Emulation of the MAVERICC (NCT01374425) trial using specialty oncology electronic health records databases

NCT07861386

20th August, 2026

1. Title Page

Title	Emulation of the MAVERICC (NCT01374425) trial using specialty oncology electronic health records databases
Research question & Objectives	Emulation of the MAVERICC (NCT01374425) trial, which compared mFOLFOX6-BV (modified leucovorin/fluorouracil/oxaliplatin plus bevacizumab) to FOLFIRI-BV (leucovorin/5-fluorouracil/irinotecan plus bevacizumab) on the risk of death in patients with previously untreated metastatic colorectal cancer.
Protocol version	V1.0
Last update date	Aug 20, 2026
Contributors	**Primary investigators contact information:** Xiangzhong Xue, Shirley V. Wang **Contributor names:** Yuxin Wang (programmer)
Study registration	**Site:** clinicaltrials.gov **Identifier: NCT07861386**
Sponsor	**Organization:** Food and Drug Administration **Contact:** n/a
Conflict of interest	SVW has been an ad hoc consultant to ICA Group, Cytel Inc, and MITRE a federally funded research and development center for the Centers for Medicare and Medicaid
Protocol repository	Clinicaltrials.gov
Analytic code repository	https://gitlab.partners.org/drugepi/encore/mavericc-nct-01374425 (access only through Mass General Brigham network for authorized personnel)
Quarto study report (including annotated code and output)	https://gitlab.partners.org/drugepi/encore/mavericc-nct-01374425/-/tree/main/public?ref_type=heads (access only through Mass General Brigham network for authorized personnel)
***encore.io*[1] version**	0.2.0 (see attached documentation *encore.io_0.2.0.pdf*)
encore.analytics[1] version	0.2.0 (https://janickweberpals.github.io/encore.analytics/)
[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

1. Title Page 1
2. Abstract 3
3. Amendments and updates 3
4. Rationale and background 3
5. Research question and objectives 3
6. Research methods 6
6.1. Data sources 6
6.1.1. Context and rationale for data sources 6
6.2. Data management 9
6.3. Quality control 9
6.4. Study design 9
6.5. Study design diagram 10
6.6. Setting 13
6.6.1. Context and rationale for definition of time 0 (and other primary time anchors) for entry to the study population 13
6.6.2. Context and rationale for study inclusion criteria 13
6.6.3. Context and rationale for study exclusion criteria ..... Error! Bookmark not defined.
6.6.4. Context and rationale for exposure(s) of interest 13
6.6.5. Context and rationale for outcome(s) of interest 14
6.6.6. Context and rationale for follow up 16
6.6.7. Context and rationale for covariates 16
6.7. Data analysis 26
6.7.1. Context and rationale for analysis plan 26
6.8. Study size and feasibility 37
7. Limitation of the methods 37
8. Protection of human subjects 37
9. References 38
10. Appendices 42
10.1. CONSORT diagrams 42
10.2. Covariate balance figures 45
10.3. Sample size/power calculations 48
10.4. Additional Figures and Tables 49

2. Abstract

This trial emulation study aims to emulate the MAVERICC trial (NCT01374425) using real-world specialty oncology electronic health record data and to investigate the concordance between the trial's original and emulated treatment effect estimates on overall survival (OS). MAVERICC was a Phase II, open-label, randomized study assessing the efficacy and safety of modified leucovorin/5-fluorouracil/ oxaliplatin plus bevacizumab (mFOLFOX6-BV) with leucovorin/5-fluorouracil/irinotecan plus bevacizumab (FOLFIRI-BV) in patients with previously untreated metastatic colorectal cancer (mCRC).

3. Amendments and updates

Version date	Version number	Section of protocol	Amendment or update	Reason
August 20, 2026,	V1.0	NA	Initial version	NA

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, which 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 MAVERICC trial. MAVERICC was a Phase II, open-label, randomized study assessing the efficacy and safety of modified leucovorin/5-fluorouracil/ oxaliplatin plus bevacizumab (mFOLFOX6-BV) with leucovorin/5-fluorouracil/irinotecan plus bevacizumab (FOLFIRI-BV) in patients with previously untreated metastatic colorectal cancer (mCRC).

In MAVERICC, the primary outcome was overall survival (OS), with a hazard ratio (HR) of 0.76 (95% confidence interval (CI), 0.56–1.04).[5] The median overall survival was 27.4 months (95% CI 24.6–36.7) in the FOLFIRI-BV arm compared to 23.9 months (95% CI 20.4-26.1) in the mFOLFOX6-BV arm. The OS endpoint was published in Clinical Cancer Research in May 2019.[5]

5. Research question and objectives

The primary and secondary research question were 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] among patients with previously untreated mCRC who initiated FOLFIRI-BV versus mFOLFOX-BV.
Hypothesis:	Initiation of FOLFIRI-BV, compared with initiation of mFOLFOX-BV, is associated with overall survival compatible with the hazard ratio of 0.76 (95% CI, 0.56 – 1.04) observed in the MAVERICC trial
Population ***(mention key inclusion-exclusion criteria):***	• Age at least 18 years • Histologically or cytologically confirmed CRC • ECOG PS 0–1
Exposure:	Initiation of FOLFIRI-BV
Comparator:	Initiation of mFOLFOX-BV
Outcome:	Time to all-cause mortality (OS)
Time ***(when follow up begins and ends):***	From the day of the end of the treatment assessment window until death or last observed clinical activity/last sign of the patient being alive or data cut-off, whichever occurred earlier
Setting:	1L metastatic CRC
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 MAVERICC is illustrated side by side in **Table 2**.

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

Protocol component	MAVERICC RCT	Emulation
Eligibility Criteria	Age ≥18 years at randomization	Age ≥18 years at treatment initiation
	Histologically or cytologically confirmed colorectal cancer	Line of therapy setting classified as “Metastatic” (EDB1, EDB3) or evidence of metastatic disease at treatment initiation (EDB4)

	ECOG performance status 0–1	ECOG 0–1 (if available/ascertainable in the database)
	Participants with treated brain metastases are eligible for study participation. Participants may not receive ongoing treatment with steroids at Screening. Anticonvulsants (at stable dose) are allowed. Treatment for brain metastases may be whole-brain radiotherapy, radiosurgery, neurosurgery, or a combination as deemed appropriate by the treating physician. Radiotherapy and stereotactic radiosurgery must be completed at least 28 days prior to randomization.	Exclude patients with evidence of brain metastases (Treatment history for brain metastases and steroid use at screening are incompletely captured; patients with any evidence of brain metastases are excluded as a conservative proxy)
	Any prior systemic treatment for mCRC	Exclude patients who received prior systemic cancer treatment for mCRC.
	Received prior adjuvant chemotherapy for CRC 12 months prior to the randomization	Exclude patients who received chemotherapy within 12 months prior to the index date.
	Malignancies other than mCRC within 5 years prior to randomization, except for adequately treated carcinoma in situ of the cervix, basal or squamous cell skin cancer, localized prostate cancer treated surgically with curative intent, and ductal carcinoma in situ treated surgically with curative intent	Exclude patients with another primary malignancy within 5 years before the index date, except carcinoma in situ of the cervix, basal or squamous cell skin cancer, localized prostate cancer, and ductal carcinoma in situ.
	Treatment with any other investigational agent, or participation in another investigational drug trial within 28 days prior to randomization	Exclude patients who received investigational agents 28 days prior to randomization
Treatment strategies	mFOLFOX-BV versus FOLFIRI-BV	mFOLFOX-BV versus FOLFIRI-BV
Assignment procedures	Random assignment to receive either mFOLFOX-BV or FOLFIRI-BV in a 1:1 ratio	Randomization was emulated via 1:1 matching of patients who initiated mFOLFOX-BV versus FOLFIRI-BV, based on a propensity score that included major risk factors for overall survival.
Follow-up period	Time from randomization until death from any cause or censoring at the last contact that the patient was known to be alive.	Time from the end of the treatment assessment window[a] 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	As started (observational analogue of intent-to-treat)

[a]Proprietary business rules to define initiation of the line of therapy (including exposure assessment window) cannot be shared

EDB1 =ENCORE DataBase 1; EDB3 =ENCORE DataBase 3; EDB4 =ENCORE DataBase 4; dMMR = mismatch-repair–deficient; MSI-H = microsatellite-instability-high; ECOG = Eastern Cooperative Oncology Group; CNS = central nervous system

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, CRC-specific data are available for EDB1, EDB3, 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[6,7]).

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 the section 7). Following a comprehensive assessment of all three databases—including their patient-identification window/sampling strategy, the availability of key eligibility criteria and confounders, feasibility counts, and covariate balance after confounding adjustment—all were determined to be fit for the purpose of emulating the MAVERICC trial and were included in the main analyses. Each database was analyzed separately, and the resulting treatment-effect estimates were pooled using a fixed-effect meta-analysis. [8,9]

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	EDB3	EDB4
Data Source(s):	EHR-derived	EHR-derived	EHR-derived
Study Period:	Patient identification period: 01/01/2013-04/30/2024 with follow-up information through data cut-off date on 04/30/2024	Follow-up information through 02/24/2026 (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:	01/01/2013-04/30/2024	On or before 02/24/2026	After 10/01/2018
Data Version (or date of last update):	Delivery: Jul 11, 2024	Delivery: Apr 23, 2026	Delivery: Oct 24, 2023 Updated (demographics): Feb 29, 2024
Data sampling/extraction criteria:	Patients are sampled if they have a confirmed diagnosis of metastatic CRC via abstraction on or after 1 Jan 2011, and at least 2 EHR visits on or after 1 Jan 2011. Both ICD-9 (153.x, 154.x) and ICD-10 codes (C18x, C19x, C20x, C21x) are used for the initial selection, and advanced diagnosis are then confirmed via abstraction (since ICD codes do not specify advanced diseases).	This database identifies patients for curation using structured ICD-10 diagnosis codes corresponding to the indication of interest — for colorectal cancer, C18.0 and C18.2–C18.9 (colon), C19 (rectosigmoid junction), and C20 (rectum); the appendix code C18.1 is not included — along with at least the year of diagnosis. Once this initial screening list is generated, patients are randomly selected for further review. Curation begins with confirmation of the diagnosis and diagnosis date, primarily based on pathology reports and other unstructured data sources. All patients must be over 18 years of age at the time of their first diagnosis. Certain colorectal cases are excluded from curation. Specifically, non-invasive lesions — carcinoma in situ (Tis),	CRC 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 CRC and with a documented visit date, within the defined reporting period, to one of the 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.

		including intramucosal carcinoma, and high-grade dysplasia confined to the mucosa without invasion through the muscularis mucosae into the submucosa — are not eligible. However, lesions demonstrating submucosal or deeper invasion (i.e., at least pT1) qualify for curation. Additionally, in-situ lesions that later progress to invasive carcinoma are eligible; in these cases, the diagnosis date should reflect the date of the original in-situ diagnosis, not the later invasive diagnosis. Patients are ultimately selected for inclusion in data products through quality control processes that assess consistency and potential conflicts in their records. These evaluations may rely on structured data alone or a combination of structured and curated data, depending on the specific data product. Importantly, CAI does not exclude patients based on data completeness, in order to prevent the introduction of selection bias.	
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.	EHR-derived. Mortality/date of death is a composite endpoint derived from structured EHR data, manual curation, and third-party sources including obituary data and the Social Security Death Index. De-identified tokens link patients across datasets using hashed PII.	Mortality/date of death is a composite endpoint of structured and unstructured EHR data, supplemented with commercially available claims data, obituary data, and

		Curated data is prioritized, followed by EHR and then third-party sources. Reported death dates are shifted to the nearest Sunday within four days to enhance privacy. Curated death information follows a source hierarchy: death certificate, obituary, or provider-reported date.	the Social Security Administration death master file.
Conversion to CDM[2]:	No	Yes	No
Software for data management:	R 4.3.2	R 4.3.2	R 4.3.2

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 ensure 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

