. The proportion of patients initiating either treatment in 2018 was low due to the cutoff date for frontline treatment initiation by March 31, 2018.

Table 4 includes estimates of rwOS at 12 months, as well as unadjusted associations between frontline therapy and rwOS, overall and stratified by key variables of interest, in the fully restricted cohorts. Twelve-month rwOS ranged from 45% to 57% in the chemotherapy group and 44% to 69% in the PD-1 combination group. Unadjusted HR for death, comparing PD-1 combination to chemotherapy, ranged from 0.79 (95% CI: 0.68, 0.92) to 1.10 (95% CI: 0.70, 1.72). In four of the five fully restricted cohorts, HR point estimates were below 1, and four confidence intervals overlapped 1.

Table 5 lists adjusted associations between frontline therapy and rwOS in the fully restricted cohorts, controlling for age (<65 vs. ≥65 years), gender, ECOG (0, 1, or unknown), and smoking status (history, no history, unknown). The adjusted HR ranged from 0.80 (95% CI: 0.69, 0.93) to 1.15 (96% CI: 0.71, 1.85). HR point estimates were below 1 in four of the five fully restricted cohorts and four confidence intervals overlapped 1.

Post hoc analyses explored the unadjusted associations between frontline therapy and rwOS during sequential exclusion step (baseline cohort to fully restricted cohort). Patient numbers ranged from 293 to 2,673 in baseline and 164 to 1,906 in fully restricted cohorts (**Figure S1**). Numerical differences in HRs in analyses with varying exclusion criteria were very small. The finding of statistically significant association or lack thereof was not altered in any of the five data sets across sequential application of criteria.

Between 37% and 44% of the chemotherapy patients in the fully restricted cohort initiated a second line containing immunotherapy prior to the data cutoff (data not shown).

DISCUSSION

This study evaluated the reproducibility and performance of rwOS across five real-world mNSCLC patient data sets receiving chemotherapy or PD-1 combination after applying selected clinical trial inclusion/exclusion criteria.[6] Specifically, the study evaluated whether the application of common inclusion/exclusion criteria and methods across different data sources would result in similar rwOS findings, and if not, whether the observed

Table 4 Unadjusted associations between use of frontline therapy and rwOS

	Keynote 189	A	B	C	D	E
Number of patients	616	346/54	1,501/405	232/36	748/132	155/125
Number of events (Chemotherapy/ PD-1 combination)	235	203/25	1094/216	163/22	421/61	164
12-month OS (Chemotherapy/ PD-1 combination)	0.49/0.69	0.57/0.67	0.45/0.53	0.53/0.44	0.56/0.62	0.51/0.56
Unadjusted Hazard ratio (HR) for death (95% CI)	0.49 (0.38–0.64)	0.99 (0.65–1.50)	0.79 (0.68–0.92)	1.10 (0.70–1.72)	0.92 (0.70–1.20)	0.97 (0.71–1.33)
Age HR (95% CI)						
<65	0.43 (0.31–0.61)	1.25 (0.71–2.20)	0.75 (0.60–0.94)	1.30 (0.68–2.49)		1.26 (0.79–2.0)
>65	0.64 (0.43–0.95)	0.80 (0.42–1.53)	0.84 (0.69–1.03)	0.94 (0.49–1.81)	1.05 (0.88–1.26)	0.74 (0.48–1.15)
Sex HR (95% CI)						
Male	0.70 (0.50–0.99)	0.81 (0.44–1.48)	0.75 (0.63–0.91)	1.08 (0.49–2.36)	0.96 (0.80–1.15)	1.31 (0.85–2.02)
Female	0.29 (0.19–0.44)	1.24 (0.70–2.22)	0.81 (0.63–1.03)	1.07 (0.62–1.87)		0.72 (0.45–1.14)
PD-L1 expression status HR (95% CI)						
<1%	0.59 (0.38–0.92)	1.84 (0.58–5.81)	1.02 (0.69–1.51)	1.17 (0.34–4.03)		NA
>1%	0.47 (0.34–0.66)	1.45 (0.57–3.67)	0.70 (0.55–0.90)	0.76 (0.33–1.74)	0.80 (0.56–1.14)	NA
1–49%	0.55 (0.34–0.90)	1.55 (0.56–4.30)	0.77 (0.57–1.05)	1.24 (0.44–3.45)	0.77 (0.52–1.13)	NA
>50%	0.42 (0.26–0.68)	0.91 (0.08–10.21)	0.69 (0.44–1.08)	0.31 (0.06–1.53)	0.70 (0.37–1.32)	NA
Brain metastases HR (95% CI)						
Evidence of	0.36 (0.20–0.62)	1.63 (0.80–3.32)	0.96 (0.67–1.39)	0.95 (0.32–2.81)		1.44 (0.73–2.84)
No evidence of	0.53 (0.39–0.71)	0.78 (0.46–1.32)	0.76 (0.65–0.90)	1.12 (0.68–1.83)	1.04 (0.80–1.34)	0.86 (0.60–1.23)
Platinum-based drug HR (95% CI)						
Carboplatin	0.52 (0.39–0.71)	0.94 (0.62–1.44)	0.78 (0.67–0.90)	NA		0.94 (0.69–1.29)
Cisplatin	0.41 (0.24–0.69)	N/A	1.34 (0.19–9.71)	1.03 (0.66–1.63)	0.58 (0.38–0.88)	NA

