1. Title Page

Title	CARE Initiative Study: Real-world emulation of the PALOMA-2 comparative effectiveness trial of palbociclib and letrozole vs. placebo and letrozole for the first-line treatment of ER+/HER2- advanced breast cancer
Research Question and Objectives	This study seeks to advance understanding of under what conditions real-world evidence studies can provide reliable conclusions about drug effectiveness. The objective is to emulate the PALOMA-2 randomized controlled trial of palbociclib and letrozole as first-line treatment in postmenopausal adult female patients with ER+/HER2- advanced breast cancer using real-world electronic health record data.
Protocol version	v1.1
Last update date	August 13, 2024
Contributors	Paige Sheridan, PhD Aetion, Inc 5 Penn Plaza 7th Floor New York, NY 10001 Natalie Levy, PhD Aetion, Inc 5 Penn Plaza 7th Floor New York, NY 10001 Nileesa Gautam, MSc Aetion, Inc 50 Congress St. Suite 1025 Boston, MA 02109 Adina Estrin, MS

	Aetion, Inc 5 Penn Plaza 7th Floor New York, NY 10001 Monica Iyer Aetion, Inc 5 Penn Plaza 7th Floor New York, NY 10001 Sarah McDonald Aetion, Inc 5 Penn Plaza 7th Floor New York, NY 10001 Ann Madsen, PhD Aetion, Inc 5 Penn Plaza 7th Floor New York, NY 10001 Ulka Campbell, PhD Aetion, Inc 5 Penn Plaza 7th Floor New York, NY 10001
Sponsor	This work was collectively sponsored by Abbvie, Amgen, AstraZeneca, Bayer, Gilead, Janssen, and Pfizer.
Conflict of interest	PS, NL, NG, AE, MI, SM, AM, and UC are shareholders and employees of Aetion, Inc.

Table of contents

2. Abstract

Background

The Coalition to Advance Real-World Evidence through Randomized Controlled Trial Emulation (CARE) Initiative is a program designed to build an empirical evidence base for the use of real-world data (RWD) in clinical and regulatory decision-making.[1] Using randomized controlled trials (RCT) as a benchmark for causal effect estimates, a series of RCT emulations will be conducted across varying trials, real world data sources, and study design elements to better understand under what conditions non-interventional studies, using data generated during routine clinical care, can provide reliable conclusions about drug effectiveness.

Research Question and Objectives

In this study, real-world electronic health record (EHR) data will be used to emulate the Palbociclib: Ongoing Trials in the Management of Breast Cancer (PALOMA-2) efficacy trial of palbociclib as first-line therapy in patients with estrogen receptor-positive (ER+)/human epidermal growth factor receptor 2-negative (HER2-) advanced breast cancer.[2] Similarly to the PALOMA-2 trial, this study will compare real-world progression-free survival (rwPFS) between patients who initiate palbociclib and letrozole and those treated with letrozole alone.

Research Methods

The inclusion and exclusion criteria applied in the PALOMA-2 trial will be operationalized in a real-world EHR data source, as closely as is feasible, to create an observational cohort similar to the trial study population. Patients who initiate first-line treatment with palbociclib and letrozole within 15 days of one another in the metastatic setting will be considered exposed. Patients who initiate letrozole only with no evidence of palbociclib within 15 days of letrozole initiation will comprise the comparator group. Inverse probability of treatment weighting (IPTW) will be used to control for measured hypothesized confounders. Kaplan-Meier methods and Cox proportional hazards models will be used to compare median rwPFS and hazards of progression and death between the exposure and comparator groups.

3. Amendments and updates

Version date	Version number	Section of protocol	Amendment or update	Reason
August 13, 2024	1.1	Tables 4, 5, 9	Amendments to some variable assessment windows	To improve emulation of trial criteria and measures

4. Milestones

Table 1. Milestones

Milestone	Date
Initial feasibility assessment	March 21, 2024
Additional data explorations	April 8, 2024
Draft 1 of protocol complete	May 30, 2024
Final protocol shared with steering committee	June 28, 2024
Amended protocol shared with steering committee	August 13, 2024

5. Rationale and background

The potential of non-interventional studies using real-world data (RWD) — healthcare data generated during routine clinical practice — to produce evidence about the effectiveness and safety of biomedical products is increasingly recognized by clinical and regulatory decision makers.[3,4] This is reflected by the growing use of RWD to support regulatory approvals.[5] Real-world evidence (RWE) studies complement randomized controlled trials (RCTs) by generating new hypotheses, producing results more quickly and at a lower cost, including broader patient populations, reflecting clinical care patterns, and assessing longer-term outcomes.[1,6] These advantages of RWE studies are of particular value in the field of oncology due to high unmet medical need, poor patient outcomes for several cancer types, a rapidly evolving treatment landscape, and the need to generate additional confirmatory evidence following accelerated regulatory approval.[7]

At the same time, causal inference from non-randomized studies leveraging RWD may be hindered by threats to internal validity. Due to a lack of randomization, RWE studies may suffer from unmeasured or inadequately controlled confounding. Key variables may be missing or misclassified in data generated from clinical practice, which may introduce information bias and limit direct comparisons with clinical trials. Therefore, successful application of RWD to support clinical and regulatory decision-making requires a thorough understanding of the circumstances under which RWD can generate valid evidence about treatment effectiveness.

The Coalition to Advance Real-World Evidence through Randomized Controlled Trial Emulation (CARE) Initiative aims to contribute to this understanding by building an empirical evidence base for the generation of RWD-based evidence of treatment effectiveness.[1] To do this, electronic health record (EHR) data collected during routine healthcare practice will be used to emulate the primary outcomes of completed RCTs for oncology therapies. The RCT results will provide a benchmark causal effect estimate against which the findings of non-randomized emulations can be compared. No standard metric has been proposed to quantify agreement between emulation and RCT results and previous

work has used a variety of measures.[8] The CARE emulations will focus on *qualitative agreement* — whether findings from a non-interventional study and RCT are in the same direction and are of similar magnitude. The choice to use a metric that is not anchored to statistical significance reflects conclusions from the CARE pilot study about specific challenges in oncology emulations (e.g., small real-world sample sizes) and the non-inferential goals of this work.[9,10] Through this effort, the CARE Initiative seeks to identify under what conditions non-interventional studies using data generated during routine clinical care can provide reliable conclusions about drug effectiveness.

6. Research question and objectives

In this emulation of the Palbociclib: Ongoing Trials in the Management of Breast Cancer (PALOMA-2) trial, real-world EHR data will be used to estimate the effectiveness of initiating first-line treatment with palbociclib and letrozole versus letrozole alone among patients with estrogen receptor-positive (ER+)/human epidermal growth factor receptor 2-negative (HER2-) advanced breast cancer.[2]

Table 2. Primary research question and objective

Objective:	The objective of this non-interventional study is to estimate the effectiveness of initiating first-line palbociclib and letrozole versus letrozole alone in a real-world emulation of the PALOMA-2 RCT.
Hypothesis:	Patients with ER+/HER2- metastatic breast cancer treated with palbociclib and letrozole will have improved real-world progression-free survival (rwPFS) compared with patients treated with letrozole alone.
Population:	Adult (≥18 years of age), postmenopausal female patients with ER+/HER2- metastatic breast cancer
Exposure:	Palbociclib and letrozole as first-line treatment for metastatic breast cancer
Comparator:	Letrozole as first-line treatment for metastatic breast cancer
Outcome:	rwPFS, defined as time from study treatment initiation to disease progression or death
Setting:	Clinical data sourced from oncology practices in the United States (U.S.)
Main measure of effect:	Hazard ratio for rwPFS in the intent-to-treat population

7. Research methods

7.1 Study design

7.1.1 Overview of key design elements of the PALOMA-2 trial

Study design: The PALOMA-2 trial (NCT01740427)[11] was an international, randomized, double-blind, placebo-controlled Phase III clinical trial comparing the efficacy and safety of palbociclib and letrozole versus placebo and letrozole for the first-line treatment of ER+/HER2- advanced breast cancer. Figure 1 displays the study design diagram for the PALOMA-2 trial.

Population: The trial study population included postmenopausal females, 18 years of age or older with histologically or cytologically confirmed ER+/HER2- adenocarcinoma of the breast and evidence of locoregionally recurrent or metastatic disease. Eligible patients had received no prior systemic therapy for advanced disease; did not have disease recurrence during or within 12 months of neoadjuvant or adjuvant treatment with a non-steroidal aromatase inhibitor; had adequate organ function; had an Eastern Cooperative Oncology Group (ECOG) performance status of 0 to 2; and had measurable disease according to the Response Evaluation Criteria in Solid Tumours (RECIST) v.1.1.[12] Patients with visceral metastasis who were at risk of short-term, life-threatening complications were excluded. Full inclusion and exclusion criteria are listed in Table 4 and Table 5.

Endpoints: The primary trial endpoint was progression-free survival (PFS), defined as the time from the date of randomization to the date of the first documentation of objective progression of disease as evaluated by study investigators according to RECIST v.1.1, or death due to any cause. Follow-up for PFS continued until the first of: documented disease progression; death; initiation of a new anti-cancer treatment; discontinuation from overall study participation due to symptomatic deterioration, unacceptable toxicity, withdrawal of consent; or loss to follow-up. Comparator patients were not able to initiate palbociclib.

Analysis: Efficacy was assessed in the intent-to-treat population defined by randomized treatment assignment. Median PFS and corresponding 95% confidence intervals were estimated after a maximum follow-up of 33 months using the Kaplan-Meier method and compared by treatment status using a log-rank test. The hazard ratio for progression was calculated using a Cox proportional-hazards model.

7.1.2 Overview of key design elements of the real-world emulation study

Study design: This new user, non-randomized active comparator cohort study will compare rwPFS between patients with records indicating initiation of palbociclib and letrozole versus letrozole alone following qualifying metastatic disease diagnosis in the EHR data source. The data originate from oncology practices in the U.S. and include both structured and curated data elements abstracted from unstructured sources including provider notes, pathology, and imaging reports (Section 7.6.1).

Population: The study population will include postmenopausal females, 18 years of age or older with a diagnosis of ER+/HER2- metastatic adenocarcinoma of the breast recorded in the EHR. Patients will be selected to reflect the PALOMA-2 trial eligibility criteria, as feasible in the RWD source, to create a trial-similar population. Eligible patients will include those with no record of previously receiving systemic anti-cancer therapy for advanced disease; without neoadjuvant or adjuvant treatment with a non-steroidal aromatase inhibitor in the 12 months prior to

metastatic diagnosis; without evidence of inadequate organ function; and without evidence of an ECOG performance status >2. Real-world operationalization of all trial inclusion and exclusion criteria are listed in Table 4 and Table 5. Study exposure groups will be ascertained within a 15-day time window ('exposure ascertainment window'), beginning on the day of the first record of a study drug. Patients initiating treatment for metastatic disease with palbociclib and letrozole within the exposure ascertainment window will be classified as exposed. Patients initiating treatment with letrozole and with no evidence of palbociclib within the exposure ascertainment window will be classified as comparator patients.

Endpoints: The study endpoint, rwPFS, will be based on curated progression information and date of death available in the data source. Patients will be followed until the first of: progression or death; initiation of a new anti-cancer therapy; the administrative end of the study period (Day 1004, or 33 months, to conform to the maximum follow-up time at which PFS was evaluated in the PALOMA-2 trial); the end of data; or loss to follow-up (the last date prior to a period of >90 days without curated EHR activity and without death).

Analysis: Inverse probability of treatment weighting (IPTW) will be used to adjust for measured baseline confounders. Median rwPFS and corresponding 95% confidence intervals will be estimated using the Kaplan-Meier method and compared between exposure groups using a log-rank test. The hazard ratio for progression will be calculated using a Cox proportional-hazards model.

