Emulation of the FLAURA (NCT02296125) trial using specialty oncology electronic health records databases

NCT06675695

30th October 2024

1. Title Page

Title	Emulation of the FLAURA (NCT02296125) trial using specialty oncology electronic health records databases
Research question & Objectives	Emulation of the FLAURA (NCT02296125) trial, which compared Osimertinib to standard of care on the risk of death in patients with non-small cell lung cancer.
Protocol version	V1.3
Last update date	November 17, 2024
Contributors	**Primary investigators contact information:** Janick Weberpals, Shirley V. Wang **Contributor names:** Thomas DeRamus (programmer)
Study registration	**Site:** n/a **Identifier:** n/a
Sponsor	**Organization:** n/a **Contact:** n/a
Conflict of interest	n/a
Protocol repository	Clinicaltrials.gov <URL>
Analytic code repository	https://gitlab-scm.partners.org/drugepi/encore/flaura-nct-02296125
Quarto study report (including annotated code and output)	https://drugepi.gitlab-pages.partners.org/encore/flaura-nct-02296125/
***encore.io*[1] version**	0.2.0 (see attached documentation *encore.io_0.2.0.pdf*)
[1] Internally-developed R package to streamline analytics across all available databases and to enhance consistency, transparency and reproducibility in variable definitions and analytic workflows across trial emulations.	

Table of contents

2. Abstract

This trial emulation study aims to emulate the FLAURA trial (NCT02296125) using real-world specialty oncology electronic health records data and to investigate the concordance between the trial's original and the emulated treatment effect estimate on overall survival (OS). FLAURA was a Phase III, double-blind, randomised study assessing the efficacy and safety of AZD9291 (= osimertinib, 80 mg orally, once daily) versus a standard of care (SoC) Epidermal Growth Factor Receptor (EGFR) Tyrosine Kinase Inhibitor (TKI) (either gefitinib [250 mg orally, once daily] or erlotinib [150 mg orally, once daily]) in patients with locally advanced or metastatic EGFR sensitizing mutation (EGFRm+) Non-small Cell Lung Cancer (NSCLC) who are treatment-naïve and eligible for first-line treatment with an EGFR-TKI.

3. Amendments and updates

Version date	Version number	Section of protocol	Amendment or update	Reason
July 22, 2024	V1.0	NA	Initial version	NA
September 26, 2024	V1.1	Context and rationale for data sources, Context and rationale for covariates, Data analysis, Context and rationale for analysis plan	• Added more details to covariate encoding • Specified that EDB1 will be used for primary analysis • Provided reasons for excluding EDB2 and EDB4 from primary analysis • Added sensitivity analyses	
October 7, 2024	V1.2	Context and rationale for analysis plan	• Added a complete case sensitivity analysis	
November 17, 2024	V1.3	Various	• Incorporated FDA team feedback	

4. Rationale and background

Randomized controlled trials (RCTs) are generally regarded as the gold-standard of evidence for establishing efficacy of medical products. However, real-world data (RWD) are increasingly used to complement evidence from RCTs. Yet, to have confidence in the accuracy of non-interventional studies medical products and their outcomes in oncology, investigators need to know what questions can be validly answered, with which non-interventional study designs, and which analysis methods

are appropriate, given the data that is available. Building on a process from the RCT DUPLICATE initiative[1–4] **EmulatioN of Comparative Oncology trials with Real-world Evidence** (**ENCORE**) is the trial emulation discussed in this protocol is part of the expansion project specific to oncology and aims to emulate 12 randomized oncology RCTs using multiple EHR data sources.

The purpose of this protocol is to describe the emulation of the **FLAURA trial**.[5,6] FLAURA was a Phase III, double-blind, randomised study assessing the efficacy and safety of AZD9291 (= osimertinib, 80 mg orally, once daily) versus a standard of care (SoC) Epidermal Growth Factor Receptor (EGFR) Tyrosine Kinase Inhibitor (TKI) (either gefitinib [250 mg orally, once daily] or erlotinib [150 mg orally, once daily]) in patients with locally advanced or metastatic EGFR sensitizing mutation (EGFRm+) Non-small Cell Lung Cancer (NSCLC) who are treatment-naïve and eligible for first-line treatment with an EGFR-TKI.

The primary trial estimate targeted for emulation is a hazard ratio (HR) for death (overall survival) of 0.80 (95% CI 0.64 to 1.00) with a median overall survival time of 38.6 months (95% CI 34.5 to 41.8) in the osimertinib group and 31.8 months (95% CI, 26.6 to 36.0) in the comparator group. Market availability of osimertinib began April 18, 2018 based on the progression-free survival (PFS) endpoint.

The PFS endpoint was published in the NEJM on November 18, 2017 (PMID 29151359).[5]

The OS endpoint was published in the NEJM on November 21, 2019 (PMID 31751012).[6]

5. Research question and objectives

The primary and secondary research question is summarized in Table 1.

A. Primary research question and objective

Table 1. Primary and secondary research questions and objective.

Objective:	To compare the overall survival [OS] in patients who initiated osimertinib versus patients who initiated erlotinib or gefitinib.
Hypothesis:	Initiation of osimertinib improves overall survival time as compared to initiation of erlotinib or gefitinib

Population ***(mention key inclusion-exclusion criteria):***	• Age at least 18 years
	• Locally advanced or metastatic Non-small Cell Lung Cancer (NSCLC) patients who receive treatment without curative intent
	  ○ The line of therapy for patients in EDB1 and EDB2 is implicitly advanced/metastatic because the line of therapy classification starts after their advanced/metastatic diagnosis in the respective database
	  ○ In EDB4, patients must explicitly have any evidence of a metastasis prior initiating Osimertinib, gefitinib or erlotinib.

	• Adenocarcinoma histology • EGFR sensitizing mutation (any EGFRm+) • Treatment-naïve and eligible for first-line (1L) treatment with an EGFR-targeted tyrosine kinase inhibitor (TKI) • ECOG 0 or 1
Exposure:	Initiation of osimertinib
Comparator:	Initiation of erlotinib or gefitinib
Outcome:	Time to all-cause mortality (OS)
Time *(when follow up begins and ends):*	From the day of treatment initiation until death or last observed clinical activity/last sign of the patient being alive or data cut-off, whichever occurred earlier
Setting:	1L EGFRm+ metastatic NSCLC
Main measure of effect:	Primary: Hazard ratio (95% CI) for overall survival Secondary: median overall survival time (difference) in % (95% CI)

The emulation of the main protocol elements of the FLAURA is illustrated side by side in **Table 2**.

Table 2. Trial emulation table summarizing the main protocol elements of the FLAURA trial and the planned emulation.

Protocol component	FLAURA RCT	Emulation
Eligibility criteria	- Age at least 18 years at time of randomization - Adenocarcinoma of the lung - Locally advanced or metastatic NSCLC, not amenable to curative surgery or radiotherapy - Tumour harbours one of the 2 common EGFR mutations known to be associated with EGFR-TKI sensitivity (Ex19del, L858R), either alone or in combination with other EGFR mutations - Patients must be treatment-naïve for locally advanced or metastatic NSCLC and eligible to receive first-line treatment with gefitinib or erlotinib - World Health Organization Performance Status of 0 to 1 - No prior treatment with any systemic anti-cancer therapy for locally advanced/metastatic NSCLC - No prior treatment with an EGFR-TKI.	- Age at least 18 years at time of treatment initiation - Mention of 'adenocarcinoma' histology ('non-squamous' in EDB1) - Line of therapy setting needs to be 'advanced' (EDB1), 'metastatic' (EDB2) or there needs to be evidence of metastatic disease at time of treatment initiation (EDB4) - Evidence of any EGFR mutation in the 180 days before/on treatment initiation - Treatment needs to be first-line for advanced (EDB1) or metastatic (EDB2) disease or patients with prior exposure to any potential antineoplastic treatment commonly used in the advanced/metastatic NSCLC setting are excluded (EDB4) - ECOG 0 or 1 in the 90 days before/on treatment initiation
Treatment strategies	osimertinib versus SoC EGFR-TKI (erlotinib or gefitinib)	osimertinib versus SoC EGFR-TKI (erlotinib or gefitinib)
Assignment procedures	Random assignment to receive either osimertinib or SoC EGFR-TKI (gefitinib or erlotinib) in a 1:1 ratio	1:1 matching of patients based on their propensity to initiate osimertinib as opposed to SoC EGFR-TKI
Follow-up period	Time from randomization until death from any cause or censoring of patients who did not die on the basis of the last recorded date that the patient was known to be alive	Time from treatment initiation until death from any cause or censoring of patients who did not die on the basis of the last recorded date that the patient was known to be alive
Outcome	Overall survival	Overall survival
Causal contrast	Intent-to-treat effect	Counterfactual comparison of initiation of osimertinib as opposed to SoC EGFR-TKI

6. Research methods

6.1. Data sources

6.1.1. Context and rationale for data sources

The overall ENCORE project uses data from a total four different oncology-specific electronic health records (EHR)-derived data sources: ConcertAI, COTA, Flatiron Health, McKesson/Ontada. For ENCORE, not all databases are available for each cancer indication and the names of the databases will henceforth be blinded and referred to as ENCORE DataBase (EDB) 1, 2, 3 and 4 (the numbering does not coincide with the above order of mention of the databases).

For this trial emulation, NSCLC-specific data are available for EDB1, EDB2 and EDB4.. The fitness-for-purpose of the data for the given trial emulation were assessed and considered for the final selection of the databases.

Reason for selection: All considered databases draw from a comprehensive national sample of patients with cancer in the US with detailed EHR-derived information on the information necessary to study medication effectiveness in oncology.

Strengths of data source(s): Size and detailed clinical information on oncology-specific variables and outcomes (validated composite all-cause mortality sourced from different data sources[7,8]).

Limitations of data source(s): General limitations across all data sources include missing data, potential lack of data continuity, heterogeneous data provenance, quality/heterogeneous ascertainment of mortality endpoint data and the variability in how line of treatment is captured and curated (a more comprehensive discussion of the data sources and approaches is provided in section 7. After a comprehensive assessment of all data sources regarding their fitness for the purpose of emulating the FLAURA trial, EDB2 and EDB4 were found insufficient to be included in the main analysis for the following reasons.

- Rationale for excluding of EDB2 from primary analysis: After applying all I/E criteria, EDB2 results in a cohort with the very small sample size even before matching (Figure 7). Compared to EDB1 and EDB4, there is also an absence of important prognostic covariates which may lead to biased effect estimates due to unmeasured confounding. In addition, EDB2 did not show sufficient balance in measured covariates (Figure 10).
- Rationale for excluding of EDB4 from primary analysis: As shown in Table 3, the patient identification period in EDB4 starts on 10/01/2018 which is after the official approval of osimertinib for 1L metastatic EGFRm+ NSCLC. This means there are limited “control” patients captured in EDB4 from the pre-osimertinib era (see Figure 15). Given the documented clinical benefit and the corresponding rapid uptake of osimertinib after April 2018, it could be assumed that patients who initiated the “old” standard-of-care after April 2018 did so for specific reasons. This means these patients may be likely to be significantly different in their baseline characteristics and prognosis for reasons that may not be measurable.

For these reasons, only EDB1 will be used for the main analysis. However, EDB2 and EDB4 will be considered as part of a sensitivity analysis in which all databases are individually analyzed and treatment effect estimates are pooled using a meta-analytic fixed effects model (see sensitivity analysis #11 in Table 9).

Data source provenance/curation: In brief, all databases provide EHR-derived oncology-specific patient-level information which are either derived directly (e.g., through structured data fields and dropdown menu selections) from EHR and/or undergo semi-automated abstraction processes from unstructured reports. The detailed data provenance, abstraction processes and implemented business rules to curate and prioritize certain variables may vary by database and can be found in legacy publications by the data partners.

Table 3. Metadata about data sources and software.

	EDB1	EDB2	EDB4
Data Source(s):	EHR-derived	EHR-derived	EHR-derived
Study Period:	Patient identification period: 01/01/2011-04/30/2024 with follow-up information through data cut-off date on 04/30/2024	Follow-up information through 02/24/2023 (there is no specific time period restrictions for patient eligibility)	Patient identification period: 10/01/2018-09/30/2023 with follow-up information through data cut-off date on 09/30/2023.
Eligible Cohort Entry Period:	Anytime at start of study drug initiation	Anytime at start of study drug initiation	Anytime at start of study drug initiation
Data Version (or date of last update):	Delivery: Jul 11, 2024	Delivery: Sep 8, 2023 Updated (LoT addition): Mar 11, 2024	Delivery: Oct 24, 2023 Updated (demographics): Feb 29, 2024
Data sampling/extraction criteria:	Patients are sampled if they have a confirmed diagnosis of advanced NSCLC via abstraction on or after 1 Jan 2011, and at least 2 EHR visits on or after 1 Jan 2011. Both ICD-9 (162.x) and ICD-10 (C34x or C39.9) codes are used for the initial selection, and advanced diagnosis are then confirmed via abstraction (since ICD codes do not specify advanced diseases).	Patients are sampled if they were diagnosed with NSCLC and do not meet any of the following exclusion criteria: patient is <18 years of age at the time of diagnosis, does not have the malignancy of interest, is not evaluated at the accessible provider site for the malignancy of interest, has concurrent primary malignancies, has no date of diagnosis in EHR, patient chart has no clinician note available in the EHR, is diagnosed with a malignancy after the diagnosis of the malignancy of interest prior to evaluation for the malignancy of interest at the accessible provider site, is only seen once at the provider site for the malignancy of interest, is initially misdiagnosed and treated for another malignancy, but was later confirmed to have the malignancy of interest, is on therapy, active surveillance, or observation for a malignancy diagnosed prior to the diagnosis of the malignancy of interest at the time of diagnosis of the malignancy of interest, is metastatic (includes leukemias and multiple myeloma) for a malignancy diagnosed prior to the malignancy of interest, with the	NSCLC Patients with an office visit in the reporting period will be included in the report with full patient history. Patients are sampled if they were diagnosed with NSCLC and with a documented visit date, within the defined reporting period, to one of the network facilities and were at least 20 years of age at the time of first diagnosis. Patients who were on a clinical trial at any point in their treatment history are excluded.

		presence of low grade (inclusive of grades 1 and 2) neuroendocrine histology	
Type(s) of data:	EHR-derived	EHR-derived	EHR-derived
Data linkage[1]:	Mortality/date of death is a composite endpoint of structured and unstructured data from the EHR, obituary data, and the social security death index	Mortality/date of death is a composite endpoint of structured and unstructured data from EHRs and commercially available obituary data including the Social Security Administration death master file	Mortality/date of death is a composite endpoint of structured EHR data and linked commercial data from clearinghouse providers hosted by Datavant.
Conversion to CDM[2]:	No	No	No
Software for data management:	R 4.3.2	R 4.3.2	R 4.3.2

[1] Mortality/date of death is a composite endpoint that is often derived from various linked sources including social security death index/ Social Security Administration death master file, obituary data and EHR records
[2] CDM = Common Data Model

6.2. Data management

Data is stored on secure Mass General Brigham corporate provisioned and backed up servers physically located in our Mass General Brigham corporate data centers. Mass General Brigham corporate data centers are designed to insure availability of the affiliated hospitals' and research applications and IT systems in the event of a disaster. The Division follows Mass General Brigham workstation requirements which include: encryption at rest, up-to-date malware protection including antivirus, spyware detection and removal tools, Crowdstrike End Point protection installed, devices enrolled in enterprise Mobile Device Management (MDM) solution as appropriate, any laptop/computer used for business purposes must not be shared with family, friends, or other unauthorized individuals, and compliance with enterprise Password Requirements. Only authorized personnel have read-only access to raw data files.

