

# Comparing prospectively assigned trial and real-world lung cancer patients

Brigham Walker<sup>\*,1,2</sup> , Herman E Ray<sup>2,3</sup> , Ping Shao<sup>2</sup> , Claudio D'Ambrosio<sup>2</sup>, Craig White<sup>2</sup> & Mark S Walker<sup>2</sup> 

<sup>1</sup>Tulane University, New Orleans, LA 70118, USA

<sup>2</sup>ConcertAI, LLC, Cambridge, MA 02138, USA

<sup>3</sup>Kennesaw State University, Kennesaw, GA 30144, USA

\*Author for correspondence: Tel.: +1 504 988 1312; [bwalker6@tulane.edu](mailto:bwalker6@tulane.edu)

Journal of **Comparative Effectiveness Research**

**Aim:** To evaluate the comparability of a probable clinical trial (CT) cohort derived from electronic medical records (EMR) data with a real-world cohort treated with the same therapy and identified using the same inclusion and exclusion criteria to emulate an external control. **Methods:** We utilized de-identified patient-level structured data sourced from EMRs. We then compared patterns of overall survival (OS) between probable CT patients with those drawn from non-contemporaneous real-world data (RWD) using a two-sided log-rank test, hazard ratios (HRs) using a Cox proportional-hazards model and Kaplan–Meier (KM) survival curves. Each regression estimate was calculated with a corresponding 95% confidence interval. We additionally conducted multiple matching methods to assess their relative performance. **Results:** Median (standard deviation) OS was 10.2 (0.7) months for the RWD arm and 11.3 (1.3) for the probable CT arm with a Log rank p-value equal to 0.4771. OS in both cohorts is longer than the reported CT median OS of 9.2 (0.6). The HRs generated under all five assessed matching methods (including without adjustment) were not statistically significant at the 95% confidence level. **Conclusion:** Our results suggest, with caveats noted, that survival patterns between real-world and CT cohorts in this NSCLC setting are not statistically significantly different.

**Plain language summary: Comparing trial & real-world lung cancer patients: What is this article about?:** Randomized controlled trials are used to evaluate whether a drug is effective by randomly assigning patients to receive that drug or a control, thereby limiting explanations other than the drug for differences in outcomes. However, due to a variety of issues including a limited number of patients available for a trial, some drugs are approved without a control group which raises questions like “how would the patient have fared without treatment or with other drugs?” One solution is to utilize data from patients who have already been treated to construct an *external* or *synthetic* control group. This may improve upon a non-controlled study but requires more evidence.

**What were the results?:** We utilized data from electronic medical records and compared survival patterns between patients treated with a drug prior to its regulatory approval (i.e., the probable clinical trial patients) to those treated with the same drug after its approval (i.e., the real-world patients). Since these patients were not randomly assigned, they differed in potentially important ways such as age or smoking habits. So, we explored different methods to account for these differences. We find that the survival patterns are not statistically significantly different between the groups, but we also highlight important limitations concerning sample size that may inform the suitability of these comparisons in other settings.

**What do the results mean?:** Survival patterns between real-world and clinical trial cohorts in this setting are not statistically significantly different, suggesting a potentially appropriate role for external controls.

**Shareable abstract:** Researchers found that survival patterns between probable randomized controlled trial and non-contemporaneous real-world non-small-cell lung cancer patients were not statistically significantly different. This finding, and its caveats, may inform emerging drug evaluation methods.

First draft submitted: 27 November 2023; Accepted for publication: 1 May 2024; Published online: 24 May 2024

**Keywords:** clinical trial • external control arm • oncology • single arm trials • synthetic control arm

Randomized controlled trials (RCTs) rightly remain the gold standard in assessing therapeutic efficacy. However, single-arm studies are increasingly used to support initial regulatory approvals, frequently due to necessity to meet equipoise requirements. This raises concerns about potential biases introduced by a lack of a control group [1–3]. This trend has run concurrently to the increased use of real-world data (RWD) in the face of growing complexity in cancer care [4]. Combined with a range of challenges with RCTs (e.g., patient and provider interest in the novel agent and disinterest in the control agent and sponsors' desire to minimize enrollment in the control arm, minimize treatment burdens and reduce trial noncompliance), there has been a consequent interest in the use of RWD to complement prospective single-arm trials [5–8].

