Using Electronic Health Records to Derive Control Arms for Early Phase Single-Arm Lung Cancer Trials: Proof-of-Concept in Randomized Controlled Trials

Gillis Carrigan[1,*], Samuel Whipple[1], William B. Capra[1], Michael D. Taylor[1], Jeffrey S. Brown[2], Michael Lu[1], Brandon Arnieri[1], Ryan Copping[1] and Kenneth J. Rothman[3]

Oncology drug development increasingly relies on single-arm clinical trials. External controls (ECs) derived from electronic health record (EHR) databases may provide additional context. Patients from a US-based oncology EHR database were aligned with patients from randomized controlled trials (RCTs) and trial-specific eligibility criteria were applied to the EHR dataset. Overall survival (OS) in the EC-derived control arm was compared with OS in the RCT experimental arm. The primary outcome was OS, defined as time from randomization or treatment initiation (EHR) to death. Cox regression models were used to obtain effect estimates using EHR data. EC-derived hazard ratio estimates aligned closely with those from the corresponding RCT with one exception. Comparing log HRs among all RCT and EC results gave a Pearson correlation coefficient of 0.86. Properly selected control arms from contemporaneous EHR data could be used to put single-arm trials of OS in advanced non-small cell lung cancer into context.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

☑ Precision oncology drugs increasingly receive accelerated or breakthrough regulatory approval based on single-arm trials. An inherent feature of these designs is that a standard-of-care control arm is not included, leading to challenges in interpretation of efficacy. Historical controls have been used in the past as comparators in oncology trials; however, underlying differences in populations, cohort period effects, and outdated standard of care have led to biased comparisons.

WHAT QUESTION DID THIS STUDY ADDRESS?

☑ How closely can contemporaneous external control (EC) arms derived from electronic health records (EHRs) mirror overall survival observed in the control arms from advanced non-small cell lung cancer (aNSCLC) randomized controlled trials (RCTs)?

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

☑ Contemporaneous EC arms derived from EHRs may offer advantages in interpreting efficacy in early phase oncology trials over previous approaches that have relied solely on outdated historical controls.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?

☑ These results make a case for the use of EHR-derived EC groups for comparison with single-arm clinical trial cancer populations. The implication of this approach relates to the potential for accelerated development and evaluation of oncology treatments and ultimately enhanced access to the same for patients with cancer.

The need to bring safe and effective oncology treatment options to patients quickly is highlighted by the 21st Century Cures Act[1–3] and the Cancer Moonshot initiative.[4] Oncology research in recent years has included novel trial designs sometimes with only a single treatment arm. Precision oncology drugs increasingly receive accelerated or breakthrough regulatory approval based on these single-arm trials. An inherent feature of these designs is that a standard-of-care control arm is not included, leading to challenges in interpretation of efficacy.[3,5–9]

[1]Genentech, South San Francisco, California, USA; [2]Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, Massachusetts, USA; [3]Research Triangle International, Boston, Massachusetts, USA. *Correspondence: Gillis Carrigan (gjmcarrigan@gmail.com)

Summary data from existing publications or historical controls from prior studies are often used as comparators in oncology trials. These approaches are beset by methodological issues, including biased comparisons resulting from underlying differences in populations, cohort period effects, and outdated standards of care. These issues have fostered skepticism about the interpretation of single-arm trials.

Curated electronic health record (EHR) datasets are now large enough, with sufficient clinical detail, to create contemporaneous external control (EC) groups. Moreover, regulators, such as the US Food and Drug Administration (FDA), as well as organizations like the National Cancer Institute are formulating their thinking around the use of real-world data (RWD) and evidence in regulatory decision making and the efficient conduct of oncology clinical trials, respectively.[10,11] Here, we report on a proof-of-concept study to assess how well these ECs approximate the standard-of-care arms in randomized controlled trials (RCTs). If ECs can consistently replicate RCT data, they may potentially serve as meaningful comparators for early phase single-arm trials, addressing many of the aforementioned biases.[9] Specifically, we assessed how closely results from RCTs in advanced non-small cell lung cancer (aNSCLC) could be approximated, by substituting EHR-based EC groups as the comparator.

RESULTS

Nine aNSCLC trials met the trial selection criteria (see **Figure 1**). One trial, NCT01328951, was excluded because we could not assemble a control group in the EHR data. This trial compared early vs. late initiation of treatment with erlotinib and required a nontreatment (i.e., placebo) maintenance interval that could not be identified in the EHR. Consequently, 8 trials with a total of 11 experimental arms were assessed. The effect on study size from applying each trial's specific eligibility criteria on the respective EC cohorts is displayed in **Figure 2**. The group sizes ranged from 381 patients in NCT01351415 to 3,200 patients in one of the ECs for trial NCT01496742.

