Protocol C4891001

A PHASE 3, RANDOMIZED, OPEN-LABEL, MULTICENTER TRIAL OF ARV-471 (PF-07850327) VS FULVESTRANT IN PARTICIPANTS WITH ESTROGEN RECEPTOR-POSITIVE, HER2-NEGATIVE ADVANCED BREAST CANCER WHOSE DISEASE PROGRESSED AFTER PRIOR ENDOCRINE BASED TREATMENT FOR ADVANCED DISEASE (VERITAC-2)

Statistical Analysis Plan
(SAP)

Version: **Version 2**

Date: **November 05, 2024**

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

APPENDICES

1. VERSION HISTORY

Table 1. Summary of Changes

Version/ Date	Associated Protocol Amendment	Rationale	Specific Changes
V1 (16Dec2022)	Protocol Amendment 1 (25Oct2022)	NA	NA
05Nov2024	Protocol Amendment 3 (04Nov2024)	1. Statistical hypothesis testing strategy was amended upon FDA recommendation based on emerging evidence of differential efficacy of SERDs in ESR1 mutation positive vs ESR1 mutation negative patients. The new hypothesis testing strategy incorporates a testing order based on clinical importance and logical relevance per FDA comments.	Section 5.1.2 sample size calculation Section 5.1.3 Multiplicity adjustment Section 7 Interim analyses
		2. Add Sensitivity Analysis 4 to reflect accuracy of data source	Section 6.1.1.2 Stratification factor using data from CRF instead of IRT
		3. Add Sensitivity Analysis 5 to address FDA request	Section 6.1.1.2 PFS analysis for participants with ESR1 mut negative (excluding ESR1 mut unknown)
		4. Add pre-specified minimum efficacy criteria of PFS for ESR1 mut negative subgroup to address FDA request	Section 5.1.3 Multiplicity Adjustment and Decision Rules. Pre-specified a minimum efficacy criteria according to Rothmann criteria.

		5. Add Supplemental Analysis 6 to address EMA comments	Section 6.1.1.3 Supplemental analysis for PFS using treatment policy (regardless new anticancer therapy, AE withdrawal, etc. for censoring)
		6. Provided more details for pharmacokinetic analysis and ECG analyses for QTc substudy.	Section 6.2.2.2 Section 6.2.2.3
		7. Provide further details on subset grouping and add additional subgroup analysis.	Section 6.3 Editing the list of subset analyses and added a variable of prior CDK4/6 inhibitor treatment
		8. Simplify Dose Intensity Calculation	Section 6.4.4.1 Specified calculations for each drug.
		9. Editorial changes to provide clarity and streamline contents.	Section 3.3.2 Section 3.4.1 Section 3.5.3 Section 5.2.2 Section 5.3.4 Section 6.2.2.1 Section 6.2.2.3 Section 6.2.2.5 Section 6.2.2.6 Section 6.4.4.2 Section 6.5.1 Section 6.5.4

2. INTRODUCTION

This statistical analysis plan (SAP) provides the detailed methodology for summary and statistical analyses of the data collected in Study C4891001.

Study C4891001 is a global Phase 3 multicenter, randomized, open label, parallel-group study aimed to demonstrate that ARV-471 is superior to fulvestrant in prolonging the

progression-free survival (PFS) in women and men with ER(+)/HER2(-) aBC who have progressed after prior endocrine treatment-based regimen(s).

The past decade has seen significant improvements in the treatment options for patients with advanced ER(+)/HER2(-) breast cancer. However, there is still an unmet need to improve clinical outcomes after progression with existing ET therapies, particularly for the second- and third-line patients where the desire is to delay treatment with chemotherapy. ARV-471 may represent a potential alternative due to its unique mechanism of action that may have the advantage of achieving complete inhibition of the ER by essentially eliminating the entire protein. This may be particularly advantageous in the setting of constitutively active ESR1 mutations.

Preliminary data from the ongoing FIH Phase 1/2 ARV-471-mBC-101 study suggest that ARV-471 has the potential to provide clinical benefit to participants with ER(+)/HER2(-) a/mBC.

ARV-471 has been found to be safe and well-tolerated in the ongoing FIH study, up to a total daily dose of 700 mg. Coupled with the data from non-clinical studies, the risk of clinically significant toxicities in this study is assessed as low. In addition, ARV-471 was active in non-clinical models harboring clinically relevant ESR1 mutations associated with resistance to ET, including superior TGI compared to fulvestrant in a Y537S ESR1 *mutant* patient-derived xenograft model. Preliminary clinical data from the ongoing FIH ARV 471-mBC-101 study demonstrated that ARV-471 has antitumor activity in pretreated patients with CDK4/6 inhibitors plus ET for their a/mBC.

The totality of data from these nonclinical pharmacology, PK and metabolism, and toxicology studies and available safety and efficacy clinical data support continued clinical development of ARV-471 in ER(+) / HER2(-) advanced breast cancer.

The purpose of this study is to demonstrate the superiority of ARV-471 in terms of progression free survival compared to fulvestrant in participants with ER(+)/HER2(-) aBC who have progressed after prior endocrine based treatment(s) for their advanced disease in two populations: 1) all randomized participants, and 2) participants with ESR1 mutation.

In this document, when describing study treatment of the experimental drug, “vepdegestrant” is used interchangeably with “ARV-471”.

2.1. Study Objectives, Endpoints, and Estimands

Table 2. Primary and Secondary Objectives and Endpoints

Objectives	Endpoints
Primary:	**Primary:**

Objectives	Endpoints
• To demonstrate that ARV-471 is superior to fulvestrant in prolonging PFS in participants with ER(+)/HER2(-) aBC (all participants and participants with ESR1 mutation-positive BC) who have received prior endocrine-based treatment for their advanced disease	• PFS, defined as the time from the date of randomization to the date of first documented disease progression, as determined by BICR per RECIST v1.1, or death due to any cause, whichever occurs first
Key Secondary:	**Key Secondary:**
• To demonstrate that ARV-471 is superior to fulvestrant in prolonging overall survival (all participants and participants with ESR1 mutation-positive BC)	• OS, defined as the time from the date of randomization to the date of death due to any cause
Secondary:	**Secondary:**
• To compare measures of tumor control between treatment arms and to evaluate the DOR within each treatment arm	• OR: confirmed CR or PR by BICR • CBR defined as confirmed CR or PR at any time or SD or non-CR/non-PD ≥24 weeks by BICR • DOR
• To evaluate safety and tolerability between the treatment arms	• Type, incidence, severity (as graded by NCI CTCAE v5.0), seriousness and relationship to study medications of AEs and any laboratory and ECG abnormalities
• To characterize the effects of ARV-471 on QTc	• QTc
• To evaluate patient reported outcomes between two treatment arms	• EORTC QLQ C30 • EORTC QLQ BR23 • EuroQol; EQ-5D-5L • BPI-SF
• To determine plasma concentrations of ARV-471 and ARV-473 after repeated dosing of ARV-471	• Plasma concentrations of ARV-471 and its epimer ARV-473
• To assess changes from baseline levels in plasma ctDNA	• ctDNA plasma quantitative changes from baseline to evaluate their associations with clinical outcomes
Tertiary/Exploratory:	**Tertiary/Exploratory:**
• To evaluate correlations of biomarkers with ARV-471 exposure, efficacy and other clinical outcomes	• Tumor tissue and/or blood biomarkers, to determine gene mutations (eg, ESR1, PIK3CA), and cell proliferation and oncologic driven proteins (eg, Ki67) involved in the biology aBC, mechanisms of resistance of ARV-471

2.1.1. Primary Estimands

Primary Estimand/Coprimary Estimands

The primary estimand of this study is based on a hypothetical strategy, which estimates the treatment effect if participants maintain their randomized treatment and adhere to the protocol. The estimand is defined according to the primary objective and is in alignment with the primary endpoint PFS.