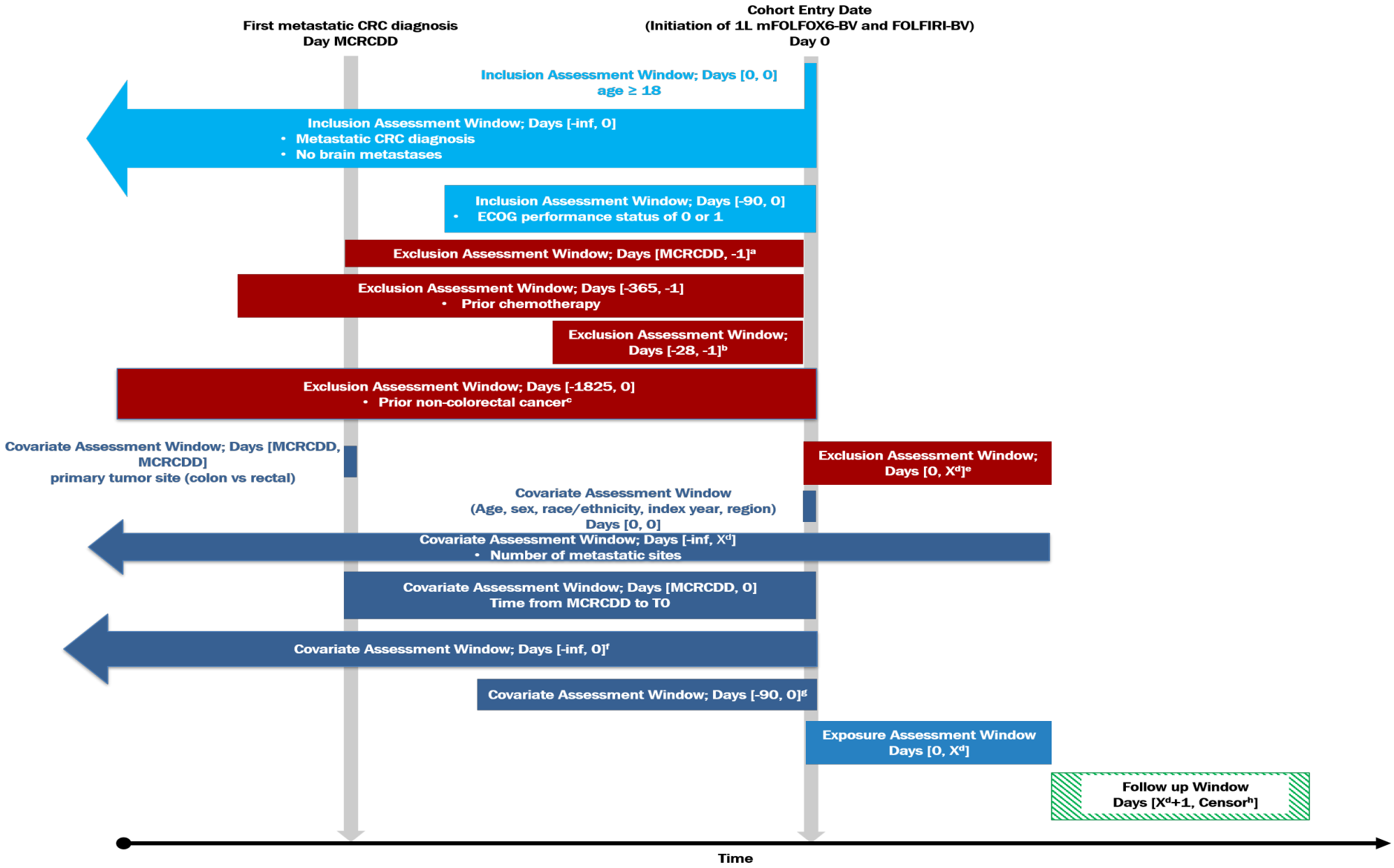
Research design (e.g. cohort, case-control, etc.): Cohort study

Rationale for study design choice: Resembles the principles of the (target) trial emulation framework.[10]

6.5. Study design diagram

Figure 1 depicts study design and variable measurement considerations for the emulation of the MAVERICC 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.[11–15]

Figure 1. Study design illustration for MAVERICC trial emulation.



a. Prior systemic treatment for metastatic CRC

b. Prior investigational agent

c. Exclude patients with another primary malignancy within 5 years before the index date, except carcinoma in situ of the cervix, basal or squamous cell skin cancer, localized prostate cancer, and ductal carcinoma in situ.

d. Proprietary business rules to define initiation of the line of therapy (including exposure assessment window) cannot be shared.

e. Systemic anticancer therapy other than mFOLFOX-BV and FOLFIRI-BV

f. Baseline biomarkers including NRAS, KRAS, BRAF, MMR/MSI, De novo metastatic status, time from initial diagnosis to T0, smoking, family history, race/ethnicity, sidedness, etc.

g. 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)

h. 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)

mFOLFOX-BV = modified fluorouracil, leucovorin, oxaliplatin, and bevacizumab; FOLFIRI-BV = fluorouracil, leucovorin, irinotecan, and bevacizumab; MCRCDD = metastatic colorectal cancer diagnosis date; CRC = colorectal cancer; 1L = first-line antineoplastic systemic therapy, NRAS = NRAS proto-oncogene; KRAS = KRAS proto-oncogene; BRAF = BRAF proto-oncogene; MMR = mismatch repair; MSI = microsatellite instability

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 FOLFIRI-BV (exposure) or mFOLFOX-BV (comparator) as 1L systemic antineoplastic treatment for mCRC. 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 reports (MAVERICC) 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.

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

The exposure and comparator were defined to emulate the agents compared in the trial—FOLFIRI-BV versus mFOLFOX-BV in a 1L advanced 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 as start and end dates. Based on this table, patients who received mFOLFOX-BV or FOLFIRI-BV treatment regimens are identified by their generic names (string match) in 1L. The LOT implicitly only considers regimens that were given in 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).
- **EDB3**: 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 therapy, based on a proprietary business rule with an exposure assessment window. That is, each patient is represented with one row per curated line of therapy AND drug name with corresponding information online number, regimens as well start and end dates. Based on this table, patients are identified who received only mFOLFOX-BV or FOLFIRI-BV within the first line of therapy using the exposure assessment window; follow-up begins after the end of the exposure assessment window. 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 EDB3' Quarto report' (access within MGB network only).

- **EDB4:** For the EDB4 database, the following logic is applied.

- Identify patients with evidence of a metastasis from the diagnosis table in which the earliest date associated with evidence of metastasis is captured as a structured field (metastasis date).
- Identify all potential antineoplastic drugs typically used in metastatic CRC (see list below*). Only these are considered.
- Identify patients who initiated the MAVERICC regimens as their 1L antineoplastic treatment after the date of metastasis. Treatment line identification is established according to the business rules set by the data vendor.

***Antineoplastic drugs considered:** fluorouracil, adagrasib, bevacizumab, bevacizumab-adcd, bevacizumab-awwb, bevacizumab-bvzr, bevacizumab-maly, bevacizumab-tnjn, capecitabine, cemiplimab-rwlc, cetuximab, dostarlimab-gxly, encorafenib, entrectinib, fam-trastuzumab deruxtecan-nxki, fruquintinib, ipilimumab, irinotecan, lapatinib, larotrectinib, leucovorin, levoleucovorin, nivolumab, nivolumab-hyaluronidase, oxaliplatin, panitumumab, pembrolizumab, pembrolizumab and berahyaluronidase alfa-pmph, pertuzumab, ramucirumab, regorafenib, repotrectinib, retifanlimab-dlwr, selpercatinib, sotorasib, tislelizumab-jsgr, toripalimab-tpzi, trastuzumab, trastuzumab-anns, trastuzumab-dkst, trastuzumab-dttb, trastuzumab-hyaluronidase-oysk, trastuzumab-pkrb, trastuzumab-qyyp, trastuzumab-strf, trifluridine and tipiracil, tucatinib, and ziv-aflibercept.

Context and rationale for outcome(s) of interest

The primary outcome for the database study was defined to emulate the overall survival outcome in the trial as time from index to death from any cause.

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. 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.
- **EDB3**: 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. Patients without evidence of death are censored at the earlier of the last recorded activity date or the data cut-off date. Activity dates are defined as documented in Table 4. Death dates are compiled from Electronic Medical Records (EMR), manual curation, and third-party death data, linked via de-identified tokens generated from hashed personally identifiable information (PII). Sources of death dates are prioritized as follows: (1) manual curation, (2) EMR, and (3) third-party sources. For privacy reasons, the final reported death date is shifted to the nearest Sunday within four days of the actual date. Manually curated death dates follow a prioritization hierarchy of source documentation: death certificate first, followed by scanned obituaries or death announcements, and lastly exact dates reported by providers without other supporting documentation. The OS endpoint calculation uses a parameterized R function (*edb3_get_os()*), with detailed documentation provided in the attached PDF.

- **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.

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

6.6.4. Context and rationale for follow-up

Only intention-to-treat (ITT) analyses will be conducted. Cross-over from the comparator to the exposure regimen can be expected, as oxaliplatin- and irinotecan-based regimens are commonly used sequentially across treatment lines in mCRC. The MAVERICC trial did not report protocol-defined on-study crossover between treatment arms or provide detailed information on subsequent anticancer treatments after discontinuation. Therefore, the comparison of first-line mFOLFOX-BV versus FOLFIRI-BV under an ITT framework should be interpreted primarily as a comparison of initial treatment strategies, rather than as the isolated treatment effect of one regimen in the absence of subsequent treatment switching.

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.

6.6.5. 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.[11] 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 the appendix **Table 10** and appendix **Table 11**.

Primary tumor sidedness was not available as a curated variable in the 3 databases. Although sidedness is a prognostic factor in metastatic colorectal cancer, it is primarily used to inform the selection of anti-EGFR therapy and is not generally expected to drive selection between mFOLFOX6 plus bevacizumab and FOLFIRI plus bevacizumab.[36–38] To assess whether imbalance in sidedness could nevertheless contribute to residual confounding, we derived a proxy measure of left-sided, right-sided, or unknown primary tumor location from ICD-9/10 diagnosis codes (Table 16).[39] In EDB1, a variable indicating primary tumor site (colon vs rectum) was additionally incorporated, with rectal tumors classified as left-sided. The proxy was not included in the propensity-score model but was evaluated as an external balance diagnostic and was well balanced after matching across all 3 databases (Figures 14-16), providing no evidence of substantial between-group imbalance in measured sidedness.

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[11]	[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	[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]

Characteristic	Harmonized analysis variable name	R function to derive covariate (see pdf in appendix)	Details[1]	Variable encoding	Assessment window
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]
Primary tumor site	crcsite	NA	Primary tumor site (available only in EDB1)	Colon/Rectum	[-inf;0]
Primary tumor sidedness[2]	side	NA	Derived from the most recently recorded primary tumor sidedness identified using ICD-9/10 codes before the index date in EDB1, EDB3, and EDB4 (Table 16). For EDB1, the primary tumor site recorded in the "crcsite" variable was additionally incorporated into the assessment of tumor	Left/Right/Unknown	[-inf;0]

Characteristic	Harmonized analysis variable name	R function to derive covariate (see pdf in appendix)	Details[1]	Variable encoding	Assessment window
			sidedness, with rectal tumors classified as left-sided.		
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[11]; due to I/E criteria ECOG is modelled as a binary (0, 1) covariate	[-90;0]
BRAF mutation status	c_braf_status	edbx_get_biomarker()	Evidence of BRAF mutation status. If multiple measurements are available, the result closest to the index date is prioritized.	Positive / negative / missing	[-inf;0]
NRAS mutation status	c_nras_status	edbx_get_biomarker()	Evidence of NRAS mutation status. If multiple measurements are available, the result closest to the index date is prioritized.	Positive / negative / missing	[-inf;0]
KRAS mutation status	c_kras_status	edbx_get_biomarker()	Evidence of KRAS mutation status. If multiple measurements are available, the result closest to the index date is prioritized.	Positive / negative / missing	[-inf;0]
MMR/MSI status	c_mmr_msi_status	edbx_get_biomarker();	Evidence of MMR or MSI status. If multiple measurements are available,	Positive if either MMR or MSI was positive; negative	[-inf;0]

Characteristic	Harmonized analysis variable name	R function to derive covariate (see pdf in appendix)	Details[1]	Variable encoding	Assessment window
			the result closest to the index date is prioritized.	if both were negative; otherwise missing	
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)[11]	[-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	c_met_pre_index_aw	edbx_get_diagnosis_solid()	Evidence of any metastasis before the start of follow-up	Binary logical (TRUE, FALSE)	[-inf;X]; Proprietary business-rule-based covariate assessment window
Number of metastatic sites	c_number_met_sites_aw	edbx_get_diagnosis_solid()	Number of metastatic sites for a given patient before/on the start of follow-up	Integer	[-inf;X]; Proprietary business-rule-based covariate assessment window
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	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]

Characteristic	Harmonized analysis variable name	R function to derive covariate (see pdf in appendix)	Details[1]	Variable encoding	Assessment window
index date					
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) [3]	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) [3]	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]
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]

Characteristic	Harmonized analysis variable name	R function to derive covariate (see pdf in appendix)	Details[1]	Variable encoding	Assessment window
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[3]	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[3]	c_eosinophils_leukocytes_ratio_cont	edbx_get_labs()	Eosinophils/100 leukocytes in blood. In case of ties, the lower	continuous	[-90;0]
Glucose[3]	c_glucose_mg_dl_cont	edbx_get_labs()	Closest glucose measurement (in serum/plasma) relative to index date in mg/dL. 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/dL.	continuous	[-90;0]

Characteristic	Harmonized analysis variable name	R function to derive covariate (see pdf in appendix)	Details[1]	Variable encoding	Assessment window
			In case of ties, the lower is selected		
Lactate dehydrogenase (LDH)[4]	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]
Lymphocyte/leukocyte ratio[3]	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[3]	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.	continuous	[-90;0]