Trial NCT02008227 (the first trial for which we applied the approach described in this paper) was a study comparing the cancer immunotherapy drug atezolizumab with docetaxel in patients with locally advanced or metastatic NSCLC who failed first-line treatment with a platinum-containing therapy; 425 patients were in the trial's docetaxel control arm. **Figure S1 (Supplementary Material S1)** shows the impact of applying study inclusion/exclusion criteria on the EHR dataset; 1,300 patients with aNSCLC were treated with docetaxel monotherapy in the second or third-line of treatment, and the final EC comprised 547 patients. **Table S1** shows the resulting balance in terms of baseline characteristics between the trial control arm and the EC (**Supplementary Material S1**). Some evidence of residual imbalances with respect to age, smoking history, race, and disease stage were observed. The original overall survival (OS) comparison from NCT02008227 (hazard ratio (HR) = 0.73, 95% confidence interval (CI) = 0.62−0.86) is in
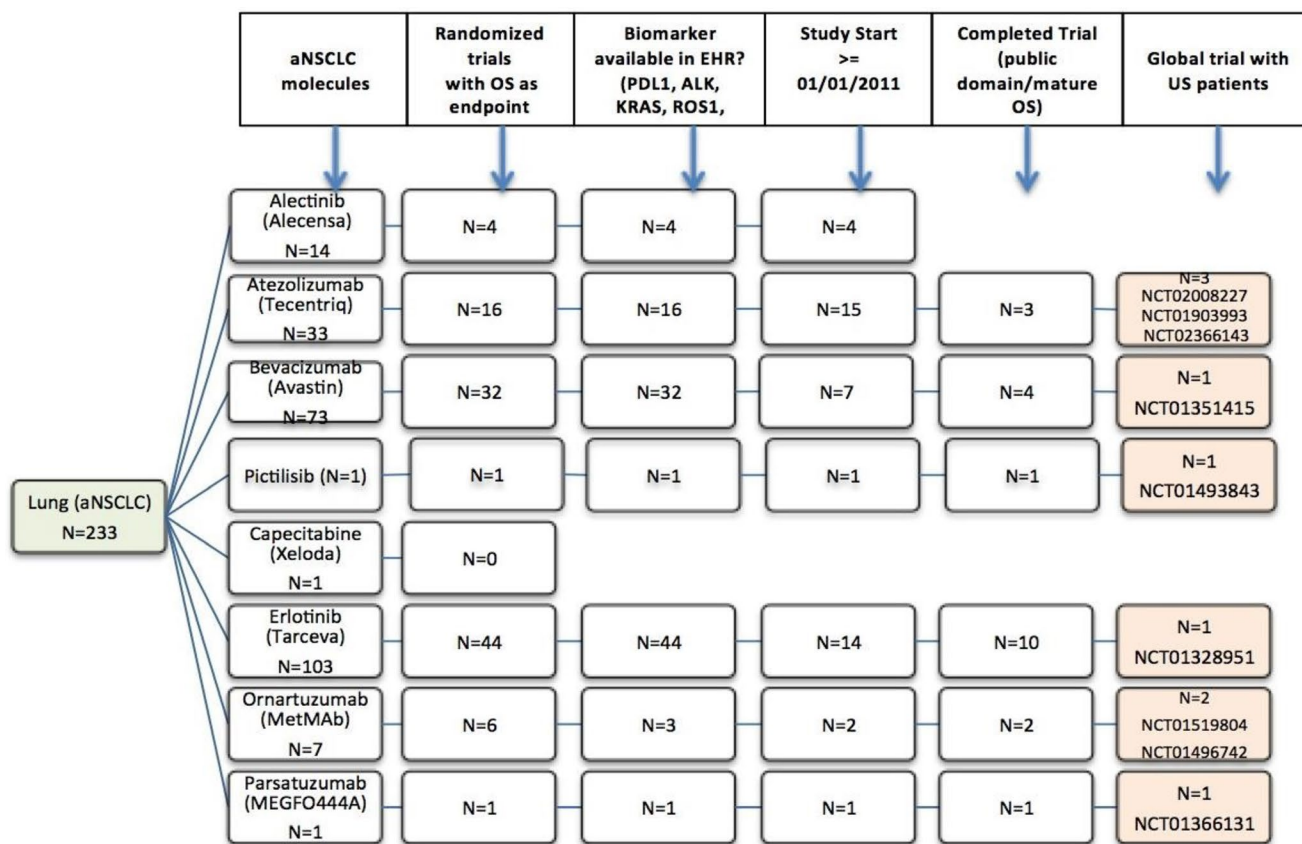