Treatment: ARV-471 vs fulvestrant

Population: Participants with ER(+)/HER2(-) advanced breast cancer who have progressed after prior endocrine-based treatment(s) for their advanced disease, regardless of tolerability,

duration of study treatment, initiation of subsequent anti-cancer therapy, or missing tumor assessments in 1) all randomized participants, and 2) participants with ESR1 mutation.

Variable: PFS defined as the time from randomization to the first documentation of objective PD determined by BICR per RECIST v1.1, or death due to any cause, whichever occurs first.

Intercurrent events: PFS data will be censored on the date of the last adequate tumor assessment for participants with the following intercurrent events

- Discontinue the study treatment due to withdrawal of consent prior to an event
- Start a new anti-cancer therapy prior to an event
- Have an event after an unacceptably long interval (2 or more missing, incomplete or non-evaluable assessments)
- Lost to follow-up

Population-level summary: A stratified log-rank test will be used to compare PFS between the two treatment arms. HR for PFS with the corresponding 2-sided 95% CI will be calculated based on Cox's proportional hazard model. PFS time associated with each treatment arm will be summarized using the Kaplan-Meier method and median PFS time with the corresponding 2-sided 95% CI will be reported.

2.1.2. Secondary Estimands

Secondary Estimands

The key secondary estimand of this study is defined according to the secondary objective for OS and is in alignment with the secondary endpoint OS. This secondary estimand is based on a treatment policy strategy, which estimates the treatment difference regardless of whether an intercurrent event occurs.

Treatment: ARV-471 vs fulvestrant

Population: Participants with ER(+)/HER2(-) advanced breast cancer who have progressed after prior endocrine-based treatment(s) for their advanced disease, regardless of tolerability, duration of study treatment, initiation of subsequent anti-cancer therapy in 1) all randomized participants, and 2) participants with ESR1 mutation.

Variable: OS defined as the time from the date of randomization until the date of death due to any cause. Survival status is expected to be collected irrespective of study treatment discontinuation or participant's request to discontinue study procedures. All participants who have not withdrawn consent for further participation in the study should be followed for survival until the end of the study.

Intercurrent events: All data will be used regardless of the occurrence of intercurrent events (discontinuation of study treatment, use of subsequent treatments, discontinuation of study, etc.)

Population-level summary: A stratified log-rank test will be used to compare OS between the two treatment arms. HR for OS with the corresponding 2-sided 95% CI will be calculated based on Cox's proportional hazard model. In order to account for the group sequential design for OS, the repeated CI (RCI) method (Jennison and Turnbull, 2000) will also be used to construct the 2-sided RCIs for the HR. OS time associated with each treatment arm will be summarized using the Kaplan-Meier method and median OS time with the corresponding 2-sided 95% CI will be reported.

2.2. Study Design

This is a global Phase 3 multicenter, randomized, open-label, parallel-group study aimed to demonstrate that ARV-471 is superior to fulvestrant in prolonging the PFS (by BICR assessment) in participants with ER(+)/HER2(-) aBC who have progressed after prior endocrine treatment-based regimen(s) for advanced disease.

The schema of the study is illustrated in Figure 1. Approximately 560 participants (of which approximately 280 are participants with ESR1 mutation) will be randomly assigned to Arm A (ARV-471, PF-07850327) or Arm B (fulvestrant).

Participants will be randomly assigned on a 1:1 basis to:

- Arm A: (Investigational Arm; n ≈ 280). Participants will receive ARV-471 200 mg orally, once daily on a 28-day continuous dosing schedule.
- Arm B: (Comparator Arm; n ≈ 280). Participants will receive fulvestrant 500 mg, intramuscularly on Days 1 and 15 of Cycle 1 and then on Day 1 of each cycle starting from C2D1 (28-day cycle).

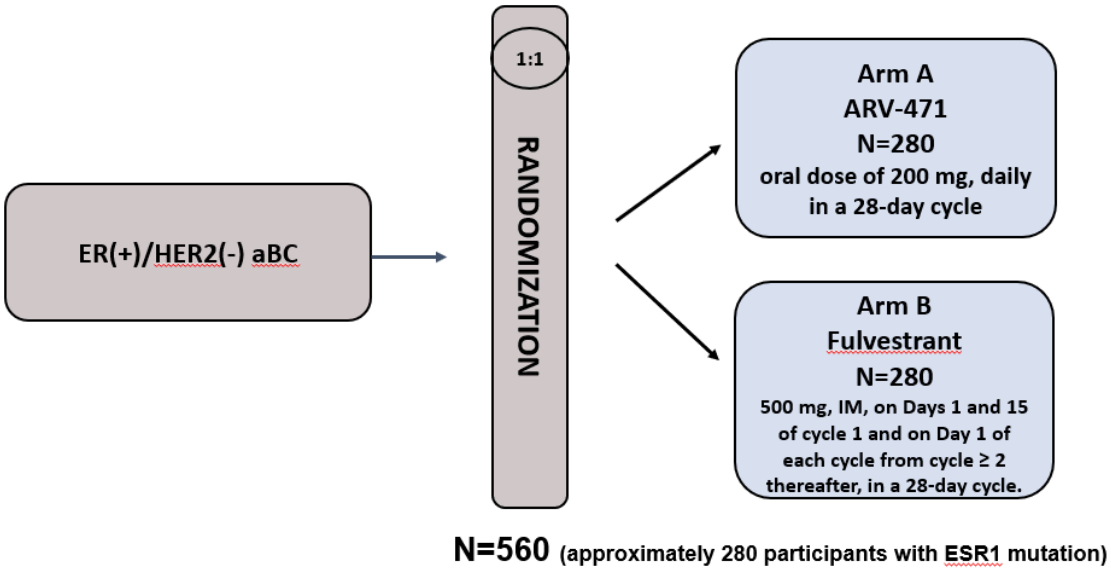
Participants will be stratified by ESR1 mutational status (Mutant, Yes/No), and visceral disease (Yes/No). Visceral refers to at least one disease site of lung, liver, brain, pleural, and peritoneal involvement, in baseline target and non-target tumor assessments.

Crossover between treatment arms will NOT be allowed.

An E-DMC will be established to review aggregate safety data to monitor safety and tolerability by treatment arm.