CI, confidence interval; NA, not available; OS, overall survival; PD-1, programmed cell death protein 1; PD-L1, programmed death-ligand 1; rwOS, real-world overall survival.

Table 5 Adjusted associations between frontline therapy and rwOS in the fully restricted cohorts

Covariate	A	B	C	D	E
Overall adjusted HR (95% CI)					
Chemotherapy	Ref.	Ref.	Ref.	Ref.	Ref.
PD-1 combination	0.99 (0.65–1.51)	0.80 (0.69–0.93)	1.15 (0.71–1.85)	0.96 (0.73–1.26)	0.95 (0.69–1.30)
Age HR (95% CI)					
<65	Ref.	Ref.	Ref.	Ref.	Ref.
>65	1.18 (0.90–1.55)	1.16 (1.04–1.30)	1.20 (0.90–1.61)	1.04 (0.86–1.25)	0.72 (0.53–0.98)
Gender HR (95% CI)					
Male	Ref.	Ref.	Ref.	Ref.	Ref.
Female	0.97 (0.74–1.26)	0.80 (0.72–0.90)	0.79 (0.59–1.06)	0.94 (0.79–1.13)	0.84 (0.62–1.15)
ECOG PS HR (95% CI)					
0	Ref.	Ref.	Ref.	Ref.	Ref.
1	1.93 (1.38–2.70)	1.40 (1.21–1.63)	1.51 (0.93–2.46)	1.31 (1.01–1.69)	1.07 (0.69–1.64)
Unknown	1.09 (0.74–1.59)	1.25 (1.08–1.45)	1.48 (0.94–2.34)	1.29 (0.96–1.75)	0.90 (0.53–1.52)
Smoking status HR (95% CI)					
Evidence of history of	Ref.	Ref.	Ref.	Ref.	Ref.
No evidence of history of	0.99 (0.67–1.47)	0.84 (0.70–1.00)	1.11 (0.73–1.68)	1.01 (0.78–1.32)	1.54 (0.61–3.90)
Unknown/missing population	NA	NA	NA	0.76 (0.39–1.48)	0.78 (0.48–1.28)

CI, confidence interval; ECOG PS, Eastern Cooperative Oncology Group performance status; HR, hazard ratio; NA, not available; OS, overall survival; PD-1, programmed cell death protein 1; Ref., reference; rwOS, real-world overall survival.

results could be explained by differences in the underlying patient populations or data-specific characteristics. While a trial replication or direct comparison was outside the scope of this work, trial-based inclusion/exclusion criteria were used to align the real-world populations and provide a relative benchmark for evaluating the performance of the real-world end point. The stepwise application of inclusion/exclusion criteria, intended to provide additional insights on factors that drove differences between estimates, did not indicate convergence or divergence of results.

The estimated treatment effects varied across data sets. Sample size limitations may have contributed to the lack of statistical significance in some of the real-world data sets compared with KEYNOTE-189. Real-World OS for patients receiving immunotherapy was shorter in real-world cohorts than in the trial, consistent with the findings from other real-world NSCLC analyses.[7–9] The range of observed results across real-world cohorts and as compared with the trial may be partially attributable to outstanding differences in cohort composition, data missingness, interpretation of end-point definitions and their completeness, and differences in routine vs. clinical trial care and data collection patterns.