Figure 1 Trial selection. aNSCLC, advanced non-small cell lung cancer; EHR, electronic health record; OS, overall survival.
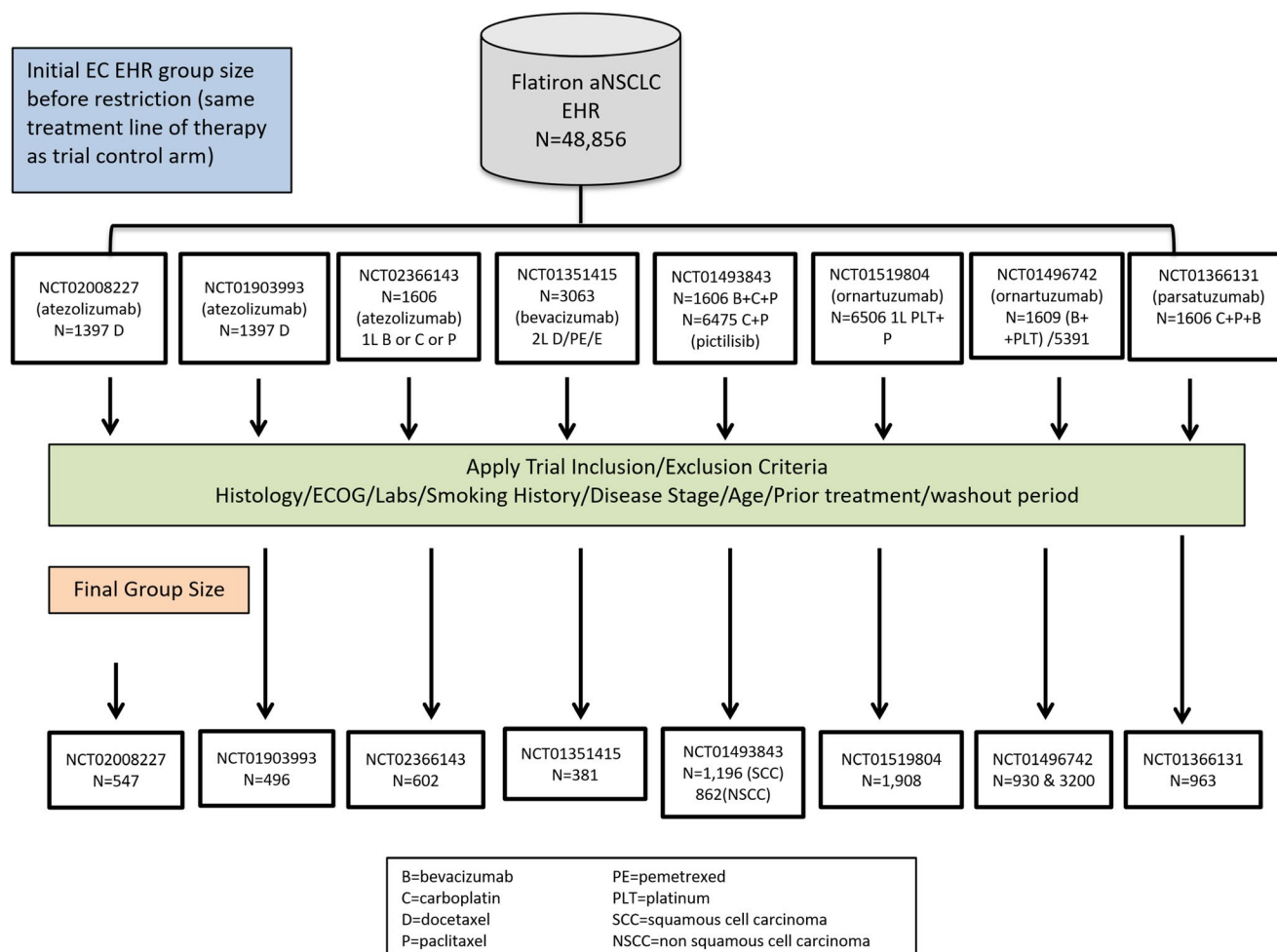
Figure 2 External control patient attrition. aNSCLC, advanced non-small cell lung cancer; EC, external control; ECOG, Eastern Cooperative Oncology Group; EHR, electronic health record.

Figure S2 on the left; for comparison, on the right is the parallel figure replacing the trial control with the EC following alignment with eligibility criteria and propensity score (PS) adjustment (HR = 0.67; 95% CI = 0.56−0.80) (**Supplementary Material S1**).

Table 1 shows all adjusted HRs from the corresponding EC Cox regression models compared with the original RCT HRs. The absolute differences in natural log HRs were < 0.2 in 10 of the 11 comparisons. One of the comparisons (NCT01519804) showed a considerable difference in HR when compared with the HR observed in the original RCT. **Figure 3** plots the HRs from the RCT (x axis) vs. HR from the EC analysis (y axis). We see from this plot that most HR comparisons are strongly correlated, with absolute differences in natural log (ln(HR)) ranging from 0.011−0.474 (see **Table 1**) and an overall weighted Pearson's correlation coefficient of r = 0.86 (95% CI = 0.54−0.96: see **Figure 3**).

In the sensitivity analysis (NCT02008227) the requirement of a laboratory value before treatment initiation reduced the size of the EC from 547 to 298. The laboratory values specified as inclusion criteria in this trial that were included in the EHR data were: white blood cell count, absolute neutrophil count, absolute lymphocyte count, platelet count, hemoglobin, aspartate aminotransferase/

alanine aminotransferase, serum creatinine, bilirubin, and serum albumin. Overall, these laboratory values ranged in frequency from ~ 65% to 90% of patients in the EHR with lower frequency than this observed in the 28-day window preceding treatment initiation. The median overall survival (mOS) was 6.93 when patients with unavailable laboratory values were excluded compared with an mOS of 6.87 months in the primary analysis when these patients were included. The resulting PS adjusted HRs were equal to 0.71 using either approach.

DISCUSSION

This study indicates that EC arms generated from curated EHR data can replicate the control arms in RCTs and could serve as meaningful comparators for single arm trials. In 10 of the 11 analyses conducted, HR estimates using EHR-derived EC groups were similar to those from the original RCTs. Three of the EC-derived HR estimates fell slightly above the line of perfect concordance with the RCT HR estimates. These EC-derived HR estimates suggest a somewhat healthier (as defined by OS) EC group relative to the RCT controls, and, therefore, biased the EC HR estimate slightly toward the null. Seven of the EC-derived HR estimates