Rationale for study design choice: The choice of study design, population, endpoint, and analysis plan are intended to emulate the PALOMA-2 trial design as closely as possible, including creating a trial-similar real-world population, controlling for confounding in the absence of randomization, and estimating the intent-to-treat effect as was done in the trial.

7.2 Study design diagram

The figures below display the study designs for the PALOMA-2 RCT (Figure 1) and this real-world emulation study (Figure 2). For simplicity, inclusion and exclusion criteria displayed correspond to those highlighted in the trial publication.[2] Full trial eligibility criteria are listed in Sections 7.3.2 and 7.3.3.

Figure 1. PALOMA-2 randomized controlled trial study design

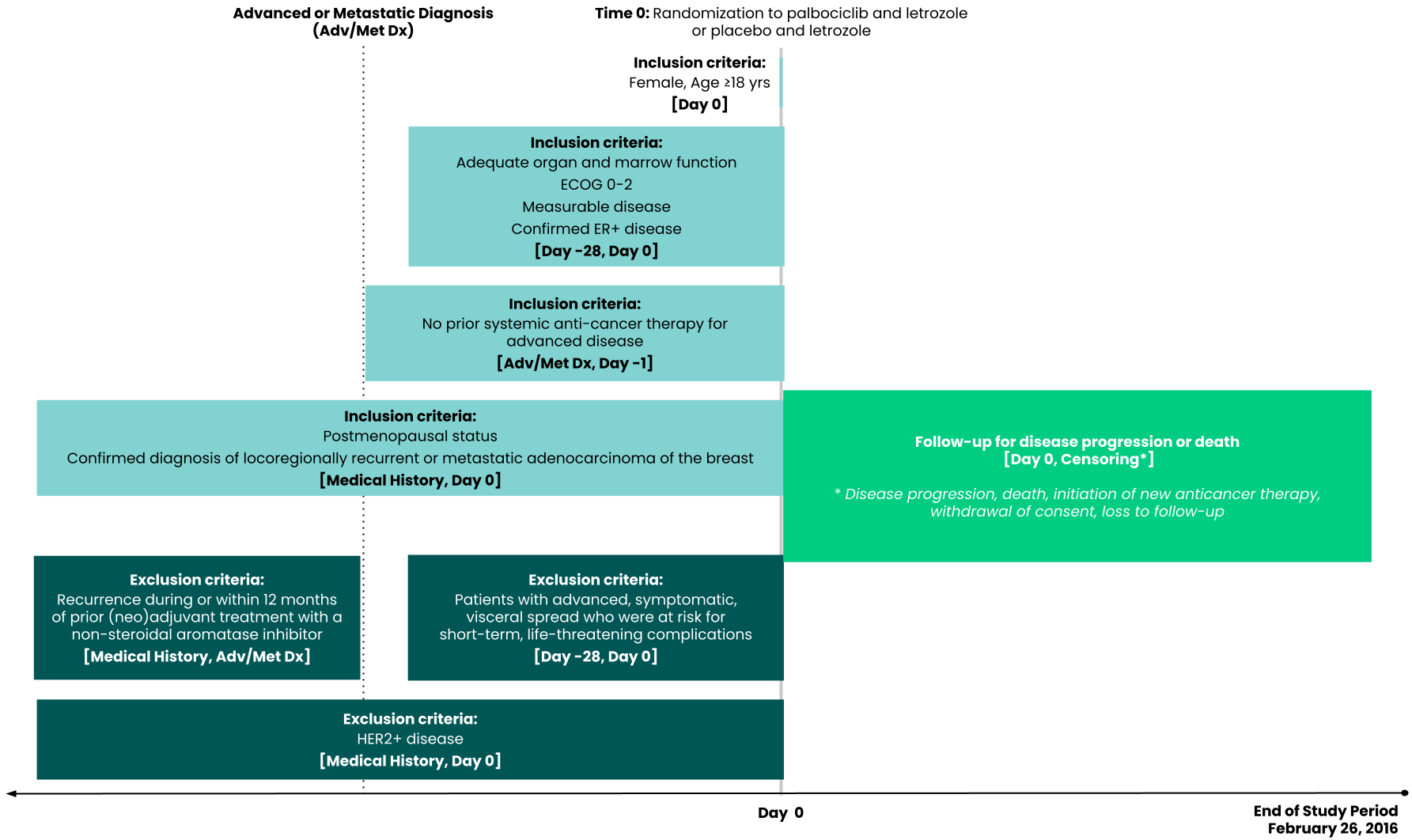
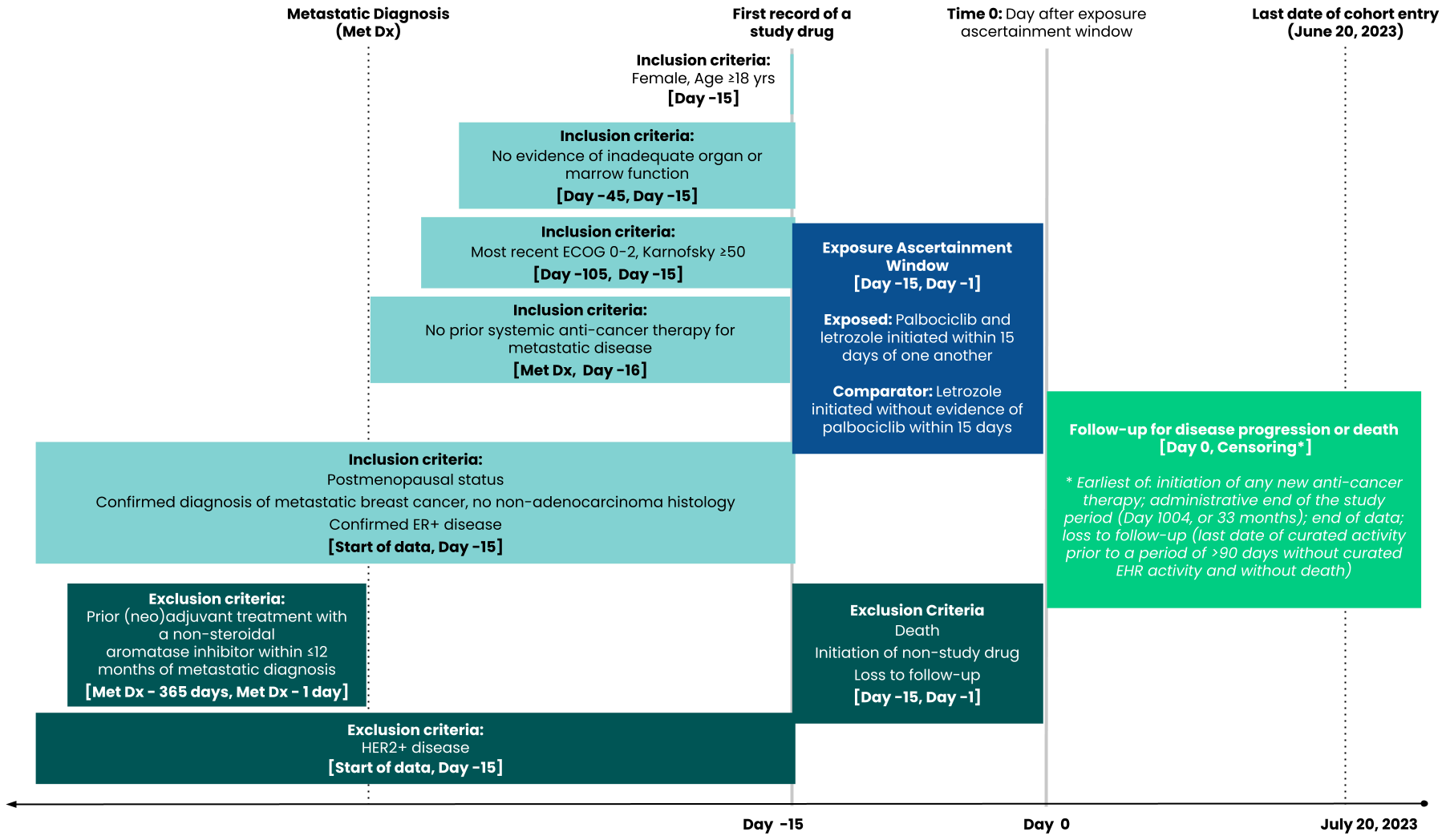


Figure 2. Real-world emulation study design



7.3 Setting

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

Figure 2 displays the study design diagram for the primary analysis. Candidate exposed and comparator patients will be identified as those initiating treatment for metastatic disease with palbociclib and/or letrozole. Study exposure group will be ascertained within a 15-day time

window (i.e., 'exposure ascertainment window'), beginning on the day of the first record of a study drug (palbociclib or letrozole) (Day -15), and ending 15 days later on Day -1. Patients experiencing an outcome or censoring event during this window will be excluded. Additional information on the exposure ascertainment window is provided in Section 7.4.1. Follow-up for progression and death will begin on Day 0 ('Time 0'). To avoid immortal time bias, Time 0 will occur after the exposure ascertainment window for all patients, irrespective of exposure status. The operational definition of Time 0 is provided in Table 3. The date of first study drug initiation will be restricted to dates from February 1, 2010 to June 20, 2023 to reflect current treatment paradigms at the time of the PALOMA-2 trial, and to allow a minimum available follow-up time of one month prior to the end of data for all patients, respectively. It should be noted that the first line of metastatic treatment can be initiated up to 30 days before the documented date of metastatic diagnosis per the data vendor's line of therapy definitions.

Table 3. Operational definition of Time 0

Study population name(s)	Time Anchor Description	Type of entry	Washout window	Incident with respect to…
Patients initiating first-line treatment for metastatic ER+/HER2- breast cancer	Time 0 will be defined as the day after the end of the 15-day exposure ascertainment window. Exposure ascertainment is described in the variables section of this protocol.	Incident	[Metastatic diagnosis date, Day -16]	Metastatic diagnosis

7.3.2 Context and rationale for study inclusion criteria

Operational definitions for the study inclusion criteria are presented in Table 4. These inclusion criteria correspond to those applied in the PALOMA-2 RCT.

As indicated below, PALOMA-2 inclusion criteria that are not relevant in a real-world clinical setting (e.g., willingness and ability to provide tumor tissues) or are not captured in routine oncology care (e.g., measurable disease) will not be applied.