The use of non-randomized control groups (e.g., 'synthetic' controls sourced from prior trials or 'external' controls sourced from RWD) has been explored in non-small-cell lung cancer (NSCLC) settings. A few studies have found that matched real-world (RW) and trial patients have similar Kaplan–Meier (KM) curves and survival outcomes using log rank test and hazard ratio (HR) methods [6,7,9,10]. While promising, the reliability of non-randomized RW control groups remains unsettled, with some instances of poor performance attributed to poor visibility into prognostic factors potentially unevenly distributed between comparison groups [9]. We add to this important emerging literature by directly comparing non-contemporaneous RW patients with probable trial patients in NSCLC. We further explore multiple matching methods to examine the reliability of RW counterfactuals. The overarching objective of the study is to compare RW patients to probable trial patients (i.e., not using data from the trial itself but rather probable trial participants identifiable in the RWD), accounting for various matching methods. The clinical trial (CT) is a motivating example with fairly simple inclusion and exclusion criteria. These results may inform researchers, regulators, providers and study sponsors as they consider RW data as a potentially suitable complement to data generated under prospective single-arm trials.

## Methods

### Reference trial

In this study, we compare probable RCT patients with non-contemporaneous RW patients. Our reference RCT is "Nivolumab versus docetaxel in advanced squamous-cell non-small-cell lung cancer" [11]. In this phase III trial, patients diagnosed with advanced (stage IIIB or IV) squamous-cell NSCLC who have disease progression during or after first-line platinum-containing chemotherapy were assigned to the intervention drug, nivolumab (n = 135), or the standard of care control, docetaxel (n = 137). The study reported that median overall survival (OS), the primary study end point, was 9.2 months for nivolumab (95% confidence interval [CI]: 7.3 to 13.3) versus 6.0 months (95% CI: 5.1 to 7.3) with docetaxel. Nivolumab – first approved and made commercially available on 22 December 2014 to treat unresectable or metastatic melanoma [12] – was approved for use in NSCLC on 4 March 2015 [13].

### Data

We used structured patient-level data sourced from electronic medical records (EMR) through ConcertAI, LLC. As with the reference trial, OS was the primary study end point. EMR-based OS has been evaluated against multiple external sources of mortality and, in the context of this study, was measured in the same way between the cohorts [14]. We defined OS as time from treatment initiation (which was assumed to be equivalent to randomization for the presumed trial patients) until death. Patients were censored at 24 months after treatment initiation, the index date in this study.

### Patient population

Given that nivolumab was first commercially available on 22 December 2014, we identified patients who were treated with nivolumab whose primary diagnosis was NSCLC prior to this date in the EMR dataset. The sample size for the probable CT cohort is 49. While we do not have flags in the data specifically denoting trial participation, these patients represent highly probable patients from the reference trial given that all 49 satisfied trial inclusion criteria observable in the data and that these patients were treated prior to commercial availability [11]. The RECIST 1.1 criteria and the availability of tumor samples for biomarker evaluation were not available and the lone trial criteria not included. We then identified NSCLC patients who started a nivolumab monotherapy regimen in the RW since approval (e.g., since 4 March 2015). Mirroring the reference trial inclusion criteria, we further only included patients with stage IIIB or IV disease aged 33 to 85 (e.g., the range from the final trial sample; though,

the youngest actual patient meeting this criterion was 39) with Eastern Cooperative Oncology Group (ECOG) 0 or 1, squamous histology, treated with a platinum-based therapy in the first-line setting (i.e., cisplatin, carboplatin, oxaliplatin, nedaplatin and lobaplatin) and treated in the second-line setting resulting in a final RW sample of  $n = 160$ . All patients in the probable CT sample were treated before 2015 while the RW comparison group was comprised of the following treatment year distributions: 2015 ( $n = 49$ ), 2016 ( $n = 62$ ), 2017 ( $n = 46$ ) and 2018 ( $n = 10$ ).

### Estimation techniques

We compared the RW patients (e.g., our treatment group) with the probable trial patients (e.g., our control group) to determine if there were statistically significant differences in OS between the groups. Following the methods utilized in the reference trial, OS was analyzed using a two-sided log-rank test. The HRs were estimated using a Cox proportional-hazards model with different matching methods. KM curves estimate the OS as well, accounting for the various weighting methodology. Details regarding the matching and propensity scores (PS) are available in the next section. Each regression estimate was calculated with a corresponding 95% CI. We adopted an alpha of 0.05 for the purpose of assessing statistical significance.