Cleaned and analysis-ready datasets, i.e., +/- imputed one-row-per-patient tables with all required exposure, outcome and covariate variables, are stored in separate sub-directories dedicated for the specific emulated trial.

6.3. Quality control

Upon delivery, data quality procedures included checks on delivered tables and variables, per table checks, descriptives on most important measures such as demographic and stage distributions by sex at time of initial diagnosis, regimen/exposure frequency counts and time-trends and overall survival benchmarks against literature and general cancer registry statistics. The R code to reproduce the quality assessments is deposited on the Mass General Brigham-provisioned GitLab server https://gitlab.partners.org/drugepi/encore/quality (repository is only accessible within the Mass General Brigham network and additionally only to authorized study personnel).

6.4. Study design

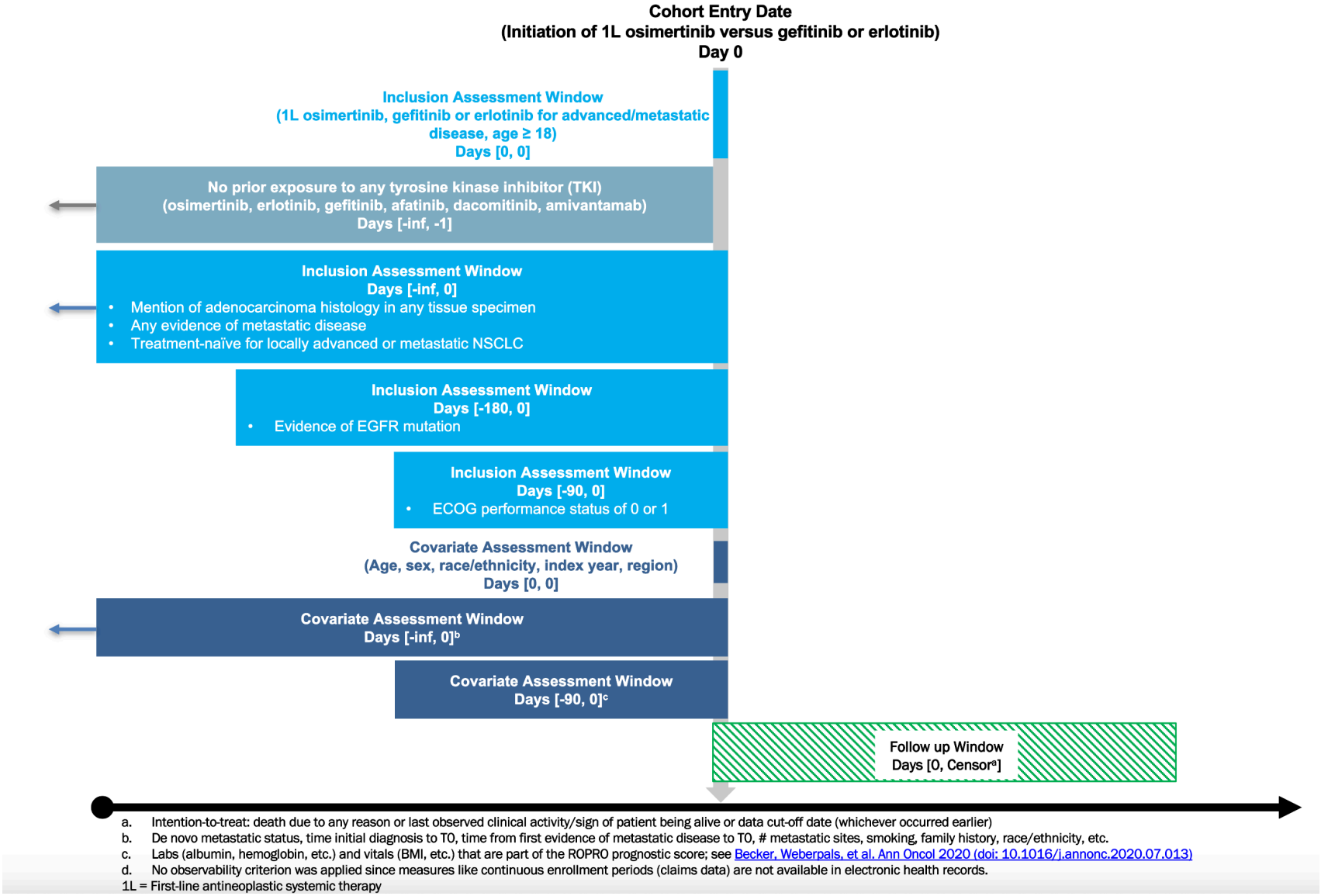
Research design (e.g. cohort, case-control, etc.): Cohort study

Rationale for study design choice: Resembles the principles of the (target) trial emulation framework.[9]

6.5. Study design diagram

Figure 1 depicts study design and variable measurement considerations for the emulation of the FLAURA trial. The selection of key confounders/prognostic factors is driven by expert knowledge and additionally based on covariates included in the real-world prognostic score (ROPRO) which is a published and validated pan-tumor and cancer-specific prognostic score framework for overall survival.[10–14]

Figure 1. Study design illustration for FLAURA trial emulation.



a. Intention-to-treat: death due to any reason or last observed clinical activity/sign of patient being alive or data cut-off date (whichever occurred earlier)
b. De novo metastatic status, time initial diagnosis to T0, time from first evidence of metastatic disease to T0, # metastatic sites, smoking, family history, race/ethnicity, etc.
c. Labs (albumin, hemoglobin, etc.) and vitals (BMI, etc.) that are part of the ROPRO prognostic score; see Becker, Weberpals, et al. Ann Oncol 2020 (doi: 10.1016/j.annonc.2020.07.013)
d. No observability criterion was applied since measures like continuous enrollment periods (claims data) are not available in electronic health records.
1L = First-line antineoplastic systemic therapy

6.6. Setting

6.6.1. Context and rationale for definition of time 0 (and other primary time anchors) for entry to the study population

Time 0 in this database study is defined as the date a patient initiated osimertinib (exposure) or erlotinib or gefitinib (comparator) as part of their 1L systemic antineoplastic treatment for advanced or metastatic NSCLC. This aims to emulate the date of randomization and cohort entry in the RCT (the time from randomization to first dose is not reported in the clinicaltrials.gov study report or the trial articles).

6.6.2. Context and rationale for study inclusion criteria

Study inclusion criteria were defined to emulate all key inclusion criteria for the trial that were deemed both clinically relevant and measurable in EHR data. See Excel appendix table 1 (Table1_I_E) for a one-by-one evaluation. A summary of the operational definitions of the inclusion criteria that were applied for each database can be found in the Excel appendix table 2. A flowchart of the study cohort assembly is provided in Appendix 3.

In the RCT, 95% of patients were metastatic at date of randomization while only the remaining 5% were locally advanced, not amenable to curative surgery or radiotherapy. Since locally advanced line settings are readily curated in EDB1 only, we will restrict the analysis in EDB2 and EDB4 to only patients with any evidence of metastases prior to or on index date (more details see below in 6.4.4).

6.6.3. Context and rationale for study exclusion criteria

Study exclusion criteria were defined to emulate all key exclusion criteria for the trial that were deemed both clinically relevant and measurable in EHR data. See Excel appendix table 1 (Table1_I_E) for a one-by-one evaluation. A summary of the operational definitions of the exclusion criteria that were applied for each database can be found in the Excel appendix table 2. A flowchart of the study cohort assembly is provided in Appendix 3.

6.6.4. Context and rationale for exposure(s) of interest

The exposure and comparator were defined to emulate the agents compared for the trial, i.e., initiation of osimertinib versus erlotinib or gefitinib in a 1L metastatic setting.

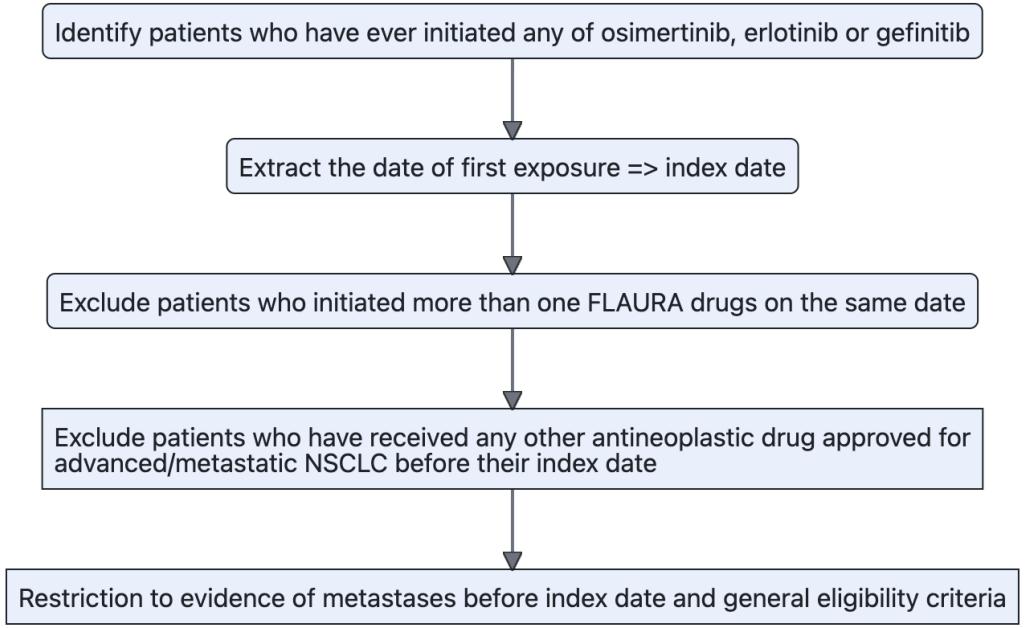
- **EDB1**: Exposure is derived using a manually curated line of therapy (LOT) table provided by the data partner that programmatically categorizes treatment regimens into a coherent line of treatment. That is, each patient is represented with one row per curated line of therapy with corresponding information on line number, regimens as well start and end dates. Based on this table, patients are identified who received osimertinib, erlotinib or gefitinib treatment regimen by their generic names (string match) in 1L. The LOT implicitly only considers regimens that were given as part of a metastatic disease setting. More details and annotated code to identify initiators can be found in the 'Derive cohort EDB1' Quarto report (access within MGB network only).
- **EDB2**: Exposure is derived using a manually curated line of therapy (LOT) table provided by the data partner that programmatically categorizes treatment regimens into a coherent line of treatment. That is, each patient is represented with one row per curated line of therapy with corresponding information on line number, regimens as well start and end dates. Based on this table, patients are identified who received osimertinib, erlotinib or gefitinib treatment

regimen by their generic names (string match) in 1L. The LOT implicitly only considers regimens that were given as part of a metastatic disease setting. More details and annotated code to identify initiators can be found in the 'Derive cohort EDB2' Quarto report' (access within MGB network only).

- **EDB4:** For the EDB4 database, the following logic is applied (see Figure 2). First, patients are identified who have ever initiated any of osimertinib, erlotinib or gefinitib. If patients initiated more than one of these three regimens simultaneously, these patients are not considered and will be excluded. To guarantee that the osimertinib, erlotinib or gefinitib are initiated in a 1L metastatic setting, patients are only considered if there is any evidence of a metastatic diagnosis before the index date. Additionally, if patients received any other antineoplastic drug before the initiation of any of the study drugs that is typically used for the 1L metastatic setting according to NCCN guidelines or expert knowledge, these patients will not be considered either (see list below*). This may decrease sensitivity as some of these drugs may be also used for the (neo)adjuvant setting but increase specificity in the selection of initiators in the 1L metastatic setting. More details and annotated code to identify initiators can be found in the 'Derive cohort EDB4' Quarto report (access within MGB network only).

  ***Antineoplastic drugs considered:** afatinib, alectinib, amivantamab, atezolizumab, bevacizumab, brigatinib, cabozantinib, capmatinib, carboplatin, cemiplimab, ceritinib, cisplatin, cobimetinib, crizotinib, dabrafenib, dacomitinib, docetaxel, durvalumab, encorafenib, erlotinib, etoposide, fam-trastuzumab deruxtecan-nxki, gefitinib, gemcitabine, ipilimumab, larotrectinib, lorlatinib, mobocertinib, necitumumab, nivolumab, osimertinib, paclitaxel, pembrolizumab, pemetrexed, pralsetinib, ramucirumab, selpercatinib, tepotinib, trametinib, trastuzumab, trastuzumab, trastuzumab-anns, trastuzumab-deruxtecan, trastuzumab-deruxtecan, trastuzumab-dkst, tremelimumab, vemurafenib, vinorelbine

Figure 2. Identification of study drug initiators in EDB4



6.6.5. Context and rationale for outcome(s) of interest

The primary outcome for the database study was defined to emulate the primary outcome for the trial, time from index to death due to any reason (overall survival). Operational definitions:

- EDB1: Time in {days, months and years} from index date to death due to any reason. The date of death is de-identified to month-level granularity or (rarely) to year-level granularity and the date of death is therefore imputed to the 15th of a month or mid-year/July 2 of the year of death, respectively. In If there is no indication that a patient died during the study period, the patient is censored. The censoring date is defined as the last visit or treatment encounter or data cut-off date, whichever occurred earlier. The overall survival endpoint is operationalized using a parameterized R function *edb1_get_os()* and more details can be found in the attached pdf documentation.
- EDB2: Time in {days, months and years} from index date to death due to any reason. If there is no indication that a patient died during the study period, the patient is censored. The censoring date is defined as the last observed activity date or data cut-off date, whichever occurred earlier. Activity dates are defined as documented in Table 4. Dates used to derive time to all-cause mortality may have some associated imprecision such that the date of death is either known completely, the year and month is known or only the year is known. The overall survival endpoint is operationalized using a parameterized R function *edb2_get_os()* and more details can be found in the attached pdf documentation.

Table 4. Relevant clinical activities considered to derive last activity date for censoring.