Table 1 Trial HRs vs. EC HRs

	RCT				EC analysis						
	Exp. arm		Control arm		Exp. arm (RCT)		Control arm (EHR)			EC adjusted	
Analysis	N	E	N	E	N	E	N	E	RCT HR (95% CI)	HR (95% CI)	Difference ln(HR)
[1] NCT02008227	425	271	425	298	396	259	482	398	0.73 (0.62–0.86)	0.71 (0.59–0.84)	0.028 (−0.132, 0.188)
[2] NCT01903993	144	78	143	95	129	74	401	314	0.72 (0.54–0.98)	0.66 (0.50–0.88)	0.087 (−0.176, 0.350)
[3] NCT02366143	356	144	336	166	319	135	556	365	0.77 (0.61–0.96)	0.75 (0.59–0.94)	0.026 (−0.179, 0.231)
[4] NCT01351415	245	194	240	193	229	182	363	295	0.88 (0.74–1.04)	0.89 (0.75–1.05)	−0.011 (−0.202, 0.179)
[5] NCT01493843: Arm A vs. B	126	79	125	60	104	63	847	612	1.03 (0.75–1.41)	0.95 (0.68–1.33)	0.081 (−0.175, 0.337)
[6] NCT01493843: Arm C vs. D	79	59	79	43	62	47	785	592	1.04 (0.72–1.50)	1.07 (0.78–1.49)	−0.028 (−0.319, 0.262)
[7] NCT01493843: Arm E vs. F	62	42	30	13	39	30	616	470	1.27 (0.75–2.15)	1.32 (0.90–1.93)	−0.039 (−0.389, 0.312)
[8] NCT01519804	55	36	54	33	47	28	1,375	735	0.89 (0.55–1.46)	1.43 (0.97–2.09)	−0.474 (−0.835, −0.114)
[9] NCT01496742: Cohort 1	69	32	70	29	59	21	834	464	1.38 (0.75–2.56)	1.26 (0.80–1.97)	0.091 (−0.310, 0.492)
[10] NCT01496742: Cohort 2	59	37	61	36	47	22	2,899	1,619	1.15 (0.68–2.56)	1.11 (0.73–1.70)	0.035 (−0.332, 0.403)
[11] NCT01366131	52	24	52	18	46	15	761	577	1.08 (0.52–2.21)	0.90 (0.53–1.51)	0.182 (−0.276, 0.640)

CI, confidence interval; E, number of events (deaths); EC, external control; EHR, electronic health record; Exp., experimental; HR, hazard ratio; ln(HR), natural log hazard ratio; N, number of patients at risk; RCT, randomized controlled trial.
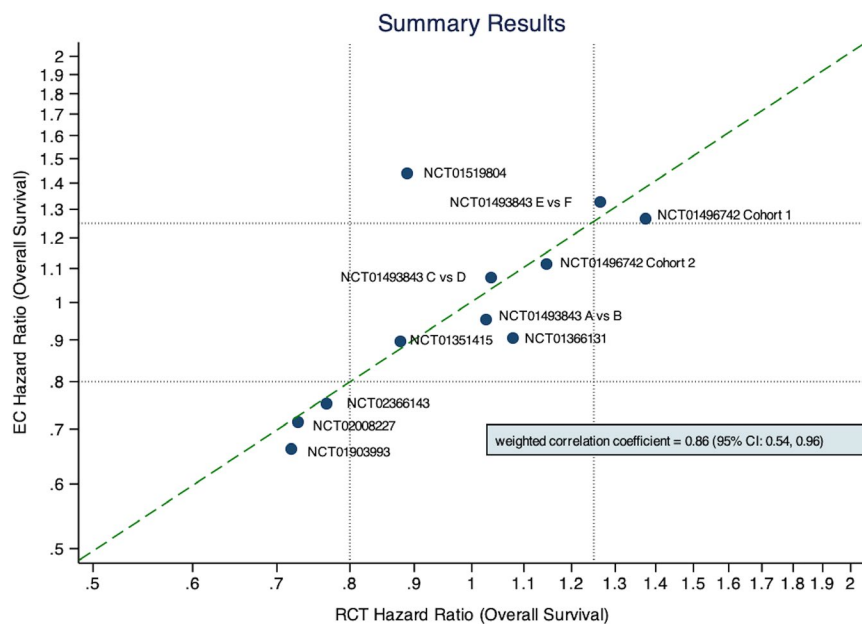

Figure 3 Comparison of log hazard ratios between the original trial results and the EHR controlled analyses. CI, confidence interval; EC, external control; EHR, electronic health record; RCT, randomized controlled trial.

differed from the RCT HR estimates in the other direction, suggesting a less healthy EC cohort and a potential bias of the EC-derived HR away from the null. In some cases, restricting the EHR on the trial eligibility criteria offered reasonable alignment with the trial cohort (see **Figure 4**), however, in many instances, simple alignment of the EC cohort with trial eligibility criteria offered only partial effectiveness of replicating RCT control arms with EHR-derived ECs, but eligibility alignment accompanied by PS adjustment provided additional bias control and was an essential aspect of the overall approach (for example, see **Figure 5**).

Equally as important as demonstrating the approach work is exploring when it does not work. The one discordant result (RCT NCT01519804) was from a small trial (RCT experimental arm size of 47 after PS trimming and 54 before). The small study size could contribute to the discordance, although there were other analyses with comparably small groups. This study comprised patients with squamous cell aNSCLC with notably lower survival rates than observed in other trials in this population (median OS = 9.1 months in trial experimental arm and 8.5 months in the trial controls), whereas the EHR-derived squamous aNSCLC EC cohort had a higher mOS that was more in line with rates observed in the literature (median OS ~ 12 months).[12–14] This was the only trial in which we observed at the outset a longer mOS in the raw unadjusted EHR cohort compared with the trial mOS. Deep review of the study's clinical trial record suggested that one plausible explanation for the discordant result could relate to enrichment for mesenchymal-to-epithelial transition (MET) diagnostic positive patients in NCT01519804, a trial of an investigational MET inhibitor. MET is a proto-oncogene that encodes for the MET tyrosine receptor kinase and its only known

ligand is hepatocyte growth factor. Although 30–40% of squamous NSCLC is expected to be classified as MET-positive,[15,16] of patients in NCT01519804 were MET-positive. High levels of receptor expression, as well as high MET gene copy number, are independent prognostic factors for poor outcome in patients with NSCLC.[17,18] MET expression assessment is not standard of care in routine clinical oncology practice and, therefore, not captured in the EHR database. It is likely the final EHR-derived cohort had considerably fewer MET-positive patients compared with the MET-positive-enriched RCT, thereby skewing the EC group to have a longer OS than the RCT population. This possibility reinforces the need for granularity of clinical trial and EHR data, including relevant biomarkers, when constructing ECs, especially for targeted agents. In hindsight, our prespecified trial selection criterion necessitating biomarker availability in the EHR for studies with a biomarker-defined population should have been expanded to also apply to trials in which the study population was biomarker enriched. Subsequent assessments of EC should continue to elucidate scenarios where ECs cannot be confidently generated. Despite the possibility of MET-positive results accounting in part for the observed discordant finding, it is unlikely that, given the observed mOS in this trial compared with the mOS in the EHR cohort, the proportion of MET-positive patients (~ 50%) would completely account for this finding.