### Matching & propensity scores

We assessed the differences in covariate distributions between the two groups using ANOVA – Equal Variances for continuous variables and Chi-square tests for categorical variables. The covariates assessed included: age at index in years as well as in categorical terms (i.e., under 65, 65 to under 75 and 75 and older), brain metastasis, age category with and without brain metastases, smoking status for current or former smokers versus never or undocumented, gender, gender interacted with smoking status, race (Black or African–American, White, or other), stage at index date (IIIB vs IV), practice region (American South vs other regions) and time from completion of most recent prior systemic regimen (less than 3 months vs 3 or more months).

We then conducted multiple matching methods to assess their relative performance in achieving balance and ultimate concordance between probable trial and RW populations. These include: Cox proportional-hazards model weighted by inverse probability of treatment weighting (IPTW), adjusting for PS in Cox proportional-hazards model, Cox proportional-hazards model stratified on PS deciles, PS 1:4 greedy nearest neighbor matching with caliper of 0.5 and directly adjusting for all covariates in Cox proportional-hazards model. We balanced on the following variables: age at index, brain metastasis, gender, race, smoking status and two interaction items – age with brain metastasis and gender with smoking status (Table 1). The variables were selected for their clinical relevance, and the interaction variables were included to improve the fit of the model.

## Results

### Covariate comparisons

The unadjusted covariate means were statistically significantly different at the 95% confidence level for two (notably, related) of the thirteen evaluated covariates: smoking (RWD  $n$  [%] = 145 [90.6%]; probable CT  $n$  [%] = 28 [57.1%]; omnibus Chi-square  $p$ -value < 0.0001) and gender interacted with smoking (as an example: RWD  $n$  (%) male current or former smoker = 100 (62.5%); probable CT  $n$  (%) = 16 (32.7%); omnibus Chi-square  $p$ -value < 0.0001). Following the greedy match procedure, while statistically closer in distributions, the distributions remained statistically significantly different for smoking (omnibus Chi-square  $p$  = 0.0010) and its interaction with gender (omnibus Chi-square  $p$  = 0.0100) (Table 1).

### Survival comparisons

Mean (Standard deviation [SD]) OS was 12.2 (0.7) for the RWD arm and 12.5 (1.3) for the probable CT arm with a Log rank  $p$ -value equal to 0.4771. The median OS (95% CIs) were 10.2 (8.1–12.4) and 11.3 (5.9 to 17.0) and overlap (Table 2). The HRs generated under all six assessed matching methods (including without adjustment), using the probable CT arm as the reference group, were not statistically significantly different at the 95% confidence level: Unadjusted HR: 0.880 (95% CI: 0.613–1.263), IPTW HR: 0.851 (95% CI: 0.591–1.224), PS HR: 0.743 (95% CI: 0.486–1.137), PS Decile HR: 0.920 (95% CI: 0.265–3.191), Greedy Match HR: 0.804 (95% CI: 0.502–1.289), and Multivariate HR: 0.739 (95% CI: 0.483–1.132) (Figure 1 & Table 3 for more detail). The similarity of these survival patterns is reflected in the KM OS curves that were generated for the unadjusted comparison and for the greedy match, IPTW and IPTW adjusted models (Figure 2).