Characteristic	Harmonized analysis variable name	R function to derive covariate (see pdf in appendix)	Details[1]	Variable encoding	Assessment window
			In case of ties, the lower is selected. Used to compute neutrophil/lymphocyte (NLR) ratio		
Neutrophil/lymphocyte ratio[3]	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[3]	c_urea_nitrogen_mg_dl_cont	edbx_get_labs()	Closest urea nitrogen measurement (in serum/plasma) relative to index date in mg/dL. In case of ties, the lower is selected	continuous	[-90;0]
Systolic blood pressure[3]	c_sbp_cont	edbx_get_vitals()	Closest systolic blood pressure (in mmHg) measurement. In	continuous	[-90;0]

Characteristic	Harmonized analysis variable name	R function to derive covariate (see pdf in appendix)	Details[1]	Variable encoding	Assessment window
			case of ties, the lower is selected		
Body mass index (BMI)[3]	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, BMI is (additionally) computed from individual height and weight measurements	continuous	[-90;0]
Heart rate[3]	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]

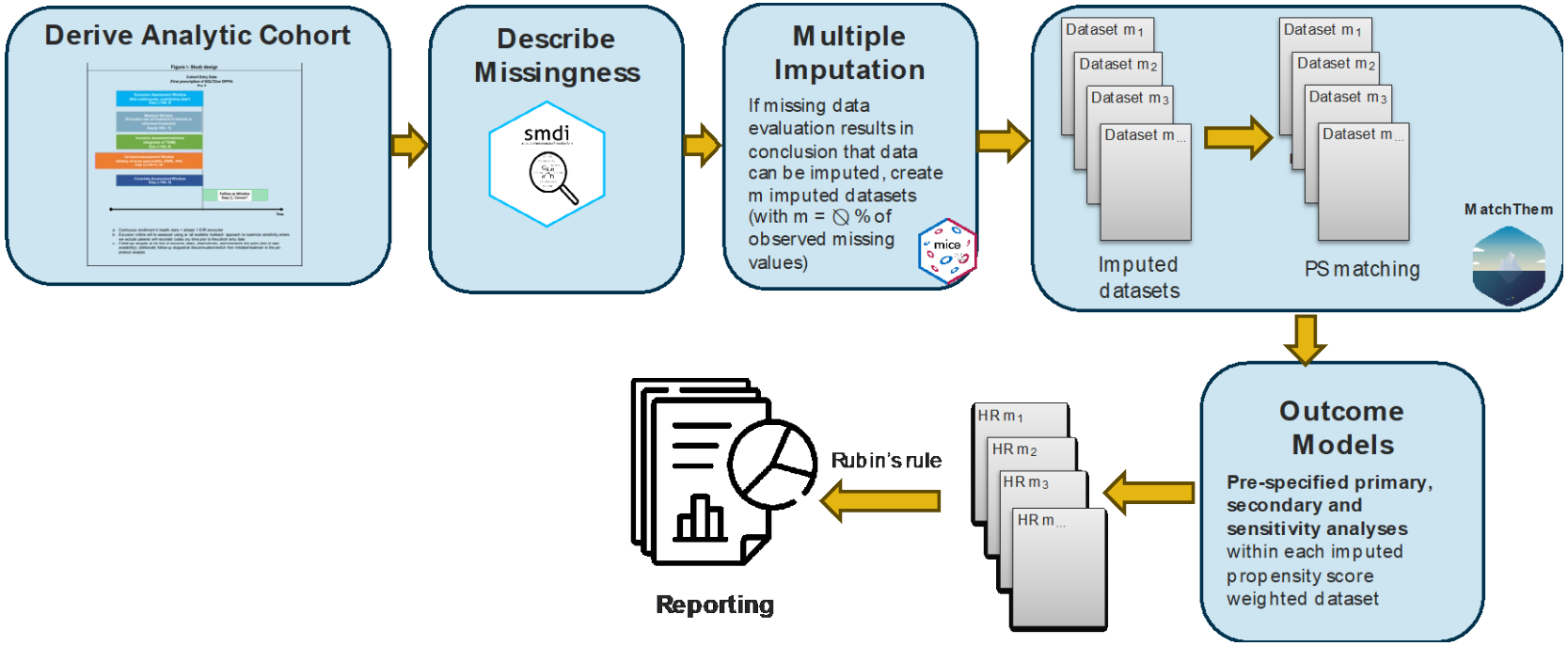
[1] x stands for the pseudonymized number of the respective database, i.e., EDB1, EDB3 or EDB4
[2] Primary tumor sidedness was not included as a covariate in the propensity score, however balance was evaluated after propensity score matching.
[3] For calculation of ROPRO prognostic score[11], this variable is log transformed.
[4] For calculation of ROPRO prognostic score[11], this variable is log-log transformed.

6.7. Data analysis

6.7.1. Context and rationale for analysis plan

To emulate the MAVERICC trial, the following analytical workflow will be used (Figure 2). 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).[16] The analytical cohort will be derived by first identifying the metastatic CRC inception cohort of initiators of FOLFIRI-BV or mFOLFOX-BV 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 2. 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]).[17] 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).[18,19] 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[20] and Little's[21] 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[22] 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[20]/ Little[21]**	**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).	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	No association in either univariate or adjusted model and no meaningful difference in the log HR after full adjustment (~MCAR).

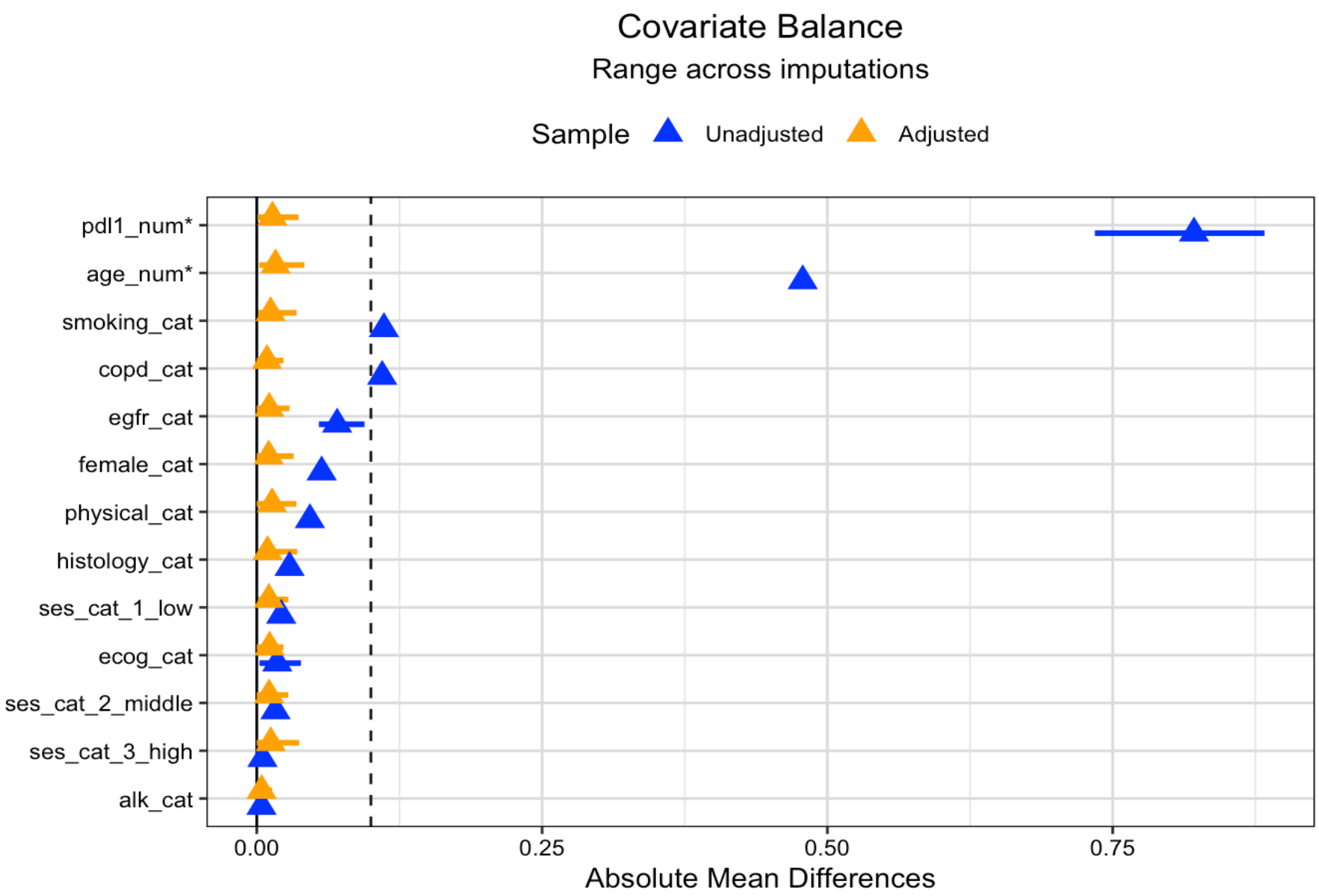
	>0.1[a]: imbalances in observed patient characteristics; missingness may be likely at random (~MAR).		relationships between covariates and missingness (~MAR).	Association in univariate but not fully adjusted model (~MAR). Meaningful difference in the log HR also after full adjustment (~MNAR).

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 patient 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.[23][a] Analogous to propensity score-based balance measures.[23]

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).[24] 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.[25,26] Hence, multiple imputation with flexible, non-parametric random forest imputation algorithms[27] (mice R package[28]) 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 3. Covariate balance across imputed datasets (simulated example).



To estimate the treatment effects for mFOLFOX-BV using propensity score matching across imputed datasets we will apply the "within" approach using the "MatchThem" R package.[29,30] 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 FOLFIRI-BV using a 1:1 nearest neighbor matching algorithm without replacement and a caliper of 0.2 of the standard deviation of the logit 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 3. 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. Confidence intervals will be estimated using cluster-robust standard errors.[31] As a secondary endpoint, we will additionally estimate the median OS survival time difference between the two treatment 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.[28,32] To arrive at a single estimate across databases, the final estimates will be summarized through a meta-analytic fixed effects model.[33] 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 FOLFIRI-BV, compared with initiation of mFOLFOX-BV, is associated with overall survival compatible with the hazard ratio of 0.76 (95% CI, 0.56–1.04) observed in the MAVERICC trial
Exposure contrast:	Initiation of FOLFIRI-BV vs. initiation of mFOLFOX-BV
Outcome:	Time to all-cause mortality
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://gitlab.partners.org/drugepi/encore/mavericc-nct-01374425 (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 0.2 of the standard deviation of the logit of the propensity score.
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 substantive model, i.e., exposure, outcome, confounders/prognostic factors, and additional auxiliary covariates.

Subgroup Analyses	*List all subgroups*
	None (The MAVERICC trial included subgroup analyses stratified by ERCC1 expression; however, ERCC1 status could not be captured in the EDB1 and therefore could not be incorporated into the emulation)

B. Secondary Analysis

Table 8. Secondary analysis specification.

Hypothesis:	Initiation of FOLFIRI-BV, compared with initiation of mFOLFOX-BV, is associated with overall survival compatible with the hazard ratio of 0.76 (95% CI, 0.56–1.04) observed in the MAVERICC trial
Exposure contrast:	Initiation of FOLFIRI-BV vs. initiation of mFOLFOX-BV
Outcome:	Median overall survival time for patients, i.e., time until 50% of the patients in each stratum becomes 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://gitlab.partners.org/drugepi/encore/mavericc-nct-01374425 (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 0.2 of the standard deviation of the logit of the propensity score.
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 substantive model, i.e., exposure, outcome, confounders/prognostic factors, and additional auxiliary covariates.

Subgroup Analyses	*List all subgroups*
	None (The MAVERICC trial included subgroup analyses stratified by ERCC1 expression; however, ERCC1 status could not be captured in the EDB1 and therefore could not be incorporated into the emulation)

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	Approach implements a more conservative censoring rule	Approach addresses ghost-time bias by censoring

		asynchronously. To account for the potential lag of updated mortality information (*ghost-time bias*[34]), 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.[35]		patients without a recorded 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 the most important covariates or those with the highest suspicion of being MNAR will be evaluated
Sensitivity #6	Re-weighting of strong risk factors and/or treatment effect modifiers distribution to match that of MAVERICC	In the presence of effect modification, treatment effect estimates may be different if the distribution of strong risk factors/effect modifiers is different in the emulated cohort versus the trial cohort	Re-weighting adjusts for differences in distributions of key risk factors and/or treatment effect modifiers (see subgroup analysis in Table 7)	Re-weighting risk factors/potential effect modifiers to match the MAVERICC trial and simultaneously balancing them across treatment groups may be challenging due to differences in measurement

Sensitivity #7	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

6.8. Study size and feasibility

See appendix 2 for covariate balance before and after the propensity score matching. Power and unstratified number of events are available in appendix 3. Propensity score distribution plots with c-statistics are available in Appendix 4. Covariate balance for primary tumor sidedness (left-sided, right-sided, or unknown) was assessed separately as an external balance diagnostic because sidedness was not included in the propensity score model, as described in Section 6.6.5. Results are presented in Appendix 5.

7. Limitations 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 #7 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.