Table / clinical activity considered	Dates considered
Adverse events	Event date
Therapy (cellular, systemic, radiation, surgery)	Start and end dates or declined intervention date, surgery date, assessed resection dates
Palliative care referral	Referral date
Visits	Contact/visit date
Vitals	Assessed date
Labs	Lab date
Biomarkers	Specimen collection date
Patient observation period	Start and end dates
Demographics	Date of most recent contact with provider, date patient was diagnosed with a second primary malignancy
Performance assessments	Documented date, reported date

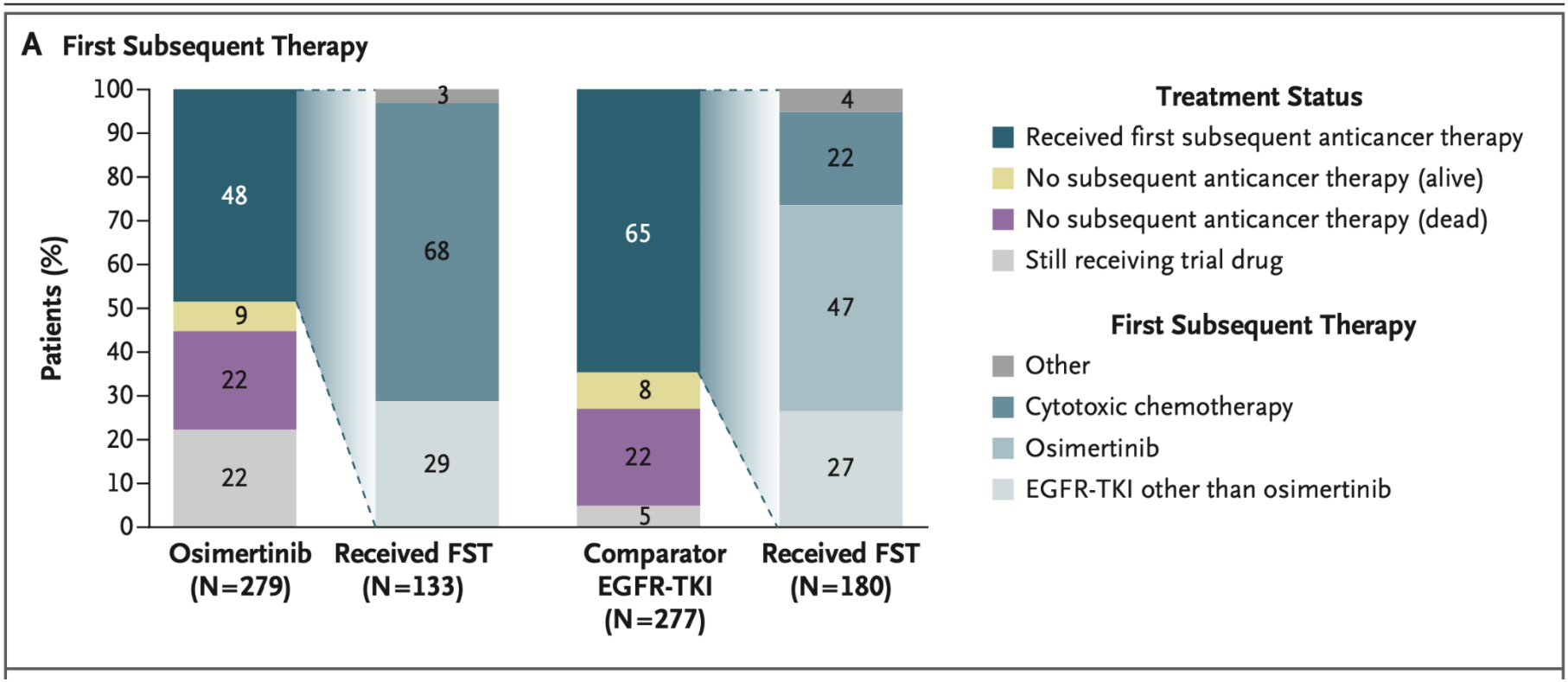
Secondary diagnoses	Diagnosis date
Progression, histology, lymphovascular invasion, metastatic sites, pancoast tumor, perineural invasion	Assessed date
Stage/TNM	Assessed date
Smoking	Assessed date

- **EDB4**: Time in {days, months and years} from index date to death due to any reason. The date of death is de-identified to month-level granularity and the day of death is therefore imputed to the 15th of a month. If there is no indication that a patient died during the study period, the patient is censored. The censoring date is defined as the last date of vital signs recorded as proof that the patient was alive at that time (de-identified to week-level granularity) or data cut-off date, whichever occurred earlier. The overall survival endpoint is operationalized using a parameterized R function *edb4_get_os()* and more details can be found in the attached pdf documentation.

6.6.6. Context and rationale for follow up

Only intention-to-treat (ITT) analyses will be conducted. Although cross-over from the exposure to the comparator can be expected which biases the exposure treatment effect more towards the null, this also applies to the RCT. According to Ramalingam et al.[6], 65% of patients allocated to the standard of care arm received a subsequent antineoplastic therapy following erlotinib or gefinitib discontinuation (Figure 3). Out of these 65%, almost half (47%) crossed over to receive osimertinib. An as-treated analysis is not considered since in the context of oncology, reasons for discontinuation usually are due to toxicity, death or progression/non-response to the current treatment, all of which are highly correlated with the outcome under study which would hence lead to bias due to informative censoring.

Figure 3. Illustration of proportion of patients who received a first subsequent therapy after discontinuation of their assigned treatment. Taken from N Engl J Med 2020;382:41-50[6]



6.6.7. Context and rationale for covariates

We identified a series of covariates that are strong prognostic factors for the outcome and auxiliary covariates which may be useful to impute missing data. Such covariates comprise demographics, covariates indicating disease-severity, cancer-specific covariates as well as pathological and genetic factors. In addition, selected labs and vitals are considered since they were shown to carry a high amount of prognostic information as described in Becker, Weberpals, et al.[10] For these variables, additional plausibility checks and transformations are carried out. In detail, labs and vitals are individually checked if they cross a certain biologically implausible threshold (e.g., a heart rate of 0) in which cases the values are set missing and imputed in a next step. These thresholds were compiled by experienced practicing physicians and medical oncologists and are listed in appendix **Table 10** and **Table 11**.

Note that not all covariates are available across all databases used for this trial emulation. In the analytical stage, the most comprehensive model will be fit for each database individually.

Table 5. Operational definitions of key covariates used for trial emulation.

Characteristic	Harmonized analysis variable name	R function to derive covariate (see pdf in appendix)	Details[1]	Variable encoding	Assessment window
Age at initial diagnosis	dem_age_initial_diagnosis	edbx_get_demographics()	Age measured at initial diagnosis of eligible primary tumor	Nominal (<60, 60-69, 70-79, 80+)	[-inf;0] at initial diagnosis of primary cancer
Age at index date	dem_age_index	edbx_get_demographics()	Age measured at index date	Binary (<60, 65+); modelled continuously in ROPRO[10]	[0;0]
Sex	dem_sex	edbx_get_demographics()	Sex	Binary (Male, Female)	[0;0]
Year of index date	c_year_index	De novo derived from dt_index	Calendar year in which patient initiated study treatment	Nominal (<2018, 2018+)	[0;0]
Family history	dem_family_history	edbx_get_demographics()	Family history of cancer	Logical (TRUE, FALSE)	[0;0] (no specific date is associated)
Race	dem_race	edbx_get_demographics()	Race categorized as in the original RCT	Binary (Asian vs non-Asian)	[0;0]
Ethnicity	dem_ethnicity	edbx_get_demographics()	Ethnicity	Hispanic, Non-Hispanic	[0;0]
Region	dem_region	edbx_get_demographics()	US region patient receives care in; if given on a state level, region is manually mapped (see **Table 12**)	Nominal (Northeast, South, West, Midwest)	[0;0]
Practice type	dem_practice	edbx_get_demographics()	Setting patient is receiving care at	Nominal (academic, community, academic & community)	[-inf;0]
Socio-economic status	dem_ses	edbx_get_demographics()	Socioeconomic status (SES) index based on residence area of patient	Nominal (from '1 - Lowest SES' through '5 - Highest SES)	[-inf;0]
Smoking	c_smoking_history	edbx_get_demographics()	History of current or former (= TRUE) or never (= FALSE) smoking on or anytime before index date; if there are multiple records per patient, any evidence of former/current smoking is prioritized	Binary logical (TRUE, FALSE)	[-inf;0]
EGFR status	C_egfr_biomarker_status	edbx_get_biomarker()	Evidence of <u>any</u> EGFR mutation present. If patient has multiple measurements, any evidence of a mutation is prioritized. In case of ties, the closest measurement relative to index date is selected	Binary (negative, positive)	[-180;0]

Characteristic	Harmonized analysis variable name	R function to derive covariate (see pdf in appendix)	Details[1]	Variable encoding	Assessment window
ECOG	c_ecog	edbx_get_ecog()	ECOG performance status measured closest to index date within assessment window. In case of ties, the lower ECOG value is selected	Nominal (0, 1, 2, 3, 4); modelled as ordinal numeric in ROPRO[10]; due to I/E criteria ECOG is modelled as a binary (0, 1) covariate	[-90;0]
Stage	c_stage_initial_dx	edbx_get_diagnosis_solid()	AJCC summary group stage at initial diagnosis	Ordinal numeric (from 0 to IV with sub-categories, e.g., IA1)[10]	[-inf;0] at initial diagnosis of primary cancer
De novo metastatic status	c_de_novo_mets_dx	edbx_get_diagnosis_solid()	Evidence of presence of one or multiple metastases at/before initial diagnosis	Binary logical (TRUE, FALSE)	[-inf;0] at initial diagnosis of primary cancer
Evidence of metastases between initial diagnosis and index date	c_met_pre_index	edbx_get_diagnosis_solid()	Evidence of any metastasis between initial diagnosis and index date (includes the initial diagnosis and index date)	Binary logical (TRUE, FALSE)	[-initial dx;0]
Number of metastatic sites	c_number_met_sites	edbx_get_diagnosis_solid()	Number of metastatic sites for a given patient before/on index date	Integer	[-inf;0]
Time between initial diagnosis to index date	c_time_dx_to_index	edbx_get_diagnosis_solid()	Time in days between initial diagnosis to index date	Continuous	[-initial dx;0]
Time between earliest evidence of a metastatic and index date	c_time_met_dx_to_index	edbx_get_diagnosis_solid()	Time in days between earliest evidence of a metastatic and index date	Continuous	[-inf;0]
Histology (adenocarcinoma)	c_histology_match	edbx_get_histology()	Evidence of adenocarcinoma histology (non-squamous cell for EDB1)	Binary logical (TRUE, FALSE)	[-inf;0]
Albumin	c_albumin_g_l_cont	edbx_get_labs()	Closest albumin measurement (in serum/plasma) relative to index date in g/L. In case of ties, the lower is selected	continuous	[-90;0]
Alkaline phosphatase (ALP) [2]	c_alp_u_l_cont	edbx_get_labs()	Closest alkaline phosphatase measurement (in serum/plasma) relative to index date in U/L. In case of ties, the lower is selected	continuous	[-90;0]
Alanine aminotransferase (ALT) [2]	c_alt_u_l_cont	edbx_get_labs()	Closest alanine transaminase measurement (in serum/plasma) relative to index date in U/L. In case of ties, the lower is selected	continuous	[-90;0]

Characteristic	Harmonized analysis variable name	R function to derive covariate (see pdf in appendix)	Details[1]	Variable encoding	Assessment window
Aspartate aminotransferase (AST)	c_ast_u_l_cont	edbx_get_labs()	Closest aspartate aminotransferase measurement (in serum/plasma) relative to index date in U/L. In case of ties, the lower is selected. Used to compute AST-ALT ratio	continuous	[-90;0]
AST/ALT ratio	c_ast_alt_ratio_cont	edbx_get_labs()	AST/ALT ratio calculated from c_ast_u_l_cont/c_alt_u_l_cont	continuous	[-90;0]
Bilirubin[2]	c_bilirubin_mg_dl_cont	edbx_get_labs()	Closest bilirubin measurement (in serum/plasma) relative to index date in mg/dL. In case of ties, the lower is selected	continuous	[-90;0]
Calcium[2]	c_calcium_mg_dl_cont	edbx_get_labs()	Closest calcium measurement (in serum/plasma) relative to index date in mg/dL. In case of ties, the lower is selected	continuous	[-90;0]
Chloride	c_chloride_mmol_l_cont	edbx_get_labs()	Closest chloride measurement (in serum/plasma) relative to index date in mmol/L. In case of ties, the lower is selected	continuous	[-90;0]
Eosinophils/100 leukocytes[2]	c_eosinophils_leukocytes_ratio_cont	edbx_get_labs()	Eosinophils/100 leukocytes in blood. In case of ties, the lower	continuous	[-90;0]
Glucose[2]	c_glucose_mg_dl_cont	edbx_get_labs()	Closest glucose measurement (in serum/plasma) relative to index date in mmol/L. In case of ties, the lower is selected	continuous	[-90;0]
Hemoglobin	c_hemoglobin_g_dl_cont	edbx_get_labs()	Closest hemoglobin measurement (in blood) relative to index date in g/L. In case of ties, the lower is selected	continuous	[-90;0]
Lactate dehydrogenase (LDH)[3]	c_ldh_u_l_cont	edbx_get_labs()	Closest LDH measurement (in serum or plasma) relative to index date in U/L. In case of ties, the lower is selected	continuous	[-90;0]
Lymphocytes	c_lymphocyte_10_9_l_cont	edbx_get_labs()	Closest lymphocytes measurement (in blood) relative to index date in 10^9/L. In case of ties, the lower is selected. Used to compute neutrophil/lymphocyte ratio	continuous	[-90;0]