**Table 1. Unadjusted versus post-greedy match baseline demographic characteristics.**

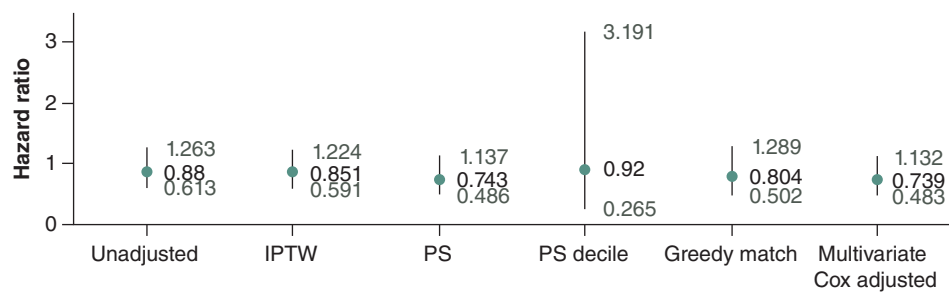
Variables	Unadjusted			Post-greedy match		
	RWD (n = 160)	Probable CT (n = 49)	p-value	RWD (n = 122)	Probable CT (n = 43)	p-value
Age in years at index, mean (SD)	67.1 (8.86)	64.5 (10.26)	0.0879 <sup>†</sup>	68.7 (8.90)	66.5 (8.40)	0.1649 <sup>†</sup>
Brain metastasis, n (%)	31 (19.4%)	16 (32.7%)	0.0514 <sup>‡</sup>	26 (21.3%)	11 (25.6%)	0.5638 <sup>‡</sup>
Age without metastasis, mean (SD)	67.6 (9.28)	66.4 (9.77)	0.5279 <sup>†</sup>	69.8 (9.12)	67.2 (8.87)	0.1630 <sup>†</sup>
Age with metastasis, mean (SD)	65.0 (6.61)	60.6 (10.42)	0.0792 <sup>†</sup>	64.7 (6.81)	64.6 (6.87)	0.9696 <sup>†</sup>
Age category, n <65 (%) <sup>¶</sup>	63 (39.4%)	24 (49.0%)	0.4820 <sup>‡</sup>	34 (27.9%)	19 (44.2%)	0.1379 <sup>‡</sup>
Age category x brain metastasis, ex: n <65 (%)	14 (8.8%)	9 (18.4%)	0.3748 <sup>§</sup>	11 (9.0%)	5 (11.6%)	0.3719 <sup>§</sup>
Smoking, n (%) current/former vs never/undocumented	145 (90.6%)	28 (57.1%)	<0.0001 <sup>‡</sup>	107 (87.7%)	28 (65.1%)	0.0010 <sup>‡</sup>
Female, n (%)	50 (31.3%)	21 (42.9%)	0.1333 <sup>‡</sup>	43 (35.2%)	17 (39.5%)	0.6151 <sup>‡</sup>
Gender x smoking, ex: n (%) male current/former	100 (62.5%)	16 (32.7%)	<0.0001 <sup>‡</sup>	69 (56.6%)	16 (37.2%)	0.0100 <sup>‡</sup>
Race, n (%) Black or African-American <sup>#</sup>	26 (16.3%)	4 (8.2%)	0.3099 <sup>‡</sup>	8 (6.6%)	4 (9.3%)	0.8004 <sup>‡</sup>
Stage at index date, n (%) IIIB vs IV (%)	42 (26.3%)	11 (22.4%)	0.5926 <sup>‡</sup>	34 (27.9%)	10 (23.3%)	0.5564 <sup>‡</sup>
Region, n (%) American South vs others	88 (55.0%)	31 (63.3%)	0.3066 <sup>‡</sup>	63 (51.6%)	29 (67.4%)	0.0728 <sup>‡</sup>
Time from completion of most recent prior systemic regimen, n (%) <3 months	135 (84.4%)	38 (77.6%)	0.2684 <sup>‡</sup>	102 (83.6%)	32 (74.4%)	0.1847 <sup>‡</sup>

Statistical test will apply only to mutually exclusive columns.  
<sup>†</sup> ANOVA-Equal Variances.  
<sup>‡</sup> Chi-square test.  
<sup>§</sup> Fisher's Exact test.  
<sup>¶</sup> Age categories ≥65 to <75 and ≤75 also included.  
<sup>#</sup> Race categories also included White and Other.  
 CT: Clinical trial; RWD: Real-world data; SD: Standard deviation.