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. Parikh AR, Lee FC, Yau L, et al. MAVERICC, a Randomized, Biomarker-stratified, Phase II Study of mFOLFOX6-Bevacizumab versus FOLFIRI-Bevacizumab as First-line Chemotherapy in Metastatic Colorectal Cancer. *Clin Cancer Res Off J Am Assoc Cancer Res*. 2019;25(10):2988-2995. doi:10.1158/1078-0432.CCR-18-1221

6. 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

7. 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.

8. Real-world uptake and effectiveness of triplet versus doublet chemotherapy with bevacizumab in patients with metastatic colorectal cancer in the United States - ClinicalKey. Accessed June 24, 2026. https://www.clinicalkey.com/#!/content/playContent/1-s2.0-S2949819824000487?returnurl=https:%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS2949819824000487%3Fshowall%3Dtrue&referrer=

9. Arnold D, Lueza B, Douillard JY, et al. Prognostic and predictive value of primary tumour side in patients with RAS wild-type metastatic colorectal cancer treated with chemotherapy and EGFR directed antibodies in six randomized trials. *Ann Oncol Off J Eur Soc Med Oncol*. 2017;28(8):1713-1729. doi:10.1093/annonc/mdx175

10. 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

11. 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

12. 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

13. 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

14. 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

15. 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

16. 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

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

18. 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

19. 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

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

21. 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

22. 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

23. 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

24. 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

25. 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

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

27. 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

28. 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

29. 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

30. 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

31. 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

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

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

34. 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

35. 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.

36. Treatment of Metastatic Colorectal Cancer: ASCO Guideline | Journal of Clinical Oncology. Accessed August 19, 2026. https://ascopubs.org/doi/10.1200/JCO.22.01690?utm_source=chatgpt.com

37. Arnold D, Lueza B, Douillard JY, et al. Prognostic and predictive value of primary tumour side in patients with RAS wild-type metastatic colorectal cancer treated with chemotherapy and EGFR directed antibodies in six randomized trials. *Ann Oncol Off J Eur Soc Med Oncol*. 2017;28(8):1713-1729. doi:10.1093/annonc/mdx175

38. Holch JW, Ricard I, Stintzing S, Modest DP, Heinemann V. The relevance of primary tumour location in patients with metastatic colorectal cancer: A meta-analysis of first-line clinical trials. *Eur J Cancer*. 2017;70:87-98. doi:10.1016/j.ejca.2016.10.007

39. Swami N, Hwang WT, Mamtani R, O'Hara MH, Chapin WJ. Second-Line Treatment Strategies for Right-Sided, RAS/RAF Wild-Type Colorectal Cancer. *JAMA Netw Open*. 2025;8(6):e2515087. doi:10.1001/jamanetworkopen.2025.15087

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

41. Solar Vasconcelos JP, Chen N, Titmuss E, et al. Transverse Colon Primary Tumor Location as a Biomarker in Metastatic Colorectal Cancer: A Pooled Analysis of CCTG/AGITG CO.17 and CO.20 Randomized Clinical Trials. *Clin Cancer Res Off J Am Assoc Cancer Res*. 2024;30(6):1121-1130. doi:10.1158/1078-0432.CCR-23-3275

10. Appendices

10.1. CONSORT diagrams

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

Figure 4. CONSORT attrition to select eligible MAVERICC-like populations in EDB1.

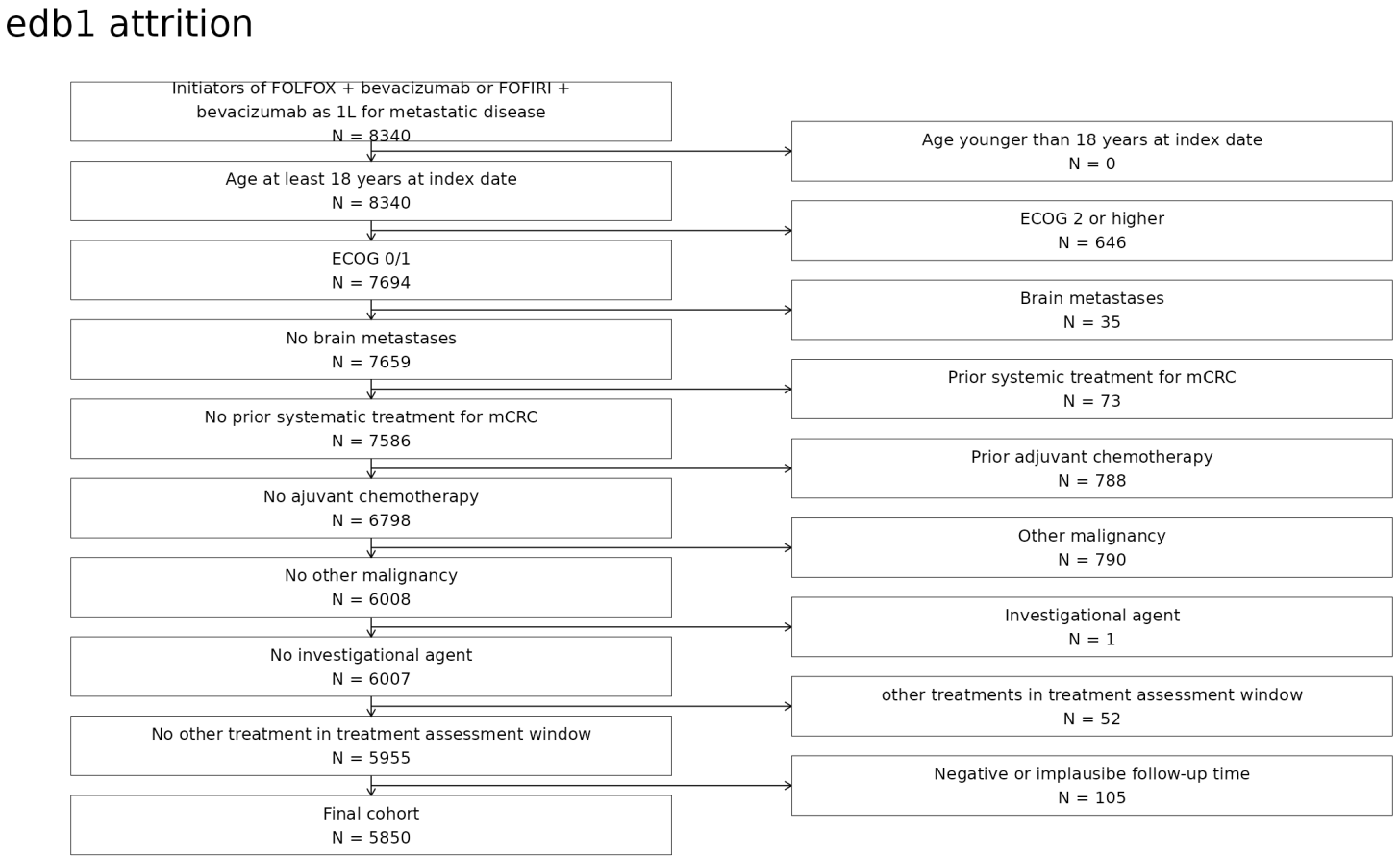


Figure 2. CONSORT attrition to select eligible MAVERICC-like populations in EDB3.

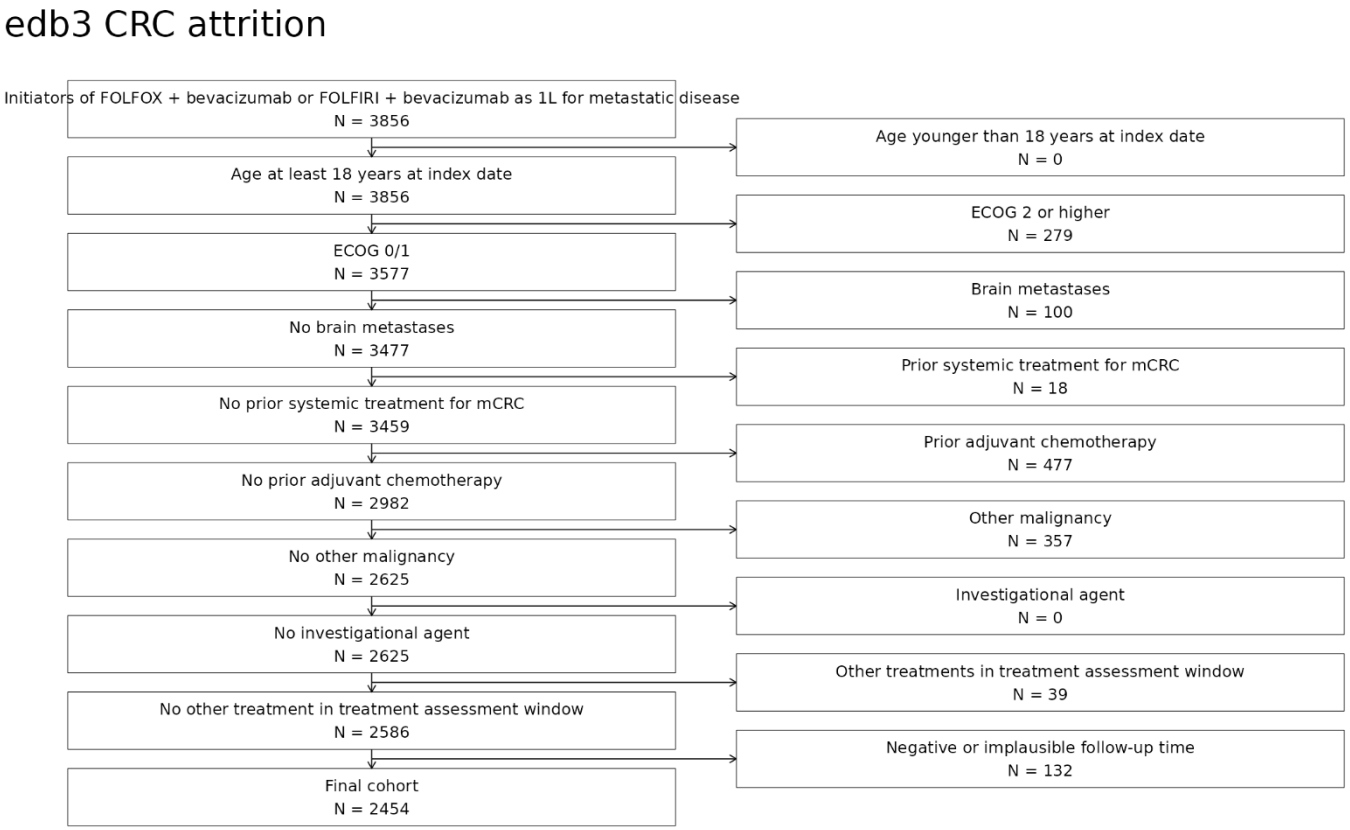
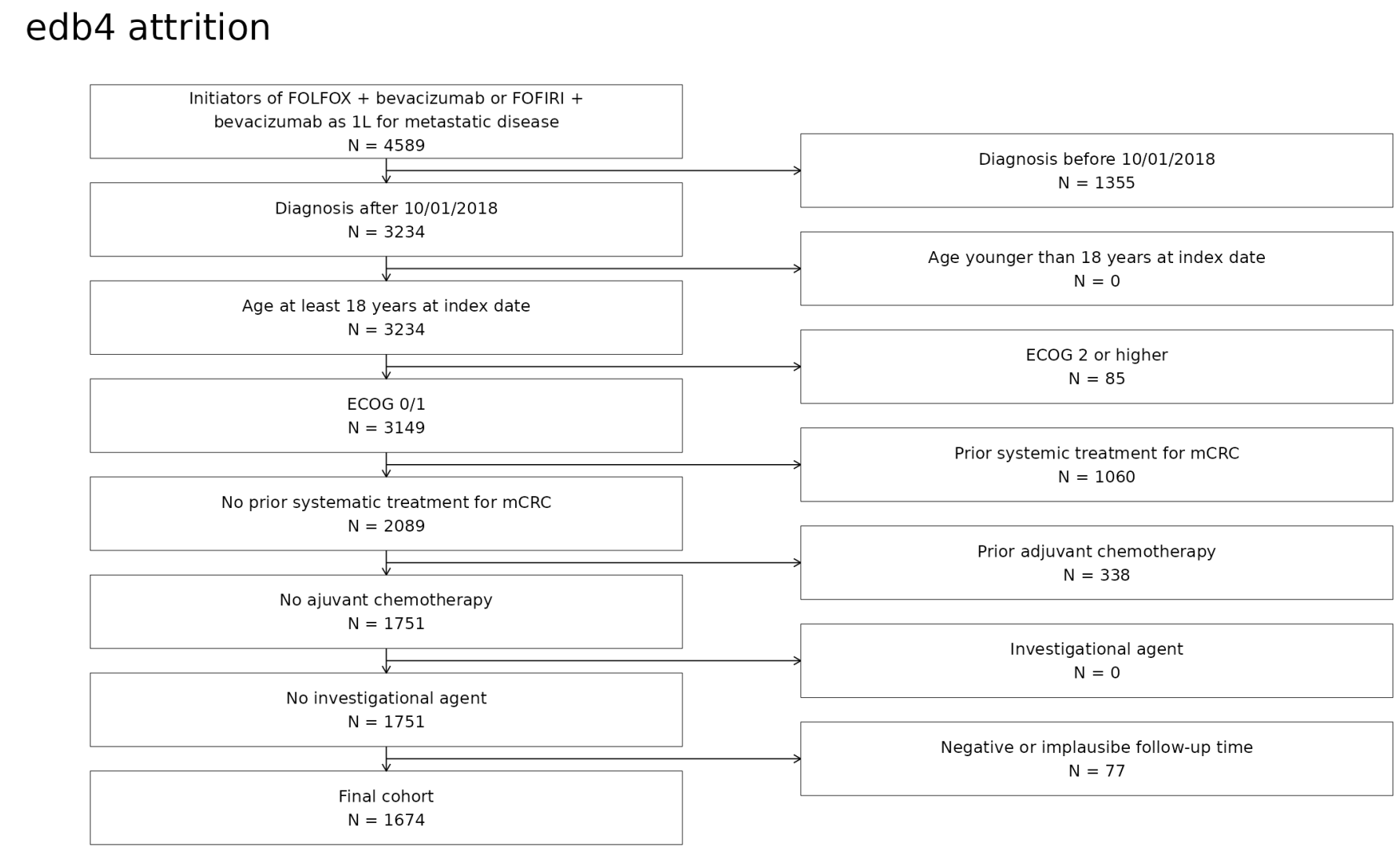


Figure 6. CONSORT attrition to select eligible MAVERICC-like populations in EDB4.