Other research efforts for establishing ECs have also shown early promise.[19–22] Jia et al.[22] examined an application in prostate cancer using a modeling approach based on existing trial data with reasonable success. Other studies drawing on aggregate findings from the clinical trial literature have been performed in acute myeloid leukemia[20] and anaplastic lymphoma kinase-targeted
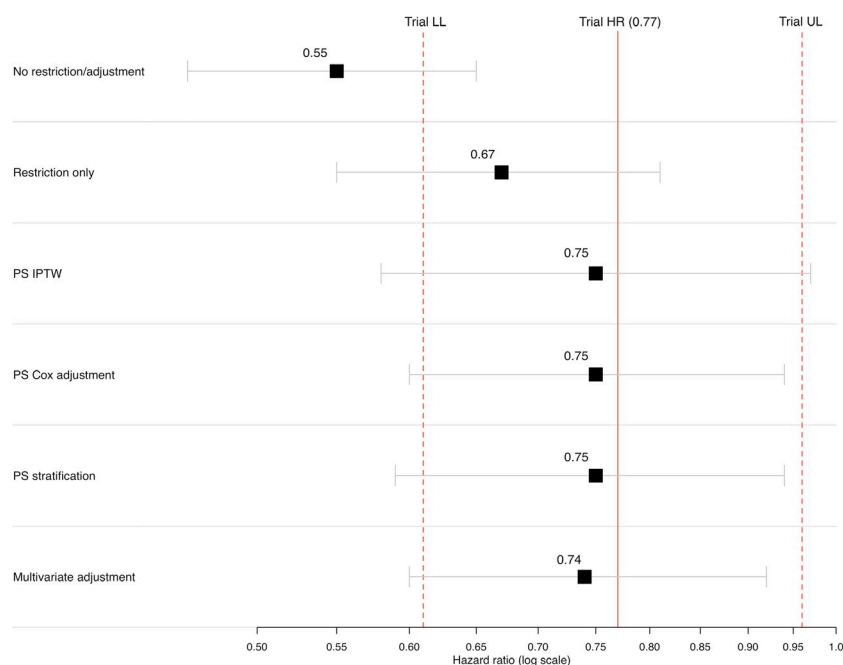

Figure 4 Estimates of relative risk across different adjustment methods comparing external control analysis to trial result: randomized controlled trial NCT0236143. To achieve greater balance among baseline covariates for PS analysis methods—"PS IPTW," "PS Cox Adjustment," and "PS Stratification"—PS were trimmed by removing nonoverlapping observations from the PS distributions of the experimental and controls arms resulting in slightly modified cohorts in both arms. HR, hazard ratio; IPTW, inverse probability of treatment weighting. LL, lower limit; PS, propensity score; UL, upper limit.
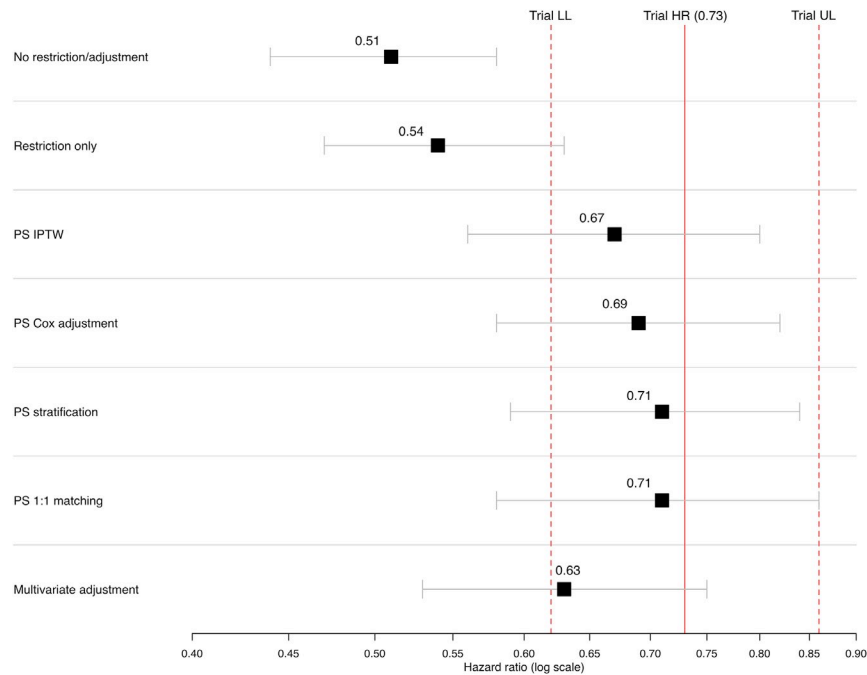
Figure 5 Estimates of relative risk across different adjustment methods comparing external control analysis to trial result: RCT NCT 02008227. To achieve greater balance among baseline covariates for propensity score (PS) methods—"PS IPTW," "PS Cox Adjustment," and "PS Stratification"—PS were trimmed by removing nonoverlapping observations from the PS distributions of the experimental and controls arms resulting in slightly modified cohorts in both arms. HR, hazard ratio; IPTW, inverse probability of treatment weighting. LL, lower limit; PS, propensity score; UL, upper limit.