Table 4. Operational Definitions of Inclusion Criteria

Trial Criterion	Real-world operationalization [a]	Assessment window	Rationale for real-world operationalization (where applicable)	Validity concerns and how they will be addressed (where applicable)
Adult women (≥18 years of age)	Recorded female sex and age at study drug initiation ≥18 years.	[Day -15]	N/A	N/A

Trial Criterion	Real-world operationalization [a]	Assessment window	Rationale for real-world operationalization (where applicable)	Validity concerns and how they will be addressed (where applicable)
Proven diagnosis of adenocarcinoma of the breast	Breast cancer diagnosis. Histology not indicative of non-adenocarcinoma histologies.	[Start of data, Day -15]	Specific histology results cannot be identified for all patients in the dataset. As a result, patients with non-adenocarcinoma histology will be excluded.	Patients with non-specific histology results who do not have adenocarcinoma and who would have been ineligible for the trial may be included. This may affect comparability between the study and trial populations. The specific histology values for all included patients will be reported descriptively.
Evidence of locoregionally recurrent or metastatic disease not amenable to resection or radiation therapy with curative intent and for whom chemotherapy is not clinically indicated	Staging information indicative of metastatic disease.	[Initial diagnosis date, Day -15]	Locoregionally recurrent disease cannot be identified in the dataset. Patients initiating study treatment will be assumed to be ineligible for curative therapy or chemotherapy.	Patients with advanced disease who would have been eligible for the trial may be excluded. While patients with advanced disease constituted <3% of the trial population, this may affect comparability between the study and trial populations.
Documentation of histologically or cytologically confirmed diagnosis of ER+ breast cancer based on local laboratory results.	ER biomarker test with a result interpretation of 'Positive.' For patients with multiple ESR1 tests, the entry closest in time to study drug initiation will be used.	[Start of data, Day -15]	N/A	N/A
Previously untreated with any systemic anti-cancer therapy for their locoregionally recurrent or metastatic ER+ disease	No first-line regimen for metastatic disease prior to study drug initiation.	[Metastatic diagnosis date, Day -16]	N/A	N/A

Trial Criterion	Real-world operationalization [a]	Assessment window	Rationale for real-world operationalization (where applicable)	Validity concerns and how they will be addressed (where applicable)
Postmenopausal women, defined as women with prior bilateral surgical oophorectomy or medically confirmed post-menopausal status	Surgery indicative of bilateral surgical oophorectomy or recorded post-menopausal status. Age >50 years if both of the above are missing.	[Start of data, Day -15]	Age >50 years will be used as a proxy for post-menopausal status for patients with missing data. This reflects the average age of menopause for females in the U.S.[13]	Patients >50 years of age but who are not postmenopausal and who would have been ineligible for the trial may be included. This may affect comparability between the study and trial populations.
Measurable disease as defined per RECIST v.1.1	This criterion cannot be operationalized.	N/A	RECIST assessments are not performed in routine oncology care.	Patients with non-measurable disease who would have been ineligible for the trial may be included. This may result in longer estimates of rwPFS than were observed in the trial due to difficulties in objectively measuring progression for these patients.
ECOG performance status 0-2	ECOG performance status 0-2 or missing Karnofsky performance status ≥50 or missing For patients with multiple performance status records, the entry closest in time to study drug initiation will be used.	[Day -105, Day -15]	Performance status will be evaluated in the 90 days before study drug initiation to balance missingness and misclassification, given infrequent real-world assessments.	Patients with missing performance status records who would have been ineligible for the trial may be included. This may affect comparability between the study and trial populations. Sensitivity analyses restricting to patients with known values will be performed, sample size permitting.

Trial Criterion	Real-world operationalization [a]	Assessment window	Rationale for real-world operationalization (where applicable)	Validity concerns and how they will be addressed (where applicable)
Adequate organ and marrow function	No lab results indicating inadequate organ or marrow function as defined in the PALOMA-2 trial protocol.	[Day -45, Day -15]	Lab tests specified in the trial are not performed for all patients in routine oncology care. As a result, patients with lab results indicative of inadequate organ or marrow function will be excluded. Lab results will be evaluated in the 30 days before study drug initiation to balance missingness and misclassification.	Patients with inadequate organ and marrow function and therefore potentially reduced survival who would have been ineligible for the trial may be included. We do not expect this to affect a large number of patients as physicians are unlikely to start treatment for patients with poor organ or marrow function.
Resolution of all acute toxic effects of prior anti-cancer therapy or surgical procedures to National Cancer Institute (NCI) Common Terminology Criteria for Adverse Events (CTCAE) version 4.0 Grade ≤1	This criterion cannot be operationalized.	N/A	Acute toxic effects and their resolution cannot be identified in the data source.	Patients with ongoing acute toxic effects who would have been ineligible for the trial may be included. We do not expect this to affect a large number of patients as physicians are unlikely to start a new treatment for patients with ongoing acute toxic effects.
Willingness and ability to comply with scheduled visits, treatment plan, laboratory tests, and other study procedures	This criterion is not relevant in a real-world clinical setting.	N/A	N/A	N/A
Willingness and ability to provide tumor tissues	This criterion is not relevant in a real-world clinical setting.	N/A	N/A	N/A
Evidence of a personally signed and dated informed consent document	This criterion is not relevant in a real-world clinical setting.	N/A	N/A	N/A

N/A = not applicable

[a] See Appendix C for code list.

7.3.3 Context and rationale for study exclusion criteria

Operational definitions for the study exclusion criteria are presented in Table 5. These exclusion criteria correspond to those applied in the PALOMA-2 RCT.

As indicated below, PALOMA-2 exclusion criteria that are not relevant outside of an RCT and in a real-world clinical setting (e.g., investigational site staff, their relatives, or Pfizer employees) will not be applied.

An additional exclusion criteria that was not relevant for the PALOMA-2 trial but is necessary for emulation using RWD is the exclusion of patients with evidence of a censoring event or death during the exposure ascertainment window (described in Section 7.4.1). Patients with evidence of a progression event during the exposure ascertainment window will not be excluded as these are likely latent progression events that preceded study drug initiation.

Table 5. Operational Definitions of Exclusion Criteria

Trial Criterion	Real-world operationalization [a]	Assessment window	Rationale for real-world operationalization (where applicable)	Validity concerns and how they will be addressed (where applicable)
HER2-positive tumor	V-erb-b2 avian erythroblastic leukemia viral oncogene homolog 2 (ERBB2) biomarker test with a result description of 'Positive'. For patients with multiple ERBB2 tests, the entry closest in time to study drug initiation will be used.	[Start of data, Day -15]	N/A	N/A
Patients with advanced, symptomatic, visceral spread, that are at risk of life-threatening complications in the short term	This criterion cannot be operationalized.	N/A	A patient's short-term risk of life-threatening complications cannot be identified in the data source.	Patients at acute risk of life-threatening complications who would have been ineligible for the trial may be included. We do not expect this to affect a large number of patients as physicians are unlikely to start a new treatment for patients at high risk of complications.

Trial Criterion	Real-world operationalization [a]	Assessment window	Rationale for real-world operationalization (where applicable)	Validity concerns and how they will be addressed (where applicable)
Known active, uncontrolled, or symptomatic central nervous system metastases, carcinomatous meningitis, or leptomeningeal disease as indicated by clinical symptoms, cerebral edema, and/or progressive growth. Patients with a history of central nervous system (CNS) metastases or cord compression are eligible if they have been definitively treated with local therapy (e.g., radiotherapy, stereotactic surgery) and are clinically stable off anticonvulsants and steroids for at least 4 weeks before randomization	Diagnosis of brain, CNS, and/or spinal cord metastases.	[Start of data, Day -15]	Active, uncontrolled, or symptomatic CNS metastases and diagnoses of carcinomatous meningitis and leptomeningeal disease cannot be identified in the data source. Definitive treatment with local therapy and clinical stability without anticonvulsants and steroids cannot be identified in the data source.	Patients with a history of CNS metastases who would have been eligible for the trial (those who received definitive treatment and were clinically stable at study drug initiation) may be excluded. This may affect comparability between the study and trial populations.
Prior neoadjuvant or adjuvant treatment with a non-steroidal aromatase inhibitor (i.e., anastrozole or letrozole) with disease recurrence while on or within 12 months of completing treatment	Neoadjuvant or adjuvant treatment with anastrozole or letrozole with time between the last date of neoadjuvant or adjuvant therapy and metastatic diagnosis ≤12 months.	[Metastatic diagnosis date -365 days, Metastatic diagnosis date – 1 day]	N/A	N/A
Prior treatment with any cyclin-dependent kinase 4/6 (CDK4/6) inhibitor	Treatment with ribociclib, abemaciclib, or palbociclib.	[Start of data, Day -16]	N/A	N/A

Trial Criterion	Real-world operationalization [a]	Assessment window	Rationale for real-world operationalization (where applicable)	Validity concerns and how they will be addressed (where applicable)
Patients treated within the last 7 days prior to randomization with: Food or drugs that are known to be Cytochrome P450 3A4 (CYP3A4) inhibitors; drugs that are known to be CYP3A4 inducers; drugs that are known to prolong the QT interval	Treatment with a CYP3A4 inhibitor or inducer or drugs known to prolong the QT interval, as specified in the PALOMA-2 trial protocol. Consumption of foods known to be CYP3A4 inhibitors (grapefruit, grapefruit juice) and use of St. John's Wort cannot be operationalized.	[Day -45, Day -15]	To account for real-world frequency of medication capture in the EHR, this exclusion criteria will be assessed in the 30 days prior to study drug initiation. Food consumption and supplement use cannot be identified in the data source.	Patients with less recent exposure to these treatments who would have been eligible for the trial may be excluded. Conversely, patients with grapefruit or St. John's Wort exposure who would have been ineligible for the trial may be included. This may affect comparability between the study and trial populations.
Major surgery, chemotherapy, radiotherapy, any investigational agents, or other anti-cancer therapy within 2 weeks before randomization. Patients who received prior radiotherapy to ≥25% of bone marrow are not eligible independent of when it was received	First-line regimen for metastatic disease prior to study drug initiation. Anti-cancer therapy other than study treatment in the same first-line therapy or neoadjuvant or adjuvant therapy. Major cancer-related surgery. Radiotherapy to ≥25% of bone marrow cannot be operationalized.	[Day -29, Day -16]	Non-cancer surgeries cannot be identified in the data source. Extent of radiotherapy cannot be identified in the data source.	Patients with recent major non-cancer surgery or radiotherapy to ≥25% of bone marrow who would have been ineligible for the trial may be included. This may affect comparability between the study and trial populations.
Diagnosis of any other malignancy within 3 years prior to randomization, except for adequately treated basal cell or squamous cell skin cancer, or carcinoma in situ of the cervix	Diagnosis of a second primary malignancy.	[Day -1110, Day -15]	Information on adequate treatment for basal cell or squamous cell skin cancer or carcinoma in situ of the cervix cannot be identified in the data source.	Patients with adequately controlled basal cell or squamous cell skin cancer or carcinoma in situ of the cervix who would have been eligible for the trial may be excluded. This may affect comparability between the study and trial populations.

Trial Criterion	Real-world operationalization [a]	Assessment window	Rationale for real-world operationalization (where applicable)	Validity concerns and how they will be addressed (where applicable)
Heart-rate corrected QT interval (QTc) >480 msec (based on the mean value of the triplicate electrocardiograms [ECGs]), family or personal history of long or short QT syndrome, Brugada syndrome or known history of QTc prolongation, or Torsade de Pointes	QTc >480 msec and family history of long or short QT syndrome cannot be operationalized. Diagnosis of long or short QT syndrome, Brugada syndrome, QTc prolongation, or Torsade de Pointes.	[Start of data, Day -15]	QTc measurements and family history of long or short QT syndrome cannot be identified in the data source.	Patients with QTc >480 msec or family history of long or short QT syndrome who would have been ineligible for the trial may be included. This may affect comparability between the study and trial populations.
Uncontrolled electrolyte disorders that can compound the effects of a QTc-prolonging drug (e.g., hypocalcemia, hypokalemia, hypomagnesemia)	Diagnosis of hypocalcemia, hypokalemia, or hypomagnesemia.	[Day -105, Day -15]	Diagnosis of electrolyte disorders will be evaluated in the 90 days before study drug initiation to reflect acute diagnoses and balance missingness and misclassification. Controlled disease status cannot be identified in the data source.	Patients with controlled disease who would have been eligible for the trial may be excluded. This may affect comparability between the study and trial populations.

Trial Criterion	Real-world operationalization [a]	Assessment window	Rationale for real-world operationalization (where applicable)	Validity concerns and how they will be addressed (where applicable)
Any of the following within 6 months of randomization: myocardial infarction, severe/unstable angina, ongoing cardiac dysrhythmias of grade ≥2, atrial fibrillation of any grade, coronary/peripheral artery bypass graft, symptomatic congestive heart failure, cerebrovascular accident including transient ischemic attack, or symptomatic pulmonary embolism	Diagnosis of myocardial infarction, angina, ongoing cardiac dysrhythmias, atrial fibrillation, congestive heart failure, cerebral infarction, transient ischemic attack, or pulmonary embolism. Coronary/peripheral artery bypass graft cannot be operationalized.	[Day -195, Day -15]	Non-cancer surgeries cannot be identified in the data source. Severe/unstable angina cannot be identified in the data source.	Patients who have undergone artery bypass graft and who would have been ineligible for the trial may be included. Patients with stable angina who would have been eligible for the trial may be excluded. This may affect comparability between the study and trial populations.
Active inflammatory bowel disease or chronic diarrhea, short bowel syndrome, or any upper gastrointestinal surgery including gastric resection	Diagnosis of inflammatory bowel disease, chronic diarrhea, or short bowel syndrome. Upper gastrointestinal surgery cannot be operationalized.	Active inflammatory bowel disease or chronic diarrhea: [Day -105, Day -15] Short bowel syndrome: [Start of data, Day -15]	Diagnosis of inflammatory bowel disease and chronic diarrhea will be evaluated in the 90 days before study drug initiation to reflect active disease and balance missingness and misclassification. Diagnosis of short bowel syndrome will be evaluated from the start of data to reflect the chronic nature of this condition. Non-cancer surgeries cannot be identified in the data source.	Patients who have undergone upper gastrointestinal surgery and who would have been ineligible for the trial may be included. This may affect comparability between the study and trial populations.