Participants will undergo efficacy assessments as outlined in schedule of activities section of study protocol. Efficacy analyses will be performed using response assessed by BICR as the primary data source. BICR will review all radiological images and clinical information. All radiographic images will be collected and objectively verified by an independent third-party core imaging laboratory as described in the Study Imaging Manual. It is anticipated that the final PFS analysis will occur at approximately 310 PFS events in all participants population and 165 PFS events in the ESR1 mutation positive subgroup population. The final OS analysis will occur at approximately 396 OS events in all participants population and 194 OS events in ESR1 mutation positive subgroup population.

Figure 1. Study Design



In Arm A, blood samples for pharmacokinetics analysis of ARV-471 and its epimer, ARV-473, will be collected. These samples will also be used for exposure/response analysis for safety and efficacy findings.

The study includes a QTc sub-study in approximately 80 participants in Arm A enrolled at selected sites which will evaluate the effect of ARV-471 on QTc interval via triplicate ECGs (central reading) time-matched with select PK draws.

Study Arms and Duration:

Intervention Name	ARV-471 (PF-07850327)	Fulvestrant
Arm Name	A (Investigational)	B (Comparator)
Unit Dose Strength(s)	100 mg	250 mg/5 mL
Route of Administration	Oral	IM
Use	Experimental	Active comparator
IMP or NIMP/AxMP	IMP	IMP

Study Arm(s)		
Arm Title	Arm A (ARV-471)	Arm B (Fulvestrant)
Arm Type	Experimental	Active comparator
Arm Description	Participants will receive two ARV-471 100 mg tablets (200 mg QD) administered orally, once daily, on Days 1-28 of each 28-day cycle.	Participants will receive fulvestrant as two 5 mL IM injections (250 mg/5 mL) on Days 1 and 15 of Cycle 1 and then on Day 1 of each cycle starting from C2D1 (28-day cycle).

Participants will continue to receive assigned treatment until objective disease progression, unacceptable toxicity, death, participant refuses further treatment.

Pre- and peri-menopausal female and male participants must receive therapy with the LHRH agonist starting at Cycle 1 Day 1 (if not already on treatment) and continued during the study.

3. ENDPOINTS AND BASELINE VARIABLES: DEFINITIONS AND CONVENTIONS

3.1. Primary Endpoint – PFS

The primary endpoint is PFS which is defined as the time from the date of randomization to the date of the first documentation of objective progression of disease (PD) determined by BICR per RECIST v1.1, or death due to any cause in the absence of PD, whichever occurs first, in each of the two populations: 1) all randomized participants, and 2) participants with ESR1 mutation. The final PFS analyses will be performed when the required number of events are reached in both testing populations.

PFS (in months) = (date of event or censoring – date of randomization +1)/30.4375

Disease assessments are performed as scheduled according to the schedule of activities to prevent the introduction of bias into the assessment of efficacy. Screening/baseline tumor assessments are carried out within 28 days of randomization (unless otherwise specified). Post-baseline tumor assessments are performed every 8 weeks for the first 48 weeks, every 12 weeks thereafter and every 24 weeks for bone scan, from randomization until disease progression. If a tumor assessment must be performed early or late, subsequent assessments are conducted according to the original schedule. Additional imaging assessments may be performed at any time if clinically indicated (eg, suspected PD, symptomatic deterioration, etc).

Radiographic imaging, for all known or suspected disease sites, is required at screening and during the active treatment phase using the same imaging techniques. The CT and MRI scans (including brain CT/MRI, if applicable) should be performed with contrast agents unless contraindicated for a medical reason. Premedication prior to contrast media administration as per local guidelines is allowed. If IV contrast is medically contraindicated, the imaging modality to be used to follow the disease (either CT without contrast or MRI) should be the modality which best evaluates the disease, and the choice should be determined by the investigator in conjunction with the local radiologist. MRI of the abdomen and pelvis can be substituted for CT if MRI adequately depicts the disease. However, CT of the chest should not be substituted by chest MRI even if IV contrast is contraindicated. In such case CT will be performed without contrast. If MRI is used to follow up bone lesion(s) it must be performed a few days before any treatment that may affect bone marrow cellularity (ie, G-CSF).

Addressing Intercurrent Events

PFS will be censored on the date of the last adequate disease assessment for participants who do not have an event (PD per RECIST v1.1 or death due to any cause), before the start of a new anticancer therapy for participants who start a new anticancer therapy (e.g. systemic therapy, surgery on the lesions, therapeutic radiation) prior to an event, or before the gap for participants with an event after a gap of 2 or more missing disease assessments. Participants who do not have an adequate post-baseline disease assessment will be censored on the date of randomization unless death occurs on or before the time of the second planned disease assessment in which case the death will be considered an event.

Addressing Missing Data

Participants who do not have an evaluation of tumor response after randomization and do not have an early death prior to the second planned disease assessment or who have inadequate baseline or post-baseline/follow-up tumor assessments will have their PFS time censored on the date of randomization with the duration of one day.

PFS data will be censored on the date of the last tumor assessment for participants who are alive and do not have objective tumor progression.

3.2. Secondary Endpoints

3.2.1. Key Secondary Endpoint - OS

OS is the key secondary endpoint. The study is also designed to perform the hypothesis testing to demonstrate that ARV-471 is superior to fulvestrant in prolonging OS in participants with ER(+)/HER2(-) advanced breast cancer whose disease has progressed on prior endocrine treatment(s) in each of the two populations: 1) all randomized participants, and 2) participants with ESR1 mutation. The final OS analyses will be performed when the required number of events are reached in both testing populations.

OS is defined as the time from date of randomization to date of death due to any cause.

OS (in months) = (date of death or censoring – date of randomization +1)/30.4375

Following the End of Treatment visit, survival data and follow-up information will be collected in all participants every 3 months from the end of treatment. In addition to survival information, the study will collect anticancer treatment information.

Addressing Intercurrent Events

All data will be used regardless of whether or not the intercurrent event (discontinuation of study treatment, use of subsequent treatments, discontinuation of study, etc.) occurs.

Addressing Missing Data

Participants lacking survival data beyond randomization will have their OS times censored at randomization with the duration of one day.

In the absence of confirmation of death, survival time will be censored to last date the participant is known to be alive. For participants lacking survival data beyond the date of their last follow up, the OS time will be censored on the last date they were known to be alive.

3.2.2. Objective Response (OR)

OR is defined as a best overall response of confirmed complete response (CR) or confirmed partial response (PR) according to RECIST v.1.1 recorded from randomization until disease progression, or death due to any cause. Only tumor assessments performed on or before the start date of any new anti-cancer therapies will be considered in the assessment of best overall response. Both CR and PR must be confirmed by repeat assessments performed no less than 4 weeks after the criteria for response are first met. Otherwise, the participant will be considered as non-responders in the OR rate analysis. Additionally, participants with inadequate data for tumor assessment (e.g., no baseline assessment or no follow-up assessments) will be considered as non-responders in the OR rate analysis.

3.2.3. Duration of Response (DOR)

DOR is defined as the time from the first documentation of objective tumor response (CR or PR) to the first documentation of objective tumor progression, or death due to any cause, whichever occurs first. DOR will only be calculated for the subgroup of participants with an OR.