While several, common inclusion/exclusion criteria were used, differences in their application and in missing data patterns across the five groups may have contributed to variability of observed results. Data missingness is a known and challenging aspect of observational data. Specifically, the proportion of patients missing ECOG PS was up to a quarter and patients lacking laboratory values to ascertain organ function comprised up to 87% of study cohorts, allowing substantial heterogeneity in cohort characteristics. Missing International Classification of Diseases (ICD) or laboratory values may have resulted in differential misclassification of entry criteria and key covariates, potentially biasing observed estimates.

The proportion of patients with evidence of brain metastases in real-world data sets also varied widely, as determined using ICD codes. Given the low sensitivity of ICD codes in identifying brain metastases, differences in this important prognostic factor could lead to variable estimates of the 12-month rwOS. Additional, unmeasured sources of heterogeneity, such as comorbidities, socioeconomic status, health-insurance coverage, and variation in care between community and academic practices may have contributed to the variability in adjusted estimates. Given high rates of crossover in the chemotherapy groups, potentially variable treatment timing, and duration and dosing may have contributed to the relatively strong performance of patients receiving chemotherapy.[10]

Differences in mortality ascertainment may have also contributed to the observed range of results. The sensitivity and specificity of mortality information, obtained from a variety of sources ranging from only structured EHRs to composite approaches of chart review, third party mortality data sources, and the Social Security Death Index data, varied across real-world data sets. Poor completeness of mortality data leads to overestimated survival,[7] and lower statistical power. Additionally, should completeness and/or accuracy of the data vary within cohorts (e.g., due to improved records over time), measures of association may be biased. Finally, granularity of available death dates and handling of partially complete dates may have also varied across groups.

Descriptive comparisons to KEYNOTE-189 should consider differences in patient care in a trial setting. Strict trial protocols dictate regular data collection at baseline and follow-up intervals for RECIST objective response and mortality assessments, whereas

real-world studies observe care as it is delivered and recorded in clinical practice. Consistent reporting of data completeness is critical to inform appropriate analysis, including potential sensitivity analyses, and interpretation of results, both within and across data sets.

The study's ability to apply further inclusion/exclusion criteria, as well as conduct sensitivity analyses, was limited by sample size considerations. Some data sources used only structured data, which limited the extent of the covariate information collected compared with using unstructured data. Since participating groups had different underlying data availability, sensitivity analyses could not be performed consistently across groups, limiting insights into the mechanism and impact of missing data, for example on ECOG PS or kidney and liver organ function. The study partially used noncontemporaneous controls, which could complicate outcome interpretation. Finally, blinding of groups precluded the discussion of data set-specific nuances which could inform observed differences.

In future work, sensitivity analyses can help elucidate data set–specific factors that may drive results. These include only selecting patients who: (i) had stage IV NSCLC at diagnosis; (ii) initiated frontline treatment in 2017 or 2018; (iii) had comparable distribution of potential follow-up across cohorts; (iv) had known ECOG PS and organ function; and (v) had known timing and duration of treatment prior to crossover. It is also important to understand missingness for core variables to inform the selection of proper analytic methods for main and/or sensitivity analyses (imputation-based or model-based approaches vs. complete case analysis). In studies evaluating multiple sources of RWD, sequential application of eligibility criteria can be considered to evaluate consistency of results across data sets and the impact of select clinical characteristics. Designing a shared RWD master protocol *a priori* may assist in understanding differences, promoting efficiency, and increasing reproducibility. Additionally, sensitivity analyses assessing the variability resulting from different mortality assessment approaches and determination of exact death date across data sets (**Table 1**) could inform the relative contribution of these factors to observed differences. Lastly, future analyses that include additional clinical demographic factors and social determinants of health merit future investigation.

This study sought to evaluate the performance of rwOS and the considerations necessary to assess directionality of treatment associations in a real-world population across five US oncology EHR RWD providers with different sources of patient data by aligning the patient population with key inclusion/exclusion criteria from the KEYNOTE-189 study. Such efforts to achieve consistency across RWD sources are necessary to distinguish true treatment effects from ones driven by methodological choices, missing data, confounding, and unmeasurable influences on treatment choice. While an association between frontline treatment and rwOS was not consistently detected, differences in methodologies, delivery of care (protocol vs. observational), capture of critical data elements (required routinely throughout RCT), and residual heterogeneity in real-world patient cohorts help contextualize the observed similarities and differences across the real-world data sets and as compared with the trial. Measuring real-world effectiveness and safety in routine care alongside clinical trials may have an important role in completing the picture of how well a therapy works, for which patients it is most effectively useful, and under what conditions in the future. Building on this research, agreement on minimum reporting and performance standards and capturing of post-baseline events (e.g., frequency and timing of treatment crossover) or subsequent treatments, as well as a process to evaluate real-world end points across data sets could inform best practices that may help unlock the potential of EHR-derived RWD.