10.2. Covariate balance figures

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

Figure 7. EDB1 covariate balance of covariates included in propensity score model before and after matching.

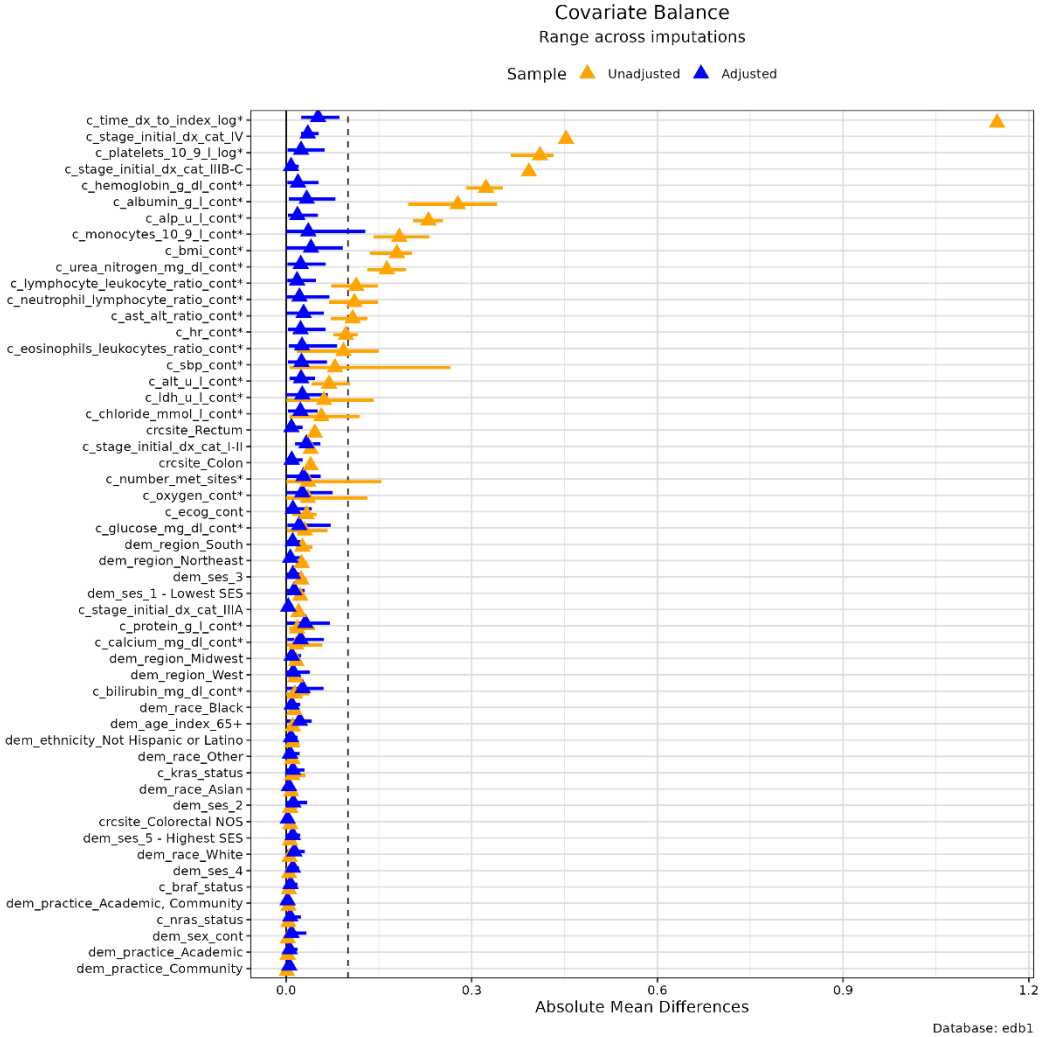


Figure 8. EDB3 covariate balance of covariates included in propensity score model before and after matching.

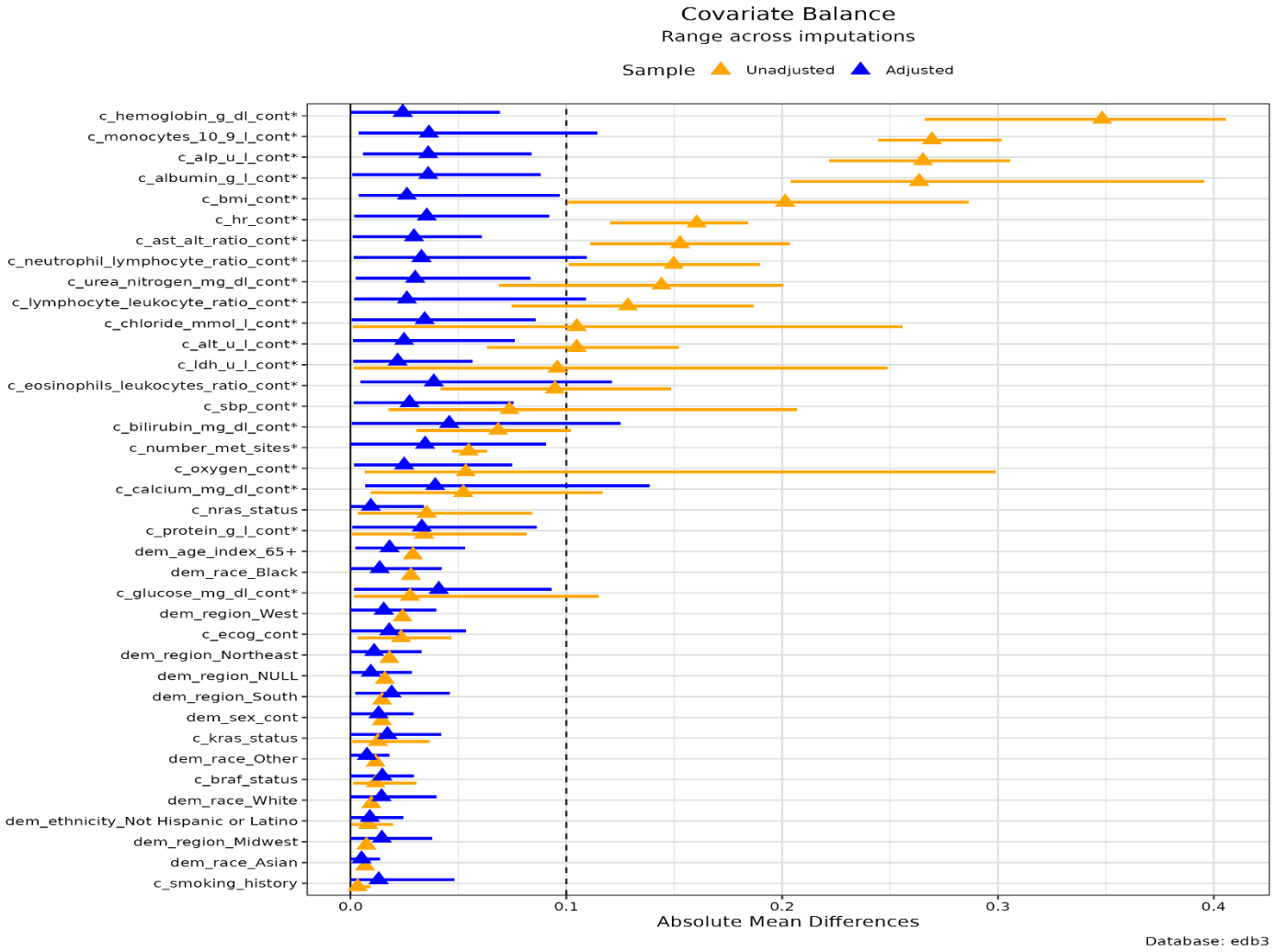
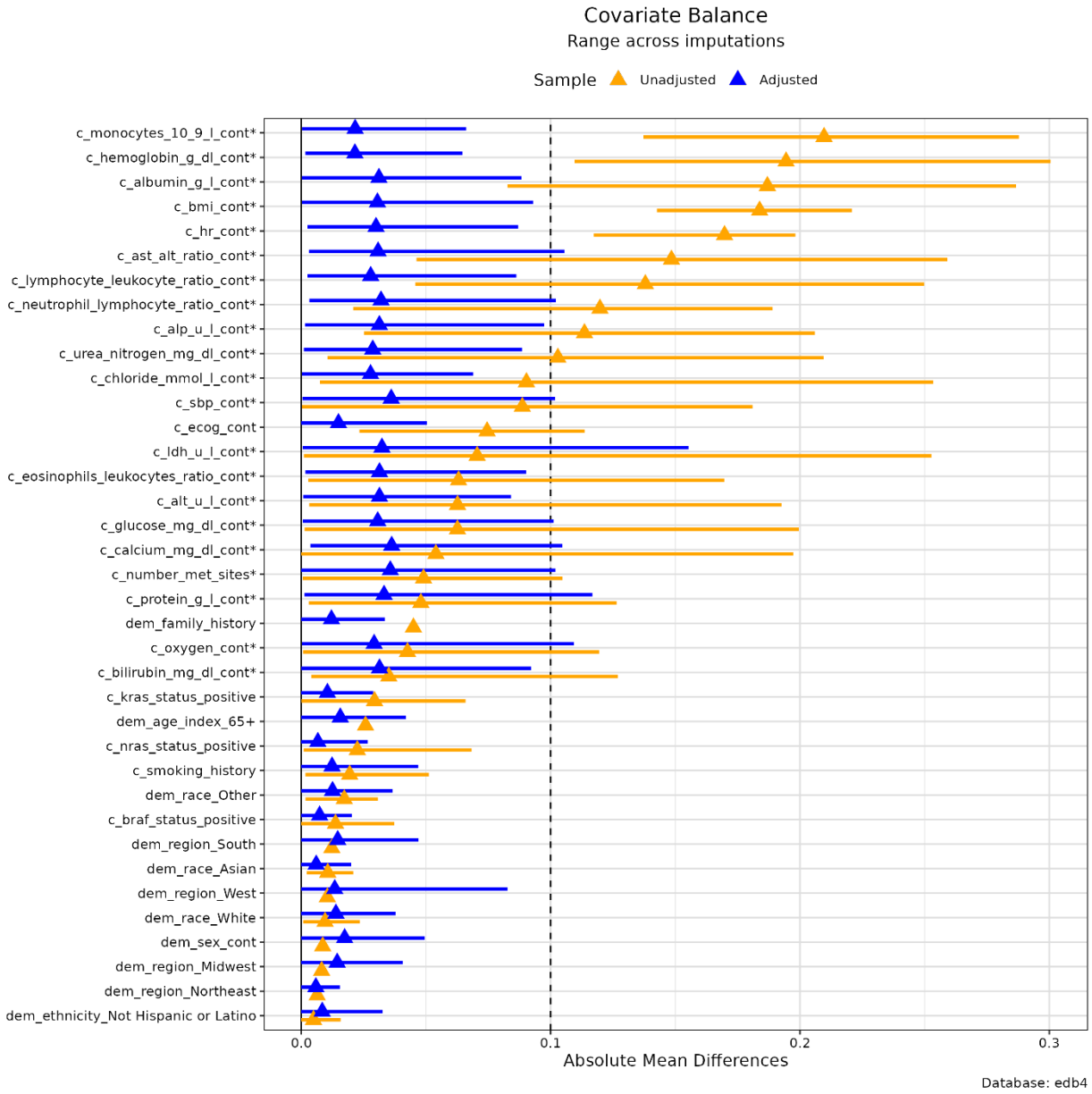