Trial Criterion	Real-world operationalization [a]	Assessment window	Rationale for real-world operationalization (where applicable)	Validity concerns and how they will be addressed (where applicable)
Known hypersensitivity to letrozole, or any of its excipients, or to any palbociclib (PD-0332991)/ placebo excipients	This criterion cannot be operationalized.	N/A	Hypersensitivity cannot be identified in the data source.	Patients with known hypersensitivity to letrozole or palbociclib who would have been ineligible for the trial may be included. We do not expect this to affect a large number of patients as physicians are unlikely to prescribe these treatments to patients with a known hypersensitivity.
Known human immunodeficiency virus infection	Diagnosis of human immunodeficiency virus infection.	[Start of data, Day -15]	N/A	N/A
Other severe acute or chronic medical or psychiatric conditions or laboratory abnormality that may increase the risk associated with study participation or investigational product administration or may interfere with the interpretation of study results and, in the judgment of the investigator, would make the patient inappropriate for entry into this study	This criterion is not relevant in a real-world clinical setting.	N/A	N/A	N/A

Trial Criterion	Real-world operationalization [a]	Assessment window	Rationale for real-world operationalization (where applicable)	Validity concerns and how they will be addressed (where applicable)
Patients who are investigational site staff members or relatives of those site staff members or patients who are Pfizer employees directly involved in the conduct of the trial	This criterion is not relevant in a real-world clinical setting.	N/A	N/A	N/A
Participation in other studies within 4 weeks before randomization	This criterion cannot be operationalized.	N/A	Dates of clinical trial participation cannot be identified in the data source.	Patients who have recently participated in a clinical trial and who may have initiated treatment for metastatic disease in that setting who would have been ineligible for the trial may be included. This may affect comparability between the study and trial populations, and may bias comparative effect estimates if trial participation is related to the study treatment received.
Recent or active suicidal ideation or behavior	This criterion cannot be operationalized.	N/A	Suicidal ideation or behavior cannot be identified in the data source.	Patients with recent or active suicidal ideation or behavior who would have been ineligible for the trial may be included. We do not expect this to affect a large number of patients given the low prevalence of suicide mortality among patients with advanced breast cancer (<1%).[14,15]

N/A = not applicable

[a] See Appendix C for code list.

7.4 Variables

7.4.1 Context and rationale for exposures of interest

Operational definitions for the two treatment strategies that will be compared are presented in Table 6. Exposure will be defined based on treatment initiated during the 15-day exposure ascertainment window. While patients in the PALOMA-2 trial received study treatments on the same day, a 15-day window was selected to allow for potential delays in data entry and treatment administration variability that may occur in routine practice. In a feasibility assessment of 944 patients in the data source who initiated both letrozole and palbociclib as first-line therapy, 93.3% started both medications within 15 days. A shorter exposure window may be too restrictive and result in few eligible patients, while longer time periods may result in treatment patterns in the study population that are vastly different from the trial. As discussed above, patients who experience a censoring event or death during the exposure ascertainment window will be excluded from the study to align the start of follow-up for both exposure groups.

Table 6. Operational Definitions of Exposures

Group name(s)	Details	Washout window	Assessment Window	Incident with respect to…	Source of algorithm	Validity concerns and how they will be addressed
Exposed	Patients initiating treatment with palbociclib and letrozole within 15 days of one another, in the metastatic setting	N/A, treatment in first-line metastatic setting	[Metastatic diagnosis date, Day -1]	Metastatic Diagnosis	Curated regimen definition; clinical experts	Real-world exposure group definitions allow for more flexibility in treatment timing than in the trial. The potential impact of this on results will be explored through a sensitivity analysis shortening the duration of the exposure ascertainment window (Table 11).
Comparator	Patients initiating treatment with letrozole with no evidence of palbociclib treatment within 15 days, in the metastatic setting	N/A, treatment in first-line metastatic setting	[Metastatic diagnosis date, Day -1]	Metastatic Diagnosis	Curated regimen definition; clinical experts	

N/A = not applicable

7.4.2 Context and rationale for outcome of interest

The operational definition of the outcome of interest, rwPFS, is presented in Table 7. This outcome corresponds to the primary outcome of the PALOMA-2 trial.

Table 7. Operational Definitions of Outcome

Outcome name	Details	Primary outcome	Type of outcome	Washout window	Source of algorithm	Validity concerns and how they will be addressed
rwPFS	Time from Day 0 to disease progression or death	Yes	Time-to-event	N/A	Progression events are curated in the data. All tumor progression events after initial cancer diagnosis are captured.	Real-world progression is not evaluated at fixed intervals as was done in the trial. The frequency and timing of real-world progression assessment results by exposure group will be reported to contextualize study findings.

N/A = not applicable

7.4.3 Context and rationale for follow-up

Follow-up will begin on Day 0, and will continue until the earliest of:

1. Date of documented progression or death (Table 7);
2. Initiation of any new anti-cancer therapy;
3. The administrative end of the study period (Day 1004, or 33 months, of follow-up), to align with the maximum follow-up time at which PFS was evaluated in the PALOMA-2 trial;
4. End of the data (July 20, 2023);
5. Loss to follow-up: The last date of curated EHR activity prior to a period of >90 days without curated EHR activity and without death.

Operational definitions for the study censoring criteria are presented in Table 8. These censoring criteria correspond to those applied in the PALOMA-2 RCT where applicable.

Table 8. Operational Definitions of Censoring Criteria

Trial Criterion	Real-world operationalization	Rationale for real-world operationalization (where applicable)	Validity concerns and how they will be addressed (where applicable)
Initiation of a new anti-cancer treatment	Treatment with any new anti-cancer therapy.	N/A	N/A

Trial Criterion	Real-world operationalization	Rationale for real-world operationalization (where applicable)	Validity concerns and how they will be addressed (where applicable)
End of study	Administrative end of study (Day 1004, or 33 months) or July 20, 2023.	Align with maximum time in PALOMA-2 or end of available data in real-world data source.	N/A
Withdrawal of consent	This criterion is not relevant in a real-world clinical setting.	N/A	N/A
Loss to follow-up	A period of >90 days without curated EHR activity and without death.	Curated activity in the real-world data source indicates points at which progression can be recorded. Metastatic breast cancer patients likely have contact with the healthcare system at least every 90 days for lab work, prescription refills, outpatient visits, or scans. Periods greater than 90 days may indicate loss to follow-up during which censoring or outcome events cannot be captured.	Patients who use the health care system less frequently will be censored. A sensitivity analysis will be conducted expanding the period without curated activity to >180 days (Table 11).
≥2 missed disease assessments	This criterion will not be operationalized.	Real-world progression is not evaluated at fixed intervals as was done in the trial. Patients with gaps in activity of greater than 90 days will be censored under the loss to follow-up definition.	The frequency of real-world progression assessments will be reported to contextualize study findings.

N/A = not applicable

7.4.4 Context and rationale for covariates (confounding variables and effect modifiers, e.g., risk factors, comorbidities, comedications)

Operational definitions for the study covariates are presented in Table 9. Covariates were chosen based on the primary trial publication and the research team's substantive knowledge of potential confounders. For time-varying characteristics, the value closest in time prior to study drug initiation will be used.

Table 9. Operational Definitions of Covariates

Characteristic	Details/Levels[a]	Type of variable	Assessment window
Patient Demographic Characteristics			
Age	Age at study drug initiation 18 to <65 yrs, ≥65 yrs	Continuous, Categorical	[Day -15]
Race	American Indian or Alaska Native, Asian, Native Hawaiian or other Pacific Islander, Black or African American, White, other or unknown race	Categorical	[Start of data, Day -15]
Ethnicity	Hispanic or Latino, not Hispanic or Latino, unknown	Categorical	[Start of data, Day -15]
Region	Northeast, south, midwest, west, missing	Categorical	[Start of data, Day -15]
Clinical Characteristics			
Oncology clinic site type	Academic, community, missing	Categorical	[Start of data, Day -15]
Performance status	ECOG performance status 0, 1, 2, or Karnofsky performance status 50, 60, 70, 80, 90, 100, or missing.	Categorical	[Day -105, Day -15]
Smoking status at initial diagnosis	Current smoker, former smoker, non-smoker, missing	Categorical	[Start of data, Initial diagnosis date]
Number of outpatient visits in past year	N/A	Continuous	[Day -380, Day -15]
Disease Characteristics and Treatment History			
Year of study treatment initiation	N/A	Categorical	[Day -15]
Prior therapy	Aromatase inhibitor, antiestrogen	Dichotomous	[Start of data, Day -15]
Disease stage at initial diagnosis	Disease stage I-IV	Categorical	[Initial diagnosis date - 30, Initial diagnosis date + 90]

Characteristic	Details/Levels[a]	Type of variable	Assessment window
Time between metastatic diagnosis and treatment initiation	N/A	Continuous	[Metastatic diagnosis date, Day -15]
Time interval between initial diagnosis and metastatic diagnosis (approximate disease-free interval)	N/A	Continuous	[Initial diagnosis date, Metastatic diagnosis date]
Prior neoadjuvant or adjuvant chemotherapy	Yes, no	Dichotomous	[Start of data, Day -16]
Prior neoadjuvant or adjuvant hormone therapy	Yes, no	Dichotomous	[Start of data, Day -16]
Surgery in the curative setting	Yes, no	Dichotomous	[Start of data, Day -16]
Radiation in the curative setting	Yes, no	Dichotomous	[Start of data, Day -16]
Recurrence type	Recurrent, de novo metastatic, missing	Categorical	[Initial diagnosis date - 30, Initial diagnosis date + 90]
Metastatic disease site	Visceral (lung, including pleura, liver), non-visceral (all others), bone only	Categorical	[Start of data, Day -15]
Number of metastatic disease sites	1, 2, ≥3	Continuous, Categorical	[Start of data, Day -15]

N/A = not applicable

[a] See Appendix C for code list.

7.5 Data analysis

7.5.1 Context and rationale for analysis plan

Analytic Population

Primary analyses will be conducted in the real-world study population based on the first study treatment initiated. This approach is intended to emulate the intent-to-treat analysis conducted in the PALOMA-2 trial, where patients were analyzed based on their randomized treatment assignment. As crossover between treatment arms was not allowed in the PALOMA-2 trial, intent-to-treat and per-protocol trial analyses would produce equivalent results. Conversely, real-world patients initially prescribed letrozole could initiate palbociclib after the end of the exposure ascertainment window. To investigate the potential for crossover between comparison groups prior to disease progression in the real-world population, descriptive analyses of all first-line treatments initiated and time between treatments will be conducted.

IPTW will be used to approximate full conditional exchangeability between the comparison groups and facilitate estimation of the population average treatment effect.[16–18] Propensity scores (PS) reflecting the conditional probability of initiating treatment with palbociclib and letrozole will be calculated via multivariable logistic regression. Exposure to palbociclib and letrozole will be regressed on *a priori* identified potential confounders (Table 9). Inverse probability of treatment weights will be calculated as the inverse of the propensity score (1/PS) for patients in the exposed group and as the inverse of one minus the propensity score (1/1-PS) for comparator patients.[19] Patients are therefore weighted by the inverse probability of initiating the treatment they actually started, conditional on the observed covariates included as independent variables in the PS model. This approach aims to create a pseudo-population with full exchangeability on measured confounders within which the treatment effect is estimated.

