

**Real-World Progression-Free Survival as an Endpoint in Advanced Non-Small-Cell Lung Cancer: Replicating Atezolizumab and Docetaxel Arms of the OAK Trial Using Data Derived From Electronic Health Records**

Shivani K. Mhatre<sup>1</sup>, Robson J. M. Machado<sup>2</sup>, Thanh G.N. Ton<sup>1</sup>, Huong Trinh<sup>1</sup>, Julien Mazieres<sup>3</sup>, Achim Rittmeyer<sup>4</sup>, Michael T. Bretscher<sup>2</sup>

<sup>1</sup>Genentech, Inc., South San Francisco, CA, USA; <sup>2</sup>F. Hoffmann-La Roche Ltd, Basel, Switzerland; <sup>3</sup>Centre Hospitalier Universitaire, Université Paul Sabatier, Toulouse, France;

<sup>4</sup>Department of Thoracic Oncology, Lungenfachklinik Immenhausen, Immenhausen, Germany

**Corresponding author information:**

Michael T. Bretscher

F. Hoffmann-La Roche Ltd

Grenzacherstrasse 124,

4070 Basel, Switzerland

Email: [michael.bretscher@roche.com](mailto:michael.bretscher@roche.com); Phone: +41 61 682 34 15

**Keywords (3-10):** cancer immunotherapy, chemotherapy, external control, non-small cell lung cancer, progression-free survival, real-world data

## ABSTRACT

**Background:** Evaluating cancer treatments in real-world data (RWD) requires informative endpoints. Due to non-standardized data collection in RWD, it is unclear if and when common oncology endpoints are approximately equivalent to their clinical trial analogues. This study used RWD to replicate both the atezolizumab and docetaxel arms of the OAK trial. Outcomes using progression-free survival (PFS) derived from abstracted physician's notes in RWD (rwPFS) were then compared against PFS outcomes derived according to Response Evaluation Criteria in Solid Tumors (RECIST) from the clinical trial (ctPFS).

**Methods:** Atezolizumab and docetaxel arms of the phase III OAK RCT (NCT02008227) were replicated in a US nationwide real-world database by applying selected OAK inclusion/exclusion criteria, followed by adjustment for baseline prognostic variables using propensity score-based methods. Multiple rwPFS definitions were characterized and a definition was chosen that was acceptable from both clinical and data analysis perspectives. Concordance of outcomes was assessed using Kaplan-Meier (KM) medians and hazard ratios (HRs).

**Results:** Overall, 133 patients receiving atezolizumab and 479 patients receiving docetaxel were selected for the RWD cohort. After adjustment, prognostic variables were balanced between RCT arms and corresponding RWD cohorts. Comparing rwPFS against ctPFS outcomes in terms of KM median and HR showed better concordance for docetaxel (2.99 vs 3.52 months; HR, 0.99, 95% CI, 0.85-1.15) than for atezolizumab (3.71 vs 2.76 months; HR, 0.8, 95% CI 0.61-1.02). The latter improved when events labelled "pseudo-progression" were excluded from the RWD (im-rwPFS) and immune-modified RECIST PFS (im-ctPFS) was used in the RCT Atezolizumab data (4.24 vs 4.14 months; HR, 0.95, 95% CI, 0.70-1.25). These findings were robust across several sensitivity analyses.

**Conclusions:** While rwPFS and ctPFS were similar under docetaxel treatment, this was only the case for atezolizumab when immune-modified progression criteria were used, suggesting

that similarity of RWD endpoints to their clinical trial analogues depends on drug category and possibly other factors. Replication of RCTs using RWD and comparison of outcomes can be used as a tool for characterizing RWD endpoints. Additional studies are needed to verify these findings and to better understand the conditions for approximate numerical equivalence of rwPFS and ctPFS endpoints.

## INTRODUCTION

Real-world data (RWD) derived from electronic health records (EHRs) can provide an opportunity to complement clinical trials (CTs) by including broader patient populations that reflect the actual standard of care and by potentially providing access to large amounts of data.[1] These data can help generate insights on all aspects of routine medical care, including the comparison of treatments.

A prerequisite for evaluating therapies using RWD is the availability of reliable and informative endpoints. The analogues of established CT endpoints such as overall survival (OS), progression-free survival (PFS) and overall response rate (ORR) may be extracted from EHRs and used for outcome comparison. However, data collection processes in clinical practice are considerably less standardized and controlled compared to CTs.[2] Therefore, CT endpoints and their respective RWD analogues are not guaranteed to be practically “equivalent”, which implies that they could lead to different numerical results.

Overall survival (OS), the time from randomization to death of any cause, is often considered the “gold standard” endpoint for evaluating cancer therapies.[3] Due to the unambiguous definition of a death event and because close tracking of patients is not required to establish the date of death, OS derived from RWD can be directly compared with OS from CTs, provided the completeness of death information in the RWD database is high.[4-6] This is not generally the case for PFS, a commonly used non-OS endpoint in cancer trials, defined as the time to either progression or death.[7] Data for PFS are collected differently in RWD (rwPFS) compared with CTs (ctPFS), as illustrated in **Figure 1A**. In clinical practice, there are no strict, standardized protocols for scans or scheduled assessments, and patient visits and imaging may occur at irregular time intervals.[8] Patient access to imaging may vary (e.g., due to health insurance coverage). Physicians are also likely to identify clinically relevant progression with the primary objective to make treatment decisions rather than comparing drug

efficacy. Hence, it is unclear to what extent real-world progression (rwP) events are captured according to the same criteria as in CTs (Response Evaluation Criteria in Solid Tumors [RECIST]).[8, 9] Continuous patient follow-up, which is required to accurately measure rwPFS, is not always guaranteed in RWD because patients may visit alternative health care providers outside of the data collection network. In addition, there are differences in how mortality data are collected: in CTs, death events are captured up to a clinical cut-off date, which is identical for every patient in the study. Since this does not exist in RWD, a decision must be made on which death events to include in rwPFS. Due to the apparent differences in data collection, a better understanding is needed of when the conditions for comparing rwPFS and ctPFS endpoints in group-level analyses are fulfilled - both in terms of data quality, as well as with respect to indication or treatment category.

RWD endpoints, such as time to next treatment, time to treatment discontinuation, time to progression, and rwPFS, have been characterized by assessing surrogacy and correlation of outcomes with OS within the same patients.[10, 11] With this approach it is possible to verify if real-world endpoints are consistent with each other and whether they predict OS. However, such an analysis does not provide any insight into numerical similarity of real-world endpoints and CT analogues. Even an endpoint that perfectly predicts OS could do so by yielding very different numerical results.

Near-equivalence of CT and analogous RWD-derived endpoints is particularly important for comparisons across different datasets, such as in the development of external control (EC) arms from RWD.[12] In this type of analysis, an EC cohort is selected by applying trial inclusion and exclusion criteria to the RWD source, followed by additional covariate adjustment using propensity-score (PS) methods.[13]

Such replication of CT arms from RWD can be used as a tool to assess the comparability of real-world endpoints and their CT analogues. If ctPFS outcomes can reliably

and consistently be replicated using rwPFS, this increases confidence in numerical equivalence of the two endpoints - provided good balance of prognostic variables was achieved between the patient populations. A previous study with four CTs for metastatic non-small lung cancer (NSCLC) demonstrated that rwPFS can replicate ctPFS for patients receiving chemotherapy.[14] Another preliminary study in patients with metastatic breast cancer found rwPFS outcomes derived from EHR RWD were similar to ctPFS.[15] The primary aim and novelty of the present analysis was to replicate the cancer immunotherapy (CIT; atezolizumab) treatment arm of the OAK Phase III RCT (using both the ctPFS and immune-modified [im]-ctPFS endpoints of OAK), in addition to the chemotherapy (docetaxel) treatment arm, and to assess concordance of outcomes between cohorts using the rwPFS and ctPFS endpoints. Finally, there are several different ways of computing rwPFS from the raw data, with little information in the literature describing which methods would most closely replicate ctPFS. Hence, a secondary aim of this analysis was to investigate the impact of varying the rwPFS definition and propose best practices for future, similar analyses.

## **METHODS**

### ***Data Source and Study Design***

Data from the global, multicenter, open-label, randomized, controlled, Phase III “OAK” trial (NCT02008227),[16, 17] were used to evaluate the ability to replicate ctPFS outcomes in a RWD-derived cohort, using the rwPFS endpoint (**Figure 1B**). Patients with advanced NSCLC (aNSCLC) who had received 1-2 prior lines of chemotherapy ( $\geq 1$  of which platinum-based) were included, irrespective of programmed death-ligand 1 (PD-L1) expression levels. Patients received either the PD-L1 inhibitor, atezolizumab, 1200 mg intravenously (IV) every 3 weeks (q3w) until progressive disease (PD) or loss of clinical benefit, or the chemotherapeutic drug, docetaxel, 75 mg/m<sup>2</sup> IV q3w until PD.

ctPFS was determined by the investigator according to RECIST v1.1 guidelines, which comprise pre-specified criteria for defining progression, standard imaging protocols, and scheduled assessments.[9] ctPFS was defined as the time between the date of randomization and the date of first documented PD or death. Participants who were alive and had not experienced PD at the time of analysis were censored at the time of the last tumor assessment. The OAK trial also included im-ctPFS as an exploratory endpoint, which used immune-modified RECIST guidelines (imRECIST) to assess progression. Per imRECIST, a progression is considered a pseudo-progression and not a true progression event if the subsequent scan ( $\geq 4$  weeks) reads a stable disease, partial response, or complete response.[17, 18]

This study used the US nationwide Flatiron Health (FH) EHR-derived de-identified database to construct external CIT and chemotherapy arms. The FH database is a longitudinal database, comprising de-identified patient-level structured and unstructured data, curated via technology-enabled manual abstraction. During the study period, the de-identified data originated from approximately 280 cancer clinics ( $\approx 800$  sites of care).[19, 20] Patients included in the database were diagnosed with advanced or metastatic disease on or after January 1, 2011 and had  $\geq 2$  visits in the FH system. Inclusion and exclusion criteria from the OAK trial [16] were applied to the data where feasible, and patients who had received atezolizumab or docetaxel and satisfied all criteria were included (**Table S1**). Additionally, patients who did not start first line therapy within 90 days of their advanced diagnosis date were excluded in order to select patients who more closely replicated those enrolled in the RCTs.