NSCLC,[19,23] all of which showed promise with respect to replicating control arms from RCTs. To date, few studies have used RWD, such as curated EHR data, to replicate trial arms. It is clear from the present findings that there are differences between EHR-based and trial populations, which underline the importance of applying RCT eligibility criteria and proper adjustment methods. Future research should explore alignment of different eligibility criteria, such as brain metastases and HIV status, as well as optimal sample weighting and adjustment techniques.

As evinced through the recent development of an FDA program on real-world evidence, and through the 21st Century Cures Act, discussion has accelerated around facilitating the use of RWD to support regulatory submissions including the use of ECs for single arm trials.[3–11,24,25] The availability of large patient-level datasets concurrent in time with trial data is moving these discussions from the conceptual to applied.

We chose the Flatiron Health observational database due to its large population of patients with aNSCLC and near real-time recency of the data, which may have mitigated confounding period effects observed in past research efforts. For example, all the RCTs had a minimum of 3 years of overlapping data with the EHR. In terms of the integrity of the OS end point, mortality capture in the Flatiron EHR aNSCLC database is > 90% complete.[26] Furthermore, the granularity of the EHR data afforded the opportunity to control for differences in the RCT and EHR populations for critical prognostic variables common to both RCT and EHR data. Finally, access to patient-level data from recent aNSCLC RCTs permitted a thorough exploration of this approach.[27]

Study limitations

Creating EHR-derived EC arms is constrained by the availability of covariates common to both the RCTs and the EHR data, a problem that is not at play in clinical trials. There was some overlap in the EC cohorts used as comparators for these trials, implying some nonindependence of the cohorts used in the EC analyses. There is a much greater opportunity for residual bias when combining EHR-derived cohorts as comparator arms to trial arms relative to RCTs. The methods applied in the current study cannot control for unmeasured confounders directly, whereas unknown or unobserved confounders can be controlled for through randomization in clinical trials. Even measured confounding can be problematic in the nonrandomized observational setting. For that reason, it is important that inferences drawn from comparisons with ECs consider consistency with other RCT and non-RCTs in similar agents or those of the same agent in other settings (e.g., a different line of therapy), as well as biological plausibility. The approach described here is not intended as a replacement for RCTs, but as a means for interpreting trials when a comparator arm is not included.

There was 1 discordant result of the 11 analyses we conducted. However, we identified a plausible and likely explanation for the discordant result, namely an imbalance in a prognostic biomarker unavailable in the EHR data. Random error, related to small study size for some of the trials, also contributes to discrepancies. Any disagreement between RCT and EC results may be partially attributable to error in the RCT result.

We used a single source of data for our EC groups and our findings may not generalize to other data sources. Approaches that

apply to one disease setting, such as aNSCLC, may not generalize to other disease settings (e.g., breast cancer).

In five of the eight trials, we used a definition of aNSCLC that incorporates patients with stages IIIB and IV disease as well as patients diagnosed at earlier stages with recurrent metastatic disease. This criterion aligned with the Flatiron database and also reflected the predominant, although shifting definition, of advanced NSCLC. Three of the trials focused solely on metastatic NSCLC (NCT02366143, NCT01493843, and NCT01366131). For these trials, we modified our aNSCLC definition to exclude patients initially diagnosed with locally advanced NSCLC (e.g., stage IIIb). Because one of the three trials was not in the first-line treatment setting (NCT01493843), it was possible that some patients were excluded who could have been included (e.g., patients with subsequent metastatic disease in the second-line treatment).

Contemporaneous data are important, but our approach leaves room for improvement; as real-world datasets get larger, EC cohorts should exactly mimic the study enrollment period of the comparison trial. Additionally, trial data tend to be captured at more regular intervals than EHR-based data; the potential for informative censoring and other sources of bias in the estimation of mOS needs to be explored. The sensitivity in measuring mortality in the EHR, which is over 90%, is nonetheless likely to be less than the mortality capture in the RCTs. The result of this would be a slight overestimation of survival in the EHR EC cohorts and modest bias toward the null in comparison with the RCT arms.[28] Further work may inform the applicability of EC methods for end points, such as progression-free survival, many of which may be affected by frequency of clinical assessments.

Although previous examples exist in the literature that have used real-world control arms to make indirect treatment comparisons,[19,23] this is the first study to examine the use of real-world EC arms across a number of RCTs using patient-level data to evaluate efficacy directly. As expected, the trial patients in this study were generally younger and healthier than the EHR-derived controls, which are consistent with findings from other studies comparing patients with cancer from RCTs to patients from routine practice settings.[29] It should be acknowledged that the inherent differences between community-based EHR-derived patients and those patients typically enrolled in clinical trials is a challenge when using the approach described here. Despite this underlying challenge, these results make a case for the use of EHR-derived EC groups for comparison with single-arm clinical trial populations when appropriate methodologies and adjustment methods are used drawing on EHR databases with sufficient granularity and recency as to minimize bias. In particular, early-phase, single-arm oncology trials may benefit from a supplemental comparison with contemporaneous EHR-derived EC groups, putting study results into clearer context. Additionally, this approach may serve as a means of augmenting randomized control groups with EHR controls in the context of later phase hybrid trial designs.

METHODS
ECs
The Flatiron Health database is a longitudinal, demographically, and geographically diverse database derived from EHR data. Data are curated to align with a prespecified common data model, with data elements in unstructured documents abstracted.[9] At the time of this study, the database included information from over 260 community-based cancer treatment clinics and 3 academic networks covering more than 2 million patients with active cancer in the United States. Eligible patients were initially diagnosed with stage IIIB or IV NSCLC between January 1, 2011, and May 31, 2018, or with earlier stage disease and subsequent metastatic recurrence during this period. Mortality data were amalgamated from commercial and governmental sources plus the EHR, benchmarked to the National Death Index.[26] Institutional review board approval of the study was obtained prior to study conduct, and included a waiver of informed consent. Data provided to third parties for this analysis were de-identified and provisions were in place to prevent re-identification in order to protect patients' confidentiality.