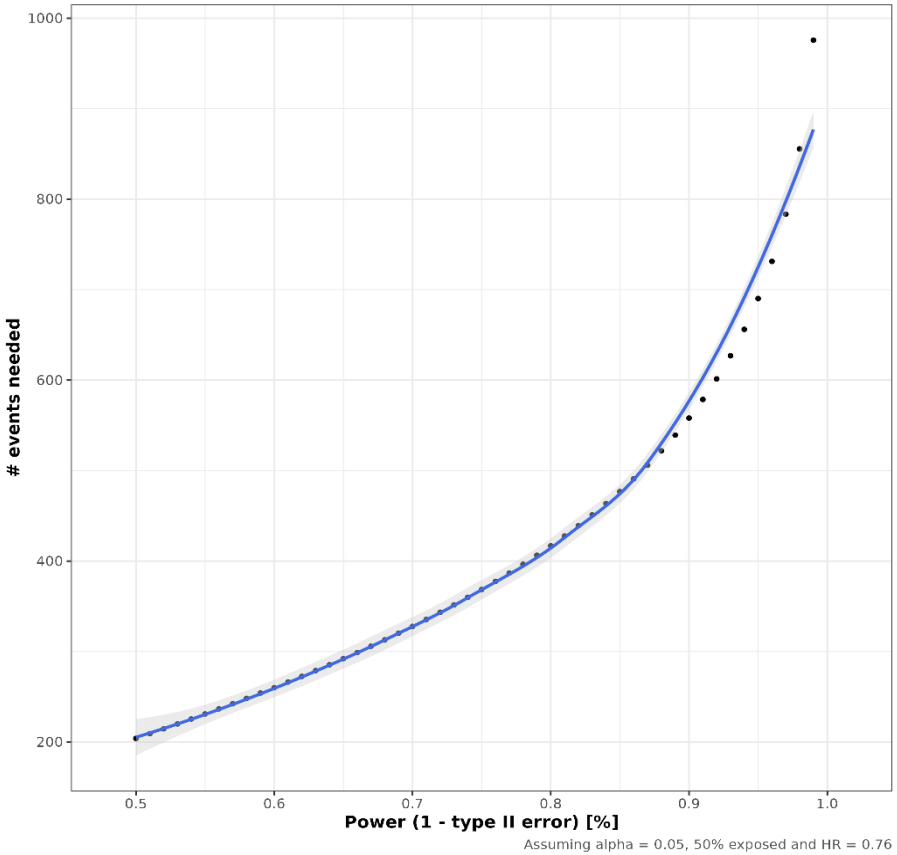
Figure 9. EDB4 covariate balance of covariates included in propensity score model before and after matching.



10.3. Sample size/power calculations

Power estimations are computed based on the average number of observed events unstratified by exposure groups, across imputed and matched datasets, according to the methodology described by Schoenfeld.[40] MAVERICC reported the HR as HR = 0.76 (95% CI, 0.56–1.04). Assuming a true HR of 0.76 and a 2-sided α of 0.05, the estimated statistical power to detect a treatment effect was 99.8% in EDB1 (1,259 events), 92.8% in EDB3 (621 events), and 85.9% in EDB4 (490 events), with 99.9% power for the pooled treatment-effect estimate from the fixed-effect meta-analysis across the 3 databases.

Figure 10. Number of events needed to achieve x% power.



10.4. Propensity Score Distribution and C-statistic

Figure 11. Propensity Score Distribution before and after matching in EDB 1

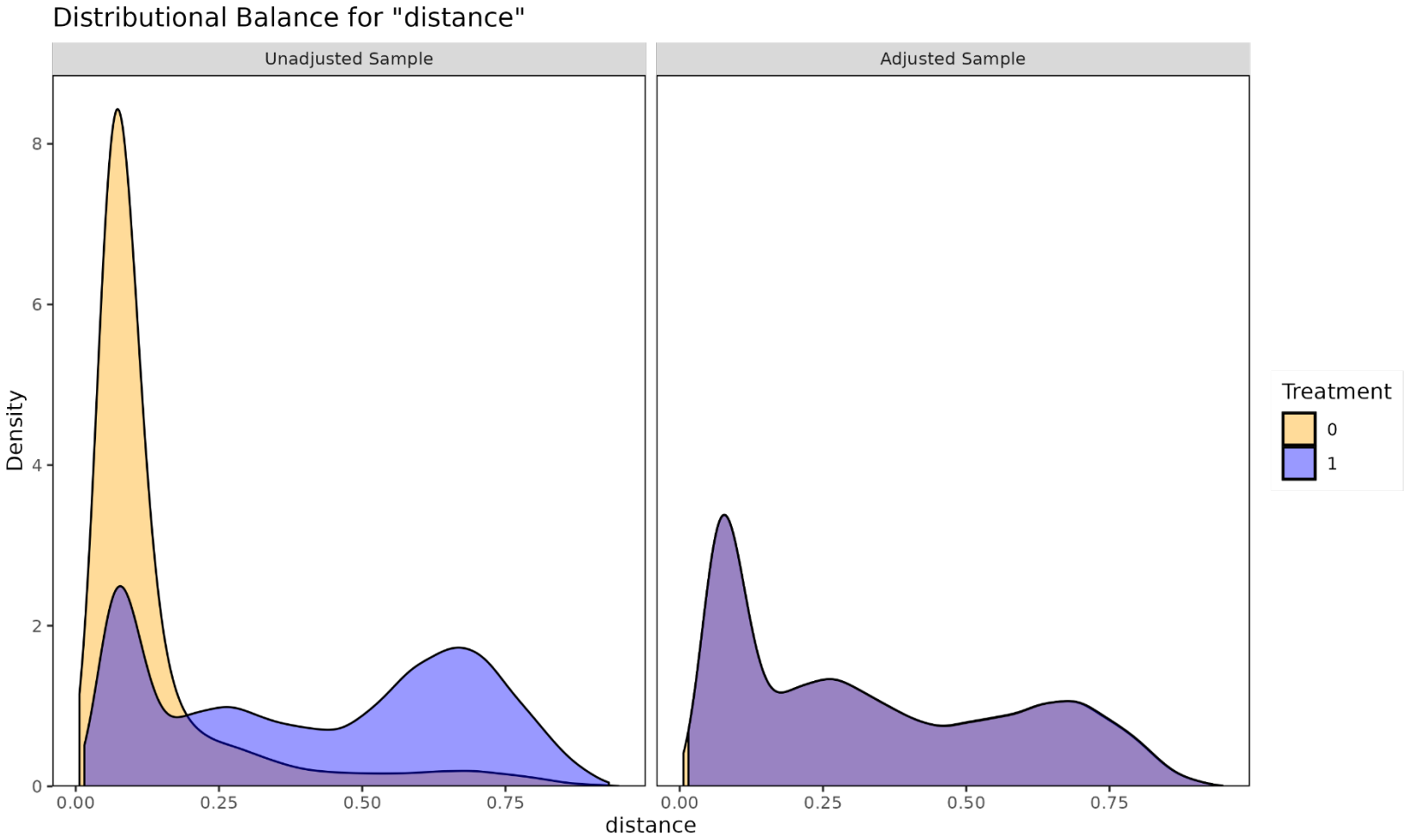


Table 10: C-statistics in the unmatched and matched patient population in EDB 1

matching_group	mean	min	max
unmatched	0.816	0.814	0.820
matched	0.500	0.499	0.501

Figure 12. Propensity Score Distribution before and after matching in EDB 3

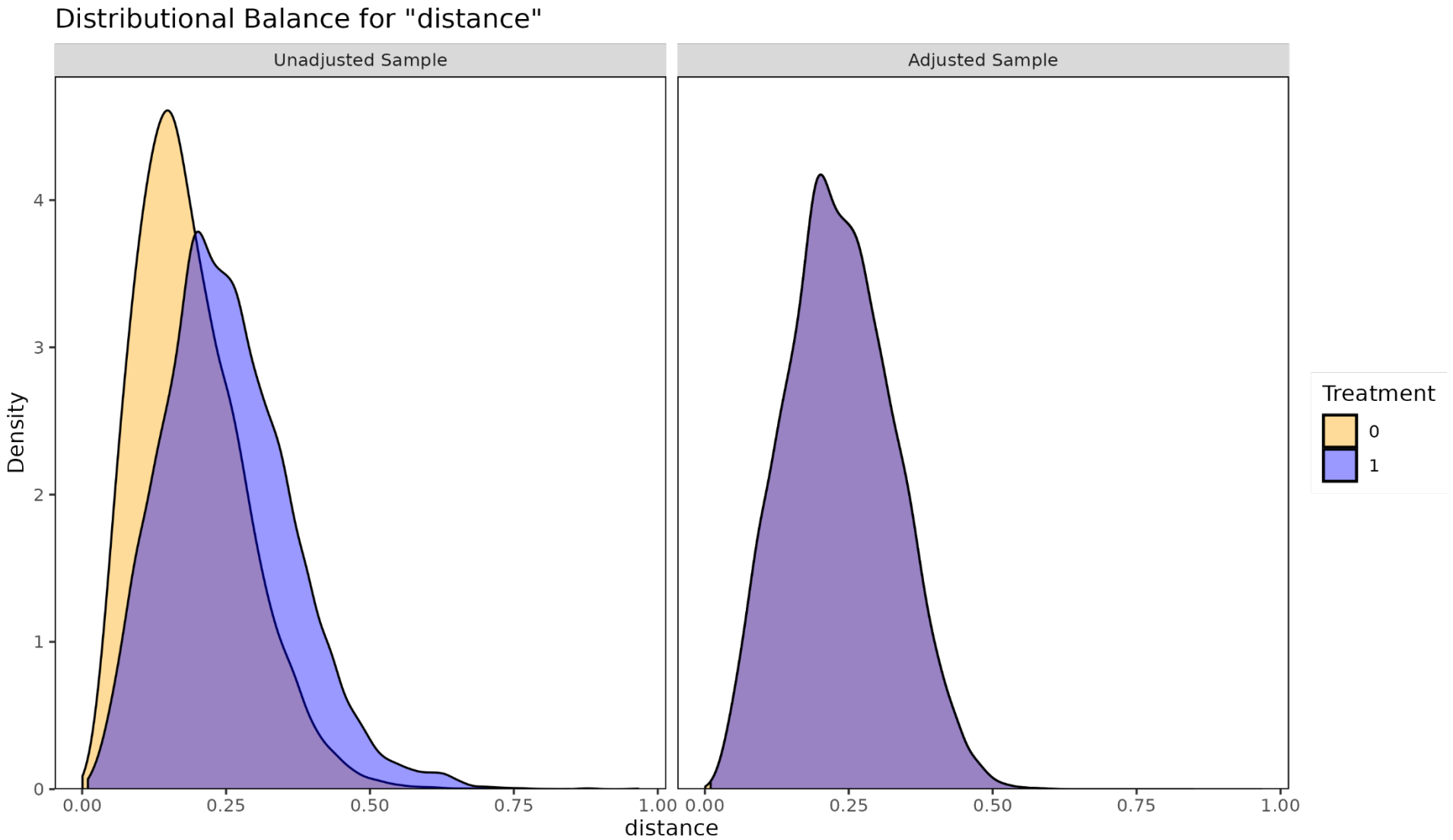


Table 11: C-statistics in the unmatched and matched patient population in EDB 3

matching_group	mean	min	max
unmatched	0.677	0.665	0.700
matched	0.500	0.499	0.500