### ***Descriptive statistics and tests***

Demographic and clinical characteristics were summarized at the time of index date or the pre-defined baseline period near the index date. Means, standard deviations (SD), medians, interquartile ranges (IQR), and ranges were used to describe continuous variables. Frequencies and proportions were used to describe categorical variables. Variables were compared between

the RCT and external CIT arms using Chi-square tests (or Fisher's exact tests in cases where two-way cell counts were low) for frequencies. For continuous variables, independent t-tests were used for normally distributed variables and Wilcoxon rank-sum tests for variables with a skewed distribution. *P*-values <0.05 were considered statistically significant for both continuous and categorical variables. All statistical analyses were performed using R (v 4.0.0).[21]

### ***Covariate adjustment and outcome comparisons***

Differences between RWD and RCT cohorts in clinical variables (age, sex, race, histology, stage at diagnosis, smoking history, and months since advanced diagnosis) were adjusted using inverse probability of treatment weighting (IPTW). PS for IPTW adjustment were obtained from a logistic regression model estimating one term for every category in the baseline covariates (no interactions) and a cubic spline term for "months since advanced diagnosis." PD-L1 status was not included in the PS model because the majority of RWD patients did not receive testing (**Figure S1**). For the primary analysis, average treatment effect (ATE) weighting was chosen with removal (trimming) of patients from non-overlapping PS-regions of both RWD and RCT. This minimizes the risk of remaining covariate imbalances after adjustment, thus allowing for a more direct comparison of the endpoints themselves. Average treatment effect for the treated (ATT) weights, with trimming of RWD patients only, were used in one sensitivity analysis to demonstrate the ability to replicate outcomes in patient populations similar to OAK. Differences between baseline characteristics of RCT arm and CIT RWD cohort were assessed by the standardized mean difference (SMD). A SMD <0.1 was an indicator of good covariate balance,[22] and a SMD <0.2 was an indicator of acceptable covariate balance. Concordance of outcomes between RWD and CT endpoints was assessed by hazard ratios (HRs) and Kaplan-Meier (KM) median time to event. Confidence intervals (CIs) were determined by bootstrapping.

The primary analysis included patients with missing Eastern Cooperative Oncology group (ECOG) score but censored patients at the start of any visit gap >90 days to avoid



missing rwP events due to interrupted follow-up. Sensitivity analyses were performed 1) using a doubly robust approach (estimation of a conditional HR by fitting multivariable Cox regression to the IPTW adjusted dataset using the same covariates as the PS model); 2) adjusting for ECOG (excluding patients with missing ECOG); 3) omitting patients who were lost to follow-up <3 months after index date; and 4) using a 60-day time window to capture death events after the end of rwP follow-up (instead of 30 days). Exploratory analyses included 1) estimating the ATT (without IPTW weighting and trimming of RCT patients); and 2) allowing for visit gaps of any length during rwP follow-up (not censoring at the start of 90-day visit gaps).

### ***Development of an rwPFS Endpoint***

While the high-level definition of PFS as time to death or progression is straightforward, several definitions are possible when considering how to compute rwPFS from RWD (**Supplemental Methods**). The impact of the exact rwPFS definition on outcomes was evaluated through tabulation of KM medians, HRs, and event frequencies. Based on the analyses performed, a “reference” definition of rwPFS was chosen that was acceptable from both clinical and data analysis points of view and minimized potential biases (**Table 1, Figure 1C**).

## **RESULTS**

Overall, 133 patients who received atezolizumab and 479 patients who received docetaxel were selected for the RWD cohorts. Following adjustments via IPTW, 134.7 and 475.1 patients were included in the RWD cohorts for atezolizumab and docetaxel, respectively. Fractional patient number occurred in the adjusted cohorts due to reweighting of data points (**Table 2**).

### ***Replication of the OAK Atezolizumab Arm***

Good balance of prognostic variables was achieved between the atezolizumab RWD cohort and the corresponding RCT arm. After IPTW ATE weighting and trimming, SMD values were <0.1 with the exception of histology (0.12) (**Table 2**). KM median time-to-event was longer in rwPFS compared with ctPFS (rwPFS: 3.71, 95% CI, 2.5-4.83 months vs ctPFS: 2.76 95% CI, 2.43-2.96 months; **Figure 2A**). The corresponding HR indicated a lower event rate for rwPFS (HR, 0.80, 95% CI, 0.61-1.02). A similar pattern was observed consistently across all sensitivity analyses (**Figure 2B**). Among patients who had a PFS event, 12.7% were death events in rwPFS, compared with 13.1% in ctPFS. When events labelled “pseudo-progression” were omitted in the RWD (im-rwPFS) and compared with RCT PFS evaluated according to imRECIST guidelines (im-ctPFS), the KM median time-to-event was similar (im-rwPFS: 4.24, 95% CI, 2.83-6.64 months vs im-ctPFS: 4.14, 95% CI, 3.48-4.63 months) and the corresponding HR was close to unity (HR, 0.95, 95% CI, 0.70-1.25; **Figure 2C**). A similar pattern was observed consistently across all sensitivity analyses (**Figure 2D**). The rwPFS and im-rwPFS endpoints assessed by KM curves were very similar (**Figure S2**). Among patients who had a PFS event, 14.0% were death events in im-rwPFS, compared with 25.8% in im-ctPFS.

### ***Replication of the OAK Docetaxel Arm***

Good balance of prognostic variables was achieved between the docetaxel RWD cohort and the corresponding RCT arm. After IPTW ATE weighting and trimming all SMD values were <0.1 (**Table 2**). KM median time-to-event was similar in rwPFS compared with ctPFS (rwPFS: 2.99, 95% CI, 2.56-3.58 months vs ctPFS: 3.52 95% CI, 2.89-4.14 months; **Figure 3A**). The corresponding HR was close to unity (HR, 0.99, 95% CI, 0.85-1.15), with a similar pattern observed consistently across all sensitivity analyses (**Figure 3B**). Among patients who had a PFS event, 16.6% were death events in rwPFS, compared with 23.5% in ctPFS.

### **Characterization of rwPFS**

Follow-up of RWD patients was close, with clinic visits occurring frequently. Gaps longer than 2-3 weeks between two visits or between last visit and rwPFS event/censoring were rare (**Figure S3**). Altering the rwPFS definition by using different subsets of rwP events and by varying the handling of line of therapy (LoT) change (**Table 3**) resulted overall in only minor outcome changes compared to the “reference” endpoint definition. The largest change in the atezolizumab cohort was observed when excluding events labelled “pseudo-progression” (KM median 4.1 months vs 3.58 months; HR 0.93; **Figure S2**). In the docetaxel cohort, the largest change was observed when imputing a progression event at the start of a new LoT (KM median 2.53 months vs 2.76 months; HR 1.17). The time window for capturing death events after the end of progression follow-up was varied from 0-60 days, in 10-day increments, and the impact on event composition, KM median time-to-event, and HR (relative to the “reference” definition) was recorded (**Table S2**). In the docetaxel cohort, the number of additional death events captured per 10-day increment was largest for window sizes up to 30-40 days and decreased thereafter. A similar pattern was observed for KM median and HR, which showed marked sensitivity to window size at small absolute values, with changes per increment decreasing at window sizes above 30-40 days. The number of death events in the atezolizumab cohort was too small to draw firm conclusions, but results appeared consistent with the pattern observed in the docetaxel cohort.

### **DISCUSSION/CONCLUSION**

This study investigated the concordance of PFS outcomes in the OAK trial with those from matched real-world cohorts in patients diagnosed with aNSCLC. A novel aspect and focus of this analysis was the assessment in patients receiving a CIT regimen, since many future experimental drugs will likely need to be compared against a CIT standard of care. The main objective was to better understand the extent to which clinical trial PFS and its real-world

analogue(s) could be considered equivalent endpoints, meaning that they would lead to the same or very similar numerical results in any group-level statistical comparison. EHR-derived cohorts were selected for both the atezolizumab and docetaxel arms of OAK. A good balance of prognostic variables was achieved after IPTW ATE and trimming (removing) patients from both trial data and RWD. This is necessary for interpreting concordance of outcomes as potentially indicative of a high similarity between real-world and trial endpoints.