Confounder balance will be assessed by comparing the absolute standardized difference (ASD) in the distribution of categorical variable levels and the mean of continuous variables between the weighted exposure groups.[16] If balance is not achieved (ASD > 0.1), alternative specifications of the PS model, e.g., including variable transformations and interaction terms, will be explored. Confounders with insufficient balance may also be included as covariates in the outcome models.

The distribution of baseline patient demographic and clinical characteristics will be compared between comparator patients who index before 2015 and on/after 2015 to understand if differences exist between those who received the comparator treatments before and after palbociclib approval in this indication. Characteristics to be compared are described in Table 9. Differences will be assessed using t-tests, chi squared tests, and accompanying 95% confidence intervals.

Descriptive Analyses

The distribution of patient baseline demographic and clinical characteristics will be compared between the unweighted and weighted real-world populations and the PALOMA-2 trial population. Characteristics to be compared are described in Table 9. Differences will be assessed using t-tests, chi squared tests, and accompanying 95% confidence intervals.

Reasons for exclusion and censoring will be presented in a Consolidated Standards of Reporting Trials (CONSORT) diagram and distributions of missingness of inclusion/exclusion criteria and potential confounders will be calculated and compared by exposure group.[20]

Treatment Effectiveness

Similar to the PALOMA-2 trial, comparative treatment effectiveness of palbociclib and letrozole versus letrozole alone will be estimated using median rwPFS and by comparing the hazards of progression and death in the two exposure groups. Follow-up will be administratively censored at 33 months to emulate maximum follow-up in the PALOMA-2 trial at the time that results were reported.[2] We will emulate subgroup analyses performed in the trial, except where sample size is less than 10 in each exposure group. Post-hoc analyses to explore effect modification will be conducted for characteristics where distributions differ between the trial and RW populations.

Details of the analytic approach are presented in Table 10.

Table 10. Primary and subgroup analysis specification

Hypothesis:	Median rwPFS is longer and the hazard of progression is lower among patients who initiated palbociclib and letrozole versus letrozole alone.
Exposure contrast:	Patients initiating palbociclib and letrozole compared with patients initiating letrozole alone in the first line metastatic setting.
Outcome:	rwPFS
Analytic software:	Aetion Substantiate Version 5.01 (or latest version)
Model(s):	Median rwPFS will be estimated using a Kaplan-Meier estimator, weighted by time-fixed inverse probability of treatment weights. The weighted survival probability $S_a(t)$ for exposure group 'A=a' at time 't' will be as follows: $$S_a(t) = \prod_t 1 - \frac{d_{ta}}{r_{ta}}$$ where $d_{ta} = \sum_{i=1}^{N} w_{it} \bullet Y_{it} \cdot I(A_{it} = a)$ denotes the weighted number of events and $r_{ta} = \sum_{i=1}^{N} w_{it} \cdot I(A_{it} = a)$ denotes the weighted risk set.[21–23] This is equivalent to calculating the Kaplan-Meier estimator in the IPTW weighted population.[22] A non-parametric bootstrap will be used to derive 95% confidence intervals. A Cox proportional hazards model will be used to estimate the hazard ratio at 33 months in the IPTW weighted population as follows: $$h\left(t, L_0\right) = h_0(t) * e^{\beta_1 A_0 + \beta_2^T L_0}$$

	where $h(t, L_0)$ is the hazard of progression at discrete time interval 't' conditional on the vector of potential confounders 'L_0' assessed at baseline; $h_0(t)$ is the baseline hazard at discrete time interval 't'; and A_0 is an indicator for treatment initiation, coded as '1' and '0' for the exposure and comparator group, respectively. Patients with an outcome or censoring event on Day 0 will be assigned a follow-up time of 0.5 days. The proportional hazards assumption will be checked using plots that display the scaled Schoenfeld residuals vs. time for each covariate; if violations are detected, a time-dependent or stratified Cox proportional hazards model will be considered. The Efron method of handling ties will be used.
Confounding adjustment method	Time-fixed inverse probability of treatment weights will be used to adjust for confounding. Individual-level weights will be estimated by the following formula: $$W^A = \frac{1}{f(A\|L_0)}$$ where A is the first study treatment that the patient initiated and L_0 is a vector of baseline confounders. The quantity in the denominator $f(A\|L_0)$—the probability of exposure to treatment A given baseline confounders L_0—will be estimated using a logistic regression model with A as the dependent variable and the vector L_0 as the independent variables. The distribution of weights will be used to identify potential extreme weights. If extreme weights are identified, weight truncation and/or stabilization will be considered. All potential confounder variables will be considered for inclusion in the weight estimation (Table 9). However, as it is not possible to predict the quantity of missing values and sparseness of the data at the time of writing this protocol, the precise functional form of the final regression model will be determined at the time of analysis. Thus, variables with high missingness will be excluded from the final model. Additionally, categorical variables may be collapsed to ensure convergence of the propensity score model.
Missing data methods	Data missingness was assessed as part of an initial feasibility assessment (Appendix B); therefore, key variables are expected to have a high degree of completeness. If substantial missingness results in an insufficient sample size for the complete analytic dataset, alternative variable specifications (e.g., changing the time frames over which variables are assessed) or exclusion of variables may be considered.
Subgroup Analyses	Cox proportional hazards models will be assessed in the following subgroups to align with the trial, as feasible. Measurable disease, international region, and histopathological classification were included as subgroups in the trial but cannot be emulated in the RWD source. 1. Age (18 to <65 yrs, ≥65 yrs) 2. Race (White, Asian) 3. Site of metastatic disease at baseline (visceral (lung, liver), non-visceral) 4. Disease-free interval (<12 months, ≥12 months, *de novo* metastatic)

	5. Prior hormonal therapy (yes, no) 6. Performance status (ECOG 0 or Karnofsky Score 100, ECOG 1-2 or Karnofsky Score 50-90) 7. Bone-only disease at baseline (yes, no) 8. Prior chemotherapy (yes, no) 9. Most recent therapy (aromatase inhibitor, antiestrogen) 10. Number of disease sites (1, ≥2)

7.5.2 Context and rationale for sensitivity analyses

Sensitivity analyses will be conducted to explore the potential impact of several key study design elements. Planned analyses and their respective goals are presented in Table 11.

Table 11. Sensitivity analyses – rationale, strengths and limitations

Description	Primary Analysis	Sensitivity Analysis	Rationale	Strengths of the sensitivity analysis compared to the primary	Limitations of the sensitivity analysis compared to the primary
Contemporaneous cohort	Study population includes patients with first study drug initiation from February 1, 2010 to June 20, 2023.	The study population will be restricted to patients with first study drug initiation from February 1, 2015 to June 20, 2023 to coincide with Food and Drug Administration (FDA) accelerated approval of palbociclib.	Removing historical comparator patients will decrease potential confounding due to changes in treatment paradigms over time.	This analysis will provide effect estimates using a control group that more closely resembles the exposed group with respect to treatment standards.	Restriction to contemporaneous comparator patients may reduce the sample size and introduce other unmeasured sources of confounding due to differences between patients who do and do not initiate newly available treatment.

Description	Primary Analysis	Sensitivity Analysis	Rationale	Strengths of the sensitivity analysis compared to the primary	Limitations of the sensitivity analysis compared to the primary
Limit eligible cohort entry dates to six months prior to the end of the data	Eligible cohort entry period limited to dates at least one month prior to the end of the data.	The eligible cohort entry period will be limited to dates at least six months before the end of the data (February 1, 2010 to February 20, 2023).	Patients identified toward the end of the available study period may not have adjudicated death data, leading to underestimates of death.	Greater minimum follow-up time and opportunity for death to be identified.	The sample size will be reduced relative to the primary analysis.
Modified exposure ascertainment window (one day)	Exposure status is ascertained over a 15-day window following first study drug initiation.	The exposure ascertainment window will be shortened to occur over a single day.	Shortening the exposure ascertainment window will exclude patients who initiate letrozole and palbociclib greater than 1 day apart from the exposed group.	This analysis will more closely align the treatment strategy in the exposed group to that of the trial.	Patients who initiate palbociclib greater than 1 day after letrozole will be eligible for inclusion in the comparator group, which may bias results towards the null.
Modified exposure ascertainment window (30 days)	Exposure status is ascertained over a 15-day window following first study drug initiation.	The exposure ascertainment window will be increased to occur over 30 days.	Increasing the exposure ascertainment window will include patients who initiate letrozole and palbociclib greater than 15 days apart in the exposed group.	This analysis will increase follow-up time for patients who initiate study drugs further apart in time and will employ a more flexible definition of the exposure ascertainment period that aligns with the data vendor line of therapy definition.	Patients who initiate palbociclib greater than 15 days after letrozole will be eligible for inclusion in the exposed group, which may less accurately reflect the treatment strategy used in the trial.

Description	Primary Analysis	Sensitivity Analysis	Rationale	Strengths of the sensitivity analysis compared to the primary	Limitations of the sensitivity analysis compared to the primary
Loss to follow-up censoring definition	Patients are censored on the last date prior to a period of >90 days without curated EHR activity and without death.	Censor patients on the last date prior to a period of >180 days without curated EHR activity and without death.	Lengthening the time period without curated EHR activity will allow patients who are using healthcare less frequently to have outcomes recorded after a gap in 180 days of activity.	Progression and death events that occur after a gap of 90 days without curated EHR activity will be included.	Lengthening the time period without curated EHR activity will increase the possibility of unobserved events or censoring reasons and potentially overestimate rwPFS.
ECOG performance status assessment window	ECOG performance status for study inclusion will be assessed within 90 days from first study treatment initiation.	ECOG performance status for study inclusion will be assessed within 30 days from first study treatment initiation.	Shortening the assessment window for ECOG performance status may more accurately reflect patients' status at the time of study treatment initiation and will provide information on the sensitivity of results to this key inclusion criteria.	Assessing ECOG status closer in time to study treatment initiation may create a study population that is more similar to the trial population.	A larger number of patients may be missing ECOG performance status during the shorter time period. Patients with missing values will be included in the analysis in line with the primary approach.
Complete case - ECOG performance status	Patients with missing ECOG or Karnofsky performance status in the 90 days prior to first study treatment initiation are included.	Patients with missing ECOG or Karnofsky performance status in the 90 days prior to first study treatment initiation will be excluded.	Removing patients with missing ECOG performance status may reduce misclassification of disease severity and will enable descriptive comparison of the study population with and without these patients.	Excluding patients with missing ECOG may create a study population that is more similar to the trial population.	The sample size will be reduced.

7.6 Data source(s)

7.6.1 Context and rationale for data source(s)

The data source used in this study obtains clinical data from multiple partnerships across the U.S. These partnerships include data from organizations that aggregate data from oncology practices. The data vendor receives the complete EHR from each practice, including unstructured notes and scanned documents attached to the EHR. Structured and unstructured data are semi-normalized using an Observational Medical Outcomes Partnership (OMOP)-based common data model to improve uniformity across original data sources. Data curation (i.e., abstraction) of unstructured data from provider notes, pathology and imaging reports, and scanned documents is performed by trained reviewers. Natural Language Processing (NLP) is used to identify relevant strings of data within the patient record, but all curated data points are reviewed, interpreted, and documented by a human curation team. Patients are not filtered on the basis of data completeness to allow for flexibility when selecting the analytic sample.

Reason for selection: The dataset was considered due to its focus on oncology EHR data and was further considered fit-for-purpose after a detailed feasibility assessment that considered available sample size and completeness and quality of key inclusion criteria, exclusion criteria, key confounders, and outcomes. (Appendix B). Linked exposure and outcome data will not be accessed prior to conducting final analyses.