3.2.4. Clinical Benefit Response (CBR)

CBR is defined as confirmed CR or PR at any time, or SD/Non-CR/Non-PD (for participants with non-measurable disease) ≥24 weeks according to the RECIST version 1.1 recorded in the time period between randomization and disease progression, or death of any cause, whichever occurs first.

3.3. Other Endpoint(s)

3.3.1. Patient Reported Outcomes (PRO)

The following patient reported outcome measurements will be collected in this study:

EORTC QLQ C30, EORTC QLQ BR23 (female and male questionnaire versions), EQ-5D-5L, BPI-SF, mBPI-SF (worst pain severity [item 3] and pain interference [item 9a]), injection site pain (Arm B only), pain medication usage via analgesic log, and patient preference questionnaire.

3.3.2. QTc

The QTc sub-study is designed to characterize the effect of ARV-471 on QTc and includes subjects from the QTc Substudy Analysis Set.

3.3.3. Biomarker

Tumor tissues and/or blood samples will be used to evaluate biomarkers (genes and proteins) involved in the biology of the participant's disease, and mechanisms of sensitivity or resistance to ARV-471, to conduct genomic, and/or molecular profiling analyses.

3.3.4. PK Endpoints

Plasma concentrations of ARV-471 and its epimer ARV-473: Blood samples will be collected for PK assessment of ARV-471 and the epimer ARV-473 (same sample draw) according to the PK sampling schedule in the study protocol.

3.4. Baseline Variables

3.4.1. Study drug, study treatment and baseline definitions

In this study, study drug refers to ARV-471or Fulvestrant and study treatment refers to one of the following:

- **Arm A**: ARV-471 at a dose of 200 mg orally once daily

- **Arm B**: Fulvestrant at 500 mg intramuscularly on Days 1 and 15 of cycle 1 and then on Day 1 of each cycle starting from C2D1 (28-day cycle)

Start and end dates of study treatment
The treatment start date for participants on both Arms is the date the participant receives the first dose of the study treatment, which should be on or shortly after the date the participant is randomized in the IRT system. The date of last dose of study treatment is the latest date of non-zero dosing of the study drug.

Definition of baseline

In this study, the last available assessments collected on or prior to randomization date will be used as baseline information for efficacy analyses.

The last measurements collected on or prior to the first dose of study treatments will be used as the baseline measurements for safety, biomarker, and PRO analyses.

If an assessment is planned to be performed prior to the first dose of study treatment in the protocol and the assessment is performed on the same day as the first dose of study treatment, it will be assumed that it was performed prior to study treatment administration. Unscheduled assessments will be used in the determination of baseline.
Participants who start and discontinue study treatment on the same day may have two different sets of data collected on study day 1 (one during study and one in the End of Treatment (EOT) visit). Data reported at the EOT visit are not eligible for baseline selection.

3.4.2. Baseline characteristics

Randomization is stratified by the following, as recorded in the IRT where randomization was performed:

- ESR1 mutation positive (yes vs. no)
- Visceral Disease (yes vs. no)

Other baseline characteristics include age at randomization, gender, race/ethnicity, country of origin, height, weight, hormone receptor and HER2 status, ECOG performance status, menopausal status, primary tumor characteristics at diagnosis, and prior treatments (surgery, radiation, chemotherapy, targeted therapy, hormonal therapy), etc.

3.5. Safety Endpoints

3.5.1. Adverse Events

Safety assessment will consist of monitoring of all AEs, including serious AEs (SAEs), regular monitoring of hematology, blood chemistry, physical examinations, and vital signs. All AEs are reported from informed consent signature through treatment completion (EOT)

and including a minimum of 28 calendar days after the last administration of the study intervention.

Overall safety profile is characterized by type, frequency, severity as graded by NCI CTCAE v.5.0, timing, and relationship to treatment on each arm, and laboratory abnormalities observed. AEs will be classified using the latest version of the Medical Dictionary for Regulatory Activities (MedDRA) classification system.

Treatment-Emergent Adverse Events (TEAEs)
A conservative case approach will be taken whereby all AEs that occur on or after the first dose of study treatment and before the EOT (i.e., 28 days after last dose of study treatment) will be considered TEAEs.

TEAEs related to study drug (ARV-471, fulvestrant) as judged by the investigator, i.e., treatment related AEs, are collected.

Adverse Events of Special Interest (AESIs)
AESIs are examined as part of routine safety data review procedures throughout the clinical trial and as part of signal detection processes. A list of AESIs will be specified before final analysis.

3.5.2. Laboratory Data

Laboratory assessment will be performed according to protocol schedule of activities (see Protocol Table 2 and 3). Lab assessment parameters (including hematology and blood chemistry) are specified in Protocol Appendix 2.

Laboratory assessments
Hematology
Hemoglobin
Platelets
WBC
Absolute Neutrophils (or % as per local practice)
Chemistry
ALT
AST
Alkaline Phosphatase
Sodium
Potassium
Magnesium
Total Calcium
Total Bilirubin[a]
BUN or Urea
Serum Creatinine
eGFR as per 2021 CKD-EPI Equations (at screening only for eligibility)
Albumin
HbA1C (fasting)
Cholesterol/Triglycerides
Glucose (fasting)

Laboratory assessments
Coagulation (only at screening)
aPTT
INR
PT
Additional Test
FSH test (at screening only) to confirm a postmenopausal status of female participants under the age of 60 and with cessation of regular menses for 12 consecutive months and with no alternative medical cause.
Pregnancy Test
For WOCBP participants only. Serum pregnancy test (at screening). Following screening pregnancy tests (urine or serum) must have sensitivity of least 25 mIU/mL

a For Hy's law potential cases, in addition to repeating AST and ALT, laboratory tests should include albumin, creatine kinase, total bilirubin, direct and indirect bilirubin, gamma-glutamyl transferase, prothrombin time /INR, and alkaline phosphatase.

3.5.3. ECG Data

ECG measurements will include PR interval, QT interval, QTcF, RR interval, QRS duration and heart rate. QTcF will be used for the primary analysis. Analyses will be conducted in a similar manner but separately for the QTc Substudy Analysis Set using centrally read ECG data and the Safety Analysis Set.

4. ANALYSIS SETS (POPULATIONS FOR ANALYSIS)

For the purposes of analysis, the following analysis sets are defined in table 3.

Table 3. Analysis Sets

Participant Analysis Set	Description
Full Analysis Set (FAS)	All enrolled participants who were randomized. Participants are analyzed according to the treatment they have been randomized to receive.
Safety Analysis Set (SAS)	All randomized participants who receive at least 1 dose of study intervention.
PK Concentration Set	All participants who are in the Safety Analysis Set and have at least 1 concentration of either ARV-471 or ARV-473.
PK Steady State Trough Concentration ($C_{trough, ss}$) Set	All participants who are in the PK concentration set and who have $C_{trough,ss}$ of ARV-471 or ARV-473. $C_{trough,ss}$ is defined as pre-dose plasma concentration collected within an allowable time window relevant to dosing following at least 7 consecutive days of 200 mg ARV-471.
QTc Substudy Analysis Set	A subset of SAS participants in Arm A, who are randomized at selected sites and have ECG measurements to evaluate the effect of ARV-471 on QTcF via serial triplicate ECGs (centrally read), must have baseline ECG measurement (Cycle 1 Day 1 pre-dose) and at least one ECG measurement on Day 1 of Cycle 2 or Cycle 3 following at least 7 consecutive days of 200 mg ARV-471

Participant Analysis Set	Description
Biomarker Analysis Set(s)	Biomarker analysis population is defined as all randomized participants with at least 1 of the Biomarkers evaluated at pre and/or post dose.