Trial selection
Individual patient-level data from all Roche-sponsored aNSCLC RCTs that met the following criteria were included in the analyses: (i) first patient enrolled on or after January 1, 2011; (ii) mOS attained, with findings presented in a journal or at a congress by March 31, 2018; (iii) including at least one US study site; and (iv) in the case of a biomarker-defined study population, availability of the biomarker within the curated EHR dataset. Detailed trial information can be found on ClinicalTrials.gov.

Alignment with eligibility criteria
In three of the eight trials (NCT02366143, NCT01493843, and NCT01366131), only patients with metastatic NSCLC were eligible for enrollment. For these trials, we excluded patients with an initial diagnosis of locally advanced NSCLC (patients with stage IIIb). For the remaining trials, all patients in the aNSCLC EHR database were eligible to serve as ECs.

The initial step was to select patients from the EHR cohort who received standard-of-care treatment as identified in the trial. This treatment includes both individual medications in a treatment regimen and line of therapy. The second step was to apply RCT inclusion/exclusion criteria available in the EHR to select EHR-based controls comparable in terms of demographic and clinical characteristics with patients in an RCT. Not all RCT eligibility criteria could be applied to the EHR-derived patients, either because of missing or incomplete data in the EHR or because an RCT criterion was not captured in the EHR (e.g., measurable disease, as defined by Response Evaluation Criteria in Solid Tumors (RECIST) version 1.1). In other cases, such as medical history and comorbidities, the oncology-specific EHR did not have exhaustive medical and disease history; therefore, some factors could not be reliably ascertained.

The main inclusion/exclusion criteria consistently applied to the EHR data were disease stage at initial diagnosis, histology, Eastern Cooperative Oncology Group (ECOG) PS score (included if missing), laboratory data at baseline (included if missing), biomarker status when applicable (e.g., epidermal growth factor receptor), and prior treatments. The values used for each of these criteria depended on the values required by the respective RCT. In addition to inclusion/exclusion criteria from the RCTs, patients selected as ECs also had to initiate first-line treatment within 90 days following advanced diagnosis to reduce the likelihood of incomplete treatment capture in the EHR dataset. Information regarding trial-specific eligibility criteria is in **Table S1 (Supplementary Material S1)** with links to ClinicalTrials.gov.

PS adjustment
For each EC and experimental arm pair, the following factors were included in the PS that captured the probability of being in the trial treatment arm: age, sex, race, smoking history, histology (unless used as trial inclusion criterion), disease stage at initial diagnosis, and time from initial diagnosis to either the start of treatment (EHR data) or randomization (trial data). Patients in the nonoverlapping tails of the PS distributions were excluded ("trimmed"; i.e., patients in the

experimental arm with PS higher than the highest PS in the control arm, and patients in the control arm with PS lower than the lowest PS of the experimental arm).

Statistical analyses

The primary outcome was OS, defined as time from randomization (trial patients) or treatment initiation (EC patients) to death. Patients were censored at the earlier of end of trial follow-up period, or EHR last contact date. We estimated survival using the Kaplan−Meier method, using inverse probability of treatment weighting to balance groups with respect to the PS. We used proportional hazards (Cox) models to calculate HRs comparing trial experimental arms with EC arms. For two trials, we compared several analytic methods, such as inverse proportional weighting, PS matching, including the PS as a covariate in the proportional hazards model, standard multivariable modeling, and using a stratified proportional hazards model with PS decile as the stratification variable. Agreement across these methods was good; for subsequent trials, we used only the stratified proportional hazards approach (see **Figures 4** and **5**).

For each study, HRs and Kaplan−Meier plots derived using the EHR EC were compared with RCT results. To assess the performance across all trials, we plotted the ln(HR) from the EC comparison against the ln(HR) from the RCTs. We assessed overall concordance between published RCT HRs and those obtained with EC arms using a Pearson correlation coefficient with inverse variance weighting of each point to account for differences in precision of the various point estimates.

All analyses were conducted using SAS version 9.3 software (SAS Institute, Cary, NC) and R 2.15.2 software (R Foundation for Statistical Computing, Vienna, Austria).

Sensitivity analysis

Given the uncertainty around, including EC patients with unavailable pretreatment laboratory values, a sensitivity analysis was performed on NCT02008227. Drawing only on patients with laboratory values available within the 28-day pretreatment window, all laboratory inclusion criteria from the trial were applied to the EHR EC patients. The mOS and HR were calculated using this alternate method for comparison with our primary approach (i.e., allowing patients with unavailable pretreatment laboratory measurements to qualify for the EC group).

Institutional review board approval

Institutional review board approval through Flatiron Health was obtained prior to study conduct. Informed consent was waived as this was a noninterventional study using routinely collected data. Flatiron Health standard methodology for data security and patient privacy were implemented.