Strengths of data source(s): The dataset includes individuals from all regions of the U.S. and both community and academic providers. Due to the oncology focus of the data vendor, this dataset provides information on important diagnostic, prognostic, and clinical characteristics among breast cancer patients. The data include several important curated fields, including ECOG, line of therapy, and progression, using a broad range of clinical documentation (e.g., physician notes, pathology reports, etc.).

Limitations of data source(s): Algorithms used by the data vendor to derive treatment regimens and certain key variables such as those used to define key inclusion and exclusion criteria, potential confounders, and mortality have not been validated. The data are also limited by the accuracy of data collection in the original EHRs, the subjective nature of data abstraction, and, for some variables, the inability to determine whether missing values indicate the true absence of a condition or missing data.

Data source provenance/curation: The data vendor programmatically tracks all inbound data throughout the data pipeline for data provenance and quality control using internal identifiers. Additionally, the data curation process is operationalized to ensure consistency and high inter-rater reliability. To reduce data entry errors following abstraction, data curation software is utilized. Quality control analyses on various subsets (e.g., a random sample, a specific cohort, only curated patients) before data are finalized for external use.

Table 12. Metadata about data sources and software

Data Source(s):	[Redacted]
Study Period:	Start of all available data-July 20, 2023
Eligible Cohort Entry Period:	February 1, 2010-June 20, 2023
Data Version (or date of last update):	Q3 2023
Data sampling/extraction criteria:	Described above
Type(s) of data:	Clinical data sourced from oncology practices in the U.S.
Data linkage:	N/A
Conversion to Common Data Model:	Observational Medical Outcomes Partnership-based Common Data Model
Software for data management:	Aetion Substantiate Version 5.01 (or latest version)

N/A = not applicable

7.7 Data management

Raw data review

At Aetion, as part of the data ingestion process, raw data review is routinely conducted to understand contents of the data table(s), establish relationships, and help inform the database connection specification. Scientific integrity checks are performed to understand if the contents of the data shipment is consistent with the expected data as laid out in the applicable data usage agreement. Some of the key characteristics explored in this process include:

- Table structure (number of rows, columns, column names etc.)
- Summary counts per table (i.e., non-missing counts, unique counts)
- Variable distribution (e.g., min, mean, median, max for numeric variables; top frequencies for categorical variables)
- Date range (min, max and distribution over a time period)

- Missingness percentage of attributes

Database connection (DBC) process

Following receipt and review of the raw data, a data connector specification is drafted by a data scientist. The specification provides a map to Engineering for transformation of raw data to the Aetion longitudinal patient timeline. It includes information such as:

- Overall schema including tables (event types), rows (events), and columns (attributes); derivation of attributes to improve data flexibility on Platform and rationale for any attributes or events that are dropped
- Event dates that define how data will be reflected on the longitudinal patient timeline, and any minimal processing rules (e.g., drop an event that does not include a start or end date)
- Skeleton structure diagram that represents the logical view of the entire database, defining how the data tables are organized and related in the longitudinal patient timeline and how the relations among them are associated
- Information for user interface and labeling
- Codes and definitions; typically used to substitute users' having to look-up multiple resources to understand/process data

Validation of the DBC is completed to ensure that the implementation of DBC logic leads to transformed data output that connects to and behaves within Aetion Substantiate exactly as intended. Raw data are never loaded as-is; rather, data are transformed (via the DBC) into a longitudinal sequence of healthcare data points for each patient. DBC validation is required to confirm that this transformation was performed correctly. This helps to ensure validity/accuracy of the connected data and its importance cannot be ignored. Validation is performed via double programming, where two different people work independently from the same DBC specification and then compare their output. The DBC is considered validated if the outputs are identical. If the outputs are not identical, then the source of the discrepancy is investigated and resolved.

Following validation, the specification files are used to create an Aetion data dictionary for the dataset. In addition, throughout the data connector spec / creation process, any issues or decisions that have to be made that are not otherwise specified in the Specification files (e.g., how missing dates are handled), are noted in the data dictionary.

7.8 Quality control

Prior to deployment on Aetion Substantiate, a manual test of certain platform features and dataset values is conducted to ensure they are visible and testable on the front-end. This test is run following any deployment activity (such as a version update and/or data/shard update). Checks include:

- Baseline values for database information (dataset name, patient counts, earliest and latest event dates)
- Database configuration (specified dataset values)

- Measure, Cohort, and Analysis Generations to confirm this functionality using the dataset
- Output from generated analysis output
- Coding Systems, if applicable

The implementation of all variables, cohorts, and analytic plans will be individually checked by two analysts. Any discrepancies will be discussed with the analysts and study lead to ensure alignment with the study design outlined in the protocol.

7.9 Study size and feasibility

The PALOMA-2 trial target sample size was 650 patients with 1:1 randomization. This sample size was based on 90% power to detect a hazard ratio for progression of 0.69 at a one-sided alpha level of 0.025 (assuming 347 events).[2] A total of 666 patients were included in the trial.

Sample size requirements to detect a range of hazard ratios relevant to the PALOMA-2 trial are presented in Table 13. If the unweighted study sample size falls below the lowest estimate, corresponding to the required sample size to detect the point estimate for the hazards of progression or death observed in the trial with 80% power, implementation will pause. The study team and CARE Steering Committee will then consider the value of continuing the study with potentially insufficient power, given the lower primacy of statistical significance in an emulation setting.

Table 13. Sample size requirements

	Power	Hazard ratio for death	Ratio of exposed to unexposed	Alpha	Prevalence of death among the unexposed[a]	Total sample size required[b]
Trial sample size calculation, point estimate	90%	0.69	1:1	5%	57.7%	530
Trial sample size calculation, point estimate	80%	0.69	1:1	5%	57.7%	396
Trial result, upper confidence limit	90%	0.72	1:1	5%	57.7%	675
Trial result, upper confidence limit	80%	0.72	1:1	5%	57.7%	505
Trial result, point estimate	90%	0.58	1:1	5%	57.7%	246
Trial result, point estimate	80%	0.58	1:1	5%	57.7%	184

[a] As reported in the trial.

[b] Calculated using the *powerSurvEpi* R package[24], based on the sample size formula for proportional hazards modeling derived by Latouche et al.[25]

In feasibility analyses (Appendix B) among adult, postmenopausal females with metastatic breast cancer and no evidence of ECOG performance status >2 in the data source, there were 944 individuals with evidence of initiating both palbociclib and letrozole, and 660 individuals who initiated letrozole only. This real-world study will include data for all individuals meeting study inclusion and exclusion criteria.

8. Strengths and Limitations

This emulation study is limited by several inherent differences between trial settings and non-interventional studies leveraging RWD.

In the absence of randomization, IPTW will be used to control for confounding. However, important confounders may be unavailable or imperfectly measured in the RWD source, which may result in residual confounding. There may be residual confounding by indication arising from causes of progression that also impact physicians' treatment decisions. Several measures of disease severity (e.g., performance status, stage at initial diagnosis, disease-free interval) will be used to generate treatment probabilities, but these may not sufficiently control for confounding by indication. In particular, real-world patients who did not initiate palbociclib after approval may systematically differ from those treated with palbociclib for reasons that are not captured in the health record. These patients may also be different from those enrolled in the PALOMA-2 RCT, limiting successful emulation of the trial results.

Several trial design elements cannot be perfectly emulated or operationalized due to limitations of the data source. First, the trial included international sites, while the RWD source is restricted to EHR data from U.S. oncology clinics. Second, while patients with locoregionally recurrent disease were included in the trial, similar patients cannot be identified in the data source. These patients constituted a small proportion of the original trial population (<3% in each arm), but restriction of the real-world population to patients with metastatic disease does limit the comparability of the two populations. Third, some study inclusion and exclusion criteria and potential effect modifiers, such as measurable disease, cannot be operationalized and others, such as performance status, lab values, and patient symptoms, may be missing in the RWD due to infrequent real-world clinical assessment or inadequate capture in the EHR. Excluding patients who have missing values for these key variables may result in systematic differences between the trial and real-world populations. A preliminary data feasibility assessment was conducted prior to protocol finalization to ensure that key study eligibility criteria and potential confounders had a high degree of completeness (Appendix B). Fourth, while trial treatments were administered on the same day, these treatment strategies must be approximated (within 15 days) due to differences in medication dosing schedules, insurance delays, and provider decision-making in routine clinical practice. Finally, progression surveillance is conducted less frequently and regularly in real-world practice than was performed in the trial, which may affect estimates of rwPFS.

At the same time, this study proposes to use a high-quality, RWD source specifically designed for conducting RWD analyses in oncology. The data source includes curated, quality-controlled data elements (e.g., ECOG, progression) unique to oncology studies and necessary to this emulation. The preliminary feasibility assessment indicated low missingness of key variables, the ability to create a trial similar population through careful operationalization of trial characteristics, and a larger available sample size than was enrolled in the trial. Differences between

the trial and real-world emulation, including rates of study treatment discontinuation and crossover, will be transparently reported and compared to contextualize final results.

9. Protection of Human Subjects

This study will use de-identified secondary data and therefore does not constitute research involving human subjects. Institutional review board exemption will be requested.

10. Reporting of Adverse Events

Detection and reporting of adverse events do not apply as this study involves secondary use of real-world data from an existing data collection infrastructure.

11. References

1. Merola D, Campbell U, Gautam N, et al. The Aetion Coalition to Advance Real-World Evidence through Randomized Controlled Trial Emulation Initiative: Oncology. *Clin Pharmacol Ther*. 2023;113(6):1217-1222. doi:10.1002/cpt.2800

2. Finn RS, Martin M, Rugo HS, et al. Palbociclib and Letrozole in Advanced Breast Cancer. *N Engl J Med*. 2016;375(20):1925-1936. doi:10.1056/NEJMoa1607303

3. National Institute for Health and Care Excellence. *NICE Real-World Evidence Framework*. National Institute for Health and Care Excellence; 2022. Accessed December 21, 2022. https://www.nice.org.uk/corporate/ecd9/resources/nice-realworld-evidence-framework-pdf-1124020816837

4. Center for Drug Evaluation and Research, Center for Biologics Evaluation, Oncology Center of Excellence. *Considerations for the Use of Real-World Data and Real-World Evidence to Support Regulatory Decision-Making for Drug and Biological Products*. U.S. Food and Drug Administration; 2023. Accessed December 21, 2023. https://www.fda.gov/media/171667/download

5. Purpura CA, Garry EM, Honig N, Case A, Rassen JA. The Role of Real-World Evidence in FDA-Approved New Drug and Biologics License Applications. *Clin Pharmacol Ther*. 2022;111(1):135-144. doi:10.1002/cpt.2474

6. Visvanathan K, Levit LA, Raghavan D, et al. Untapped Potential of Observational Research to Inform Clinical Decision Making: American Society of Clinical Oncology Research Statement. *J Clin Oncol*. 2017;35(16):1845-1854. doi:10.1200/JCO.2017.72.6414

7. Castellanos EH, Wittmershaus BK, Chandwani S. Raising the Bar for Real-World Data in Oncology: Approaches to Quality Across Multiple Dimensions. *JCO Clin Cancer Inform*. 2024;(8):e2300046. doi:10.1200/CCI.23.00046

8. Franklin JM, Pawar A, Martin D, et al. Nonrandomized Real-World Evidence to Support Regulatory Decision Making: Process for a Randomized Trial Replication Project. *Clin Pharmacol Ther*. 2020;107(4):817-826. doi:10.1002/cpt.1633

9. Merola D, Campbell UB, Lenis D, Schneeweiss S, Wang SV. Calibrating real-world evidence against randomized trials: A pilot to inform the CARE Initiative. *Manuscr Submitt Publ*. Published online 2024.