Very high concordance was observed between rwPFS and ctPFS outcomes in docetaxel-treated patients (**Figure 3**). This is consistent with the PFS results of a similar study in aNSCLC where the chemotherapy control arms of four clinical trials were replicated using RWD.[14] Together, these results increase the confidence that rwPFS and ctPFS endpoints are largely equivalent in aNSCLC patients under a chemotherapy regimen. A different observation was made for patients receiving atezolizumab: median rwPFS was approximately one month longer compared to median ctPFS. Interestingly, very good concordance was observed in the same patients when im-rwPFS and im-ctPFS endpoints were compared, using immune-modified progression criteria in both RWD and RCT data. This change can largely be attributed to the difference between im-ctPFS vs ctPFS because the difference between im-rwPFS and rwPFS appeared small in comparison (**Figure 2 A-B, Figure S2, Table 3**). These observations raise the possibility that rwPFS under CIT treatment may correspond more closely to im-RECIST PFS, but it remains unclear whether this is a general pattern or an observation unique to this analysis. If a general pattern, this would imply that caution should be exercised when using rwPFS to compare RWD of CIT treatments with CT data, or to compare a CIT against other drug categories within RWD. However, if the present RWD Atezolizumab cohort were used as external control arm for a single arm clinical trial, the direction of observed bias (longer time to event) suggests a conservative estimate of PFS benefit for the experimental treatment. Importantly, these findings indicate that the similarity of rwPFS and ctPFS endpoints may be

drug dependent. Therefore, concordance of outcomes may need to be assessed separately for different drug categories. Similarly, concordance of rwPFS and ctPFS outcomes should also be verified empirically in other cancers.

Ideally, confidence in real-world endpoints could be fully established via a bottom-up approach, similar to clinical trials, where a thorough understanding of data generation provides reassurance of outcome validity. Due to the complex and poorly understood processes that generate RWD, this is likely not possible, and empirical evidence similar to the present study will likely continue to play an important role. Nevertheless, attempts should be made to characterize and understand data generation in RWD as well as possible, through detailed analysis of tumor imaging data, temporal patterns of clinic visits, or qualitative research where physicians describe what led to the decisions recorded in the data. Such analyses may help understand whether apparent equivalence of endpoints on a group level implies equivalence at the individual level. Despite differences in data collection between CT and RWD, it is possible that rwPFS and ctPFS may yield similar results. This could be the result of a shared primary objective, where both processes in routine health care and in clinical trials were designed to identify clinically relevant progression events. Different secondary objectives may be present, such as physicians making treatment decisions rather than comparing effectiveness, or clinical trials being designed to be auditable. Additionally, minimizing cost and reducing unnecessary discomfort for patients through frequent imaging may be other factors considered in routine clinical care.

During the planning of this study, it became apparent that there are multiple ways of computing rwPFS from RWD, and no consensus could be reached on a single “correct” rwPFS definition (**Supplemental information**). Instead, a pragmatic choice from a range of acceptable definitions was made, and the implications of varying the definition on rwPFS outcomes were documented (**Table 3, Table S2**) in line with FDA draft guidelines.[23] The endpoint definition

did not affect outcomes strongly enough to substantially alter the study results, with the following exceptions: 1) the omission of events labelled “pseudo-progression” for patients receiving atezolizumab as discussed above, 2) the imputation of a progression event at LoT change for patients receiving docetaxel, and 3) the size of the time window for capturing death events after the end of progression follow-up. From a clinician’s point of view, a new LoT is initiated after the patient has progressed, and consequently imputing a progression event at LoT change should not have made a substantial difference. It is unclear why there was an effect among the docetaxel patients, but it may suggest an inconsistency between the LoT and progression information in the database. The decision to choose a 30-day time window for capturing death events was motivated by the observation that extending the window size from 0 to 30 days had a substantial effect on HRs (in line with a hypothesis of patients leaving the FH network shortly before death), while extending the window size further resulted in minimal change. Software in the form of an R package ([supplemental methods](#)) is provided that simplifies calculation of rwPFS, as well as the comparison of different definitions.[24]

Some limitations of this study are worth noting. Because the study was restricted to patients with aNSCLC who were mostly from community settings, results may not be generalizable to other tumor types, stages, and patient populations. PD-L1 status was not used for PS weighting due to a limited number of real-world patients who received any PD-L1 test. Patients in both RWD atezolizumab and docetaxel treatment groups did not start their treatment during the same period in which trial patients were recruited, which has the potential to introduce calendar time bias due to changes in the standard of care. Any conclusions regarding rwPFS in atezolizumab should be considered preliminary as they are based on the replication of a single trial arm. Common limitations of RWD studies apply, where unmeasured confounding factors and missing data can still exist [13].

In conclusion, ctPFS outcomes could successfully be replicated using rwPFS for docetaxel, while replication was better for atezolizumab when immune-modified progression criteria were used in both RWD and RCT data. Additional studies are needed to verify these findings across different CITs approved for aNSCLC and in other indications.

## **ACKNOWLEDGMENTS**

We thank the patients who participated in this study and their families. This manuscript was sponsored by F. Hoffmann-La Roche Ltd. and Genentech, Inc. Support for third-party writing assistance, furnished by Jessica Swanner, PhD, of Health Interactions, Inc., was provided by F. Hoffmann-La Roche Ltd. We thank many of our colleagues for their helpful discussions surrounding this work.

## **DATA SHARING STATEMENT**

Qualified researchers may request access to individual patient level data through the clinical study data request platform (<https://vivli.org/>). Further details on Roche's criteria for eligible studies are available here (<https://vivli.org/members/ourmembers/>). For further details on Roche's Global Policy on the Sharing of Clinical Information and how to request access to related clinical study documents, see here ([https://www.roche.com/research\\_and\\_development/who\\_we\\_are\\_how\\_we\\_work/clinical\\_trials/our\\_commitment\\_to\\_data\\_sharing.htm](https://www.roche.com/research_and_development/who_we_are_how_we_work/clinical_trials/our_commitment_to_data_sharing.htm)).

The data that support the findings of this study originated from Flatiron Health, Inc. These de-identified data may be made available upon request and are subject to a license agreement with Flatiron Health; interested researchers should contact [dataaccess@flatiron.com](mailto:dataaccess@flatiron.com) to determine licensing terms.

## **AUTHOR CONTRIBUTION STATEMENTS**

All authors had full access to the data in the study and take responsibility for the integrity and the accuracy of the data analysis. All authors contributed substantially to the study design (SKM, TGNT, MTB, RJMM, HT), data analysis (MTB, RJMM, HT), interpretation of results (SKM, MTB, TGNT, HT, JM, AR), and the development of the manuscript (MTB, SKM, TGNT, HT, JM, AR). All authors read and approved the final manuscript.

## **CONFLICTS OF INTEREST**

SKM and HT are employees of Genentech, Inc. MTB is an employee of F. Hoffmann-La Roche Ltd. JM reports personal fees from Roche, Astra Zeneca, Pierre Fabre, Takeda, BMS, MSD, Jiangsu Hengrui, Blueprint, Daiichi, Novartis, and Amgen; grants from Roche, Astra Zeneca, Pierre Fabre, and BMS.



## REFERENCES

1. Rahman R, Vantz S, McDunn J, Louv B, Reyes-Rivera I, Polley MC, Merchant F, Abrey LE, Allen JE, Aguilar LK *et al*: Leveraging external data in the design and analysis of clinical trials in neuro-oncology. *Lancet Oncol* 2021, 22(10):e456-e465.
2. Sheffield KM, Dreyer NA, Murray JF, Faries DE, Klopchin MN: Replication of randomized clinical trial results using real-world data: paving the way for effectiveness decisions. *J Comp Eff Res* 2020, 9(15):1043-1050.
3. U.S. Department of Health and Human Services, Food and Drug Administration, Oncology Center of Excellence. Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics Guidance for Industry. 2018.
4. Bretscher MT, Sanglier T: Quantifying the impact of mortality underreporting on analyses of overall survival. *34th International Conference for Pharmacoepidemiology and Risk Management, Prague, Czech Republic*. 2018.
5. Carrigan G, Whipple S, Taylor MD, Torres AZ, Gossai A, Arniere B, Tucker M, Hofmeister PP, Lambert P, Griffith SD *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* 2019, 28(5):572-581.
6. Zhang Q, Gossai A, Monroe S, Nussbaum NC, Parrinello CM: Validation analysis of a composite real-world mortality endpoint for patients with cancer in the United States. *Health Serv Res* 2021, 56(6):1281-1287.
7. Lakdawalla DN, Shafrin J, Hou N, Peneva D, Vine S, Park J, Zhang J, Brookmeyer R, Figlin RA: Predicting Real-World Effectiveness of Cancer Therapies Using Overall Survival and Progression-Free Survival from Clinical Trials: Empirical Evidence for the ASCO Value Framework. *Value Health* 2017, 20(7):866-875.