Characteristic	Harmonized analysis variable name	R function to derive covariate (see pdf in appendix)	Details[1]	Variable encoding	Assessment window
Lymphocyte/leukocyte ratio[2]	c_lymphocyte_leukocyte_ratio_cont	edbx_get_labs()	Closest lymphocyte/leukocyte ratio measurement (in blood) relative to index date. In case of ties, the lower is selected. Used to compute neutrophil/lymphocyte ratio	continuous	[-90;0]
Monocytes[2]	c_monocytes_10_9_l_cont	edbx_get_labs()	Closest monocytes measurement (in blood) relative to index date in 10^9/L. In case of ties, the lower is selected.	continuous	[-90;0]
Neutrophils	c_neutrophil_10_9_l_cont	edbx_get_labs()	Closest neutrophils measurement (in blood) relative to index date in 10^9/L. In case of ties, the lower is selected. Used to compute neutrophil/lymphocyte (NLR) ratio	continuous	[-90;0]
Neutrophil/lymphocyte ratio[2]	c_neutrophil_lymphocyte_ratio_cont	edbx_get_labs()	Neutrophil/lymphocyte (NLR) ratio calculated from c_neutrophil_10_9_l_cont/ c_lymphocyte_10_9_l_cont	continuous	[-90;0]
Platelets	c_platelets_10_9_l_cont	edbx_get_labs()	Closest platelets measurement (in blood) relative to index date in 10^9/L. In case of ties, the lower is selected	continuous	[-90;0]
Protein	c_protein_g_l_cont	edbx_get_labs()	Closest protein measurement (in serum/plasma) relative to index date in g/L. In case of ties, the lower is selected	continuous	[-90;0]
Urea nitrogen[2]	c_urea_nitrogen_mg_dl_cont	edbx_get_labs()	Closest urea nitrogen measurement (in serum/plasma) relative to index date in mg/L. In case of ties, the lower is selected	continuous	[-90;0]
Systolic blood pressure[2]	c_sbp_cont	edbx_get_vitals()	Closest systolic blood pressure (in mmHg) measurement. In case of ties, the lower is selected	continuous	[-90;0]
Body mass index (BMI)[2]	c_bmi_cont	edbx_get_vitals()	Closest BMI measurement (in kg/m^2) relative to index date. In case of ties, the lower is selected. For ROPRO and in EDB2, BMI is (additionally) computed from individual height and weight measurements	continuous	[-90;0]

Characteristic	Harmonized analysis variable name	R function to derive covariate (see pdf in appendix)	Details[1]	Variable encoding	Assessment window
Heart rate[2]	c_hr_cont	edbx_get_vitals()	Closest heart rate measurement (in bpm) relative to index date. In case of ties, the lower is selected	continuous	[-90;0]
Oxygen saturation	c_oxygen_cont	edbx_get_vitals()	Closest heart rate measurement (in bpm) relative to index date. In case of ties, the lower is selected	continuous	[-90;0]

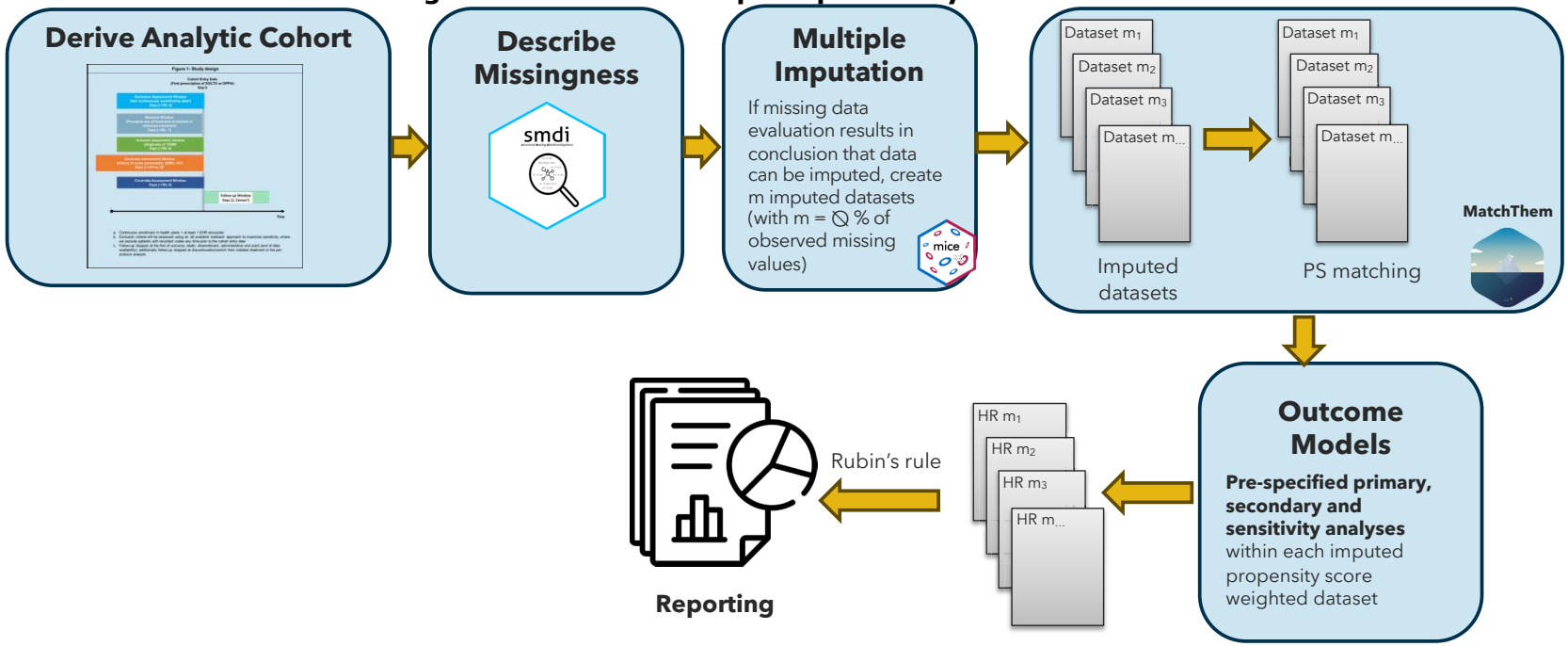
[1] x stands for the pseudonymized number of the respective database, i.e., EDB1, EDB2 or EDB4
[2] For calculation of ROPRO prognostic score[10], this variable is log transformed.
[3] For calculation of ROPRO prognostic score[10], this variable is log-log transformed.

6.7. Data analysis

6.7.1. Context and rationale for analysis plan

To emulate the FLAURA trial, the following analytical workflow will be used (Figure 4). First, an analytical cohort with covariates on key eligibility criteria and prognostic factors will be derived across all databases. To ensure reproducibility and consistency throughout the entire ENCORE project, the internally developed *encore.io* R package streamlines this process using the functions referenced in Table 5. Operational definitions of key covariates used for trial emulation.Table 5 (code documentation see appendix).[15] The analytical cohort will be derived by first identifying an advanced NSCLC inception cohort of initiators of osimertinib, erlotinib or gefinitib in the first line setting as described in section 6.6.4. Next, key eligibility criteria will be applied in which patients with missing values are considered eligible in the respective attrition steps to allow thorough missing data investigations.

Figure 4. Illustration of principled analytical workflow.



Once a full analytic cohort is built, principled missing data investigations will be employed to empirically assess assumptions on potentially underlying missingness mechanisms according to Rubin's classification of missing data (i.e., missing completely at random [MCAR], missing at random [MAR] and missing not at random [MNAR]).[16] To that end, we will adopt a principled process on missing data that was developed as part of a FDA Sentinel Innovation Center causal inference workstream that empirically evaluates different aspects across partially observed covariates based on three group diagnostics (Table 6).[17,18] In brief, the first group diagnostics computes distributions and absolute standardized mean differences (ASMD) between patients with and without an observed value for a given partially observed covariates. If missingness can be explained by observed covariates such as in MAR mechanisms, patient characteristics will significantly differ which will (in analogy to propensity scores) be indicated by ASMDs > 0.1. In addition, Hoteling's[19] and Little's[20] tests additionally provide formal hypothesis tests for such comparisons in which high test statistics and a rejection of the null hypothesis would provide evidence for differences in the distribution of patient characteristics and suggest the underlying mechanism is not MCAR or MNAR. Group 2 diagnostics assess the ability to predict missingness based on observed covariates by fitting a classification model to predict the missingness indicator of the partially observed covariate. To that end, we will fit a random forest (RF) classification model using observed covariates with a 70/30 train-test split of the complete cohort. A sufficiently high area under the receiver operating characteristic curve (AUC) metric of the test dataset may demonstrate that missingness can be predicted well and could point towards MAR as a likely mechanism as opposed to an AUC~0.5 which would suggest MCAR or MNAR. Group 3 diagnostics evaluates the association between the missingness indicator of the partially observed covariates and the outcome (OS). If the missingness of a confounder cannot be explained or approximated by observed covariates and a difference in the outcome is observed depending on the missingness indicator (e.g., $HR_{\text{missingness indicator}} \neq 1$), this may be indicative of an underlying MNAR mechanism. These empirical diagnostics will be implemented through the smdi R package[21] and be further enhanced by clinical expert knowledge.

Table 6. Diagnostics to empirically differentiate and characterize missing data mechanisms.

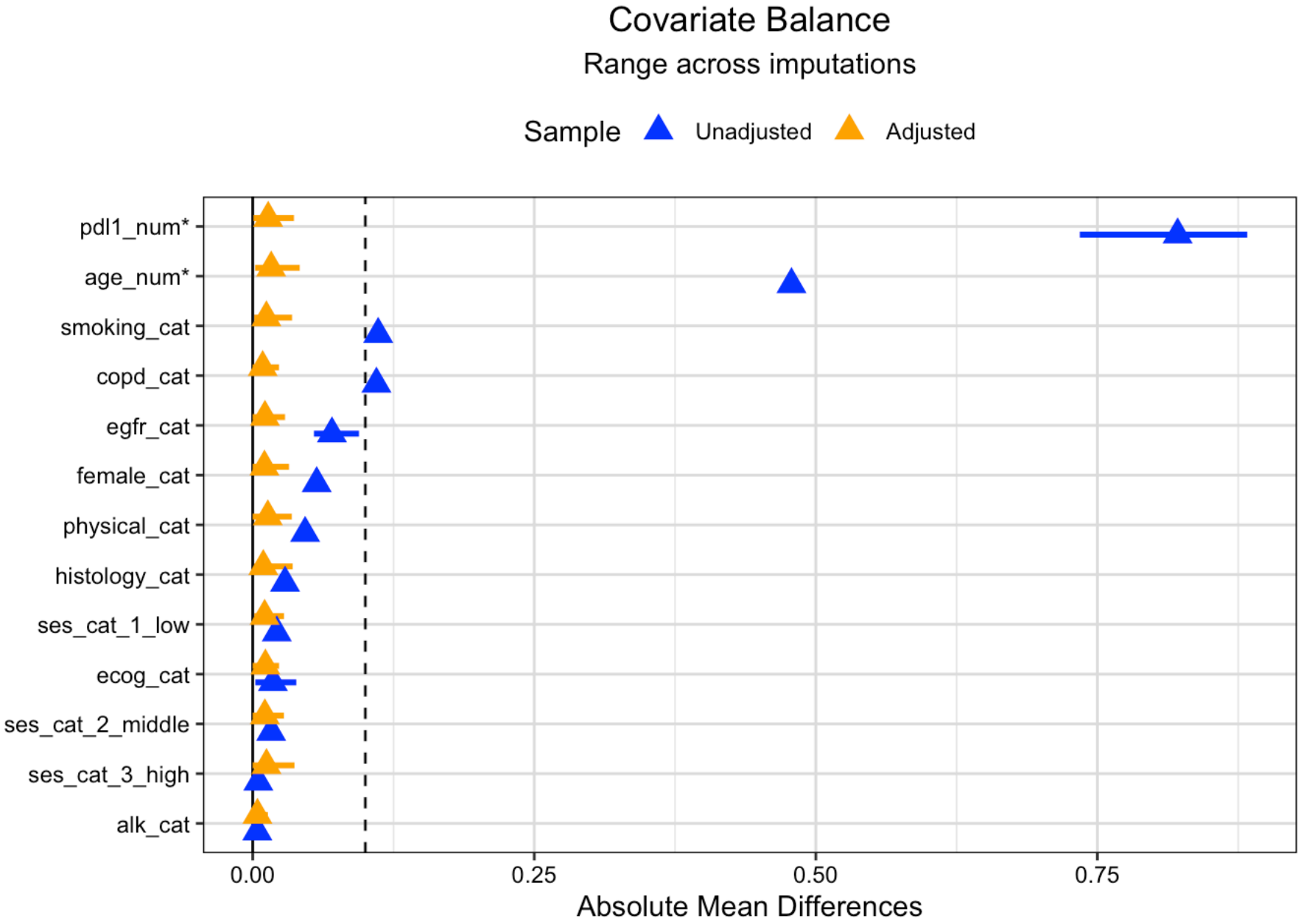
	Group 1 Diagnostics		Group 2 Diagnostics	Group 3 Diagnostics
Diagnostic metric	**Absolute standardized mean difference (ASMD)**	**P-value Hoteling[19]/ Little[20]**	**Area under the receiver operating curve (AUC)**	**Log HR (missingness indicator)**
Purpose	Comparison of distributions between patients with vs. without observed value of the partially observed covariate.		Assessing the ability to predict missingness based on observed covariates.	Check whether missingness of a covariate is associated with the outcome (differential missingness).
Example value	ASMD = 0.1	p-value < 0.001	AUC = 0.5	log HR = 0.1 (0.05 to 0.2)
Interpretation	≤0.1[a]: no imbalances in observed patient characteristics; missingness may be likely completely at random or not at random (~MCAR, ~MNAR). ≥0.1[a]: imbalances in observed patient characteristics; missingness may be likely at random (~MAR).	High test statistics and low p-values indicate differences in baseline covariate distributions and null hypothesis would be rejected (~MAR).	AUC values ~ 0.5 indicate completely random or not at random prediction (~MCAR, ~MNAR). Values meaningfully above 0.5 indicate stronger relationships between covariates and missingness (~MAR).	No association in either univariate or adjusted model and no meaningful difference in the log HR after full adjustment (~MCAR). Association in univariate but not fully adjusted model (~MAR). Meaningful difference in the log HR also after full adjustment (~MNAR).

Abbreviations: ASMD = Median absolute standardized mean difference across all covariates, AUC = Area under the curve, CI = Confidence interval, MAR = Missing at random mechanism in which the missingness probability depends on observed covariates, MCAR = Missing completely at random mechanism in which each patients has the same missingness probability, MNAR(unmeasured) = Missing not at random mechanism in which the missingness can only be explained by a covariate which is not observed in the underlying dataset, MNAR(value) = Missing not at random mechanism in which the missingness just depends on the actual value of the partially observed confounder of interest itself.

[a] Analogous to propensity score-based balance measures.[22a] Analogous to propensity score-based balance measures.[22]

While the MAR assumption is a strong assumption to hold across all considered covariates, it was shown that especially in the context of partially observed covariate data (as opposed to missing exposure and outcome data), only mechanisms in which a covariate causes its own missingness leads to critical bias (MNAR).[23] In such situations, multiple imputation can have significant advantages over a complete case analysis since additional information (auxiliary covariates and missing indicator variables) can be included in imputation algorithms which can make the MAR assumption holding more plausible and increase efficiency in treatment effect estimates since all patients and critical covariates can be retained and variances can be realistically estimated, considering both the general sampling error and the error introduced by missing data.[24,25] Hence, multiple imputation with flexible, non-parametric random forest imputation algorithms[26] (mice R package[27]) will be used for this trial emulation. The number of imputed datasets (*m*) will be determined for each database separately based on the average proportion missingness observed in the analytic cohort and results from the above-referenced missing data investigations will inform the choice of appropriate sensitivity analyses.