5. GENERAL METHODOLOGY AND CONVENTIONS

This study is designed to demonstrate that ARV-471 is superior to fulvestrant in prolonging the PFS (by BICR assessment) in participants with ER(+)/HER2(-) aBC who have progressed after prior endocrine treatment-based regimen(s) for advanced disease, in each of the two populations: 1) all randomized participants, and 2) participants with ESR1 mutation.

Approximately 560 participants (of which approximately 280 are participants with ESR1 mutation) will be randomly assigned to Arm A (ARV-471, PF-07850327) or Arm B (fulvestrant) on a 1:1 ratio.

The final PFS analyses will be performed when the required number of events are reached in both testing populations. It is anticipated that the final PFS analysis will occur at approximately 310 PFS events in all participants population and 165 PFS events in the ESR1 mutation positive subgroup population. No interim analysis is planned for PFS.

The primary analyses for progression free survival (PFS) will include all data up to and including the cutoff date which is determined by the number of events required for the analyses of PFS. Corresponding data snapshots will be taken to allow cleaning of data up to and including the cutoff date.

The final OS analysis will occur at approximately 396 OS events in all participants population and 194 OS events in ESR1 mutation positive subgroup population.

All primary and secondary endpoints based on radiological assessments of tumor burden (ie, PFS, OR, DoR, CBR) will be based on BICR assessments per RECIST v1.1.

5.1. Hypotheses and Decision Rules

5.1.1. Hypotheses

PFS:

The primary objective of this study is to demonstrate that ARV-471 is superior to fulvestrant in prolonging the PFS (by BICR assessment) in participants with ER(+)/HER2(-) aBC who have progressed after prior endocrine treatment-based regimen(s) for advanced disease, in each of the two populations: 1) all randomized participants, and 2) participants with ESR1 mutation.

PFS assessed by BICR as per RECIST v1.1 will be the primary endpoint for the study.

The following statistical hypotheses will be tested to address the primary objective:

$H_{0\ \text{all participants}}$: $HR_{PFS} \geq 1$ vs $H_{1\ \text{all participants}}$: $HR_{PFS} < 1$

$H_{0\ \text{ESR1 } mutant\ \text{participants}}$: $HR_{PFS} \geq 1$ vs $H_{1\ \text{ESR1 } mutant\ \text{participants}}$: $HR_{PFS} < 1$

where HR_{PFS} is the hazard ratio (ARV-471 vs fulvestrant) of PFS.

OS:

The key secondary objective of the study is to demonstrate that ARV-471 is superior to fulvestrant in prolonging the OS in participants with ER(+)/HER2(-) aBC who have progressed after prior endocrine treatment-based regimen(s) for advanced disease, in each of the two populations: 1) all randomized participants, and 2) participants with ESR1 mutation.

The following statistical hypotheses will be tested to address this key secondary objective for OS:

$H_{0\ \text{all participants}}$: $HR_{OS} \geq 1$ vs $H_{1\ \text{all participants}}$: $HR_{OS} < 1$

$H_{0\ \text{ESR1 } mutant\ \text{participants}}$: $HR_{OS} \geq 1$ vs $H_{1\ \text{ESR1 } mutant\ \text{participants}}$: $HR_{OS} < 1$

where HR_{OS} is the hazard ratio (ARV-471 vs fulvestrant) of OS.

5.1.2. Sample Size Determination

The study is designed to test the null hypothesis vs. the alternative hypothesis noted below in each of the two participant populations: 1) all randomized participants, and 2) participants with ESR1 mutation. The following statistical hypotheses will be tested to address the primary objective:

$H_{0\ \text{all participants}}$: $HR_{PFS} \geq 1$ vs $H_{1\ \text{all participants}}$: $HR_{PFS} < 1$

$H_{0\ \text{ESR1 } mutant\ \text{participants}}$: $HR_{PFS} \geq 1$ vs $H_{1\ \text{ESR1 } mutant\ \text{participants}}$: $HR_{PFS} < 1$

The 1-sided α level of 1.875% was used for sample size calculation according to the statistical hypothesis testing strategy.

For PFS in all participants population, based on a 1:1 randomization allocation ratio, a total of 310 PFS events will provide 92.5% power for a 1-sided log-rank test at a 0.01875 significance level to detect a HR <0.67. For PFS in participants with ESR1 mutation, a total of 165 PFS events will provide 88% power for a 1-sided log-rank test at a 0.01875 significance level to detect a HR <0.60.

It is estimated that approximately 560 participants will be needed to observe the 310 PFS events in all participants population, and out of which, approximately 280 ESR1 mutation positive participants are required to observe 165 PFS events in the subgroup population. Participants will be randomly assigned on a 1:1 basis to:

- Arm A: ARV-471, Investigational Arm – n ≈ 280 or

- Arm B: fulvestrant Comparator Arm – n ≈ 280

Participants will be stratified by ESR1 mutational status (*Mutant*, Yes/No), and visceral disease (Yes/No).

The sample size described above will also allow the assessment of differences in the secondary endpoint of OS. For OS, approximately 396 death events will provide 80% power for a 1-sided log-rank test at 0.025 significance level to detect a HR < 0.75 in all participants population; and 194 death events provide 80% power to test a HR <0.667 in ESR1 mutation positive population at 0.025 significance level.

The sample size calculation is based on the following assumptions:

- The median PFS and OS for participants receiving fulvestrant are 4 months and 25 months, respectively.
- The median PFS for participants receiving ARV-471 is 6 months in overall randomized population, which corresponds to a HR of 0.67 under an exponential assumption; and the median PFS is 6.7 months in ESR1 mutation positive population, which corresponds to a HR of 0.60 under an exponential assumption.
- The median OS for participants receiving ARV-471 is 33.2 months in overall randomized population, which corresponds to a HR 0.75 under an exponential assumption; and the median OS is 37.5 months in ESR1 mutation positive population, which corresponds to a HR of 0.67 under an exponential assumption.
- Participants dropout rate of 10% in each arm by the 20th month for PFS.
- Participant dropout rate of 10% in each arm by the 40th month for OS.

Sample Size for QTc sub-study:
The study includes a QTc sub-study in approximately 80 participants in Arm A enrolled at selected sites which will characterize the effect of ARV-471 on QTc interval via triplicate ECGs (central reading).

At each ECG measurement time point, a total of 60 evaluable participants is estimated to provide an upper limit of 90% confidence interval of 8 ms for a mean increase of 5 ms, or 10 ms for a mean increase of 7 ms in change from baseline QTcF, based on an estimated standard deviation of 14 ms for the change in QTcF from Study ARV-471-mBC-101. Approximately 80 participants will be enrolled to ensure at least 60 participants are evaluable for the QTc analysis.