1. United States Congress. 21st Century Cures Act <https://www.congress.gov/bill/114th-congress/house-bill/34/> (2016).
2. US Food and Drug Administration. Sentinel initiative <http://www.fda.gov/Safety/FDAsSentinelInitiative/ucm149340.htm> (2017). Accessed March 24, 2018.
3. Armstrong, D. *et al*. The role of non-randomized trials for the evaluation of oncology drugs. Conference on Clinical Cancer Research Issue Brief, November 2014.
4. Cancer Moonshot <https://www.cancer.gov/research/key-initiatives/moonshot-cancerinitiative>. Accessed June 10, 2018.
5. Jarow, J.P., LaVange, L. & Woodcock, J. Multidimensional evidence generation and FDA regulatory decision making defining and using "real-world" data. *JAMA* **318**, 703–704 (2017).
6. Khozin, S., Blumenthal, G.M. & Pazdur, R. Real-world data for clinical evidence generation in oncology. *J. Natl. Cancer Inst.* **109**, djx187 (2017).
7. Khozin, S., Kim, G. & Pazdur, R. From big data to smart data: FDA's INFORMED initiative. *Nat. Rev. Drug Discov.* **16**, 306 (2017).
8. Eichler, H.-G. *et al*. "Threshold-crossing": a useful way to establish the counterfactual in clinical trials? *Clin. Pharmacol. Ther.* **100**, 699–712 (2016).
9. Berger, M.L., Curtis, M.D., Smith, G., James, H. & Abernethy, A.P. Opportunities and challenges in leveraging electronic health record data in oncology. *Future Oncol.* **12**, 1261–1274 (2016).
10. US Food and Drug Administration. Framework for FDA's Real-World Evidence Program <www.fda.gov> (2018). Accessed December 10, 2018.
11. Sharpless, N.E. & Doroshow, J.H.. Modernizing clinical trials for patients with cancer. *JAMA Oncol.* E1–E2 (2019). https://jamanetwork.com/journals/jama/article-abstract/2723062. Accessed January 24, 2019.
12. Scagliotti, G.V. *et al*. Phase III study comparing cisplatin plus gemcitabine with cisplatin plus pemetrexed in chemotherapy-naive patients with advanced-stage non-small-cell lung cancer. *J. Clin. Oncol.* **26**, 3543–3551 (2008).
13. Ciuleanu, T. *et al*. Maintenance pemetrexed plus best supportive care versus placebo plus best supportive care for non-small-cell lung cancer: a randomised, double-blind, phase 3 study. *Lancet* **374**, 1432–1440 (2009).
14. Okamoto, I. *et al*. Phase III trial comparing oral S-1 plus carboplatin with paclitaxel plus carboplatin in chemotherapy-naïve patients with advanced non-small-cell lung cancer: results of a west Japan oncology group study. *J. Clin. Oncol.* **28**, 5240–5246 (2010).
15. Ichimura, E. *et al*. Expression of c-met/HGF receptor in human non-small cell lung carcinomas in vitro and in vivo and its prognostic significance. *Jpn. J. Cancer Res.* **87**, 1063–1069 (1996).
16. Spigel, D.R. *et al*. Final efficacy results from OAM4558 g, a randomized phase II study evaluating MetMAb or placebo in

combination with erlotinib in advanced NSCLC. *J. Clin. Oncol.* **29**(suppl.) (2011). Abstract 7505.

17. Park, S. *et al*. High MET copy number and MET overexpression: poor outcome in non-small cell lung cancer patients. *Histol. Histopathol.* **27**, 197–207 (2012).

18. Cappuzzo, F. *et al*. Increased MET gene copy number negatively affects survival of surgically resected non-small-cell lung cancer patients. *J. Clin. Oncol.* **27**, 1667–7410 (2009).

19. Tan, D.S.-W. *et al*. Comparative efficacy of ceritinib and crizotinib as initial ALK–targeted therapies in previously treated advanced NSCLC: an adjusted comparison with external controls. *J. Thorac. Oncol.* **11**, 1550–1557 (2016).

20. Berry, D.A. *et al*. Creating a synthetic control arm from previous clinical trials: application to establishing early end points as indicators of overall survival in acute myeloid leukemia (AML). *J. Clin. Oncol.* **35**(suppl. 15) (2017). Abstract 7021.

21. Patel, K., Ouwens, M., Shire, N. & Khosla, S. The application of electronic medical records (EMRs) as a virtual comparator arm in a lung cancer clinical trial: a case study. *J. Clin. Oncol.* **35**(suppl. 15) (2017). Abstract e18098.

22. Jia, Z. *et al*. Generation of ''virtual'' control groups for single arm prostate cancer adjuvant trials. *PLoS One* **9**, e85010 (2014).

23. Davies, J. *et al*. Comparative effectiveness from a single-arm trial and real-world data: alectinib versus ceritinib. *J. Comp. Eff. Res.* **7**, 855–865 (2018).

24. Jarow, J.P. Use of external controls in regulatory decision-making. *Clin. Pharmacol. Ther.* **101**, 595–596 (2017).

25. Hatswell, A.J. *et al*. Regulatory approval of pharmaceuticals without a randomised controlled study: analysis of EMA and FDA approvals 1999–2014. *BMJ Open* **6**, e011666 (2016).

26. Curtis, M.D. *et al*. Development and validation of a high-quality composite real-world mortality endpoint. *Health Serv. Res.* **53**, 4460–4476 (2018).

27. Weiskopf, N.G., Bakken, S., Hripcsak, G. & Chunhua, W. A data quality assessment guideline for electronic health record data reuse. *EGEMS (Wash. DC)* **5**, 1–18 (2017).

28. Carrigan, G. *et al*. An evaluation of the impact of missing deaths on overall survival analyses conducted in an electronic health records database in advanced non-small cell lung cancer patients. *Pharmacoepidemiol. Drug Saf.* **28**, 572–581 (2019).

29. Templeton, A.J. *et al*. Translating clinical trials to clinical practice: outcomes of men with metastatic castration resistant prostate cancer treated with docetaxel and prednisone in and out of clinical trials. *Ann. Oncol.* **24**, 2972–2977 (2013).