Figure 5. Covariate balance across imputed datasets (simulated example).



To estimate the treatment effects for osimertinib using propensity score matching across imputed datasets we will apply the "within" approach using the "MatchThem" R package.[28,29] That is, propensity score matching and the estimation of the treatment effect are performed in each imputed dataset separately and resulting treatment effect estimates are combined using Rubin's rule. In this study, this will be implemented by matching eligible patients on their propensity to initiate osimertinib using a 1:1 nearest neighbor matching algorithm without replacement and a caliper of 1% of the standard deviation of the propensity score. The resulting covariates balance will be assessed by computing and visualizing ASMDs before and after matching across datasets. As compared to a single dataset matching approach, this can lead to a range of ASMDs per covariate due to random variation across imputed datasets for which an example (using simulated data) is illustrated in Figure 5. If sufficient balance can be established, a Cox proportional hazards regression model will be fit to estimate the marginal average treatment effect in the matched population. Since in most databases there are more osimertinib patients than patients in the control arm, the estimand will rather correspond to the ATC than the ATT. Confidence intervals will be estimated using cluster-robust standard errors.[30] As a secondary endpoint, we will additionally estimate the median OS survival time difference between the two exposure groups using the Kaplan-Meier method. It should be noted that due to administrative and de-identification purposes, the date of death is often only available at the month- or year-granularity level, in which case the date of death will be imputed to the 15th of a month or July 2nd of a year, respectively (depending on the database). In rare cases, this can lead to negative/implausible follow-up times if the date of death is very close to the index date. These patients will be excluded from the analysis.

The final hazard ratio and median OS survival time difference estimates for each database will then be combined using Rubin's rule.[27,31] To arrive at a single estimate across databases, the final estimates will be summarized through a meta-analytic fixed effects model.[32] A summary of the analytic approach is summarized in Table 7 and an example workflow with simulated data and annotated code can be found here.

A. Primary analysis

Table 7. Primary and subgroup analysis specification

Hypothesis:	Initiation of osimertinib decreases the hazard of all-cause mortality as compared to initiation of erlotinib or gefitinib
Exposure contrast:	Initiation of osimertinib vs. initiation of erlotinib or gefitinib
Outcome:	Time to all-cause mortality (overall survival)
Databases used:	EDB1
Analytic software:	R 4.3.2. Version control of code and R packages will be established through git and Posit package manager, respectively. All packages are frozen to their most recent version as of April 24, 2024.
Model(s): *(provide details or code)*	See example code here. The annotated code for the trial emulation will be hosted at https://drugepi.gitlab-pages.partners.org/encore/flaura-nct-02296125/ (access only through MGB network for authorized personnel)
Confounding adjustment method	***Name method and provide relevant details, e.g. bivariate, multivariable, propensity score matching (specify matching algorithm ratio and caliper), propensity score weighting (specify weight formula, trimming, truncation), propensity score stratification (specify strata definition), other.***
	1:1 propensity score nearest neighbor matching without replacement and a caliper of 1% of propensity score standard deviation
Missing data methods	***Name method and provide relevant details, e.g. missing indicators, complete case, last value carried forward, multiple imputation (specify model/variables), other.***
	Multiple imputation by chained equations using a random forest imputation model across all covariate types. The number of imputed datasets will be determined by the average proportion of missing values across all partially observed covariates. Imputation models will include all variables of the substantiative model, i.e., exposure, outcome, confounders/prognostic factors and additional auxiliary covariates.
Subgroup Analyses	***List all subgroups***
	In subgroup analysis, multiple imputation, propensity score matching and balance assessment will be conducted within each subgroup separately. The treatment effect will be estimated in the combined dataset by fitting a Cox proportional-hazards model that will include the exposure, the subgroup covariate of interest, and the treatment-by-subgroup interaction term. 1. Sex (Male, Female) 2. Age (<65, ≥65)* 3. Race/ethnicity (Asian, Non-Asian)* 4. Smoking history (current/former, never) 5. ECOG (0, 1)* * indicates that a statistically significant differences were found in FLAURA

B. Secondary Analysis

Table 8. Secondary analysis specification.

Hypothesis:	Initiation of osimertinib increases overall survival as compared to initiation of erlotinib or gefitinib
Exposure contrast:	Initiation of osimertinib vs. initiation of erlotinib or gefitinib
Outcome:	Median overall survival time, i.e., time until 50% of the patients in each stratum deceased
Databases used:	EDB1
Analytic software:	R 4.3.2. Version control of code and R packages will be established through git and Posit package manager, respectively. All packages are frozen to their most recent version as of April 24, 2024.
Model(s): *(provide details or code)*	See example code here. The annotated code for the trial emulation will be hosted at https://drugepi.gitlab-pages.partners.org/encore/flaura-nct-02296125/ (access only through MGB network for authorized personnel)
Confounding adjustment method	***Name method and provide relevant details, e.g. bivariate, multivariable, propensity score matching (specify matching algorithm ratio and caliper), propensity score weighting (specify weight formula, trimming, truncation), propensity score stratification (specify strata definition), other.***
	1:1 propensity score nearest neighbor matching without replacement and a caliper of 1% of propensity score standard deviation
Missing data methods	***Name method and provide relevant details, e.g. missing indicators, complete case, last value carried forward, multiple imputation (specify model/variables), other.***
	Multiple imputation by chained equations using a random forest imputation model across all covariate types. The number of imputed datasets will be determined by the average proportion of missing values across all partially observed covariates. Imputation models will include all variables of the substantiative model, i.e., exposure, outcome, confounders/prognostic factors and additional auxiliary covariates.
Subgroup Analyses	***List all subgroups***
	None; no subgroup-specific median overall survival estimates are reported for FLAURA.

Table 9. Sensitivity analyses – rationale, strengths and limitations.

	What is being varied? How?	Why? (What do you expect to learn?)	Strengths of the sensitivity analysis compared to the primary	Limitations of the sensitivity analysis compared to the primary
Sensitivity #1	Caliper matching on ROPRO prognostic score instead of propensity score	Matching patients on validated prognostic score may be more beneficial to control for (unmeasured) confounding	Matches patients on validated prognostic score that incorporates weights of key prognostic factors	Limited experience on how to optimally use prognostic scores and should be seen as an experimental sensitivity analysis
Sensitivity #2	ATO weighting instead of matching	Weights that resemble the average treatment effect in the overlap population (ATO) create a clinical equipoise population which is comparable to an RCT	ATO weighting usually results in excellent balance and clinical equipoise	Estimates the average treatment effect among the overlap patients which may not be comparable to target population anymore
Sensitivity #3	SMR/ATT weighting instead of matching. Here symmetric trimming (i.e., setting all weights lower/higher than that at a given quantile to the weight at the quantile) of extreme weights may be considered with the quantiles chosen based on weight distribution and resulting balancing performance.	SMR weighting retains all patients and resembles the same estimand as matching	ATT weighting retains all patients	Patients with extreme weights after trimming may bias the analysis
Sensitivity #4	Censoring date is changed to 3 months before data cut-off date	For all databases, information on mortality comes from different data sources which are updated asynchronously. To account for the potential lag of updated mortality information (*ghost-time*	Approach implements a more conservative censoring rule	Approach addresses ghost-time bias by censoring patients without a recorded

		bias[33]), the censoring date for patients without mortality event in the whole patient identification period will be moved to last sign of patients being alive/visit or 3 months before data cut-off date, whichever occurred earlier.[34]		death event earlier
Sensitivity #5	Delta imputation models for MNAR (tipping point analysis)	Primary multiple imputation analysis assumes MAR which may not hold for every covariate	Estimates impact of deviations from MAR assumption on final treatment effect estimates for key covariates	Delta parameters must be assumed and results are complex to interpret in multivariate missingness settings; just must important covariates or those with highest suspicion of being MNAR will be evaluated
Sensitivity #6	Restrict to patients with Exon19 deletion only (databases for which this information is available)	FLAURA only included patients with sensitizing mutation (Exon19 deletion) but this information is not consistently provided across databases which is why in primary analyses we include all patients with any EGFR mutation assuming that it is sensitizing given that these patients initiated osimertinib, gefitinib or erlotinib	Target population is genetically more identical and increases specificity of biomarker measurement	Information not consistently available
Sensitivity #7	Re-weighting of strong risk factors and/or treatment effect modifiers distribution to match that of FLAURA	In the presence of effect modification, treatment effect estimates may be different if	Re-weighting adjusts for differences in distributions of key risk factors and/or	Re-weighting risk factors/potential

		the distribution of strong risk factors/effect modifiers is different in the emulated cohort versus the trial cohort	treatment effect modifiers (see subgroup analysis in Table 7)	effect modifiers to match the FLAURA trial and simultaneously balancing them across treatment groups may be challenging due to differences in measurement
Sensitivity #8	Including patients who have had at least 1 visit 90 days prior to treatment initiation	EHR are often lacking data continuity, and this analysis uses the requirement of 1 visit as a proxy for continuous observation periods	Considers aspect of data continuity	There may be patients who are put on treatment immediately in which case they are falsely excluded
Sensitivity #9	Exclusion of patients with osimertinib exposure before April 18, 2018	There are a few patients who are observed to have received osimertinib before its official approval (potentially due to off-label use in 1L, compassionate use, etc.)	Limits chances of exposure misclassification	May exclude truly eligible patients
Sensitivity #10	Restriction of eligible calendar time period from 2015 to 2022	Descriptive analyses indicate that there was a rapid uptake of new initiators of osimertinib immediately after approval in 2018 and a sudden decrease of SoC TKIs. This means there is hardly any overlap of patients pre/post 2018 and calendar cannot be sufficiently balanced	This sensitivity aims to mitigate the potential effects by imbalance in calendar time	Calendar time will overall remain imbalanced
Sensitivity #11	Addition of EDB2 and EDB4	EDB2 and EDB4 are not considered in primary analyses for reasons given in section 6.1.1.	Increased sample size and potentially broader coverage of general US cancer population	See limitations listed in section 7.

Sensitivity #12	Missingness is handled by restricting to patients with complete observations on a subset of the most important confounders ("complete cases")	Instead of imputing data, this sensitivity analysis restricts the analysis cohort to patients with complete observations on key confounders	Data will not be imputed and missingness is assumed to be missing completely at random	The restriction to complete cases will significantly decrease sample size. To limit the attrition of patients with partially observed covariates, it won't be possible to use all covariates used in the main analysis propensity score model, but only consider key covariates with overall low proportions of missingness (age, sex, etc.)

6.8. Study size and feasibility

See appendix 4 (first feasibility matching on age, sex, comorbidity score) and appendix 5 (second feasibility matching on all predefined propensity score variables). Propensity score distribution plots with c-statistics are available in appendix 6. Table 1 showing prevalence of baseline characteristics and standardized differences in distribution are available in appendix 7.

7. Limitation of the methods

- Missingness in prognostic factors is a major challenge which is addressed in this emulation by multiple imputation using a non-parametric imputation algorithm. Multiple imputation usually assumes that missingness can be explained by observed characteristics, which may be empirically evaluated using principled missingness diagnostics, but the true underlying missingness mechanisms are usually unknown. Nevertheless, multiple imputation makes use of additional information (auxiliary covariates) which can render the underlying missingness assumptions more plausible. In addition, assumptions for alternative missing data approaches like complete case analysis or the "missing indicator approach" come with even stronger assumptions and additionally have the limitation of significantly reduced sample sizes, especially when comprehensively adjusting for known confounders and prognostic factors.
- Data continuity is a major challenge in EHR databases since "guaranteed" observable periods (such as continuous enrolment periods in administrative claims data) do not exist which may lead to measurement error in key covariates and exposure misclassification. Sensitivity analysis #8 tries to address this requiring patients to have had at least one visit before the index date which increases the likelihood that a patients was not only diagnosed at the respective center but is also regularly seen.
- Balancing patients on calendar year is not possible since calendar year shows instrumental variable-like behaviours (see Figure 13), i.e., it perfectly predicts treatment assignment and does not have any association with the outcome other than through the exposure. This assumption is formally untestable but clinically reasonable since there has not been any other significant change in treatment paradigms for FLAURA-like EGFR-mutated NSCLC populations other than introduction of osimertinib. The improvements of radiation of brain metastases may be the only exception, but it is expected that this may be of negligible significance for the scope of this emulation.

- Proportions of Asian race and patients with smoking history is different in the used US-only EHR-derived databases and the FLAURA trial population which can lead to altered treatment effect estimates since race was found to be an effect modifier in the trial's subgroup analyses. This limitation is address in sensitivity analysis #7 by re-weighting the distribution in these covariates to match that of the FLAURA populations.

8. Protection of human subjects

This study has been approved by the Brigham and Women's Hospital Institutional Review Board.

9. References

1. Jm F, A P, D M, et al. Nonrandomized Real-World Evidence to Support Regulatory Decision Making: Process for a Randomized Trial Replication Project. *Clin Pharmacol Ther*. 2020;107(4). doi:10.1002/cpt.1633

2. Franklin JM, Patorno E, Desai RJ, et al. Emulating Randomized Clinical Trials With Nonrandomized Real-World Evidence Studies: First Results From the RCT DUPLICATE Initiative. *Circulation*. 2021;143(10):1002-1013. doi:10.1161/CIRCULATIONAHA.120.051718

3. Franklin JM, Glynn RJ, Suissa S, Schneeweiss S. Emulation Differences vs. Biases When Calibrating Real-World Evidence Findings Against Randomized Controlled Trials. *Clin Pharmacol Ther*. 2020;107(4):735-737. doi:10.1002/cpt.1793

4. Wang SV, Schneeweiss S, RCT-DUPLICATE Initiative. Emulation of Randomized Clinical Trials With Nonrandomized Database Analyses: Results of 32 Clinical Trials. *JAMA*. 2023;329(16):1376-1385. doi:10.1001/jama.2023.4221

5. Soria Jean-Charles, Ohe Yuichiro, Vansteenkiste Johan, et al. Osimertinib in Untreated EGFR-Mutated Advanced Non–Small-Cell Lung Cancer. *N Engl J Med*. 2018;378(2):113-125. doi:10.1056/NEJMoa1713137

6. Ramalingam Suresh S., Vansteenkiste Johan, Planchard David, et al. Overall Survival with Osimertinib in Untreated, EGFR-Mutated Advanced NSCLC. *N Engl J Med*. 2020;382(1):41-50. doi:10.1056/NEJMoa1913662

7. Curtis MD, Griffith SD, Tucker M, et al. Development and Validation of a High-Quality Composite Real-World Mortality Endpoint. *Health Serv Res*. 2018;53(6):4460-4476. doi:10.1111/1475-6773.12872

8. Dong S, Kansagra AJ, Kaur G, et al. Validation of a Composite Real-World Mortality Variable Among Patients with Hematologic Malignancies Treated in the United States. *Blood*. 2023;142:5145.