8. Adamson BJS, Ma X, Griffith SD, Sweeney EM, Sarkar S, Bourla AB: Differential frequency in imaging-based outcome measurement: Bias in real-world oncology comparative-effectiveness studies. *Pharmacoepidemiol Drug Saf* 2021.
9. Schwartz LH, Litière S, de Vries E, Ford R, Gwyther S, Mandrekar S, Shankar L, Bogaerts J, Chen A, Dancey J *et al*: RECIST 1.1-Update and clarification: From the RECIST committee. *Eur J Cancer* 2016, 62:132-137.
10. Stewart M, Norden AD, Dreyer N, Henk HJ, Abernethy AP, Chrischilles E, Kushi L, Mansfield AS, Khozin S, Sharon E *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* 2019, 3:1-15.
11. Kehl KL, Riely GJ, Lepisto EM, Lavery JA, Warner JL, LeNoue-Newton ML, Sweeney SM, Rudolph JE, Brown S, Yu C *et al*: Correlation Between Surrogate End Points and Overall Survival in a Multi-institutional Clinicogenomic Cohort of Patients With Non-Small Cell Lung or Colorectal Cancer. *JAMA Netw Open* 2021, 4(7):e2117547.
12. Ventz S, Lai A, Cloughesy TF, Wen PY, Trippa L, Alexander BM: Design and Evaluation of an External Control Arm Using Prior Clinical Trials and Real-World Data. *Clin Cancer Res* 2019, 25(16):4993-5001.
13. Carrigan G, Whipple S, Capra WB, Taylor MD, Brown JS, Lu M, Arnieri B, Copping R, Rothman KJ: 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* 2020, 107(2):369-377.
14. Ton TG, Pal N, Trinh H, Mahrus S, Bretscher MT, Machado RJ, Sadetsky N, Chaudhary N, Lu MW, Riely GJ: Replication of overall survival, progression-free survival and response in chemotherapy arms of first-line non-small cell lung cancer trials using real-world data external controls. Submitted 2022.

15. Huang Bartlett C, Mardekian J, Cotter MJ, Huang X, Zhang Z, Parrinello CM, Bourla AB: Concordance of real-world versus conventional progression-free survival from a phase 3 trial of endocrine therapy as first-line treatment for metastatic breast cancer. *PLoS One* 2020, 15(4):e0227256.
16. Rittmeyer A, Barlesi F, Waterkamp D, Park K, Ciardiello F, von Pawel J, Gadgeel SM, Hida T, Kowalski DM, Dols MC *et al*: Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. *Lancet* 2017, 389(10066):255-265.
17. Gandara DR, von Pawel J, Mazieres J, Sullivan R, Helland Å, Han JY, Ponce Aix S, Rittmeyer A, Barlesi F, Kubo T *et al*: Atezolizumab Treatment Beyond Progression in Advanced NSCLC: Results From the Randomized, Phase III OAK Study. *J Thorac Oncol* 2018, 13(12):1906-1918.
18. Hodi FS, Ballinger M, Lyons B, Soria JC, Nishino M, Tabernero J, Powles T, Smith D, Hoos A, McKenna C *et al*: Immune-Modified Response Evaluation Criteria In Solid Tumors (imRECIST): Refining Guidelines to Assess the Clinical Benefit of Cancer Immunotherapy. *J Clin Oncol* 2018, 36(9):850-858.
19. Ma X, Long L, Moon S, Adamson BJS, Baxi SS: Comparison of Population Characteristics in Real-World Clinical Oncology Databases in the US: Flatiron Health, SEER, and NPCR. *medRxiv* 2020.
20. Birnbaum B, Nussbaum N, Seidle-Rathkopf K, Agrawal M, Estevez M, Estola E, Haimson J, He L, Larson P, Richardson P: Model-assisted cohort selection with bias analysis for generating large-scale cohorts from the EHR for oncology research. *arXiv* 2020.
21. R Core Team 2021: A language and environment for statistical computing. In.: R Foundation for Statistical computing, Vienna Austria; 2021.

22. Austin PC: Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat Med* 2009, 28(25):3083-3107.
23. U.S. Food and Drug Administration. Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products - Guidance for Industry. In.; 2021.
24. Bretscher MT. RwPFS. 2021. <https://phcanalytics.github.io/RwPFS/>

## TABLES

**Table 1.** Elements of ctPFS and corresponding adaptations for RWD. CCOD, clinical cutoff date; EHR, electronic health record; FH, Flatiron Health; FU, follow-up; PD, progressive disease; RCT, randomized controlled trial; RWD, real-world data; rwP, real-world progression.

<sup>a</sup> Sensitivity analysis was conducted using rwP from radiographic evidence only.

Criteria	OAK RCT	Adaptation in RWD	Rationale for adaptation
Time of censoring in absence of an event	Time of last tumor assessment	End of gapless rwP FU. Earliest of: - last recorded visit - start of a visit gap >90 days - last date of rwP abstraction	<ul style="list-style-type: none"> <li>● To avoid missing rwP events, it is necessary to ensure patients have continuous FU and rwP is abstracted from the EHR data during the entire FU period.</li> </ul>
PD event	First PD event by radiographic evaluation	First rwP event (any) during rwP FU	<ul style="list-style-type: none"> <li>● Most progression events are radiographically confirmed</li> <li>● Risk of missing rwP events if restricting to radiographically confirmed rwP events only<sup>a</sup></li> </ul>
Death event	Death between randomization and CCOD	Death between index date and 30 days after end of rwP FU	<ul style="list-style-type: none"> <li>● Patients often leave EHR network shortly before death; hence, death must be captured during a short period after last visit</li> <li>● Pragmatic choice of 30 days captures a large proportion of excess deaths shortly after end of progression FU while still being short (to minimize bias)</li> </ul>
Absence of event	Alive and have not had PD between randomization and CCOD	Alive at end of rwP FU (+30 days) and have not had rwP	

**Table 2.** Baseline characteristics of the atezolizumab and docetaxel RCT arms and respective RWD cohorts before and after covariate adjustment via inverse probability weighting and trimming. RCT, randomized controlled trial; RWD, real-world data; SMD, standardized mean difference. <sup>a</sup> Fractional patient numbers may occur after propensity score adjustments due to reweighting of data points.

Treatment Arm		Atezolizumab						Docetaxel					
		Unadjusted		SMD	Adjusted		SMD	Unadjusted		SMD	Adjusted		SMD
Variable, n (%)	Categories	RCT n=425	RWD n=133		RCT n=388.4 <sup>a</sup>	RWD n=134.7 <sup>a</sup>		RCT n=425	RWD n=479		RCT n=413.8 <sup>a</sup>	RWD n=475.1 <sup>a</sup>	
Age group	<65	248.0 (58.4)	49.0 (36.8)	0.48	201.7 (51.9)	67.6 (50.2)	0.05	236.0 (55.5)	199.0 (41.5)	0.35	202.3 (48.9)	231.5 (48.7)	0.01
	65-74	137.0 (32.2)	50.0 (37.6)		136.4 (35.1)	50.3 (37.3)		159.0 (37.4)	195.0 (40.7)		162.7 (39.3)	186.4 (39.2)	
	≥75	40.0 (9.4)	34.0 (25.6)		50.2 (12.9)	16.9 (12.5)		30.0 (7.1)	85.0 (17.7)		48.8 (11.8)	57.2 (12.0)	
Sex	Female	164.0 (38.6)	64.0 (48.1)	0.18	158.8 (40.9)	51.8 (38.4)	0.05	166.0 (39.1)	212.0 (44.3)	0.09	173.2 (41.8)	194.0 (40.8)	0.02
	Male	261.0 (61.4)	69.0 (51.9)		229.6 (59.1)	83.0 (61.6)		259.0 (60.9)	267.0 (55.7)		240.7 (58.2)	281.1 (59.2)	
Histology	Non-squamous	313.0 (73.6)	101.0 (75.9)	0.03	294.0 (75.7)	108.9 (80.9)	0.12	315.0 (74.1)	360.0 (75.2)	0.03	307.8 (74.4)	356.1 (75.0)	0.01
	Squamous	112.0 (26.4)	32.0 (24.1)		94.3 (24.3)	25.8 (19.1)		110.0 (25.9)	119.0 (24.8)		106.0 (25.6)	119.0 (25.0)	
Stage at diagnosis	I	24.0 (5.6)	16.0 (12.0)	0.35	28.0 (7.2)	9.0 (6.6)	0.03	34.0 (8.0)	31.0 (6.5)	0.23	30.1 (7.3)	33.9 (7.1)	0.01
	II	31.0 (7.3)	5.0 (3.8)		23.3 (6.0)	8.6 (6.4)		23.0 (5.4)	22.0 (4.6)		19.5 (4.7)	22.2 (4.7)	
	IIIA	61.0 (14.4)	5.0 (3.8)		34.2 (8.8)	11.2 (8.3)		52.0 (12.2)	30.0 (6.3)		36.9 (8.9)	42.7 (9.0)	
	IIIB	53.0 (12.5)	20.0 (15.0)		54.2 (14.0)	19.1 (14.2)		55.0 (12.9)	53.0 (11.1)		46.7 (11.3)	54.3 (11.4)	
	IV	247.0 (58.1)	87.0 (65.4)		248.7 (64.0)	86.9 (64.5)		257.0 (60.5)	337.0 (70.4)		276.4 (66.8)	317.3 (66.8)	
	Unknown	9.0 (2.1)	0.0 (0.0)		0	0		4.0 (0.9)	6.0 (1.3)		4.3 (1.0)	4.8 (1.0)	