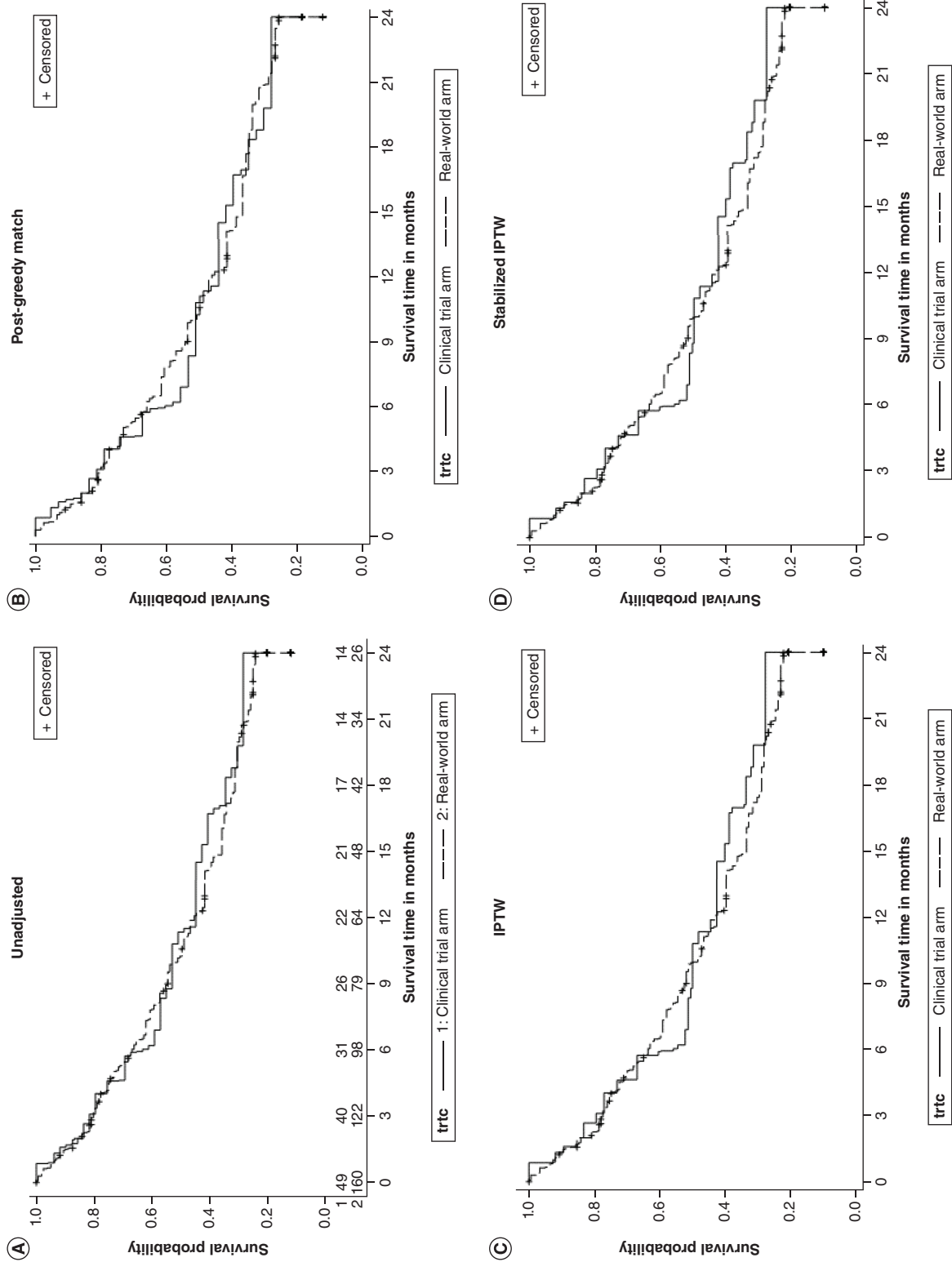
**Table 2. Kaplan–Meier descriptive statistics (censored at Month 24).**

Overall survival	RWD arm (n = 160)	Probable CT arm (n = 49)
Events/patients, n	123/160	39/49
Mean in months (SD)	12.2 (0.7)	12.5 (1.3)
Median in months (95% CI)	10.2 (8.1, 12.4)	11.3 (5.9, 17.0)
Min, max (months)	0.00, 24.00	0.87, 24.00
Log rank Chi-square		0.5056
Log rank p-value		0.4771

Time origin: Start date of nivolumab monotherapy line.  
 All death dates were counted as events.  
 CI: Confidence interval; CT: Clinical trial; Max: Maximum; Min: Minimum; RWD: Real-world data; SD: Standard deviation.



**Figure 1. Hazard ratio comparisons across multiple matching methods.** Hazard ratio is plotted with a 95% confidence interval in the whisker plot. IPTW: Inverse probability of treatment weighting; PS: Propensity score.