SUPPORTING INFORMATION

Supplementary information accompanies this paper on the *Clinical Pharmacology & Therapeutics* website (www.cpt-journal.com).

ACKNOWLEDGMENTS

We thank Eric Hua and Grace Collins for their assistance with data management and figure creation for this manuscript.

FUNDING

No funding was received for this work.

CONFLICT OF INTEREST

O.T. and S.B. own stock in Roche. E.G.-M. has received consulting/advisory fees from Deciphera and TYME. A.J.B. and E.H. have ownership in COTA, Inc. J.B.C. owns stock in IQVIA. A.B.C. owns equity in Flatiron Health, a subsidiary of Roche and stock in Roche. J.L.E. owns stock in McKesson. M.A.I. and C.S. own stock in Syapse. N.J.R. holds a leadership position in McKesson; stock/ownership in Johnson & Johnson, McKesson, and Oncolytics Biotech; holds honoraria with Bristol-Myers Squibb and Roche; and consulting roles for ADVI, Boehringer Ingelheim, Bristol-Myers Squibb, and New Century Health. J.W. owns stock in IQVIA and Merck. Y.N. owns stock in Syapse and Concertai and received travel/accommodations by Syapse. All other authors declared no competing interests for this work.

AUTHOR CONTRIBUTIONS

L.L., O.T., E.G.-M., S.B., A.J.B., M.B., J.B.C., A.B.C., J.L.E., N.J.R., M.A.I., M.D.S., Y.N., D.R.R., and J.A. designed the research. O.T., E.H., C.S., M.S., M.I., and J.W. performed the research. L.L., O.T., E.G.-M., J.B.C., A.J.B., A.B.C., J.L.E., C.W., N.J.R., M.S., M.A.I., J.W., and Y.N. analyzed the data. L.L., O.T., E.G.-M., S.B., A.J.B., J.B.C., A.B.C., J.L.E., C.S., N.J.R., M.D.S., M.S., M.A.I., Y.N., and D.R.R. wrote the manuscript.

DISCLAIMERS

The authors assume full responsibility for analyses and interpretation of these data.

DATA AVAILABILITY STATEMENT

The data underlying this article will be shared on reasonable request to the corresponding author.

1. Elting, L.S. *et al*. Generalizability of cancer clinical trial results: prognostic differences between participants and nonparticipants. *Cancer* **106**, 2452–2458 (2006).

2. Bolislis, W.R., Fay, M. & Kühler, T.C. Use of real-world data for new drug applications and line extensions. *Clin. Ther.* **42**, 926–938 (2020).
3. National Academies of Sciences, Engineering, and Medicine. *Examining the Impact of Real-world Evidence on Medical Product Development: Proceedings of a Workshop Series* (The National Academies Press, Washington, DC, 2019).
4. Stewart, M. *et al*. An exploratory analysis of real-world end points for assessing outcomes among immunotherapy-treated patients with advanced non-small-cell lung cancer. *JCO Clin. Cancer Inform.* **3**, 1–15 (2019).
5. Gandhi, L. *et al*. Pembrolizumab plus chemotherapy in metastatic non–small-cell lung cancer. *N. Engl. J. Med.* **378**, 2078–2092 (2018).
6. Curtis, M.D. *et al*. Development and validation of a high-quality composite real-world mortality endpoint. *Health Serv. Res.* **53**, 4460–4476 (2018).
7. La, J. *et al*. Real-world outcomes for patients treated with immune checkpoint inhibitors in the veterans affairs system. *JCO Clin. Cancer Inform.* **4**, 918–928 (2020).
8. Khozin, S., *et al*. Real-world outcomes of patients with metastatic non-small cell lung cancer treated with programmed cell death protein 1 inhibitors in the year following US regulatory approval. *Oncologist* **24**, 648–656 (2019).
9. Khozin, S. *et al*. Real-world progression, treatment, and survival outcomes during rapid adoption of immunotherapy for advanced non-small cell lung cancer. *Cancer* **125**, 4019–4032 (2019).
10. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. ICH Harmonised Guideline Addendum on Estimands and Sensitivity Analysis in Clinical Trials to the Guideline on Statistical Principles for Clinical Trials. Final Version. E9(R1);2019 <https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf> (2019).