Smoking history	History	341.0 (80.2)	114.0 (85.7)	0.13	317.9 (81.9)	107.2 (79.6)	0.06	353.0 (83.1)	434.0 (90.6)	0.20	361.5 (87.3)	409.1 (86.1)	0.04
	No history	84.0 (19.8)	19.0 (14.3)		70.5 (18.1)	27.5 (20.4)		72.0 (16.9)	45.0 (9.4)		52.3 (12.7)	66.0 (13.9)	
Race	White	302.0 (71.1)	101.0 (75.9)	0.17	282.8 (72.8)	101.3 (75.2)	0.05	296.0 (69.6)	334.0 (69.7)	0.16	291.8 (70.5)	329.6 (69.4)	0.03
	Others	103.0 (24.2)	22.0 (16.5)		84.3 (21.7)	26.6 (19.8)		115.0 (27.1)	110.0 (23.0)		102.1 (24.7)	123.1 (25.9)	
	Unknown	20.0 (4.7)	10.0 (7.5)		21.3 (5.5)	6.8 (5.1)		14.0 (3.3)	35.0 (7.3)		19.9 (4.8)	22.5 (4.7)	
Months since advanced diagnosis, mean (SD)		15.47 (14.49)	12.26 (10.31)	0.17	13.82 (12.68)	14.87 (13.48)	0.08	13.90 (13.47)	9.53 (5.79)	0.37	11.01 (7.79)	11.27 (8.17)	0.03

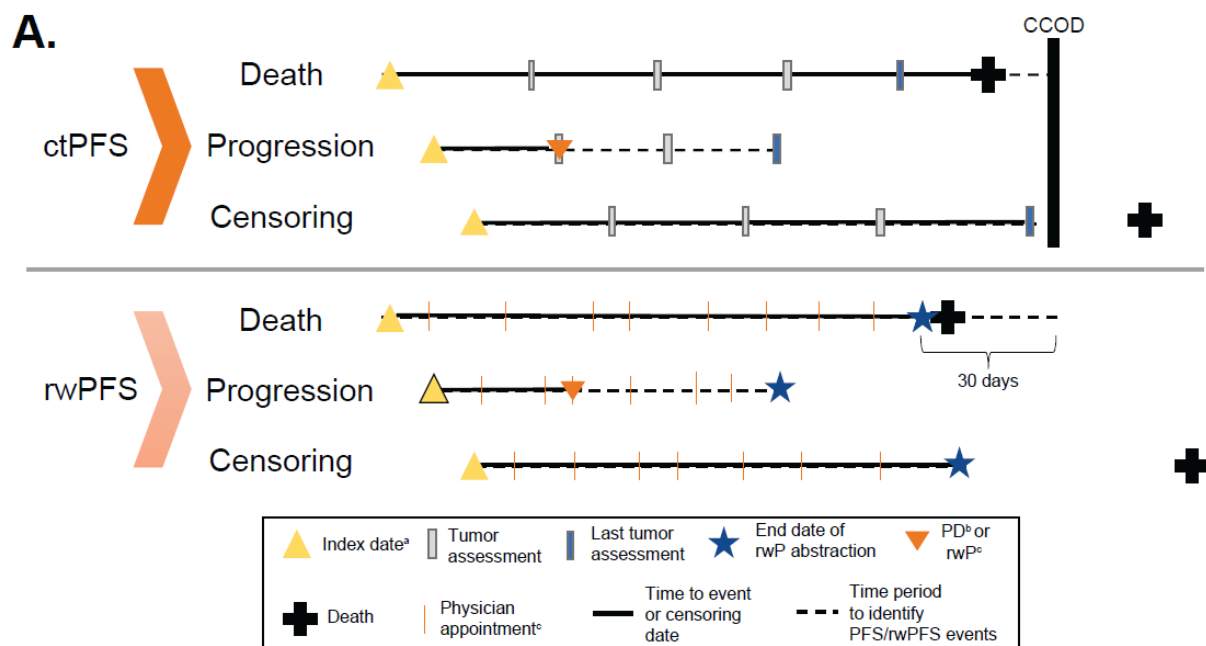
**Table 3.** Effects of varying the event and censoring definitions. CI, confidence interval; HR, hazard ratio; KM, Kaplan Meier; rwP, real-world progression; rwPFS, real-world progression-free survival.

rwPFS definition	rwP, n	Deaths, n	Censored, n	KM median rwPFS (95% CI)	HR (95% CI)
<b>Atezolizumab arm</b>					
All progression events	89	13	31	3.58 (2.73-4.34)	1 (reference)
Excluding pseudoprogression	86	14	33	4.1 (3.25-4.73)	0.93 (0.7-1.22)
Radiographic only	88	14	31	3.58 (2.73-4.34)	1 (0.76-1.31)
Progression after 14 days only	89	13	31	3.58 (2.73-4.34)	1 (0.76-1.32)
Censor at next line	88	14	31	3.48 (2.66-4.27)	1.02 (0.78-1.34)
Impute progression at next line	91	13	29	3.42 (2.53-4.14)	1.04 (0.8-1.37)
<b>Docetaxel arm</b>					
All progression events	330	65	84	2.76 (2.53-3.29)	1 (reference)
Excluding pseudoprogression	330	65	84	2.76 (2.53-3.29)	1 (0.87-1.15)
Radiographic only	321	68	90	2.79 (2.53-3.38)	0.98 (0.85-1.12)
Progression after 14 days only	325	69	85	2.79 (2.56-3.38)	0.99 (0.86-1.13)
Censor at next line	309	60	110	2.66 (2.5-3.19)	1.05 (0.91-1.21)
Impute progression at next line	348	60	71	2.53 (2.23-2.79)	1.17 (1.02-1.34)



## FIGURES

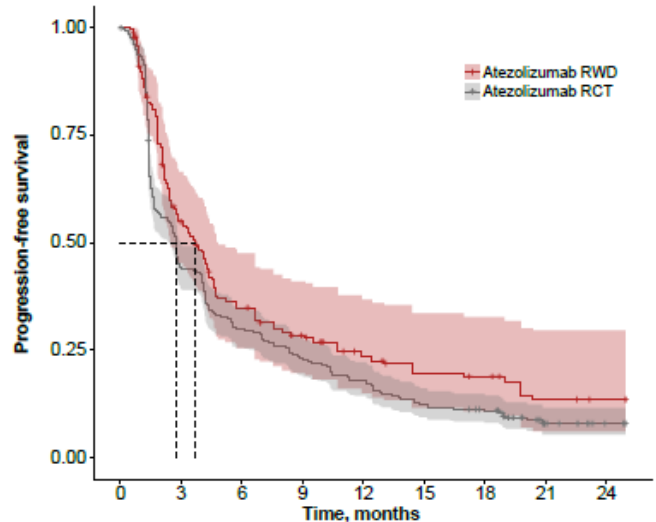
**Figure 1.** Data collection, study design, and rwPFS definition. A. Data collection processes for CTs and RWD. B. Study design of the OAK RCT and external RWD. C. Proposed definition of rwPFS in this analysis. CCOD, clinical cutoff date; I/E, inclusion/exclusion; im-ctPFS, immune-modified-clinical trial progression-free survival; im-rwPFS, immune-modified-real-world progression-free survival; IV, intravenous; LoT, line of therapy; NSCLC, non-small cell lung cancer; PD, progressive disease; PD-L1, programmed death-ligand 1; PFS, progression-free survival; PS, propensity score; q3w, every 3 weeks; RCT, randomized controlled trial; CT, clinical trial; RWD, real-world data; rwP, real-world progression; rwPFS, real-world progression-free survival. <sup>a</sup> Randomization for CT and start of treatment for RWD. <sup>b</sup> In CT. <sup>c</sup> In RWD.





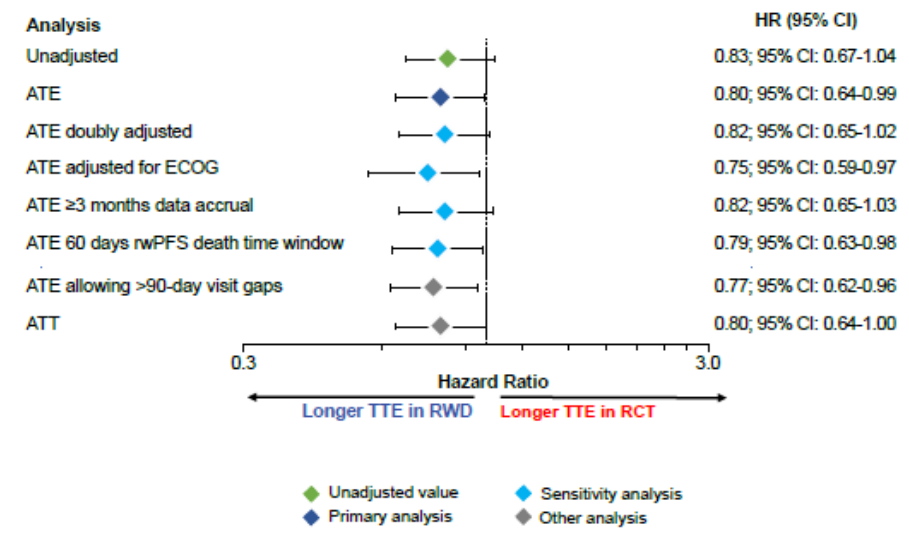
**Figure 2.** Replication of the OAK RCT atezolizumab arm. A. KM curves of ctPFS and rwPFS and B. corresponding HR values. C. KM curves of im-ctPFS and im-rwPFS and D. corresponding HR values. ATE, average treatment effect; ATT, average treatment effect in the treated; ECOG, Eastern Cooperative Oncology Group performance score; HR, hazard ratio; PFS, progression-free survival; im-ctPFS, immune-modified clinical trial progression-free survival; im-rwPFS, immune-modified real-world progression-free survival; RCT, randomized controlled trial; ctPFS, clinical trial progression-free survival; rwPFS, real-world progression-free survival; TTE, time to event