**Figure 2. Kaplan-Meier survival curve comparisons. (A)** Unadjusted, **(B)** post-greedy match, **(C)** IPTW and **(D)** stabilized inverse probability of treatment weighting. The stabilized IPTW adjusts IPTW based on the percentage of each treatment group ( $IPTW * N1 / (N1 + N2)$ ). PS, PS Decile and Multivariate Cox Adjusted models were excluded since they were not generating different cohorts (as with greedy match) or simple re-weightings (as with IPTW). IPTW: inverse probability of treatment weighting; PS: Propensity score.

Table 3. Comparison of Cox regression estimates with various matching and control methods.

Type	Estimate	Hazard ratio (95% CI)	Pr > Chi-square	Standard error
Unadjusted	-0.12739	0.880 (0.613 to 1.263)	0.4895	0.18431
IPTW	-0.16183	0.851 (0.591 to 1.224)	0.3830	0.18551
PS	-0.29698	0.743 (0.486 to 1.137)	0.1714	0.21712
PS decile	-0.08389	0.920 (0.265 to 3.191)	0.8949	0.63488
Greedy match	-0.21784	0.804 (0.502 to 1.289)	0.3656	0.24076
Multivariate Cox adjusted	-0.30197	0.739 (0.483 to 1.132)	0.1651	0.21754

CI: Confidence interval; IPTW: Inverse probability of treatment weighting; PS: Propensity score.

## Discussion

Our results generally suggest that survival patterns among a RW cohort behave like a CT cohort in the NSCLC setting as we find no statistically significant differences at the 95% confidence level between the cohorts in any assessed scenario.

These results also qualitatively highlight several themes concerning matching, including the range of possible results that arise from different methods and how underlying baseline covariate differences can translate into these results. In comparison to our unadjusted HR of 0.880, our adjusted HR estimates tended to deviate further from a value of 1.0 with four ranging from 0.739 to 0.851 compared with one at 0.920. This pattern is reflected in our survival curve results using greedy match and IPTW-based approaches. The unadjusted covariate means were significantly different at the 95% confidence level for only the smoking-related covariates. However, at the 90% confidence level, five of thirteen covariates were statistically significantly different; the RW samples were slightly older with higher instances of smoking (which may suggest relatively sicker RW patients), but with fewer cases involving brain metastases (which suggests the opposite). This underscores that covariate imbalance may not uniformly be present in the same direction. In this analysis, adjustments of brain metastasis appear to have dominated the model, given that most adjusted HRs for brain metastasis were smaller after rebalancing.

## Limitations

Our study has several limitations, starting with our CT cohort. Since we did not have direct access to this group in its entirety, we instead identified probable CT patients based on their pre-market access to nivolumab in our database. It is possible that some of these patients received care through other pre-market channels such as compassionate use via expanded access programs or Right to Try [15]. We utilized inclusion and exclusion criteria based on the CT's criteria, and 30 patients were excluded (21 due to failing criteria and 9 due to missing variables necessary to determine inclusion). While we believe that those included are likely to have received pre-market nivolumab through the CT, we cannot completely rule out that some in-sample patients did not participate in the reference CT. Relatedly, having only 49 cases in one of the groups challenges our PS modelling and highlights the potential difficulties of this approach with few observations, limits or risks overfitting the classification model and raises the possibility of our overall sample of 209 being insufficiently powered.

Though likely drawn from it, this sample is also likely non-representative of the underlying CT population whose trial-reported median OS duration is roughly two months shorter than our CT sub-population (though, the 95% CIs of these populations overlap). This potential upward bias in median OS among our probable CT population may be reflected in our results comparing survival against the RW population where, qualitatively speaking as none were statistically significant, all HRs for the RW population were less than one. We similarly find that our probable CT population smoked less and was more likely to be female than the overall published CT population. Relatedly, while both probable trial and RW patient data were generated from the same EMR environment, the probable CT patients were likely managed under CT settings that are generally more tightly controlled and monitored than RW settings. This could result in different patient management and care patterns between groups stemming from differences in the clinical environment rather than differences in the underlying clinical benefits stemming from the assigned treatment.