10. Stang A, Poole C, Kuss O. The ongoing tyranny of statistical significance testing in biomedical research. *Eur J Epidemiol*. 2010;25(4):225-230. doi:10.1007/s10654-010-9440-x

11. Pfizer. *A Randomized, Multicenter, Double-Blind Phase 3 Study of PD-0332991 (Oral CDK 4/6 Inhibitor) Plus Letrozole Versus Placebo Plus Letrozole For The Treatment Of Postmenopausal Women With ER (+), HER2 (-) Breast Cancer Who Have Not Received Any Prior Systemic*

Anti Cancer Treatment For Advanced Disease. clinicaltrials.gov; 2023. Accessed December 31, 2023. https://clinicaltrials.gov/study/NCT01740427

12. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *Eur J Cancer*. 2009;45(2):228-247. doi:10.1016/j.ejca.2008.10.026

13. Menopause - Symptoms and causes. Mayo Clinic. Accessed May 22, 2024. https://www.mayoclinic.org/diseases-conditions/menopause/symptoms-causes/syc-20353397

14. Gaitanidis A, Alevizakos M, Pitiakoudis M, Wiggins D. Trends in incidence and associated risk factors of suicide mortality among breast cancer patients. *Psychooncology*. 2018;27(5):1450-1456. doi:10.1002/pon.4570

15. Zaorsky NG, Zhang Y, Tuanquin L, Bluethmann SM, Park HS, Chinchilli VM. Suicide among cancer patients. *Nat Commun*. 2019;10(1):207. doi:10.1038/s41467-018-08170-1

16. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat Med*. 2015;34(28):3661-3679. doi:10.1002/sim.6607

17. Hernán MA, Robins JM. *Causal Inference: What If*. Chapman & Hall/CRC; 2020.

18. 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

19. 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

20. Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. Published online 2010.

21. Xie J, Liu C. Adjusted Kaplan–Meier estimator and log-rank test with inverse probability of treatment weighting for survival data. *Stat Med*. 2005;24(20):3089-3110. doi:10.1002/sim.2174

22. Westreich D, Cole SR, Tien PC, et al. Time Scale and Adjusted Survival Curves for Marginal Structural Cox Models. *Am J Epidemiol*. 2010;171(6):691-700. doi:10.1093/aje/kwp418

23. Cole SR, Hernán MA. Adjusted survival curves with inverse probability weights. *Comput Methods Programs Biomed*. 2004;75(1):45-49. doi:10.1016/j.cmpb.2003.10.004

24. Qui W, Chavarro J, Lazarus R, Rosner B, Ma J. powerSurvEpi: Power and Sample Size Calculation for Survival Analysis of Epidemiological Studies. Published online 2021. https://CRAN.R-project.org/package=powerSurvEpi

25. Latouche A, Porcher R, Chevret S. Sample size formula for proportional hazards modelling of competing risks. *Stat Med*. 2004;23(21):3263-3274. doi:10.1002/sim.1915

26. Huang Y, Alsabbagh MhdW. Estimates of population-based incidence of malignant arrhythmias associated with medication use—a narrative review. *Fundam Clin Pharmacol*. 2020;34(4):418-432. doi:10.1111/fcp.12578

27. Blanch B, Sweeting J, Semsarian C, Ingles J. Routinely collected health data to study inherited heart disease: a systematic review (2000–2016). *Open Heart*. 2017;4(2):e000686. doi:10.1136/openhrt-2017-000686

28. Murthi M, Shaka H, El-amir Z, et al. Association of hypocalcemia with in-hospital mortality and complications in patients with acute pulmonary embolism: results from the 2017 Nationwide Inpatient Sample. *BMC Pulm Med*. 2021;21(1):410. doi:10.1186/s12890-021-01784-0

29. Basnet S, Dhital R, Tharu B, Ghimire S, Poudel DR, Donato A. Influence of abnormal potassium levels on mortality among hospitalized heart failure patients in the US: data from National Inpatient Sample. *J Community Hosp Intern Med Perspect*. 2019;9(2):103-107. doi:10.1080/20009666.2019.1593778

30. Zipursky J, Macdonald EM, Hollands S, et al. Proton Pump Inhibitors and Hospitalization with Hypomagnesemia: A Population-Based Case-Control Study. *PLOS Med*. 2014;11(9):e1001736. doi:10.1371/journal.pmed.1001736

31. Elixhauser A, Steiner C, Palmer L. Clinical Classifications Software (CCS). U.S. Agency for Healthcare Research and Quality. Published 2018. Accessed May 29, 2024. https://hcup-us.ahrq.gov/tools_software.jsp

32. Tanaka Y, Shah NS, Passman R, Greenland P, Lloyd-Jones DM, Khan SS. Trends in Cardiovascular Mortality Related to Atrial Fibrillation in the United States, 2011 to 2018. *J Am Heart Assoc*. 2021;10(15):e020163. doi:10.1161/JAHA.120.020163

33. Tischer TS, Schneider R, Lauschke J, et al. Prevalence of Atrial Fibrillation in Patients with High $CHADS_2$ - and $CHA_2DS_2VAS_c$ -Scores: Anticoagulate or Monitor High-Risk Patients? *Pacing Clin Electrophysiol*. 2014;37(12):1651-1657. doi:10.1111/pace.12470

34. Andrade S, Harrold L, Tjia J, Gurwitz J. Mini-Sentinel Systematic Evaluation of Health Outcome of Interest Definitions for Studies Using Administrative Data: Cerebrovascular Accident/Transient Ischemic Attack Report. Published online 2011. https://www.sentinelinitiative.org/sites/default/files/surveillance-tools/validations-literature/Mini-Sentinel-HOI-Evidence-Review-CVA-TIA-Report_0.pdf

35. Keller K, Hobohm L, Münzel T, Ostad MA. Impact of symptomatic atherosclerosis in patients with pulmonary embolism. *Int J Cardiol*. 2019;278:225-231. doi:10.1016/j.ijcard.2018.12.019

36. Krogager ML, Søgaard P, Torp-Pedersen C, Bøggild H, Gislason G, Kragholm K. Impact of Plasma Potassium Normalization on Short-Term Mortality in Patients With Hypertension and Hyperkalemia. *J Am Heart Assoc*. Published online December 15, 2020. doi:10.1161/JAHA.120.017087

37. Glasheen WP, Cordier T, Gumpina R, Haugh G, Davis J, Renda A. Charlson Comorbidity Index: ICD-9 Update and ICD-10 Translation. *Am Health Drug Benefits*. 2019;12(4):188-197.

12. Appendices

Appendix A: List of abbreviations

Abbreviation	Definition
ASD	Absolute standardized difference
CARE	Coalition to Advance Real-World Evidence through Randomized Controlled Trial Emulation
CDK4/6	Cyclin-dependent kinase 4/6
CNS	Central nervous system
CONSORT	Consolidated Standards of Reporting Trials
CTCAE	Common Terminology Criteria for Adverse Events
CYP3A4	Cytochrome P450 3A4
DBC	Database connection
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
EHR	Electronic health record
ER+	Estrogen receptor-positive
ERBB2	V-erb-b2 avian erythroblastic leukemia viral oncogene homolog 2
FDA	Food and Drug Administration
HER2-	Human epidermal growth factor receptor 2-negative
IPTW	Inverse probability of treatment weighting
NCI	National Cancer Institute
NLP	Natural Language Processing
OMOP	Observational Medical Outcomes Partnership

Abbreviation	Definition
PD-0332991	Palbociclib
PFS	Progression-free survival
PS	Propensity score
QTc	Heart-rate corrected QT interval
RCT	Randomized controlled trials
RECIST	Response Evaluation Criteria in Solid Tumours
RWD	Real-world data
RWE	Real-world evidence
rwPFS	Real-world progression-free survival
TdP	Torsade de Pointes
U.S.	United States

Appendix B: Preliminary feasibility assessment

STEP 1: STATE RESEARCH AIM, QUESTION, AND OBJECTIVES					
STEP 1a: Overarching research aim					
To emulate the PALOMA-2 randomized controlled trial of letrozole + palbociclib as first-line treatment in postmenopausal women with estrogen receptor-positive/human epidermal growth factor receptor 2-negative advanced breast cancer using real-world data.					
STEP 1b: Trial research question					
Among adult postmenopausal women with estrogen receptor-positive/human epidermal growth factor receptor 2-negative advanced breast cancer, does initial treatment with letrozole + palbociclib compared with treatment with letrozole alone result in longer progression-free survival?					
STEP 1c: Trial primary objective(s)					
Among adult postmenopausal women with estrogen receptor-positive/human epidermal growth factor receptor 2-negative advanced breast cancer, compare progression-free survival for patients treated with palbociclib and letrozole and patients treated with letrozole alone.					
DESIGN ELEMENTS	**STEP 2: DESCRIBE ORIGINAL CLINICAL TRIAL**	**STEP 3: DESCRIBE REAL-WORLD DATA STUDY EMULATION OF ORIGINAL CLINICAL TRIAL**			
		3a. Minimal criteria for valid operationalization in real-world data source	**3b. Criteria ranking with regard to uniqueness and importance**	**Data Source**	
OVERALL RATING				**5**	
GENERAL					
Sample size	**Trial sample size**	**1.5x trial sample size[2]**	**Must Have**	Sample size among adult patients with metastatic breast cancer treated with first-line letrozole or letrozole + palbociclib[3]	
Treated	444	666		1,148	**5**
Comparator	222	333		1,086	**5**
Length and frequency of follow-up[1]	Median reported follow-up: 23 months (range not reported)	Sufficient time coverage in dataset to identify outcome after receipt of treatment	**Must Have**	Earliest metastatic diagnosis date: Q3 1980 End of data cut: Q3 2023	**Not Applicable[4]**
VARIABLE-RELATED					

Variable	Trial criterion	Minimal criteria for valid operationalization in any real-world data source based on routine clinical care	Criteria ranking with regard to uniqueness and importance	Operationalization and coverage in data source	Rating
Treatment	125 mg of palbociclib per day, administered orally in 4-week cycles (3 weeks of treatment followed by 1 week off); 2.5 mg of letrozole per day, administered orally (continuous treatment)	Date of palbociclib and letrozole treatment	**Must Have**	Date of palbociclib and letrozole treatment is available. Lines of therapy are available.	**5**
Comparator	Placebo once per day, administered orally in 4-week cycles (3 weeks of treatment followed by 1 week off); 2.5 mg of letrozole per day, administered orally (continuous treatment)	Date of letrozole treatment	**Must Have**	Date of letrozole treatment is available. Lines of therapy are available.	**5**
Inclusion Criterion 1	Women, 18 years of age or older	Year of birth; sex	**Must Have**	Year of birth and sex are available.	**5**
Inclusion Criterion 2	Estrogen receptor-positive (ER+), human epidermal growth factor receptor 2-negative (HER2-) advanced breast cancer	Diagnosis of breast cancer; date of advanced/metastatic diagnosis; date and result of biomarker tests	**Must Have**	Breast cancer diagnosis, date of metastatic diagnosis, and dates and results of biomarker tests are available. Date of advanced diagnosis is not available.	**4**

Variable	Trial criterion	Minimal criteria for valid operationalization in any real-world data source based on routine clinical care	Criteria ranking with regard to uniqueness and importance	Operationalization and coverage in data source	Rating
Inclusion Criterion 3	No prior systemic therapy for advanced disease	Names/types and dates of antineoplastic treatment; date of advanced/metastatic diagnosis	**Must Have**	Dates of antineoplastic treatment are available. Lines of therapy are available.	**5**
Inclusion Criterion 4	Postmenopausal status: Prior bilateral oophorectomy, spontaneous cessation of menses for 12 consecutive months or more, or follicle-stimulating hormone and estradiol levels in postmenopausal ranges without an alternative cause	Diagnosis of menopause or related symptoms; curated post-menopausal status; date of last menstrual period	**Nice to Have**	Post-menopausal status is available.	**5**
Inclusion Criterion 5	Adequate organ function	Dates and results of laboratory tests for hematologic, hepatic, and renal function.	**Nice to Have**	Laboratory test results and dates are available.	**5**
Inclusion Criterion 6	ECOG performance status of 0 to 2	ECOG performance status result	**Must Have**	ECOG or Karnofsky performance status is available.	**5**
Inclusion Criterion 7	Measurable disease according to RECIST v1.1, or lesions only in the bone	RECIST assessment is not performed in routine care	**Not Applicable**	RECIST is not used to assess progression or response in a real-world setting. Progression will be assessed with available real-world information (see below).	**Not Applicable**