9. Hernán MA, Wang W, Leaf DE. Target Trial Emulation: A Framework for Causal Inference From Observational Data. *JAMA*. Published online December 12, 2022. doi:10.1001/jama.2022.21383

10. Becker T, Weberpals J, Jegg AM, et al. An enhanced prognostic score for overall survival of patients with cancer derived from a large real-world cohort. *Ann Oncol*. 2020;31(11):1561-1568. doi:10.1016/j.annonc.2020.07.013

11. Becker T, Mailman M, Tan S, Lo E, Bauer-Mehren A. Comparison of overall survival prognostic power of contemporary prognostic scores in prevailing tumor indications. *Med Res Arch*. 2023;11(4). doi:10.18103/mra.v11i4.3638

12. Arkenau HT, Barriuso J, Olmos D, et al. Prospective validation of a prognostic score to improve patient selection for oncology phase I trials. *J Clin Oncol Off J Am Soc Clin Oncol*. 2009;27(16):2692-2696. doi:10.1200/JCO.2008.19.5081

13. Loureiro H, Becker T, Bauer-Mehren A, Ahmidi N, Weberpals J. Artificial Intelligence for Prognostic Scores in Oncology: a Benchmarking Study. *Front Artif Intell*. 2021;4:625573. doi:10.3389/frai.2021.625573

14. Loureiro H, Roller A, Schneider M, Talavera-López C, Becker T, Bauer-Mehren A. Matching by OS Prognostic Score to Construct External Controls in Lung Cancer Clinical Trials. *Clin Pharmacol Ther*. n/a(n/a). doi:10.1002/cpt.3109

15. Weberpals J, Wang SV. The FAIRification of research in real-world evidence: A practical introduction to reproducible analytic workflows using Git and R. *Pharmacoepidemiol Drug Saf*. 2024;33(1):e5740. doi:10.1002/pds.5740

16. RUBIN DB. Inference and missing data. *Biometrika*. 1976;63(3):581-592. doi:10.1093/biomet/63.3.581

17. Weberpals J, Raman SR, Shaw PA, et al. A Principled Approach to Characterize and Analyze Partially Observed Confounder Data from Electronic Health Records. *Clin Epidemiol*. 2024;16:329-343. doi:10.2147/CLEP.S436131

18. Sondhi A, Weberpals J, Yerram P, et al. A systematic approach towards missing lab data in electronic health records: A case study in non-small cell lung cancer and multiple myeloma. *CPT Pharmacomet Syst Pharmacol*. 2023;12(9):1201-1212. doi:10.1002/psp4.12998

19. Hotelling H. The Generalization of Student's Ratio. *Ann Math Stat*. 1931;2(3):360-378. doi:10.1214/aoms/1177732979

20. Little RJA. A Test of Missing Completely at Random for Multivariate Data with Missing Values. *J Am Stat Assoc*. 1988;83(404):1198-1202. doi:10.1080/01621459.1988.10478722

21. Weberpals J, Raman SR, Shaw PA, et al. smdi: an R package to perform structural missing data investigations on partially observed confounders in real-world evidence studies. *JAMIA Open*. 2024;7(1):ooae008. doi:10.1093/jamiaopen/ooae008

22. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivar Behav Res*. 2011;46(3):399-424. doi:10.1080/00273171.2011.568786

23. Moreno-Betancur M, Lee KJ, Leacy FP, White IR, Simpson JA, Carlin JB. Canonical Causal Diagrams to Guide the Treatment of Missing Data in Epidemiologic Studies. *Am J Epidemiol*. 2018;187(12):2705-2715. doi:10.1093/aje/kwy173

24. Bartlett JW, Seaman SR, White IR, Carpenter JR. Multiple imputation of covariates by fully conditional specification: Accommodating the substantive model. *Stat Methods Med Res*. 2015;24(4):462-487. doi:10.1177/0962280214521348

25. Weberpals J, Shaw PA, Lin KJ, et al. High-dimensional multiple imputation (HDMI) for partially observed confounders including natural language processing-derived auxiliary covariates. Published online May 17, 2024. doi:10.48550/arXiv.2405.10925

26. Shah AD, Bartlett JW, Carpenter J, Nicholas O, Hemingway H. Comparison of random forest and parametric imputation models for imputing missing data using MICE: a CALIBER study. *Am J Epidemiol*. 2014;179(6):764-774. doi:10.1093/aje/kwt312

27. Buuren S van, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *J Stat Softw*. 2011;45:1-67. doi:10.18637/jss.v045.i03

28. Leyrat C, Seaman SR, White IR, et al. Propensity score analysis with partially observed covariates: How should multiple imputation be used? *Stat Methods Med Res*. 2019;28(1):3-19. doi:10.1177/0962280217713032

29. Pishgar F, Greifer N, Leyrat C, Stuart E. MatchThem: Matching and Weighting after Multiple Imputation. *R J*. 2021;13(2):292-305. doi:10.32614/RJ-2021-073

30. Austin PC. The use of propensity score methods with survival or time-to-event outcomes: reporting measures of effect similar to those used in randomized experiments. *Stat Med*. 2014;33(7):1242-1258. doi:10.1002/sim.5984

31. Rubin DB. Multiple imputation. In: *Flexible Imputation of Missing Data, Second Edition*. Chapman and Hall/CRC; 2018:29-62.

32. Schwarzer G, Carpenter JR, Rücker G. *Meta-Analysis with R*. Vol 4784. Springer; 2015.

33. Jacobs EJ, Newton CC, Wang Y, Campbell PT, Flanders WD, Gapstur SM. Ghost-time bias from imperfect mortality ascertainment in aging cohorts. *Ann Epidemiol*. 2018;28(10):691-696.e3. doi:10.1016/j.annepidem.2018.06.002

34. Chen L, Fajardo O, Huntley M, Meyer AM, Taylor M. Use of last clinical activity date in overall survival analysis with real world data. In: *PHARMACOEPIDEMIOLOGY AND DRUG SAFETY*. Vol 30. WILEY 111 RIVER ST, HOBOKEN 07030-5774, NJ USA; 2021:116-116.

35. Schoenfeld DA. Sample-size formula for the proportional-hazards regression model. *Biometrics*. 1983;39(2):499-503.

10. Appendices

10.1. CONSORT diagrams

The following CONSORT attrition diagrams depict the process to select eligible FLAURA-like populations in EDB1, EDB2 and EDB4 for the main analysis, respectively.

Figure 6. CONSORT attrition to select eligible FLAURA-like populations in EDB1.

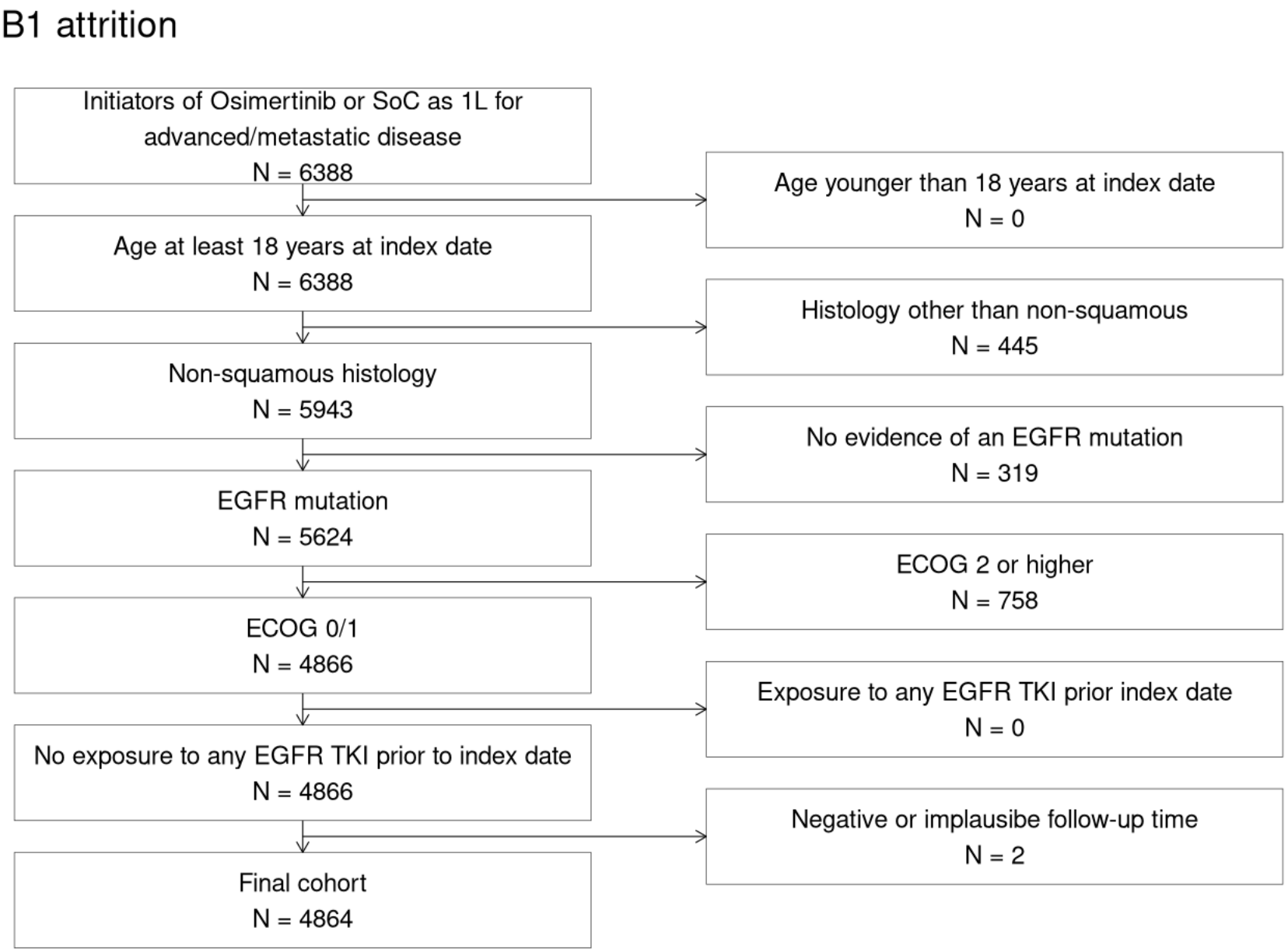


Figure 7. CONSORT attrition to select eligible FLAURA-like populations in EDB2.

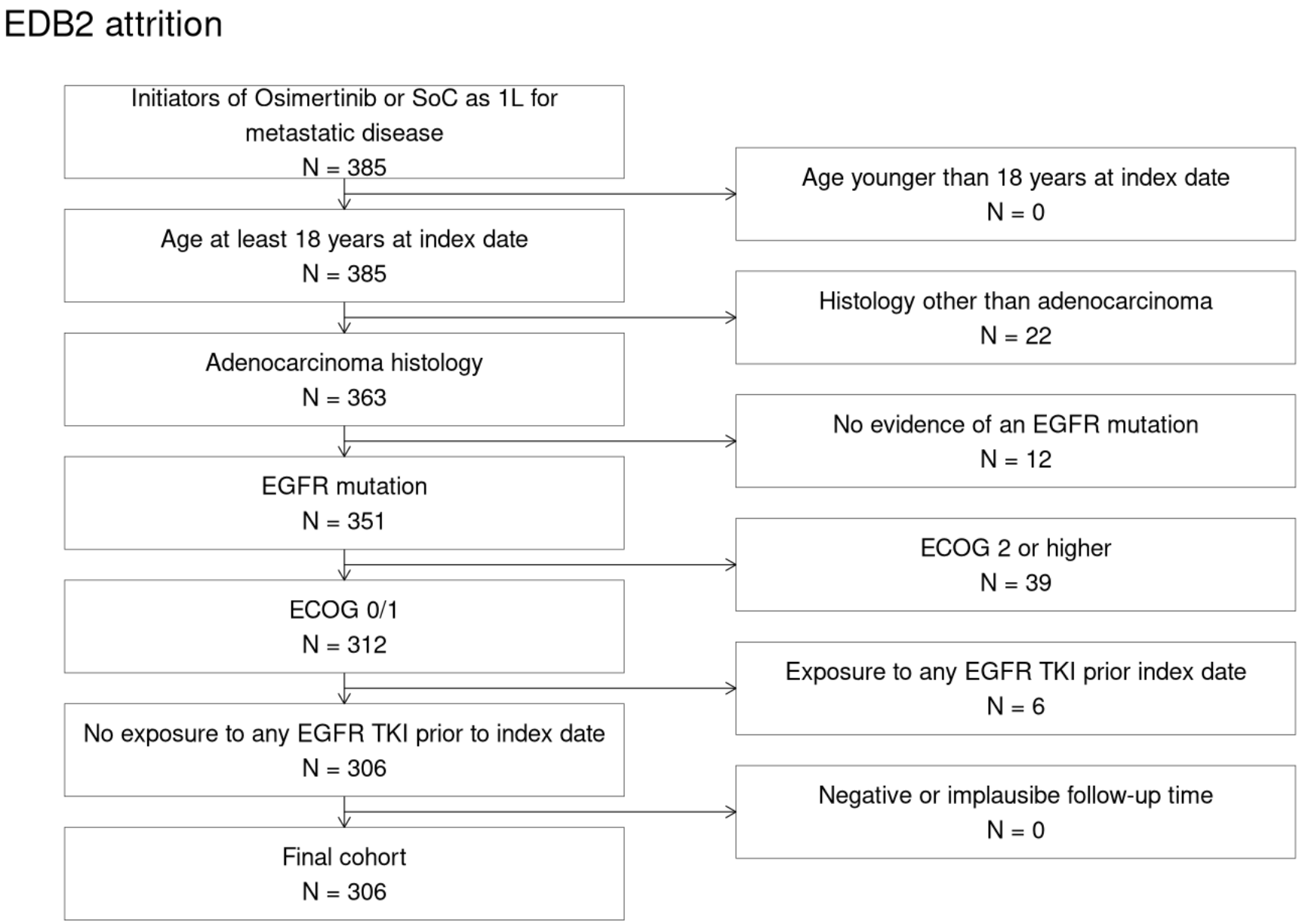
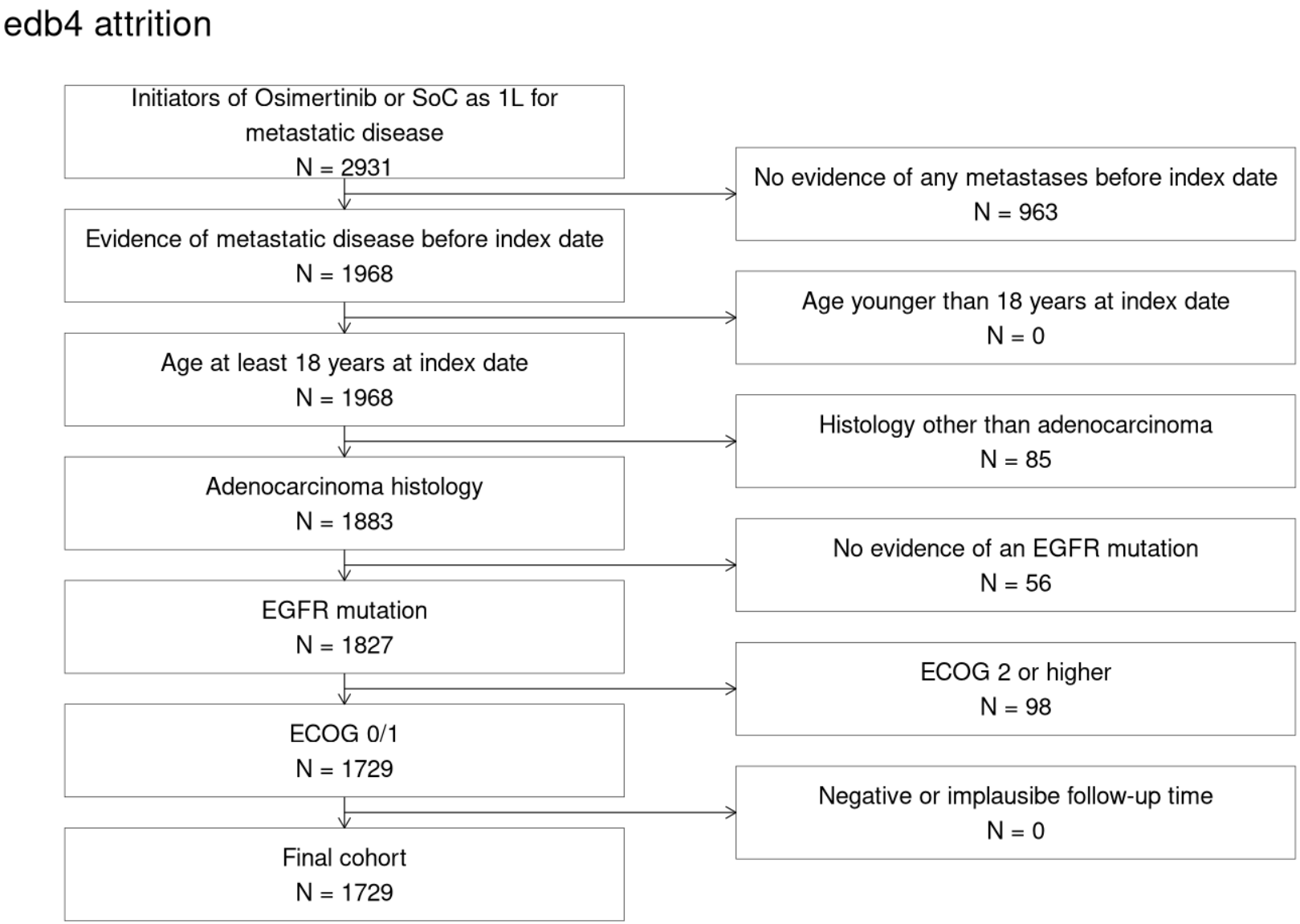