Figure 13. Propensity Score Distribution before and after matching in EDB 4

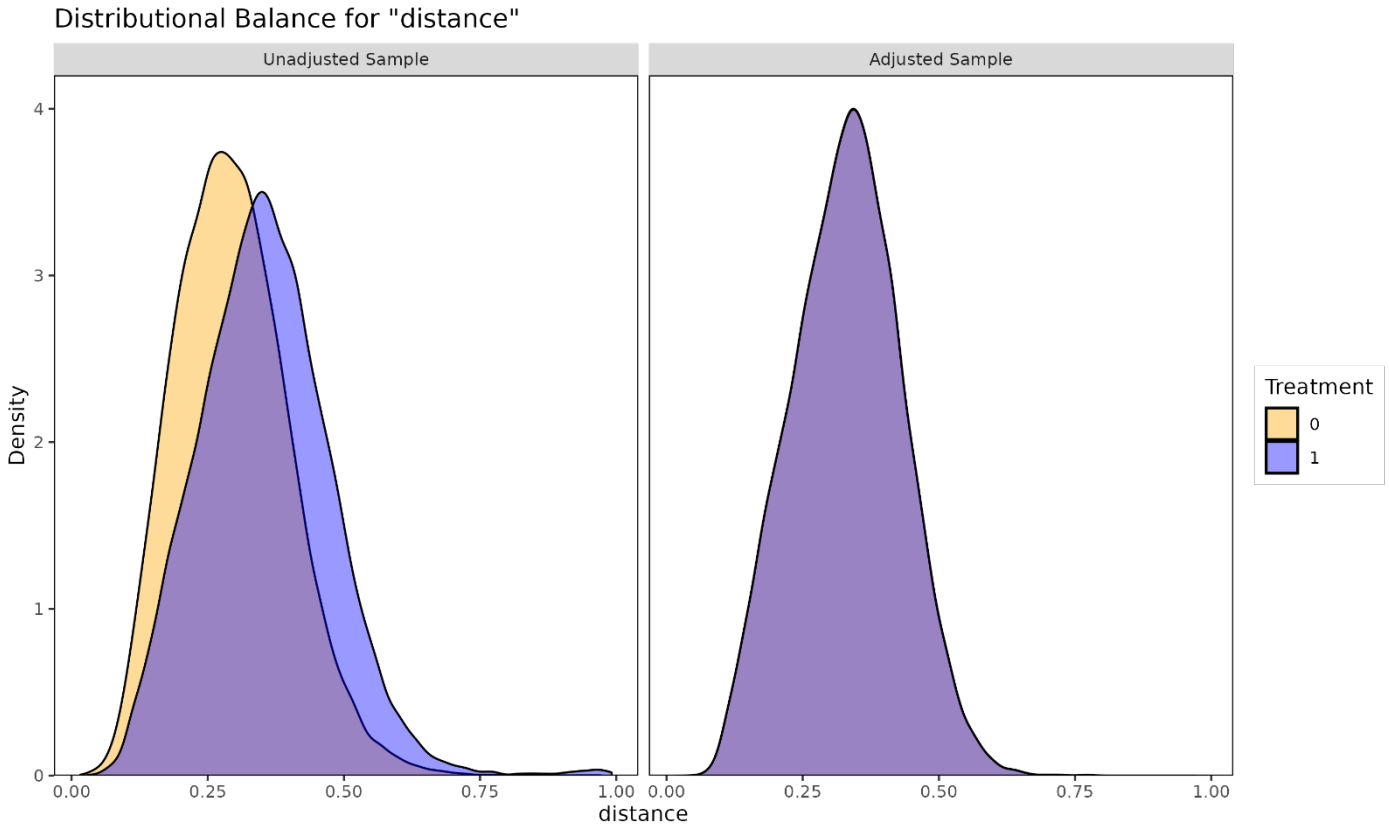

Distributional Balance for "distance"
Unadjusted Sample
Adjusted Sample
Density
distance
0.00
0.25
0.50
0.75
1.00
0
1
2
3
4
Treatment
0
1

Table 12: C-statistics in the unmatched and matched patient population in EDB 4

matching_group	mean	min	max
unmatched	0.650	0.624	0.668
matched	0.500	0.499	0.501

10.5. Sidedness balance figures

Figure 14. EDB1 covariate balance of sidedness

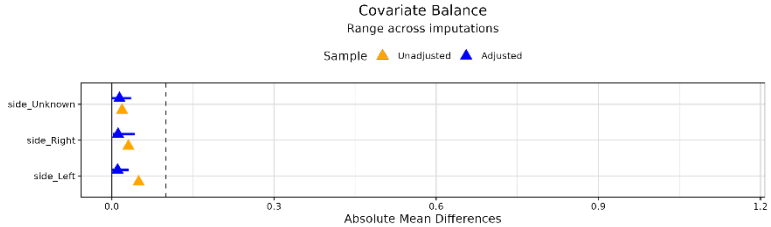


Figure 15. EDB3 covariate balance of sidedness

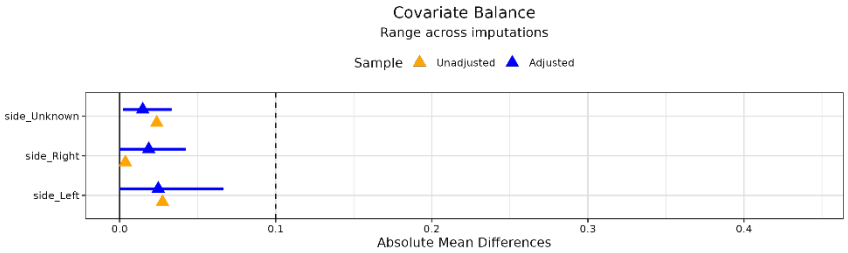
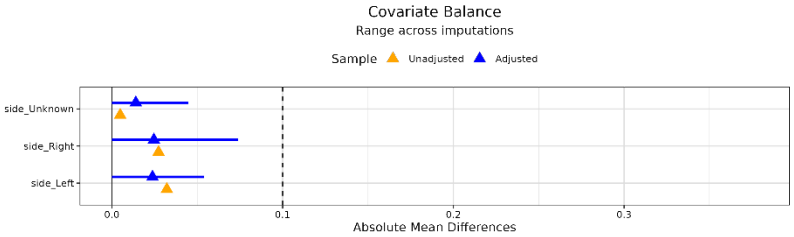


Figure 16. EDB4 covariate balance of sidedness



10.6. Additional Figures and Tables

Table 13. 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 14. 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 15. 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

Table 16. ICD-10 Code–Based Classification of Primary Colorectal Tumor Sidedness[39]

Anatomic subsite	ICD-9-CM	ICD-10-CM	Side
Cecum	153.4	C18.0	Right
Ascending colon	153.6	C18.2	Right
Hepatic flexure	153.0	C18.3	Right
Transverse colon	153.1	C18.4	Right[1]
Splenic flexure	153.7	C18.5	Left
Descending colon	153.2	C18.6	Left
Sigmoid colon	153.3	C18.7	Left
Rectosigmoid junction	154.0	C19	Left
Rectum	154.1	C20	Left
Colon, overlapping / unspecified / parent code	153.8 / 153.9 / 153	C18.8 / C18.9 / C18	Unknown

[1] Although the transverse colon spans the conventional right- and left-sided boundary, with the proximal two-thirds considered right-sided and the distal one-third left-sided, survival outcomes for transverse colon cancer appear to more closely resemble those of right-sided colon cancer.[41]