**A.**

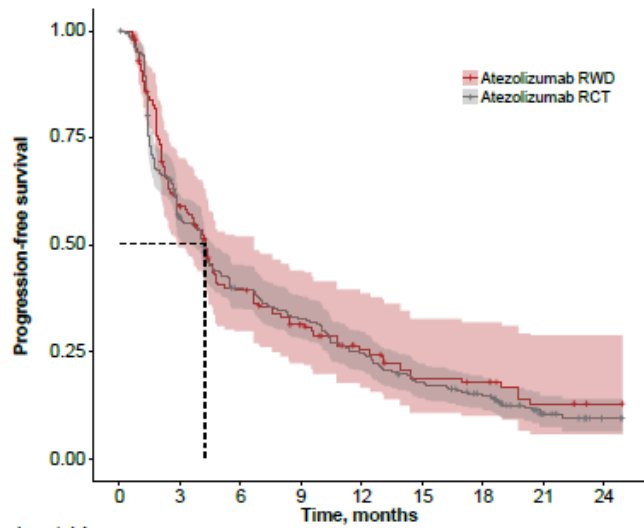


Number at risk		0	3	6	9	12	15	18	21	24
Atezolizumab RWD (n=135)	135	70	42	31	16	9	7	5	3	
Atezolizumab RCT (n=388)	388	170	115	89	69	47	39	10	2	

**B.**

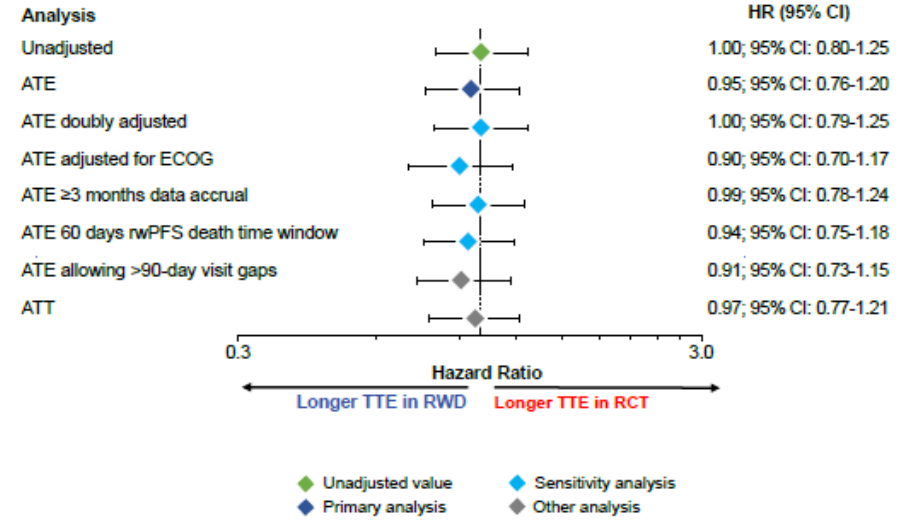


C.

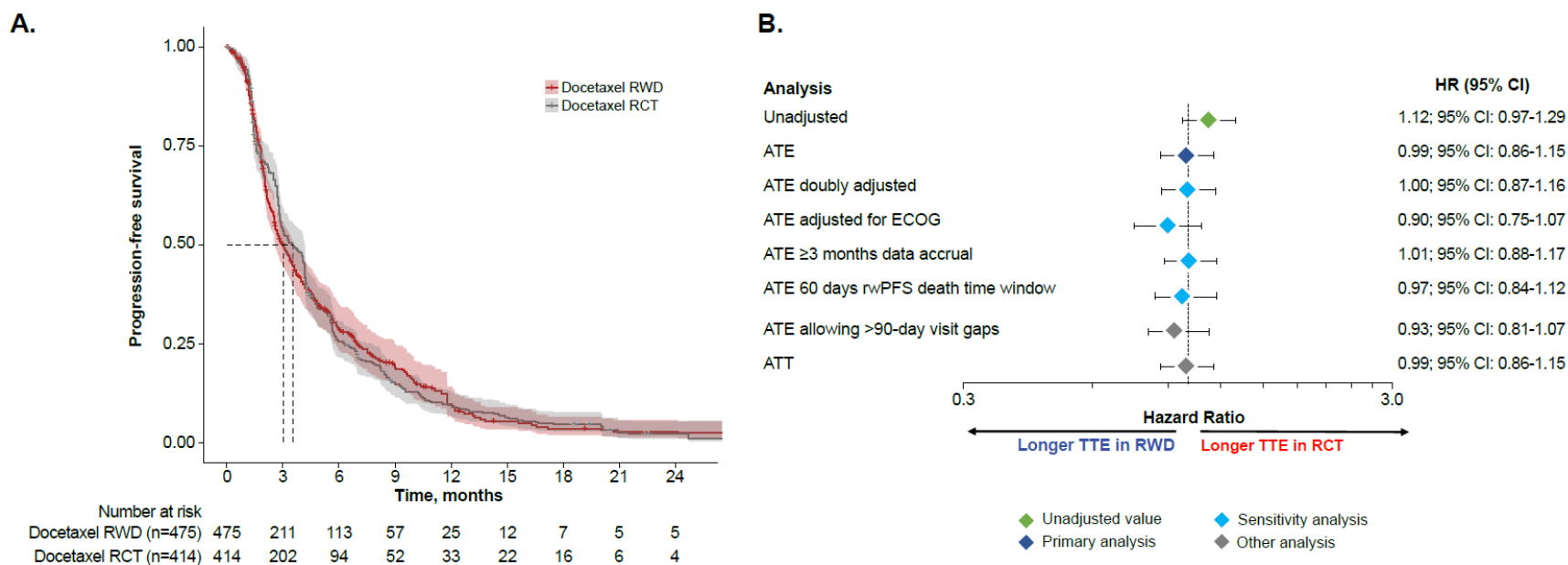


Number at risk		0	3	6	9	12	15	18	21	24
Atezolizumab RWD (n=135)		135	75	47	34	18	9	7	5	3
Atezolizumab RCT (n=388)		388	206	143	117	88	62	47	13	2

D.



**Figure 3.** Replication of the OAK docetaxel arm. A. KM curves of RWD rwPFS and OAK PFS. B. HRs by analysis. ATE, average treatment effect; ATT, average treatment treated; ECOG, Eastern Cooperative Oncology Group; HR, hazard ratio; KM, Kaplan Meier; PFS, progression-free survival; RCT, randomized controlled trial; RECIST, response evaluation criteria in solid tumors; rwPFS, real-world progression-free survival; TTE, time to event.



## SUPPLEMENTAL INFORMATION

### Methods

The following considerations illustrate the range of possible real-world progression-free survival (rwPFS) definitions:

- rwP events: i) Should all types of real-world progression (rwP) events in the Flatiron Health (FH) database be used?, or ii) should events flagged as “pseudo-progression” be omitted?, or iii) should only radiographically confirmed events be used (in line with the clinical trial data)?, and should v) events happening within 14 days of index date be discarded (as per FH recommendation)?
- Line of Therapy (LoT) change: From a clinician’s point-of-view, a new LoT is associated with progression. However, the reason for LoT change is usually not available in real-world data (RWD). Questions: i) should patients be followed for rwP beyond LoT change?, or ii) should patients be censored at LoT change?, or iii) should a progression event be imputed at LoT change (in line with the clinical LoT concept)?
- Continuous follow-up: Unlike overall survival (OS), progression-free survival (PFS) requires continuous follow-up. In RWD this may not always be the case, as patients may leave the FH network and seek treatment elsewhere. Hence, it may be beneficial to censor patients at the start of long visit gaps (indicating absence) to avoid missing rwP events. Question: how long should the maximally permissible gap in recorded visits be?
- Death events: Typically, patients leave the FH network shortly before death (i.e., transferred to hospice or a hospital). This requires capturing death events for some period after end of progression follow-up. How large should the time window be to avoid informative censoring? A larger time window captures more death events but is

also more prone to the following two biases: i) missing rwP events, leading to underestimation of the event rate, and ii) differential treatment (censoring at end of rwP follow-up vs recording of a death event) of patients depending on experiencing a death event or not.<sup>19</sup>

[RwPFS R package](#) to calculate and compare a range of rwPFS endpoints with varying definitions.



**Table S1.** Inclusion/Exclusion Criteria From the OAK Trial Applied to the FH aNSCLC EDM. 2L, second-line; 3L, third-line; ALK, anaplastic lymphoma kinase; ALP, alkaline phosphatase; a/m, advanced/metastatic; aNSCLC, advanced non-small cell lung cancer; ALT, alanine aminotransferase; ANC, absolute neutrophil count; aPPT, activated partial thromboplastin time; AST, aspartate aminotransferase; CD137, tumor necrosis factor receptor superfamily, member 9; CNS, central nervous system; CTLA4, Cytotoxic T lymphocyte-associated antigen-4; EDM, enhanced data mart; EGFR, epidermal growth factor receptor; FH, Flatiron Health; ICD, international classification of disease; INR, international normalized ratio; PD-1, programmed death 1; PD-L1, programmed death-ligand 1; q3w, every 3 weeks; TKI, tyrosine kinase inhibitor; UNL, upper limit of normal; WBC, white blood cell count; X, Application of the criteria to RWD was not feasible.