Evaluable participants are defined as those who:

- Receive at least 7 consecutive daily doses of ARV-471 200 mg prior to each post-baseline ECG measurement

5.1.3. Multiplicity Adjustment and Decision Rules

The study will perform the following statistical hypotheses tests. Figure 2 illustrates the strategy of multiplicity adjustment.

Primary endpoint family:

$H_{0\ \text{all participants}}$: $HR_{PFS} \geq 1$ vs $H_{1\ \text{all participants}}$: $HR_{PFS} < 1$

$H_{0\ \text{ESR1 } mutant\ \text{participants}}$: $HR_{PFS} \geq 1$ vs $H_{1\ \text{ESR1 } mutant\ \text{participants}}$: $HR_{PFS} < 1$

And secondary endpoint family:

$H_{0\ \text{all participants}}$: $HR_{OS} \geq 1$ vs $H_{1\ \text{all participants}}$: $HR_{OS} < 1$

$H_{0\ \text{ESR1 } mutant\ \text{participants}}$: $HR_{OS} \geq 1$ vs $H_{1\ \text{ESR1 } mutant\ \text{participants}}$: $HR_{OS} < 1$

In this study, there are two hypothesis tests in each of the two endpoint families. Multiplicity adjustments are needed to control the FWER for the overall study.

The four null hypotheses are denoted by H_1 through H_4 and are grouped into two families (primary efficacy endpoint PFS, and key secondary efficacy endpoint OS). The null hypotheses H_1 and H_3 correspond to the comparisons in ESR1 mutation positive population; while H_2 and H_4 correspond to the comparisons in ITT.

	ESR1-mut	ITT
PFS family	H_1	H_2
OS family	H_3	H_4

A graphical multiple testing strategy will be used to strongly control the family-wise error rate of the whole study at the 1-sided 0.025 level. The full α (1-sided 0.025) will initially be split into 3/4 α (0.01875) and 1/4 α (0.00625) between the PFS family and OS family. The alpha splitting for primary endpoint PFS and the key secondary endpoint OS is to ensure a chance to test OS in ESR1 mutation positive population if PFS is positive in this population.

Within each family, the testing of PFS will be conducted in a hierarchical manner using a gatekeeping procedure. Specifically:

The ESR1 mutation positive population will be tested first with 3/4 α (0.01875).

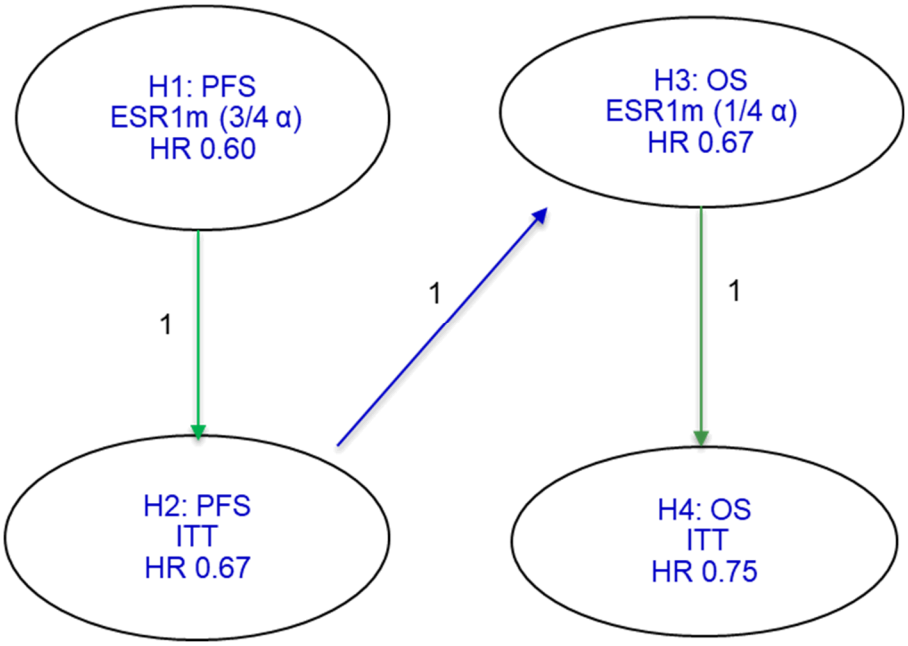
- If H1 test is positive, the statistical significance of PFS in ESR1 mutation positive population can be claimed, and the H2 will then be tested at the same 3/4 α (0.01875) level.
- If H1 test is negative, the statistical significance of PFS in ESR1 mutation positive population cannot be claimed, and H2 will not be tested.
- If H2 test is positive, the statistical significance of PFS in ITT can be claimed.

The significance level for OS family depends on the results from the PFS family. If the PFS tests are positive in both testing populations, 3/4 α used for PFS testing will be reallocated to OS endpoint family and OS will be tested with full α (0.025). If PFS test is positive in ESR1 mutation positive population but negative in ITT population, the original allocated 1/4 α (0.00625) will be used for OS test in the ESR1mutation positive population. If the PFS test is negative in the ESR1 mutation positive population, no hypothesis testing will be performed for OS.

For the tests in OS family, the ESR1 mutation positive population will be tested first. Only if the OS test is positive in ESR1 mutation positive population, the OS in ITT population can be tested. The illustration of the graphical multiple testing strategy is outlined in

Figure 2.

Figure 2. Graphical Multiple Testing Strategy



Note: value 1 in the figure between the hypotheses (circles) represent the full amount of alpha to be passed from one hypothesis test to the next when the hypothesis test is positive.

Two interim analyses are planned for OS. A Lan-DeMets alpha-spending function that approximates O'Brien-Fleming stopping boundaries will be used to control the overall type I error rate at 0.00625/0.025 level (1-sided). See section 7 for details of the interim analyses of OS.

Evaluation of treatment effect in participants without ESR1 mutation:

The study is not designed for a formal hypothesis testing in participants without ESR1 mutation. However, to address emerging evidence of differential efficacy of SERDs in ESR1 mutation positive vs ESR1 mutation negative participants, a minimum efficacy effect is pre-specified following Rothmann criteria (Clin Trials. 2023 August; 20(4): 341–350) for the subgroup of participants who are ESR1 mutation negative. Specifically, the minimum effect size of HR for PFS is set to be ≤0.789, which is the critical HR boundary of the hypothesis testing for all participants population with 3/4 α level and 310 PFS events.

5.2. General Methods

5.2.1. Analyses for Time-to-Event Endpoints

The time-to-event data include PFS, OS, and DOR.

For PFS and OS analyses, the log-rank test will be used for comparing treatments. Hazard ratios and the associated 2-sided 95% CIs are estimated by Cox proportional hazards models, including treatment as a covariate. Ties in the event times will be handled by the *exact* method.