Figure 17. Treatment initiation trends by calendar year and treatment in EDB1

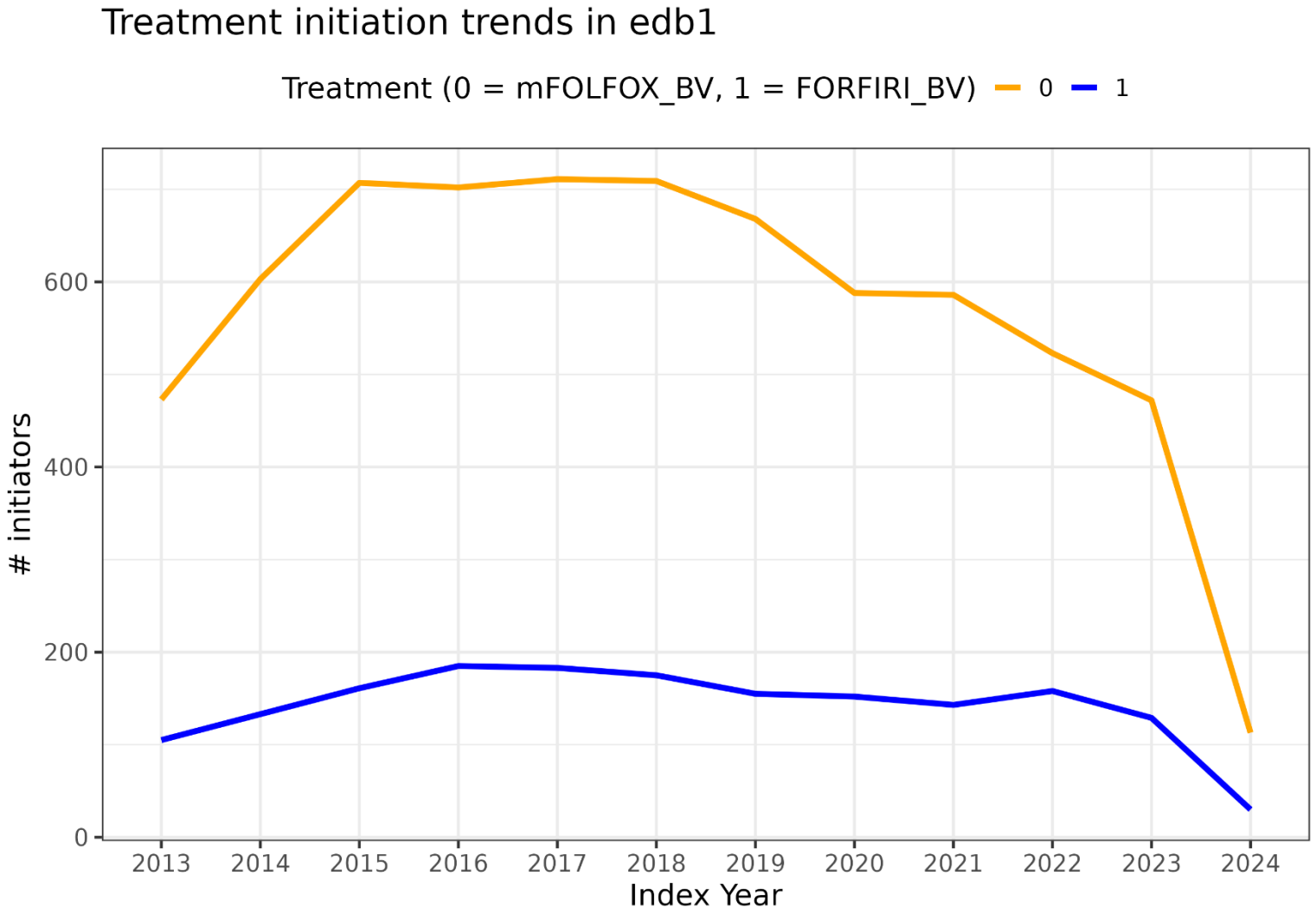


Figure 18. Treatment initiation trends by calendar year and treatment in EDB3

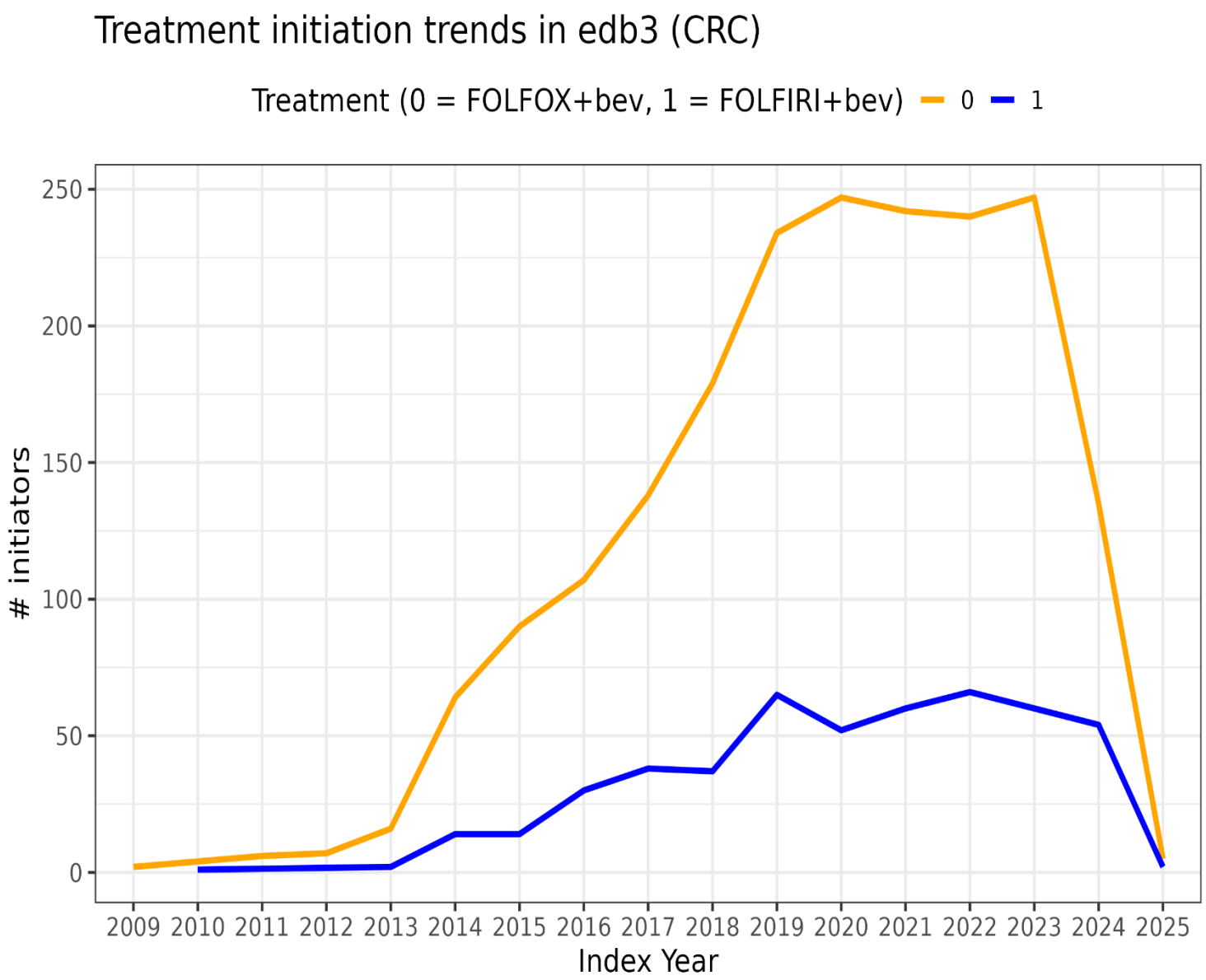
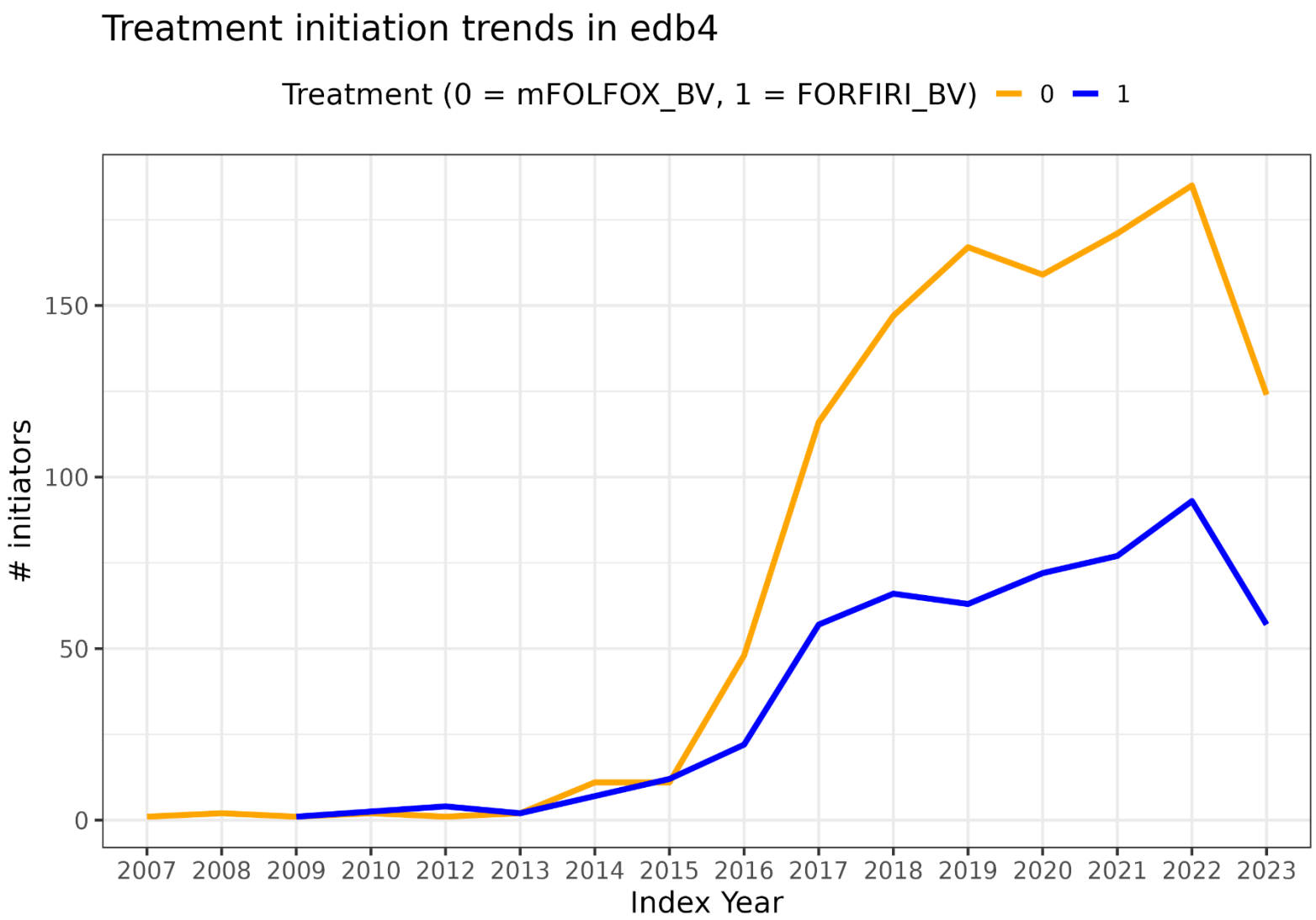


Figure 19. Treatment initiation trends by calendar year and treatment in EDB4*



*Patients were included only if they initiated mFOLFOX-BV or FOLFIRI-BV between 10/01/2018 and 09/30/2023.

NCTID	NCT01374425
Acronym	**MAVERICC**
Protocol	https://clinicaltrials.gov/study/NCT01374425#contacts-and-locations
SAP	https://clinicaltrials.gov/study/NCT01374425#contacts-and-locations
PMID	https://pmc.ncbi.nlm.nih.gov/articles/PMC9022943/
Indication	previously untreated metastatic colorectal cancer
Line of Therapy	1
Exposures	Leucovorin/5-fluorouracil/irinotecan plus bevacizumab (FOLFIRI-BV)
Comparisons	Modified leucovorin/5-fluorouracil/oxaliplatin plus bevacizumab (mFOLFOX6-BV)
Emulated outcome	Overall survival (secondary end point)

Key Inclusion/Exclusion Criteria					Comment
Criteria	**RCT eligibility criteria (taken from the original trial protocol)**	**Clinical relevance**	**Emulation [EDB1]**	**Emulation [EDB4]**	
Inclusion 1	18 Years and older (Adult, Older Adult)	Relevant	Possible	Possible	To protect privacy, most databases only provide month- or year-level granularity of dates
Inclusion 2	Histologically or cytologically confirmed colorectal cancer (CRC) with at least one measurable metastatic lesion by RECIST Version 1.1	Relevant	Possible	Possible	
Inclusion 3	Archival tumor tissue sample must be requested and available prior to study entry. If no archival tumor tissue sample is available, a fresh biopsy tissue sample must be obtained but should be discussed first with the medical monitor. A copy of the local pathology report must be submitted along with the specimens.	Limited relevance	Not implementable	Not implementable	
Inclusion 4	Eastern Cooperative Oncology Group (ECOG) performance status of 0 or 1	Relevant	Possible	Possible	
Inclusion 5	Participants with treated brain metastases are eligible for study participation. Participants may not receive ongoing treatment with steroids at Screening. Anticonvulsants (at stable dose) are allowed. Treatment for brain metastases may be whole-brain radiotherapy, radiosurgery, neurosurgery, or a combination as deemed appropriate by the treating physician. Radiotherapy and stereotactic radiosurgery must be completed at least 28 days prior to randomization.	Relevant	Limited	Limited	Brain metastasis treatment history and steroid use at screening are incompletely captured in the data. Patients with evidence of brain metastases are excluded.
Inclusion 6	Female participants should not be pregnant or breastfeeding. Female participants with childbearing potential should agree to use effective, non-hormonal means of contraception during the study and for a period of at least 6 months following the last administration of study drugs. Female participants with an intact uterus (unless amenorrheic for the last 24 months) must have a negative serum pregnancy test within 7 days prior to randomization into the study.	Limited relevance	Not implementable	Not implementable	This criterion is assumed to be met among treated patients.
Inclusion 7	Male participants must agree to use effective contraception during the study and for a period of at least 6 months following the last administration of study drugs, even if they have been surgically sterilized.	Limited relevance	Not implementable	Not implementable	This criterion is assumed to be met among treated patients.
Exclusion 1	Any prior systemic treatment for mCRC	Relevant	Possible	Possible	
Exclusion 2	Adjuvant chemotherapy for CRC completed <12 months	Relevant	Possible	Possible	
Exclusion 3	Evidence of Gilbert's syndrome or of homozygosity for the UGT1A1*28 allele	Limited relevance	Not implementable	Not implementable	
Exclusion 4	Known positivity for human immunodeficiency virus (HIV)	Limited relevance	Not implementable	Not implementable	HIV status is incompletely captured. This criterion is therefore not implemented.
Exclusion 5	Malignancies other than mCRC within 5 years prior to randomization, except for adequately treated carcinoma in situ of the cervix, basal or squamous cell skin cancer, localized prostate cancer treated surgically with curative intent, and ductal carcinoma in situ treated surgically with curative intent	Relevant	Limited	Limited	Exclude patients with evidence of malignancies other than mCRC within 5 years before index, except for carcinoma in situ of the cervix, basal or squamous cell skin cancer, localized prostate cancer, and ductal carcinoma in situ.
Exclusion 6	Radiotherapy to any site for any reason within 28 days prior to randomization, except for palliative radiotherapy to bone lesions within 14 days prior to randomization	Relevant	Not implementable	Not implementable	Radiotherapy and timing are not reliably captured.
Exclusion 7	Clinically detectable third-space fluid collections that cannot be controlled by drainage or other procedures prior to study entry	Limited relevance	Not implementable	Not implementable	
Exclusion 8	Treatment with any other investigational agent, or participation in another investigational drug trial within 28 days prior to randomization	Relevant	Possible	Possible	Exclude patients who received an investigational agent within 28 days before first-line treatment initiation.

NCTID NCT01374425
Acronym **MAVERICC**
Protocol https://clinicaltrials.gov/study/NCT01374425#contacts-and-locations
SAP https://clinicaltrials.gov/study/NCT01374425#contacts-and-locations
PMID https://pmc.ncbi.nlm.nih.gov/articles/PMC9022943/
Indication previously untreated metastatic colorectal cancer
Line of Therapy 1
Exposures Leucovorin/5-fluorouracil/irinotecan plus bevacizumab (FOLFIRI-BV)
Comparisons Modified leucovorin/5-fluorouracil/oxaliplatin plus bevacizumab (mFOLFOX6-BV)
Emulated outcome Overall survival (secondary end point)

Measurement eligibility criteria							
Criteria	**Criteria in trial protocol**	**Criteria rule as defined in original protocol**	**Time point/period of emulated measurement [days]**	**Emulation [EDB1]**	**Emulation [EDB4]**	**Comment**	**encore.io function**
Inclusion	Inclusion 1	18 Years and older (Adult, Older Adult)	[0;0]	Age at index date (year granularity level)	Age at index date (year granularity level)		edbx_get_demographics()
Inclusion	Inclusion 2	Histologically or cytologically confirmed colorectal cancer (CRC) with at least one measurable metastatic lesion by RECIST Version 1.1	[-inf;0]	Line of therapy needs to be for "Metastatic" setting (LoT table)	Any evidence of at least one distant metastasis at any time before the index date (inclusive). This captures both de novo metastatic patients and those who progressed/developed metastases before/on the index date.		EDB1: Part of the exposure definition EDB4: Derived from edbx_get_diagnosis()
Inclusion	Inclusion 4	Eastern Cooperative Oncology Group (ECOG) performance status of 0 or 1	[-90;0]	ECOG = 0 | 1	ECOG = 0 | 1		edbx_get_ecog()
Inclusion	Inclusion 5	Participants with treated brain metastases are eligible for study participation. Participants may not receive ongoing treatment with steroids at Screening. Anticonvulsants (at stable dose) are allowed. Treatment for brain metastases may be whole-brain radiotherapy, radiosurgery, neurosurgery, or a combination as deemed appropriate by the treating physician. Radiotherapy and stereotactic radiosurgery must be completed at least 28 days prior to randomization.	[-inf;0]	Patients with evidence of brain metastases are excluded.	Patients with evidence of brain metastases are excluded.	Because treatment history for brain metastases and steroid use at screening are incompletely captured, patients with any evidence of brain metastases are excluded as a conservative proxy.	
Exclusion	Exclusion 1	Any prior systemic treatment for mCRC	[MCRCDD;-1]	Exclude patients who received prior cancer treatment for mCRC.	Exclude patients who received prior cancer treatment for mCRC		
Exclusion	Exclusion 2	Adjuvant chemotherapy for CRC completed <12 months	[-365;-1]	Exclude patients who received chemotherapy within 12 months before the index date.	Exclude patients who received chemotherapy within 12 months before the index date.	Cannot reliably determine whether prior chemotherapy was adjuvant because of incomplete treatment setting information in the data.	
Exclusion	Exclusion 5	Malignancies other than mCRC within 5 years prior to randomization, except for adequately treated carcinoma in situ of the cervix, basal or squamous cell skin cancer, localized prostate cancer treated surgically with curative intent, and ductal carcinoma in situ treated surgically with curative intent	[-1825;0]	Exclude patients with another primary malignancy within 5 years before the index date, except carcinoma in situ of the cervix, basal or squamous cell skin cancer, localized prostate cancer, and ductal carcinoma in situ.	Exclude patients with another primary malignancy within 5 years before the index date, except carcinoma in situ of the cervix, basal or squamous cell skin cancer, localized prostate cancer, and ductal carcinoma in situ.		
Exclusion	Exclusion 8	Treatment with any other investigational agent, or participation in another investigational drug trial within 28 days prior to randomization	[-28;0]	Exclude patients who received investigational agents 28 days prior to randomization	Exclude patients who received investigational agents 28 days prior to randomization		

MCRCDD Metastatic colorectal cancer diagnosis date