Figure 8. CONSORT attrition to select eligible FLAURA-like populations in EDB4.



10.2. Covariate balance figures

The following figures illustrate the balance of key covariates included in propensity score models among eligible FLAURA-like populations in EDB1, EDB2 and EDB4, respectively.

Figure 9. EDB1 covariate balance of covariates included in propensity score model before and after matching.

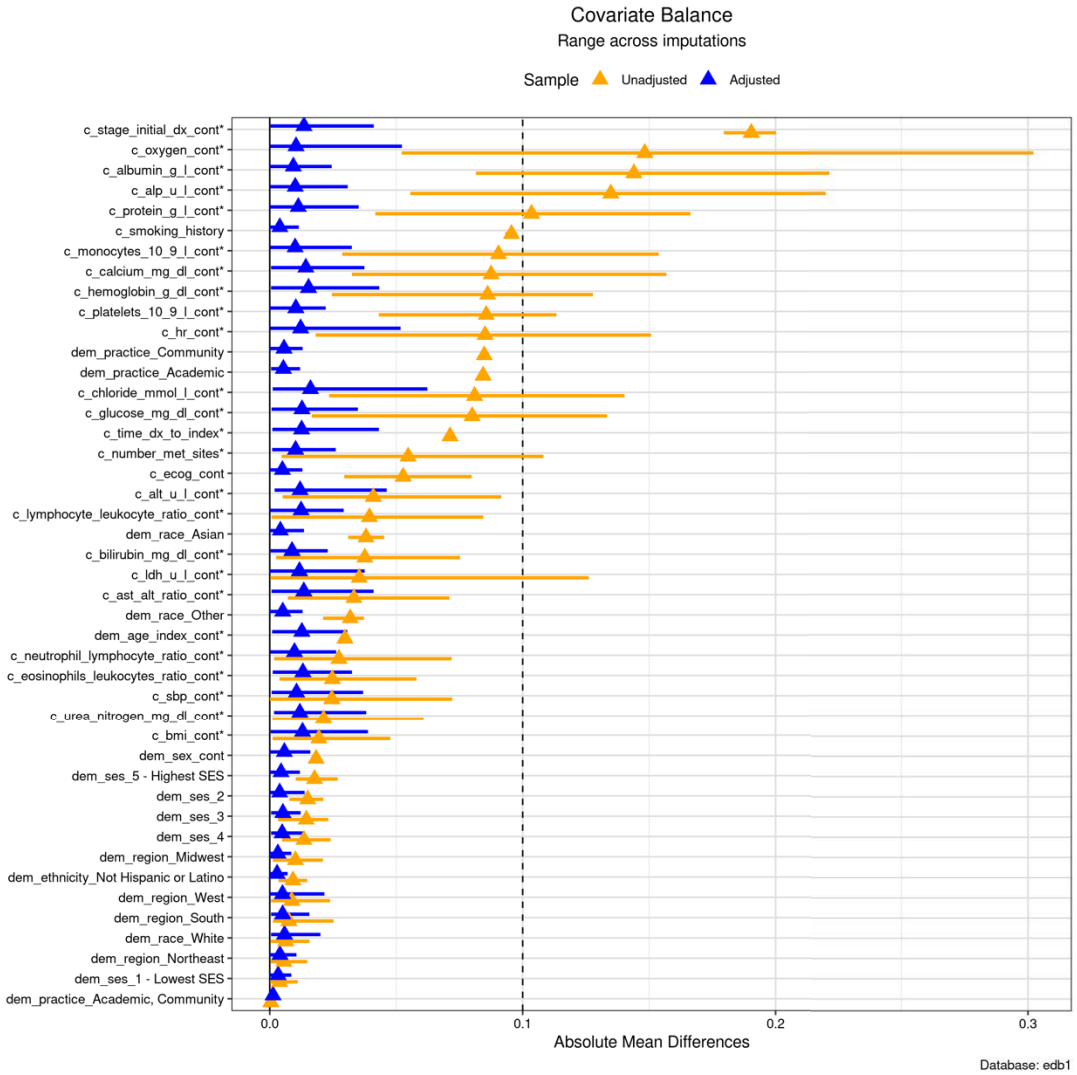


Figure 10. EDB2 covariate balance of covariates included in propensity score model before and after matching.

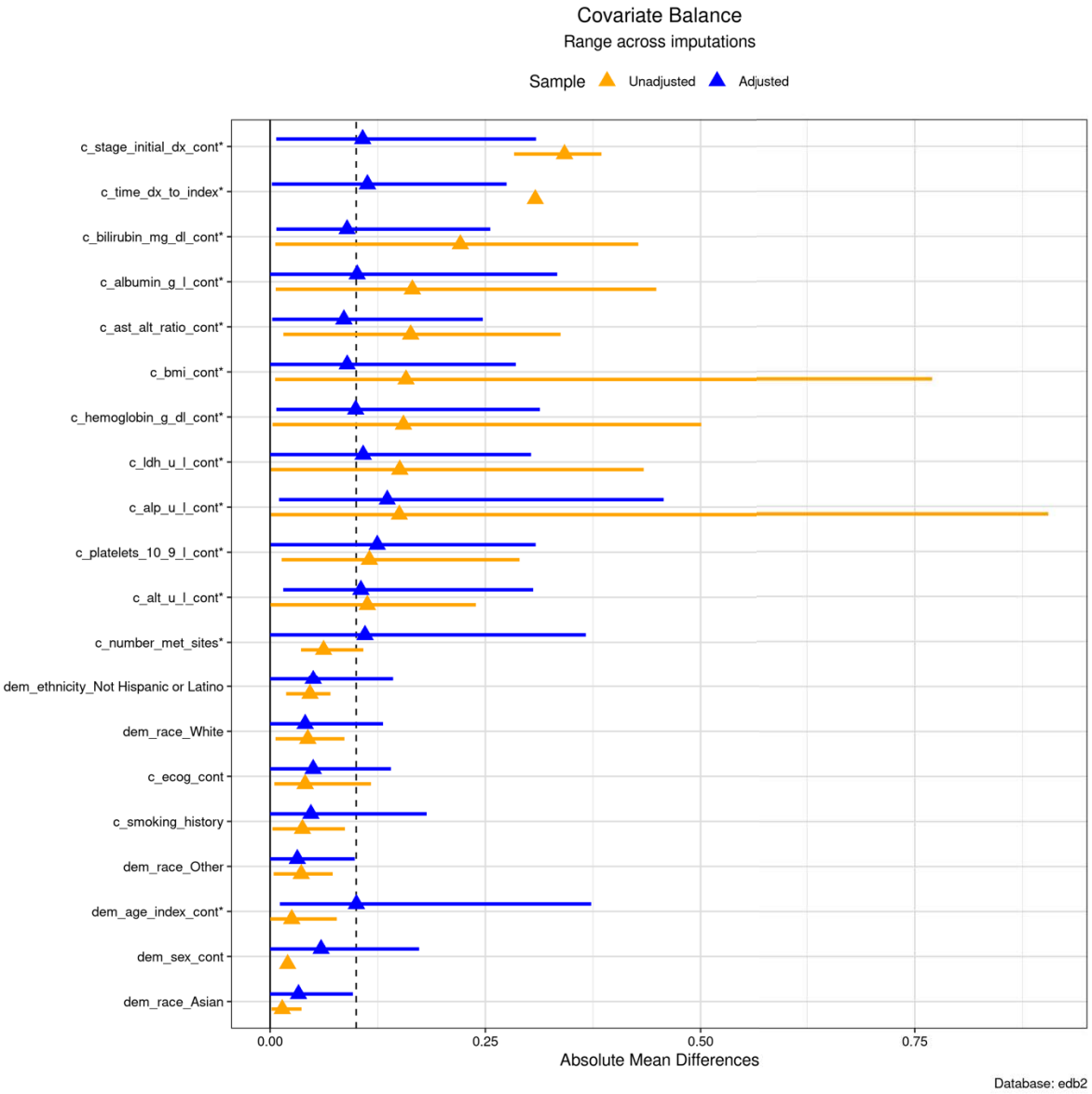
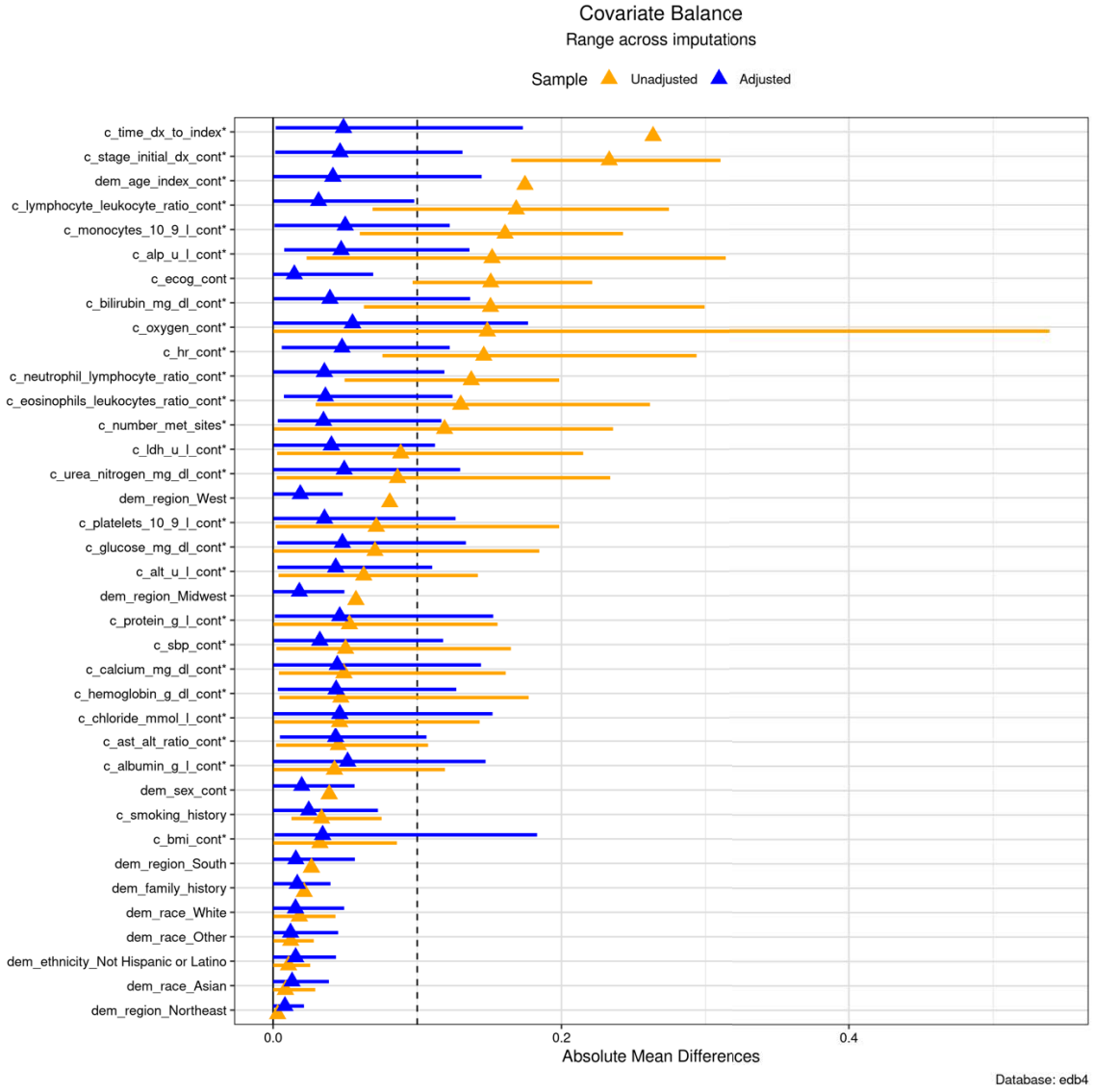