Finally, our aim in this study was to evaluate the comparability between probable CT patients and RW patients treated with the *same* therapy. This focuses our attention on evaluating null results and on discussing the limits of one's ability to reject a null hypothesis. Were we able to compare the relative performance of a probable CT treatment arm against *either* a probable CT control arm or a RW control arm, we would be discussing statistically

significant differences instead of nonsignificant results. However, this was not possible in this setting as we were unable to identify probable CT control arm cohorts given that the standard of care control arm therapy, docetaxel, was widely commercially available during our study timeframe.

## Conclusion

Overall, there has been growing interest in comparing trial and RW samples. Our study both highlights the potential merits associated with this approach (e.g., that the RWD and probable CT arms were not statistically significantly different from each other in terms of OS under any of the unmatched and matched scenarios) as well as underscores caution when there is limited sample (e.g., risk of overfitting, limited covariate balance achievement and potentially underpowering a comparison). The extent to which these potential merits outweigh these potential cautions depends upon the research context of a specific trial. For instance, the degree of uncertainty surrounding the reliability of an RWD control arm may be more salient when considering this approach in lieu of an RCT given a default inclusion of a prospectively randomly assigned control arm. However, in the context of a single-arm study in which no comparison group is considered in the assessment of clinical efficacy, these reliability concerns may be outweighed by the benefits gained in including a control group.

### Summary points

- Comparing non-small-cell lung cancer (NSCLC) cohorts generated from retrospective real-world data sources to those from prospective clinical trial data has generally shown similar survival outcomes between the groups.
- This study provides direct comparison between non-contemporaneous real-world and probable trial NSCLC patients sourced from the same platform.
- We further include an assessment of the relative performance of alternative matching methods.
- We build on the literature by using non-contemporaneous data sourced from the same platform, by including multiple matching strategies and by leveraging these analyses to highlight potential merits and limitations to the method of comparing trial and real-world patients generally.
- We illustrate the potential merits associated with an external control approach (e.g., relative comparability in survival measures across real-world and probable clinical trial arms).
- We also emphasize caution when there is limited sample (e.g., risk of overfitting, limited covariate balance achievement and potentially underpowering a comparison).
- These assessments may help researchers to evaluate the relative trade-offs of this method given the research content in which they operate.
- Survival patterns between real-world and clinical trial cohorts in this NSCLC setting are not statistically significantly different, suggesting a potentially appropriate role for external controls.

### Author contributions

B Walker: concept and design; analysis and interpretation of data; drafting of the manuscript; critical revision of the paper for important intellectual content; administrative, technical or logistic support; and supervision. HE Ray: concept and design; acquisition of data; analysis and interpretation of data; critical revision of the paper for important intellectual content; and supervision. P Shao: concept and design; acquisition of data; analysis and interpretation of data; critical revision of the paper for important intellectual content; statistical analysis; and provision of study materials or patients. C D'Ambrosio: concept and design; analysis and interpretation of data; and critical revision of the paper for important intellectual content. C White: acquisition of data; analysis and interpretation of data; and critical revision of the paper for important intellectual content. MS Walker: concept and design; analysis and interpretation of data; and critical revision of the paper for important intellectual content.

### Financial disclosure

ConcertAI, LLC, provided the data and financial support for this study. All authors are current or former employees at ConcertAI, LLC, which provided the data and financial support for this study. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

### Competing interests disclosure

The authors have no competing interests or relevant affiliations with any organization or entity with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

### Writing disclosure

No writing assistance was utilized in the production of this manuscript.

### Ethical conduct of research

Institutional review board (IRB) review was not applicable for the research described in this article. This research study was conducted retrospectively from data obtained for clinical purposes. The data used in this study were all collected originally for clinical use in the practice setting and are therefore considered secondary data. Only de-identified data were utilized in the analysis, and unique patient identifiers were completely removed from the analytical dataset.

### Data availability statement

ConcertAI, LLC does not make datasets publicly available because study data are used under license from source practices and other data providers. ConcertAI, LLC will consider requests to access study datasets on a case-by-case basis. Please contact us with any inquiries at <https://www.concertai.com/contact-us/>.

### Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc-nd/4.0/>

## References

Papers of special note have been highlighted as: ● of interest; ●● of considerable interest

1. Cavallo J. A look ahead: how the FDA is adapting in the era of precision medicine. *The ASCO Post*. Available at: <https://ascopost.com/issues/november-1-2013/a-look-ahead-how-the-fda-is-adapting-in-the-era-of-precision-medicine/> (2013).
2. Kesselheim AS, Wang B, Franklin JM, Darrow JJ. Trends in utilization of FDA expedited drug development and approval programs, 1987–2014: cohort study. *BMJ* 351, h4633 (2015).
3. Lamont EB, Hayreh D, Pickett KE *et al*. Is patient travel distance associated with survival on phase II clinical trials in oncology? *J. Natl Cancer Inst.* 95(18), 1370–1375 (2003).
4. Miksad RA, Abernethy AP. Harnessing the power of real-world evidence (RWE): a checklist to ensure regulatory-grade data quality. *Clin. Pharmacol. Ther.* 103(2), 202–205 (2018).
5. Davi R, Chandler M, Elashoff B *et al*. non-small-cell lung cancer (NSCLC) case study examining whether results in a randomized control arm are replicated by a synthetic control arm (SCA). *J. Clin. Oncol.* 37(Suppl. 15), 9108–9108 (2019).
  - **Describes the use of a synthetic control arm (e.g., data sourced from prior clinical trials) that replicated the performance of a trial control with commentary on the importance of covariate balance.**
6. Davi R, Mahendraratnam N, Chatterjee A, Dawson CJ, Sherman R. Informing single-arm clinical trials with external controls. *Nat. Rev. Drug Discov.* 19(12), 821–822 (2020).
  - **This commentary discusses the potential merits of external control arms.**
7. Davi R, Yin X, Stewart M. Exploring the validity of a synthetic control arm (SCA) for augmentation or replacement of a randomized control in difficult-to-study indications: a case study in relapsed or refractory multiple myeloma (R/R MM). *J. Clin. Oncol.* 38(Suppl. 15), e20521–e20521 (2020).
  - **This abstract describes the use of a synthetic control arm that also replicated the performance of a trial control with commentary on the importance of covariate balance.**
8. Menefee ME, Gong Y, Mishra-Kalyani PS *et al*. Project switch: docetaxel as a potential synthetic control in metastatic non-small-cell lung cancer (mNSCLC) trials. *J. Clin. Oncol.* 37(Suppl. 15), 9105–9105 (2019).
  - **This abstract describes the use of a synthetic control arm that also replicated the performance of a trial control with commentary on the importance of covariate balance.**
9. Carrigan G, Whipple S, Capra WB *et al*. Using electronic health records to derive control arms for early phase single-arm lung cancer trials: proof-of-concept in randomized controlled trials. *Clin. Pharmacol. Ther.* 107(2), 369–377 (2020).
  - **Describes the use of multiple external control arms that also replicated the performance of their respective trial controls.**
10. Carrigan G, Whipple S, Taylor MD *et al*. An evaluation of the impact of missing deaths on overall survival analyses of advanced non-small-cell lung cancer patients conducted in an electronic health records database. *Pharmacoepidemiol. Drug Saf.* 28(5), 572–581 (2019).
11. Brahmer J, Reckamp KL, Baas P *et al*. Nivolumab versus docetaxel in advanced squamous-cell non-small-cell lung cancer. *N. Engl. J. Med.* 373(2), 123–135 (2015).
  - **Describes the evidence supporting the reference drug's approval and outlines relevant inclusion/exclusion criteria.**

12. Raedler LA. Opdivo (Nivolumab): second PD-1 inhibitor receives FDA approval for unresectable or metastatic melanoma. *Am. Health Drug Benefits* 8(Spec. Feature), 180–183 (2015).
13. US Food & Drug Administration. OPDIVO (nivolumab) injection. Available at: [https://www.accessdata.fda.gov/drugsatfda\\_docs/nda/2015/125527Orig1s000TOC.cfm](https://www.accessdata.fda.gov/drugsatfda_docs/nda/2015/125527Orig1s000TOC.cfm) (Accessed: 13 April 2023).
14. Shao P, Tepsick JG, Walker B, Ray HE. Improving real-world mortality data quality in oncology research: augmenting electronic medical records with obituary, social security death index, and commercial claims data. *JCO Clin. Cancer Inform.* 7, e2300014 (2023).
15. US Food & Drug Administration. Right to Try. Available at: <https://www.fda.gov/patients/learn-about-expanded-access-and-other-treatment-options/right-try> (Accessed: 27 March 2024).