Variable	Trial criterion	Minimal criteria for valid operationalization in any real-world data source based on routine clinical care	Criteria ranking with regard to uniqueness and importance	Operationalization and coverage in data source	Rating
Exclusion Criterion 1	Prior adjuvant or neoadjuvant treatment with a nonsteroidal aromatase inhibitor was allowed unless disease had recurred while the patient was receiving the therapy or within 12 months after completing therapy	Date of surgery with curative intent; dates of antineoplastic treatment	**Nice to Have**	Dates of (neo)adjuvant and adjuvant treatment are available, but completeness could not be ascertained.	**4**
Exclusion Criterion 2	Advanced, symptomatic, visceral spread (i.e., spread to the viscera or main organs of the body) who were at risk for short-term, life-threatening complications	Life threatening complication in the short term is a qualitative assessment made by a physician and is not routinely available in medical records.	**Not Applicable**	Receipt of a study treatment will be assumed to indicate the clinician determined patient is not at risk of life-threatening complication in the short term. Per National Comprehensive Cancer Network (NCCN) guidelines, study treatments are not approved for palliative intent, but may be used in the real-world setting.	**Not Applicable**
Primary Outcome 1 (Definition & Ascertainment)	Progression-free survival	Date of death; curated progression variable; imaging results; dates of healthcare interactions	**Must Have**	Curated progression information is available. Date of death is available. Date of last activity is available.	**4**
Confounding Variable 1	Not applicable in a randomized setting	Age	**Must Have**	Year of birth is available.	**5**
Confounding Variable 2	Not applicable in a randomized setting	Sex	**Must Have**	Sex is available.	**5**
Confounding Variable 3	Not applicable in a randomized setting	Race/ethnicity	**Must Have**	Race/ethnicity is available.	**5**

Variable	Trial criterion	Minimal criteria for valid operationalization in any real-world data source based on routine clinical care	Criteria ranking with regard to uniqueness and importance	Operationalization and coverage in data source	Rating
Confounding Variable 4	Not applicable in a randomized setting	Performance status	**Must Have**	ECOG or Karnofsky performance status is available.	**5**
Confounding Variable 5	Not applicable in a randomized setting	Smoking status	**Must Have**	Smoking status is available.	**5**
Confounding Variable 6	Not applicable in a randomized setting	Progression/disease free-interval (Time from initial diagnosis to advanced/metastatic diagnosis)	**Must Have**	Date of initial and metastatic diagnosis are available. Date of advanced diagnosis is not available.	**4**
Confounding Variable 7	Not applicable in a randomized setting	Year of study treatment initiation	**Must Have**	Year of study treatment initiation is available.	**5**
Confounding Variable 8	Not applicable in a randomized setting	Number and/or location(s) of metastatic sites	**Must Have**	Metastatic site location is available. Number of metastatic sites can be determined.	**5**

Abbreviations: ECOG = Eastern Cooperative Oncology Group; ER+ = estrogen receptor positive; HER2- = human epidermal growth factor receptor 2-negative; NCCN = National Comprehensive Cancer Network; Q1/Q2/Q3/Q4 = quarter of year (Q1: January - March, Q2: April - June, Q3: July - September, Q4: October - December); RECIST = Response Evaluation Criteria in Solid Tumors.

Footnotes:

1. Follow-up time is stated as reported in the trial publication. Maximum available observation time is reported for the real-world data source. These are not directly comparable.
2. The minimum sample size for feasibility analyses was selected to account for expected attrition when all eligibility criteria are applied.
3. Sample sizes were calculated only among patients with metastatic breast cancer because curated dates of advanced breast cancer diagnosis are not available and line of therapy is defined relative to metastatic diagnosis in these data sources.
4. The final study period would be defined in the study protocol based on the date of treatment approval and relevant updates to treatment guidelines.

Keys for Ranking	
Scoring for Data Sources by Data Elements	
Scoring	**Description**
1	Data Requirements are not met
2	

3	Some data requirements are met
4	
5	All or nearly all data requirements are met

Appendix C: Code lists

Design Element	Variable	Code Type	Code
Inclusion	Non-adenocarcinoma histology	Diagnosis name	Intraductal carcinoma, non infiltrating, NOS Infiltrating duct mixed with other types of carcinoma (C50) Inflammatory carcinoma (C50) Comedocarcinoma, NOS (C50) Cribriform carcinoma, NOS Cribriform carcinoma in situ (C50) Papillary carcinoma, NOS Intraductal micropapillary carcinoma (C50) Comedocarcinoma, non infiltrating (C50) Paget disease, mammary (C50) Infiltrating lobular mixed with other types of carcinoma (C50) Neuroendocrine carcinoma, NOS Encapsulated papillary carcinoma with invasion (C50)
Inclusion	Bilateral surgical oophorectomy	Surgery name	Bilateral salpingo-oophorectomy with omentectomy, total abdominal hysterectomy and radical dissection for debulking Bilateral salpingo-oophorectomy with omentectomy, total abdominal hysterectomy and radical dissection for debulking; with pelvic lymphadenectomy and limited para-aortic lymphadenectomy Bilateral salpingo-oophorectomy with total omentectomy, total abdominal hysterectomy for malignancy Total abdominal hysterectomy with bilateral salpingo-oophorectomy Resection (initial) of ovarian, tubal or primary peritoneal malignancy with bilateral salpingo-oophorectomy and omentectomy; with radical dissection for debulking (ie, radical excision or destruction, intra-abdominal or retroperitoneal tumors) Resection (initial) of ovarian, tubal or primary peritoneal malignancy with bilateral salpingo-oophorectomy and omentectomy; with total abdominal hysterectomy, pelvic and limited para-aortic lymphadenectomy Laparoscopic bilateral salpingo-oophorectomy
Inclusion	Absolute neutrophil count (ANC)	Lab test name	Neutrophils [#/volume] in Blood Neutrophils [#/volume] in Blood by Automated count Neutrophils [#/volume] in Blood by Manual count
		Unit	x10(3)/mcL cells/uL 10*9 cells/L

Design Element	Variable	Code Type	Code
			10*9/L
Inclusion	Platelets	Lab test name	Platelets [#/volume] in Blood Platelets [#/volume] in Blood by Automated count Platelets [#/volume] in Blood by Estimate Platelets [#/volume] in Blood by Manual count
		Unit	10*9 cells/L x10(3)/mcL
Inclusion	Hemoglobin	Lab test name	Hemoglobin [Mass/volume] in Blood Hemoglobin [Mass/volume] in Blood by calculation Hemoglobin [Mass/volume] in Arterial blood Hemoglobin [Mass/volume] in Venous blood
		Unit	g/dL
Inclusion	Creatinine clearance	Lab test name	Creatinine [Mass/volume] in Blood Creatinine [Mass/volume] in Serum or Plasma
		Unit	mg/dL
Inclusion	Serum total bilirubin	Lab test name	Bilirubin.total [Mass/volume] in Serum or Plasma Bilirubin.total [Mass/volume] in Blood
		Unit	mg/dL
Inclusion	Gilbert's disease	ICD-10 code	E80.4
Inclusion	AST (SGOT)	Lab test name	Aspartate aminotransferase [Enzymatic activity/volume] in Serum or Plasma Aspartate aminotransferase [Enzymatic activity/volume] in Serum or Plasma by With P-5'-P
		Unit	U/L IU/L
Inclusion	ALT (SGPT)	Lab test name	Alanine aminotransferase [Enzymatic activity/volume] in Blood Alanine aminotransferase [Enzymatic activity/volume] in Serum or Plasma by With P-5'-P Alanine aminotransferase [Enzymatic activity/volume] in Serum or Plasma by No addition of P-5'-P Alanine aminotransferase [Enzymatic activity/volume] in Serum or Plasma
		Unit	U/L IU/L
Inclusion	Alkaline phosphatase	Lab test name	Alkaline phosphatase [Enzymatic activity/volume] in Serum or Plasma Alkaline phosphatase [Enzymatic activity/volume] in Blood
		Unit	U/L

Design Element	Variable	Code Type	Code
			IU/L
Exclusion	CYP3A4 inhibitors	Generic name	Amprenavir Atazanavir Boceprevir Clarithromycin Conivaptan Delavirdine Diltiazem Erythromycin Fosamprenavir Indinavir Itraconazole Ketoconazole Lopinavir Mibefradil Miconazole Nefazodone Nelfinavir Posaconazole Ritonavir Saquinavir Telaprevir Telithromycin Verapamil Voriconazole
Exclusion	Major surgery	Surgery name	Modified radical mastectomy Partial mastectomy Partial mastectomy with axillary lymphadenectomy Simple mastectomy Prophylactic mastectomy Excision of internal mammary lymph nodes Excision of sentinel lymph node Radical mastectomy Subcutaneous mastectomy Total mastectomy and axillary clearance Hysterectomy Ligation of fallopian tube Simple mastectomy with axillary lymph node sampling

Design Element	Variable	Code Type	Code
Exclusion	Long or short QT syndrome[26]	ICD-10 code	I45.8
Exclusion	Brugada syndrome[27]	ICD-10 code	I49.8
Exclusion	Torsade de Pointes (TdP)	ICD-10 code	I47.21
Exclusion	Hypocalcemia[28]	ICD-10 code	E83.51
Exclusion	Hypokalemia[29]	ICD-10 code	E87.6
Exclusion	Hypomagnesemia[30]	ICD-10 code	E83.42
Exclusion	Myocardial infarction	ICD-10 code	I21 I22 I25.2
Exclusion	Severe/unstable angina	ICD-10 code	I20.0
Exclusion	Cardiac dysrhythmias[31]	ICD-10 code	I47.0 I47.1 I47.2 I47.9 I48.0 I48.1 I48.2 I48.3 I48.4 I48.91 I48.92 I49.1 I49.2 I49.3 I49.40 I49.49 I49.5 I49.8 I49.9 R00.0 R00.1 R00.2
Exclusion	Atrial fibrillation[32]	ICD-10 code	I48
Exclusion	Symptomatic congestive heart failure[33]	ICD-10 code	I50.9
Exclusion	Cerebrovascular accident[34]	ICD-10 code	I60 I61

Design Element	Variable	Code Type	Code
			I62 I63 H34.1 G45
Exclusion	Symptomatic pulmonary embolism[35]	ICD-10 code	I26
Exclusion	Inflammatory bowel disease or chronic diarrhea[36]	ICD-10 code	K50
Exclusion	Inflammatory bowel disease or chronic diarrhea	ICD-10 code	K51
Exclusion	Short bowel syndrome	ICD-10 code	K90.82
Exclusion	Human immunodeficiency virus infection[37]	SNOMED code	91947003 445945000
Covariate	Disease site: visceral	Diagnosis name	Metastatic malignant neoplasm to liver Metastatic malignant neoplasm to left lung Metastatic malignant neoplasm to pleura Metastatic malignant neoplasm to lung Metastatic malignant neoplasm to right lung Metastatic malignant neoplasm to respiratory tract
Covariate	Prior adjuvant or neoadjuvant chemotherapy	Therapy name	Capecitabine Carboplatin Cyclophosphamide Docetaxel Doxorubicin Epirubicin Etoposide Fluorouracil Gemcitabine Methotrexate Oxaliplatin Paclitaxel Pemetrexed
Covariate	Prior adjuvant hormone	Therapy name	Anastrozole Exemestane Fulvestrant Goserelin Letrozole

Design Element	Variable	Code Type	Code
			Leuprolide Megestrol Tamoxifen Toremifene Triptorelin