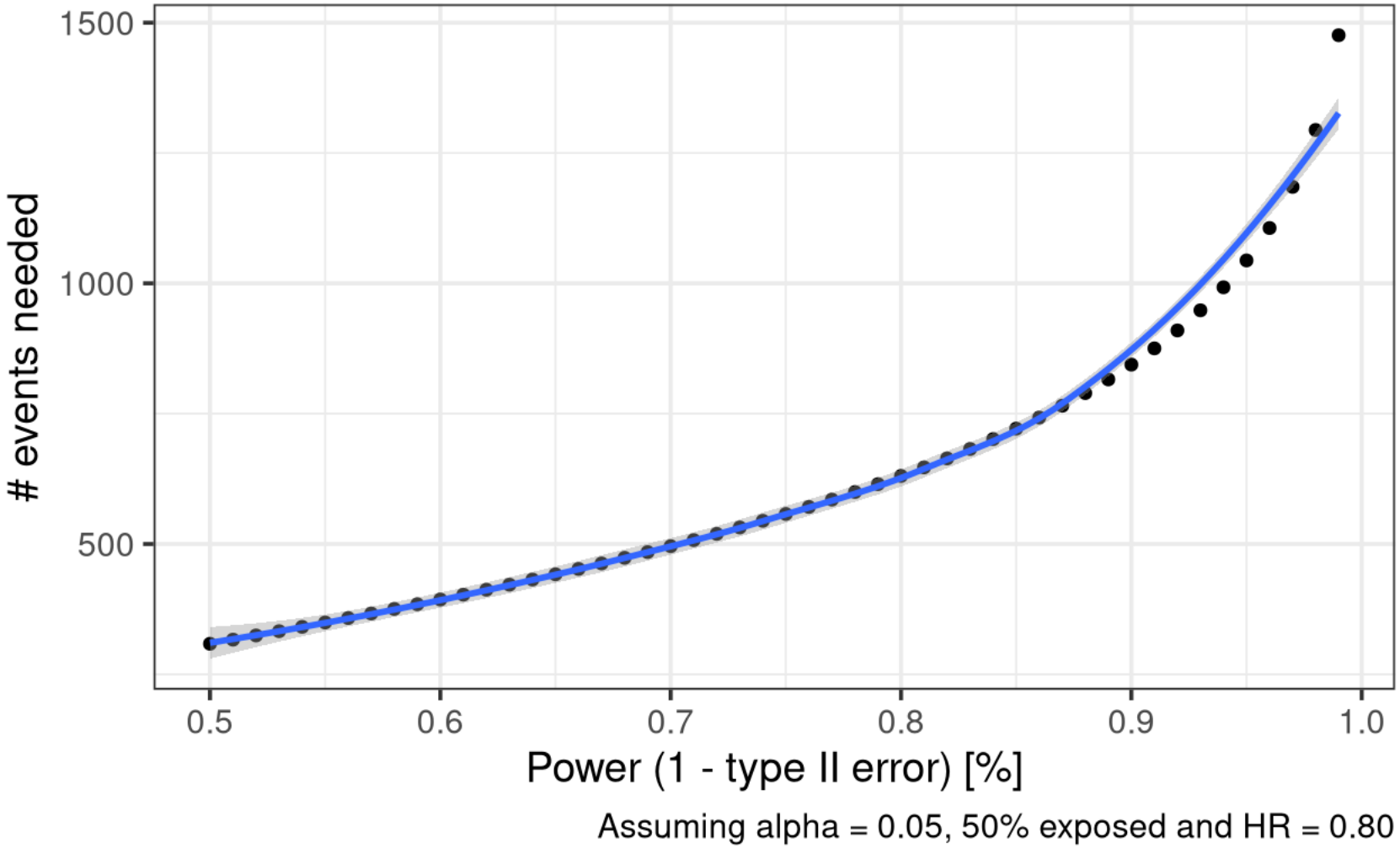
Figure 11. EDB4 covariate balance of covariates included in propensity score model before and after matching.



10.3. Sample size/power calculations

Power estimations are computed based on the average number of observed events across imputed and matched datasets according to the methodology described by Schoenfeld.[35] Considering EDB1 only for primary analysis and assuming a HR = 0.8 and an alpha = 0.05 (2-sided), this results in an estimated statistical power (1-estimated type II error) of 80%.

Figure 12. Number of events needed to achieve x% power.



10.4. Additional Figures and Tables

Table 10. Lab measurement plausibility thresholds.

Lab and standardized unit	Lower plausibility threshold	Upper plausibility threshold
c_albumin_g_l	10	200
c_alp_u_l	1	2000
c_alt_u_l	1	90000
c_ast_u_l	1	90000
c_bilirubin_mg_dl	0.1	80
c_calcium_mg_dl	0.1	20
c_chloride_mmol_l	0.1	200
c_eosinophils_leukocytes_ratio	0	100
c_glucose_mg_dl	0.1	2000
c_granulocytes_leukocytes_ratio	0	100
c_hemoglobin_g_dl	0.1	20
c_ldh_u_l	0.1	Inf
c_lymphocyte_10_9_l	0	1e+06
c_lymphocyte_leukocyte_ratio	0	100
c_monocytes_10_9_l	0	1e+06
c_neutrophil_10_9_l	0	1e+06
c_platelets_10_9_l	0	5000
c_protein_g_l	1	300
c_urea_nitrogen_mg_dl	0.1	250

Table 11. Vital sign measurement plausibility thresholds.

Vital sign	Lower plausibility threshold	Upper plausibility threshold
c_sbp	50	250
c_dbp	30	150
c_bmi	10	80
c_bsa	0.5	3.5
c_height	0.5	3
c_oxygen	50	100
c_pain	0	10
c_hr	20	250
c_resp	5	50
c_temp	86	113
c_weight	20	300

Table 12. Mapping from State to Region.

State	Region
CT	Northeast
ME	Northeast
MA	Northeast
NH	Northeast
RI	Northeast
VT	Northeast
DE	Northeast
NJ	Northeast
NY	Northeast
PA	Northeast
IL	Midwest
IN	Midwest
MI	Midwest
OH	Midwest
WI	Midwest
IA	Midwest
KS	Midwest
MN	Midwest
MO	Midwest
NE	Midwest
ND	Midwest

SD	Midwest
FL	South
GA	South
MD	South
NC	South
SC	South
VA	South
DC	South
WV	South
AL	South
KY	South
MS	South
TN	South
AR	South
LA	South
OK	South
TX	South
AZ	West
CO	West
ID	West
MT	West
NV	West
NM	West

UT	West
WY	West
AK	West
CA	West
HI	West
OR	West
WA	West

Figure 13. Treatment initiation trends by calendar year and treatment in EDB1.

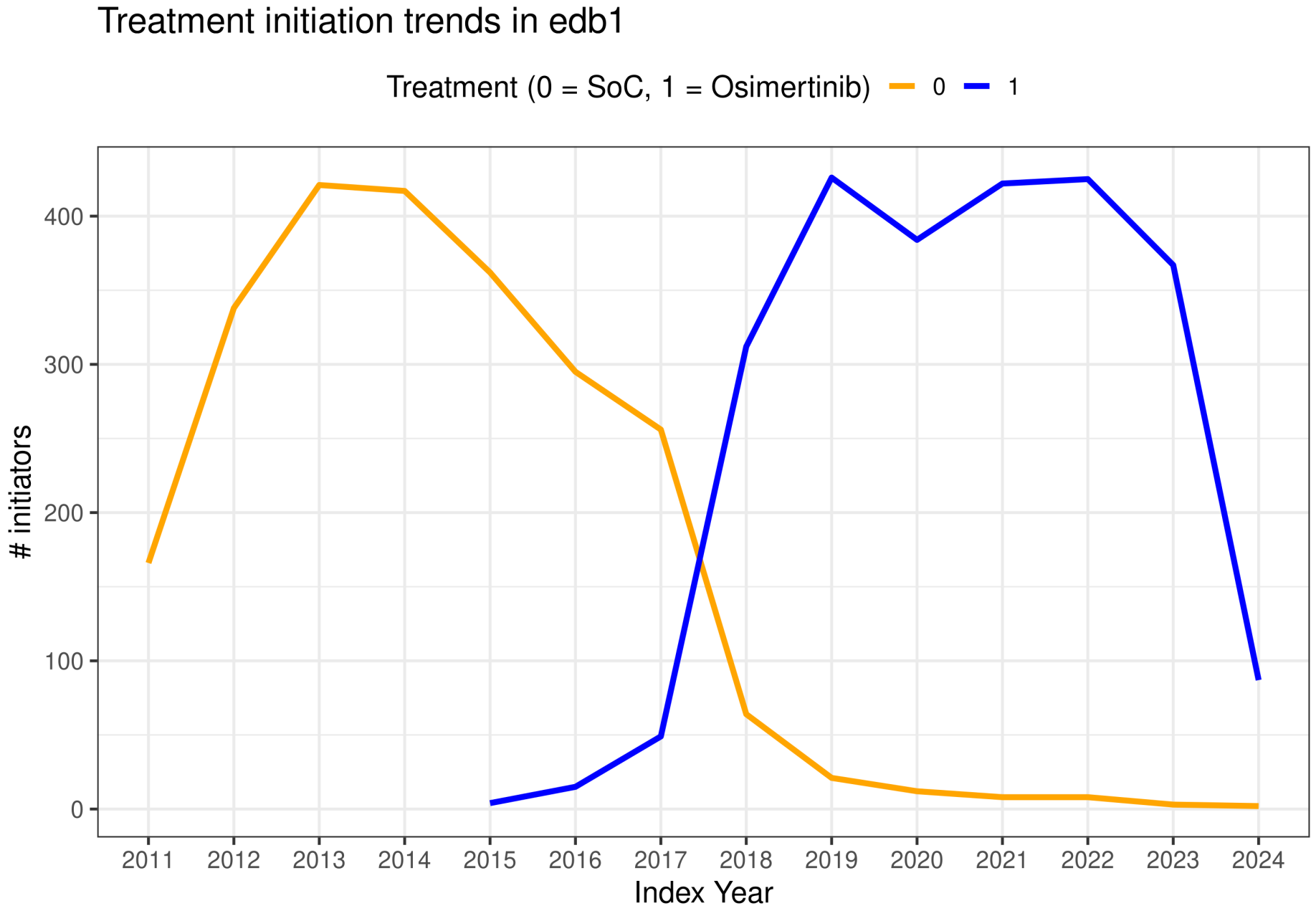


Figure 14. Treatment initiation trends by calendar year and treatment in EDB2.

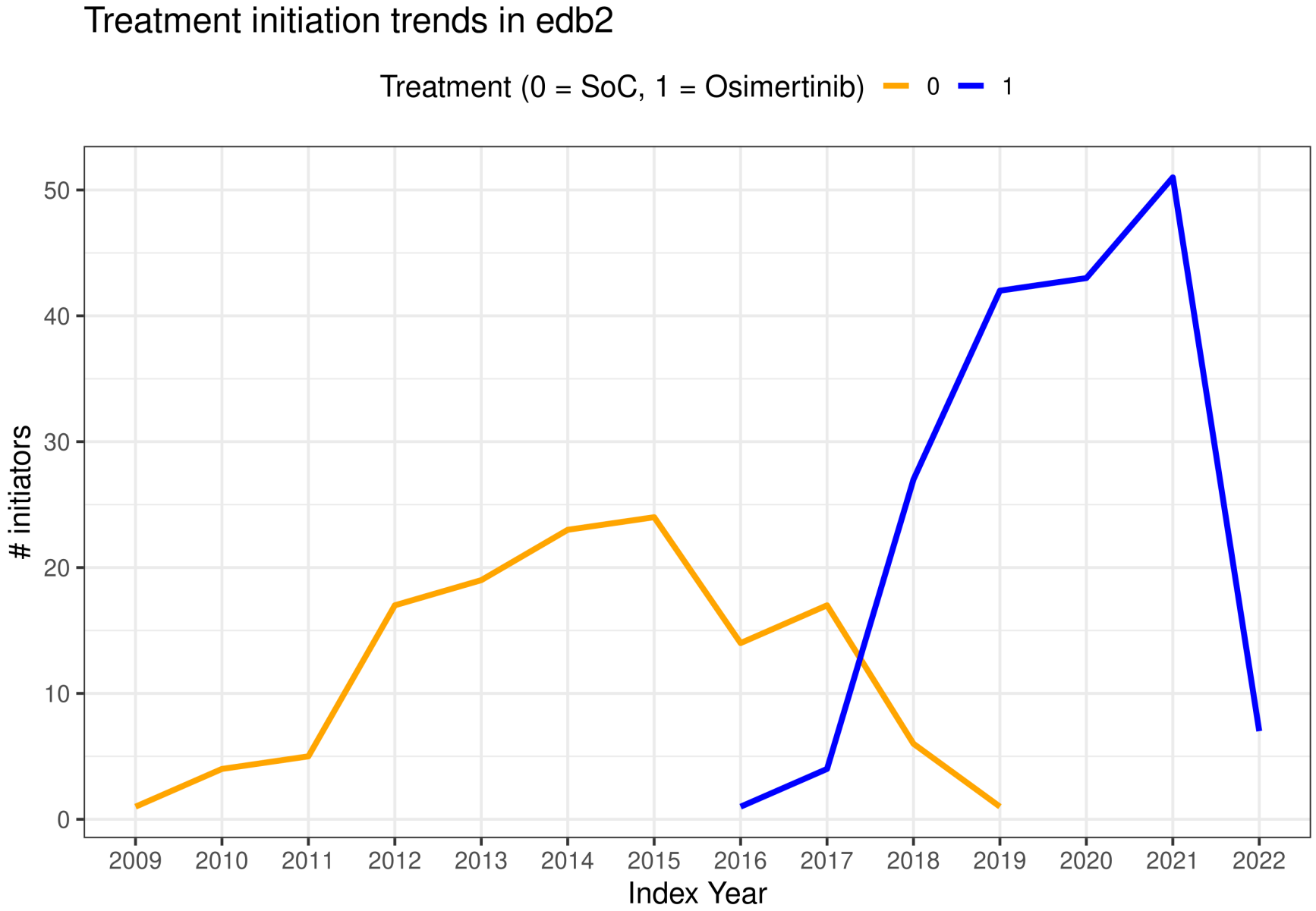


Figure 15. Treatment initiation trends by calendar year and treatment in EDB4.