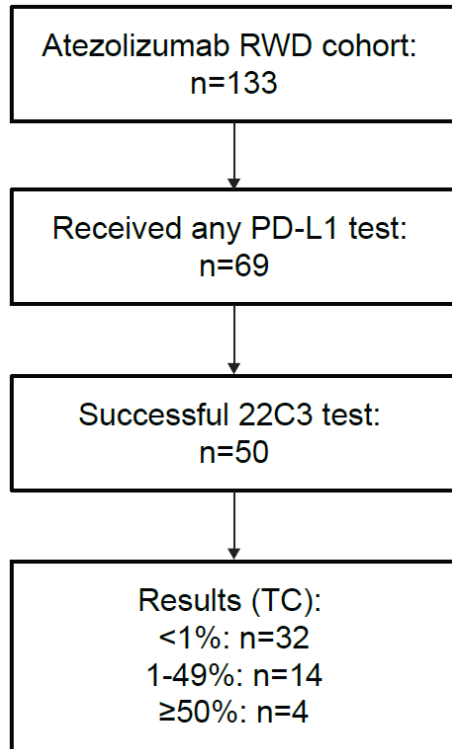
<b>Data</b>	<b>OAK trial</b>	<b>FH aNSCLC EDM</b>
aNSCLC classification	Locally advanced or metastatic (Stage IIIB, Stage IV, or recurrent)	Locally advanced or metastatic (Stage IIIB, Stage IV, or recurrent)
Investigational drug	Atezolizumab 1200 mg	Atezolizumab monotherapy
Comparator arm	Docetaxel 75 mg/m <sup>2</sup> q3w	Docetaxel monotherapy
Age, years	≥18	≥18
Histology	Squamous or non-squamous	Squamous or non-squamous
ECOG	0-1	0-1 or missing
Prior treatment requirement	Prior platinum-containing regimen for a/m NSCLC or disease recurrence within 6 months of treatment with a platinum-based adjuvant/neoadjuvant regimen (<6 months gap between last dose of neo/adjuvant treatment and a/m NSCLC diagnosis date) or combined modality (e.g., chemoradiation) regimen. Patients with a previously detected EGFR or ALK fusion oncogene must additionally have experienced PD on EGFR or ALK TKI.	2L or 3L atezolizumab monotherapy patients after prior treatment with platinum, EGFR or ALK TKI (note: Patients who progress within 6 months of ending (neo)adjuvant treatment will not be included since these treatments are not abstracted by FH)
Current line of therapy for trial	2L or 3L	
Stage of disease	IIIB - IV	IIIB-IV
Life expectancy	≥12 weeks	X
Labs	Must have adequate hematologic and end-organ function, defined by the following laboratory test results obtained within 14 days prior to randomization	Exclude patients who do not meet any of the following criteria (missing are included):

ANC	≥1500 cells/mL	≥1500 cells/mL
WBC	>2500/mL	>2500/mL
Lymphocytes	≥500/mL	≥500/mL
Serum albumin	≥2.5g/dL	≥2.5g/dL
Platelets	≥100,000/mL without transfusion	≥100,000/mL without transfusion
Hemoglobin	≥9 g/dL	≥9 g/dL
INR or aPTT	≤1.5 x ULN	≤1.5 x ULN
AST, ALT, ALP	AST or ALT ≤2.5× upper limit of normal (ULN), with ALP ≤2.5×ULN or AST and ALT ≤1.5×ULN, with ALP > 2.5×ULN	AST or ALT ≤2.5× upper limit of normal (ULN), with ALP ≤2.5×ULN or AST and ALT ≤1.5×ULN, with ALP > 2.5×ULN
Serum bilirubin	≤1.0 x ULN	≤1.0 x ULN
Serum creatinine	Creatinine clearance ≥30 mL/min	Creatinine clearance ≥30 mL/min
Exclusion	CNS metastasis	X
	Other cancers	Exclude patients who are part of other Flatiron EDMs as well
	Autoimmune disease	X
	Idiopathic pulmonary fibrosis, pneumonitis	X
	Active Hepatitis B or C	Exclude patients with Hepatitis B/C ICD codes
	Prior treatment with CD137 agonists, anti-CTLA4, anti-PD-1, or anti-PD-L1 therapeutic antibody or pathway-targeting agents	Exclude patients treated prior with CD137 agonists, anti-CTLA4, anti-PD-1, or anti-PD-L1 therapeutic antibody or pathway-targeting agents or unknown clinical trial drug

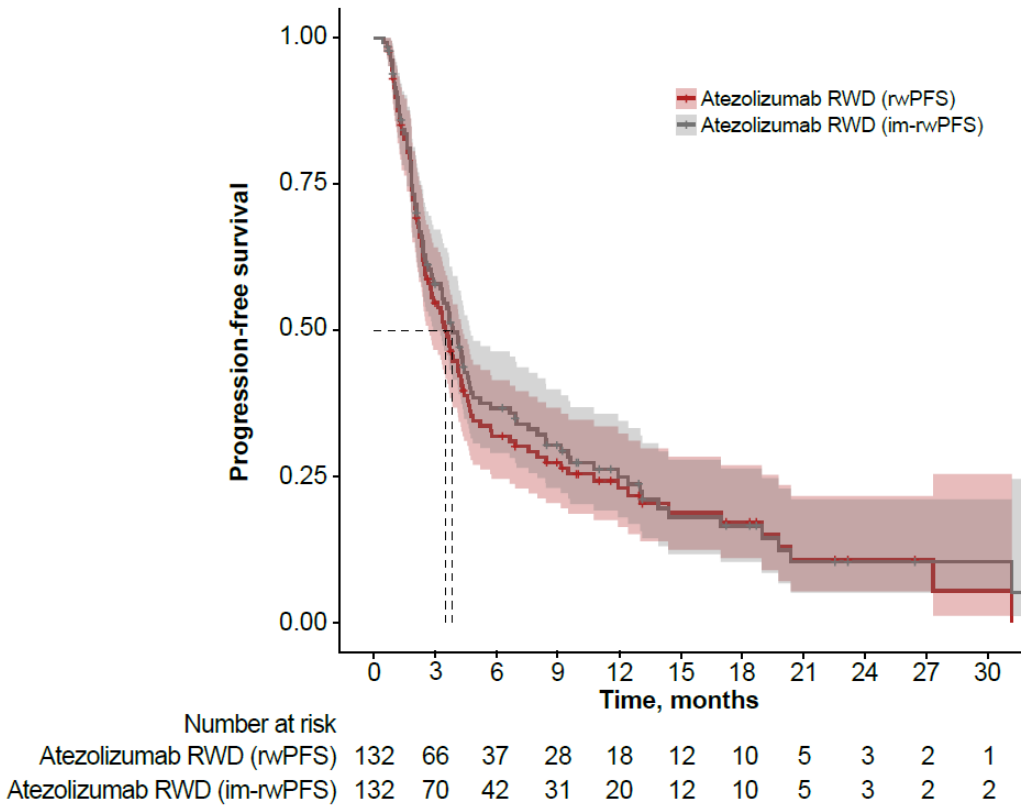
**Table S2.** Effects of varying the time window for capturing death after end of rwP follow-up. CI, confidence interval; HR, hazard ratio; KM, Kaplan Meier; rwP, real-world progression; rwPFS, real-world progression-free survival.

Death time window, days	rwP, n	Deaths, n (%)	Censored, n	Additional deaths, n	KM median rwPFS (95% CI)	HR (95% CI)
<b>Atezolizumab arm</b>						
0	89	1 (1.1)	43		4.07 (3.35-4.83)	0.89 (0.67-1.19)
10	89	5 (5.3)	39	4	3.81 (3.25-4.67)	0.93 (0.7-1.23)
20	89	8 (8.2)	36	3	3.68 (2.89-4.63)	0.96 (0.73-1.27)
30	89	13 (12.7)	31	5	3.58 (2.73-4.34)	1 (reference)
40	89	15 (14.4)	29	2	3.42 (2.66-4.27)	1.01 (0.77-1.33)
50	89	17 (16.0)	27	2	3.42 (2.66-4.27)	1.03 (0.78-1.35)
60	89	18 (16.8)	26	1	3.35 (2.53-4.24)	1.03 (0.79-1.36)
<b>Docetaxel arm</b>						
0	330	9 (2.7)	140		3.35 (2.79-3.98)	0.87 (0.76-1.01)
10	330	25 (7.0)	124	16	3.09 (2.69-3.71)	0.91 (0.79-1.05)
20	330	44 (11.8)	105	19	2.89 (2.56-3.45)	0.96 (0.83-1.1)
30	330	65 (16.5)	84	21	2.76 (2.53-3.29)	1 (reference)
40	330	78 (19.1)	71	13	2.66 (2.5-3.19)	1.02 (0.89-1.17)
50	330	84 (20.3)	65	6	2.66 (2.5-3.09)	1.03 (0.9-1.18)
60	330	93 (22.0)	56	9	2.66 (2.5-3.09)	1.04 (0.91-1.2)

**Figure S1.** PD-L1 testing in the atezolizumab RWD arm. PD-L1, programmed death-ligand 1; RWD, real-world data; TC, tumor cell.