In order to account for the group sequential design for the key secondary endpoint (OS), the repeated CI (RCI) method (Jennison and Turnbull, 2000) will be used to construct the 2-sided RCIs for the HR at the interim and the final analyses of OS. In addition, the unadjusted 2-sided 95% CIs for the HR will also be reported at the interim and the final analyses for OS.

Time-to-event endpoints will be summarized using the Kaplan-Meier method and displayed graphically when appropriate. Graphs will describe the number of participants at risk over time. 2-sided 95% CIs for medians and quartiles are based on the Brookmeyer-Crowley method. The survival probability at the specific time point will be estimated using the Kaplan-Meier method and a 2-sided 95% CI will be calculated using the log-log transformation with back transformation to a CI on the untransformed scale. The estimate of the standard error will be computed using Greenwood's formula.

Duration of follow-up for time-to-event endpoints in each treatment arms will be summarized using the reverse Kaplan-Meier method (Schemper and Smith, 1996) by reversing time-to-event endpoints' censoring and event indicators.

The proportional hazards (PH) assumption will be tested using Schoenfeld s residual test (Therneau and Grambsch, 1994). The departures from PH could be evidenced by a p-value <0.05. If there is evidence of departures from proportional hazards, then PFS and OS may be analyzed based on restricted mean survival time (RMST) differences as a supplementary analysis (Zhang, 2013).

When RMST method is used, to avoid arbitrary selection of the common cut-off τ for both treatment arms, three sets of analyses will be performed:

- τ_1 = minimum of (largest observed survival time for experimental arm, largest observed survival time for control arm).
- τ_2 = minimum of (largest survival event time for experimental arm, largest survival event time for control arm).
- $\tau 3$ = midpoint between $\tau 1$ and $\tau 2$

Since participants in both treatment arms may receive other available treatments after disease progression, the treatment effect on OS may not be able to estimate properly by above defined methods because of the confounding factors. Therefore, the proper testing statistics such as Wilcoxon test and methods like Rank-Preserving Structural Failure Time (RPSFT) approach proposed by Robins and Tsiatis may be applied to the OS analysis to assess the impact of subsequent therapy.

Frequency (number and percentage) of participants with each event type (progression of disease, and death) and censoring reasons will be presented by treatment arm. Reasons for censoring will be summarized according to the categories in Table 4 for PFS and Table 5 for OS following the hierarchy shown.

Table 4. PFS Censoring Reasons and Hierarchy

Hierarchy	Condition	Censoring Reason
1	No or inadequate baseline tumor assessment	No adequate baseline tumor assessment
2	Start a new anti-cancer therapy prior to an event	Start a new anti-cancer therapy
3	Event after 2 or more consecutive missing or inadequate post-baseline disease assessments	Unacceptable gap between event and disease assessment
4	Withdrawal of consent (for all follow-up or for clinical follow-up only) before or after beginning protocol therapy and prior to an event	Withdrawal of consent for all or clinical follow-up
5	Lost to follow-up	Lost to follow-up
6	No event and no adequate tumor assessment after randomization	No adequate post-baseline tumor assessment
7	None of the conditions in the prior hierarchy are met and prior to an event	In follow-up for disease progression

Table 5. OS Censoring Reasons and Hierarchy

Hierarchy	Condition	Censoring Reason
1	Withdrawal of consent after randomization and prior to an event	Withdrawal of consent for all follow-up
2	Lost to follow-up	Lost to follow-up
3	None of the conditions in the prior hierarchy are met and prior to an event	Alive in survival follow-up

5.2.2. Analyses for Binary Endpoints

In general, binary endpoints will be summarized by frequency counts and percentages. Where applicable, the point estimate of the odds ratio between two treatment groups will be estimated using Mantel-Haenszel method, exact 95% CI will be calculated, and p-value will be based on Cochran-Mantel-Haenszel test.

5.2.3. Analyses for Continuous Endpoints

Continuous variables will be summarized using descriptive statistics, including the non-missing number, mean, standard deviation, median, minimum, and maximum.

5.2.4. Analyses for Categorical Endpoints

Categorical variables will be summarized by frequency counts and percentages. Unless otherwise specified, the calculation of percentages will include the missing category. Therefore, counts of missing observations will be included in the denominator and presented as a separate category.

5.2.5. Analyses for PRO Endpoints

The EORTC QLQ C30, EORTC QLQ BR23, EQ-5D-5L, and BPI-SF will be completed by the participants while in the clinic prior to any Day 1 procedure at Cycles 1, 2, 3, 4, 5, 6 and on at odd cycles. In case of treatment delay, participants will complete the PRO questionnaires twice (at planned cycle and at actual treatment administration). The planned assessment will serve for the main analysis, whereas actual assessment will be analyzed in sensitivity analysis which will be specified in PRO SAP.

Analyses of EORTC QLQ-C30, QLQ-BR23 (for women participants only), EQ-5D-5L and BPI-SF will be presented by treatment groups. Analyses of the patient-reported outcomes will be presented on all participants population, and will also be presented separately for the participants with ESR1 mutation subgroup.

A PRO data completeness table will provide the numbers and percentages of participants who completed each instrument, including both the available data rate (FAS is used as a fixed denominator) and the completion rate (denominator is the number of patients expected to complete the PRO measure at the designated PRO assessment timepoint). Data completeness at the domain level will be specified in a separate SAP specific to the PRO endpoints (PRO SAP).

5.2.6. Evaluation of the Discordance Rate between Investigator and BICR on Assessing PFS data

PhRMA Method

Potential evaluation bias between the investigator and BICR assessments with respect to either the progression status of the participant or the timing at which progression occurs will be evaluated using two measures, the early discrepancy rate (EDR) and late discrepancy rate

(LDR). The agreement between the investigator and the BICR within a treatment arm is represented in a tabular form below:

Table 6. Concordance of Disease Assessment between Investigator and BICR

	Blinded Third-party Core Imaging Laboratory	
Investigator	PD	No PD
PD	a=a1+a2+a3	b
No PD	c	d

a1: number of agreements on timing and occurrence of PD.
a2: number of times investigator declares PD later than BICR
a3: number of times investigator declares PD earlier than BICR
PD: short for PFS event (PD, or death for investigator assessments and PD or death for BICR),

The EDR is defined as:

$$EDR = (b + a3) / (a + b)$$

The EDR represents a predictive value of investigator assessment and quantifies the frequency with which the investigator declares progression early relative to BICR within each arm as a proportion of the total number of investigators assessed PDs.

The LDR is defined as:

$$LDR = (c + a2) / (b + c + a2 + a3)$$

The LDR quantifies the frequency that investigator declares progression later than BICR as a proportion of the total number of discrepancies within the arm. If the distribution of discrepancies is similar between the arms, it suggests the absence of evaluation bias favoring a particular arm.

The EDR and LDR will be calculated for each treatment arm and the differential discordance around each measure will be summarized as the rate on the experimental arm minus the rate on the control arm. A negative differential discordance for the EDR and/or positive differential discordance for the LDR are suggestive of a bias in the investigator favoring the experimental arm.

5.3. Methods to Manage Missing Data

Participants may have missing data due to several reasons (regulatory reasons, missed measurement at a clinical visit, missed visit, treatment/study discontinuation, etc.). In general data points will not be imputed, including PRO data and missing tumor assessments.