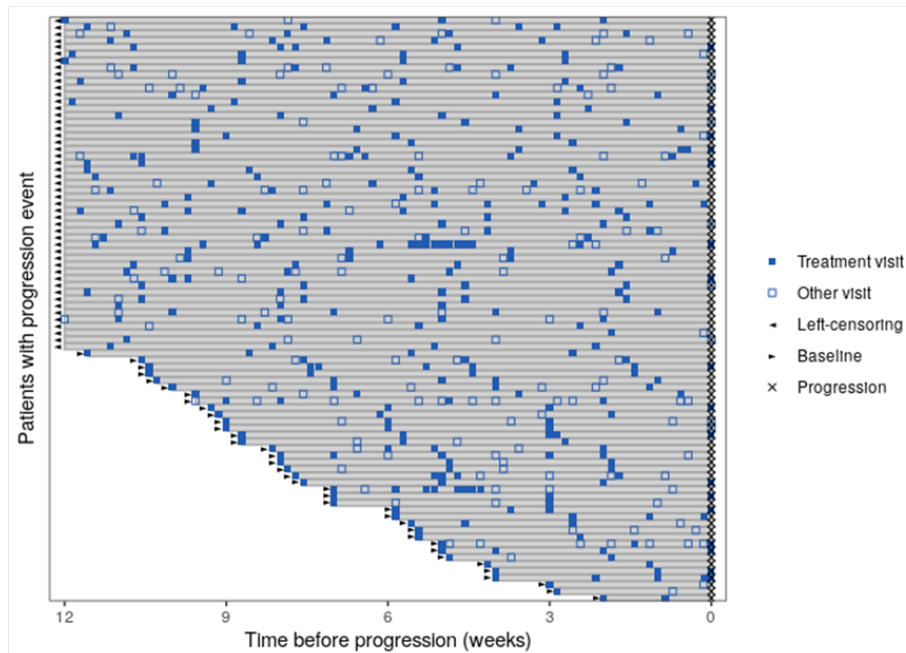


**Figure S2.** KM curves of atezolizumab rwPFS vs im-rwPFS. im-rwPFS, immune-modified-real-world progression-free survival; KM, Kaplan Meier; rwPFS, real-world progression-free survival.

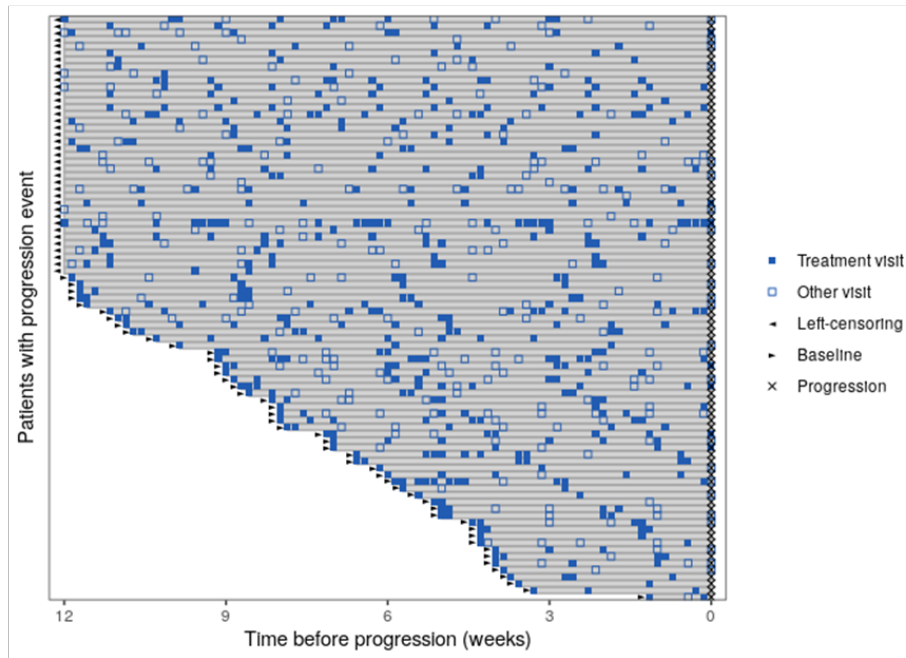


**Figure S3.** Frequency of patient visits for patients with a progression event in the A. atezolizumab arm and the B. docetaxel arm. Frequency of patient visits for patients with a death event in the C. atezolizumab arm or the D. docetaxel arm or a censoring event in the E. atezolizumab arm or F. docetaxel arm.

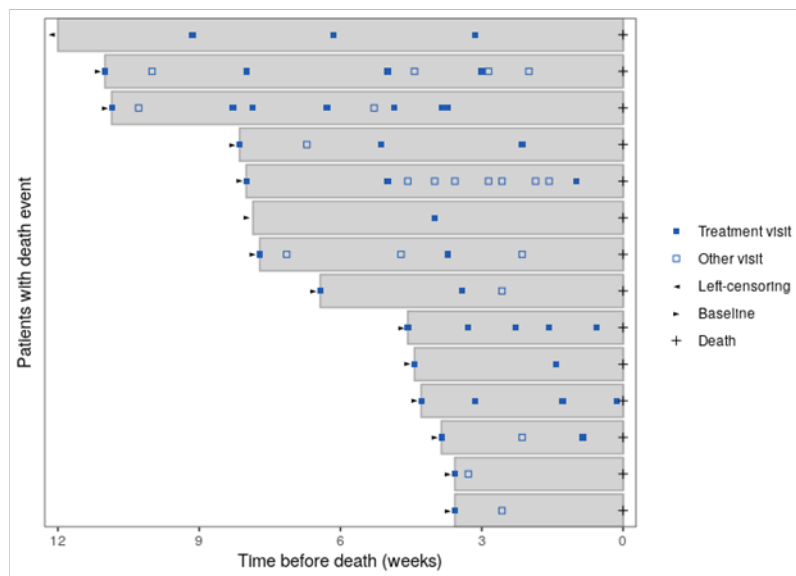
**A.**



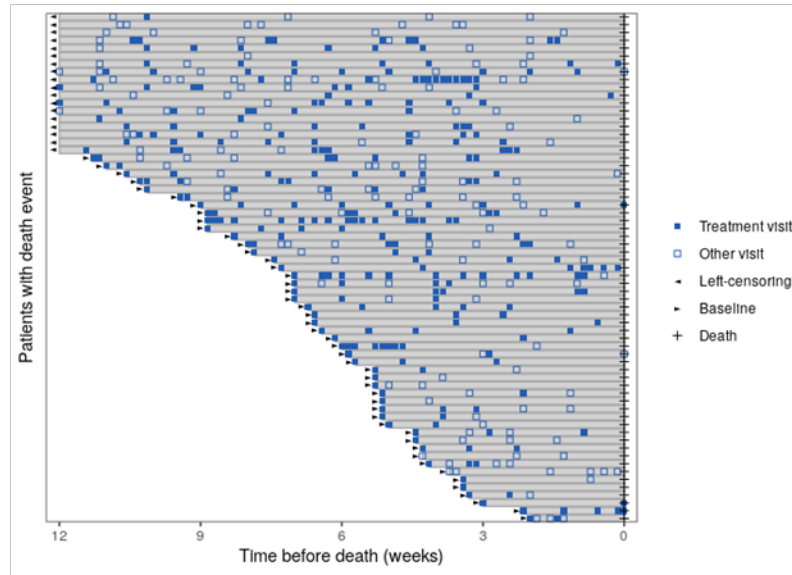
**B.**



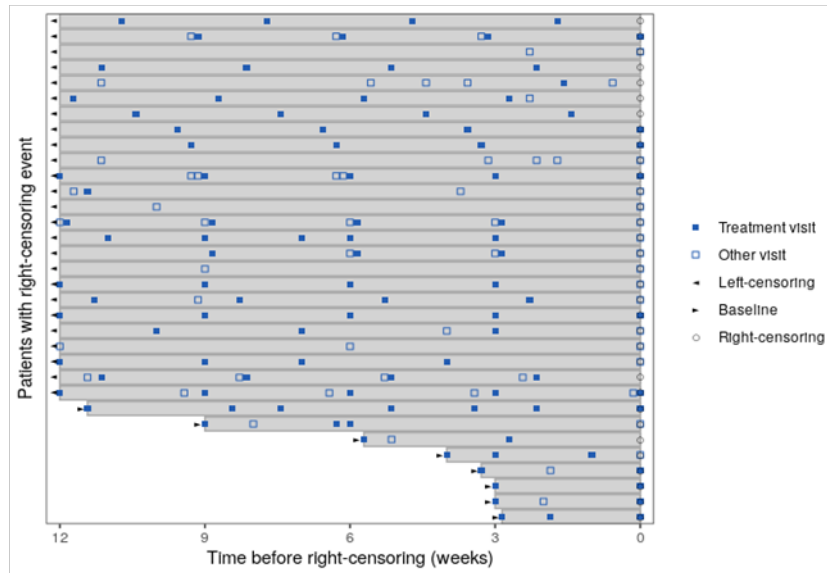
**C.**



**D.**



**E.**





**F.**