Missing statistics, e.g., when they cannot be calculated, should be presented as 'NA'. For example, if N=1, the measure of variability (SD) cannot be computed and should be presented as 'NA'.

5.3.1. Missing Dates

Attempts will be made to obtain missing date information for the evaluation of baseline characteristics, primary and secondary endpoints. If all attempts to obtain full dates fail, partially incomplete dates (day or day/month missing) will be imputed.

If start dates for adverse events or concomitant medications are completely missing a worst-case approach will be taken whereby the events will be considered treatment emergent and the medications will be considered concomitant. If only partial information is available (e.g. only a month and year or only a year) and the partial information provide sufficient information to indicate the dates are prior to the start of study treatment (e.g. month/year less than month/year of first dose) then these will be considered to have started prior to treatment; otherwise a similar worst case approach will apply and these will be considered to have started after treatment.

Date of Last Dose of Study Treatment

No imputation will be done for first dose date. Date of last dose of study treatment, if unknown or partially unknown, will be imputed as follows:

- If the last date of study treatment is completely missing and there is no Disposition CRF page for the treatment phase and no death date, the participant should be considered to be ongoing and use the data cutoff date for the analysis as the last dosing date; or
- If the last date of study treatment is completely or partially missing and there is EITHER a Disposition CRF page for the treatment phase OR a death date available (on or prior to the data cutoff date), then impute this date as the last dose date:

= 31DECYYYY, if only Year is available and Year < Year of min (Date of Completion/Discontinuation from the Disposition CRF page for the treatment phase, death date),

= Last day of the month, if both Year and Month are available and Year = Year of min (Date of Completion/Discontinuation from the Disposition CRF page for the treatment phase, death date) and Month < the month of min (EOT date, death date), or

= min (Date of Completion/Discontinuation from the Disposition CRF page for the treatment phase, death date), for all other cases.

Missing or Partial Death Dates

Missing or partial death dates will be imputed based on the last contact date:

- If the entire date is missing it will be imputed as the day after the date of last contact; or

- If the day or month is missing, death will be imputed to the maximum of the full (non-imputed) day after the date of last contact and the following:

Missing day: 1st day of the month and year of death, or

Missing day and month: January 1st of the year of death.

Date of Start of New Anti-cancer Therapy

Incomplete dates for new anti-cancer therapy will be imputed as follows and will be used to determine censoring dates for efficacy analyses:

- The end date of new anti-cancer therapy will be included in the imputation for start date of new anti-cancer therapy. If the end data of new anti-cancer therapy is:
  - completely missing then it will be ignored in the imputations below,
  - partially missing with only year available then the imputations below will consider 31DECYYYY as the end date of the new anti-cancer therapy, or
  - partially missing with only month and year available then the imputations below will consider the last day of the month for MMMYYYY as the end date of the new anti-cancer therapy.
- For participants who have not discontinued study treatment at the time of the data cutoff date, last dose of study treatment is set to the data cutoff date in the imputations below.
- If the start date of new anti-cancer therapy is completely or partially missing then the imputed start date of new anti-cancer therapy is:

= 31DECYYYY, if only Year is available and Year < Year of min [max (PD date + 1, last dose of study treatment + 1), end date of new anti-cancer therapy]

= Last day of the month, if both Year and Month are available and

Year = Year of min [max (PD date + 1, last dose of study treatment + 1), end date of new anti-cancer therapy]

Month < Month of min [max (PD date + 1 day, last dose of study treatment + 1 day), end date of new anti-cancer therapy]

= min [max (PD date + 1, last dose of study treatment + 1), end date of new anti-cancer therapy], for all other cases.

AE Onset Date

The following imputation rules apply if the event is unique for a participant or it is the first of a series of similar events; otherwise, the AE Onset Date will not be imputed:

- If the AE Collection Date is not missing, is less than the Date of First Exposure to Treatment, and is less than the AE Stop Date, then AE Onset Date is set to the Date of AE Collection.
- If the Previous Visit Date is greater than the Date of First Exposure to Treatment and less than the AE Stop Date, the AE Start Date is set to the previous visit date.
- If the Date of First Exposure to Treatment is greater than the previous visit date and less than the AE Stop Date, the AE Onset Date is set to the Date of First Exposure to Treatment.
- Otherwise AE Onset date is set to the AE Stop date.

AE Stop Date

Ongoing events will have the AE Stop Date set to one of the following values:

- Date of Death, if the participant died and a date of death exists.
- Maximum of (Participant Withdraw date, AE Onset Date, AE Collection Date) if the participant withdrew from the study and a date of withdraw exists.
- Maximum of (AE Onset Date, Participant Summary Collection Date, AE Collection Date) if the Disposition CRF page for the long-term follow-up phase exists but a date of completion/discontinuation does not exists.
- Maximum of (Last Treatment Date, AE Onset Date) if no Disposition CRF page for the long-term follow-up phase exists.

Imputation will only occur if event is unique for the participant, or it is the last of a series of similar events; otherwise the Stop Date will not be imputed. Adverse Events are deemed similar if they have the same verbatim term.

Resolved events will have the AE Stop Date set to the maximum of the AE collection date and the AE Onset date.

Other Missing or Partial Dates

Imputation methods for other partial dates as follows:

- If the day of the month is missing for a start date used in a calculation, the first day of the month will be used to replace the missing date.
- If both the day and month are missing for a start date, the first day of the year is used.
- For stop dates, the last day of the month, or last day of the year is used if the day or day and month are missing, respectively.

If the date is completely missing, no imputation will be performed.

5.3.2. Missing Data in Time-To-Event Endpoints

For all time to event endpoints unless otherwise specified, if the day of the month is missing for any date used in a calculation, the first day of the month will be assigned. However, if the month and the year of the event are same as the month and year when the participant was randomized, then the date of the randomization will be assigned. The time of a time-to-event endpoint cannot be negative.

Drop-outs without treatment are defined as those participants who have been randomized in the study but withdraw their consent or are withdrawn by the investigator from the study immediately thereafter but prior to first administration of study drug. The reasons will be collected and reported. These participants will not be replaced and will be included in the FAS analysis set with censorship defined in Section 3.

In case of missing disease assessments before the last scheduled disease assessment or before death, PFS event will be defined based on the actual disease assessment and not the scheduled disease assessments. Therefore, for the primary analysis, PFS will be reported according to the actual reporting time and not the scheduled time.

All other missing values will be reported explicitly.

5.3.3. Missing Patient Reported Outcome Data

Developers' guidance will be followed in determining the observed value of any particular endpoint which is made of multiple items.

For all questionnaires, in cases where two answers are given to one item, the more severe answer will be counted. For QLQ-C30 and QLQ-BR23, if at least half of the constituent items for the multi-item functional or symptom scale have been answered, then the score for that scale may be pro-rated based on the non-missing items. For the EQ-5D, the index score is considered missing if the answer to any of the 5 items is missing. For the BPI pain severity score, if any of the 4 questions are missing, then the calculated mean pain severity score will be considered missing. For the BPI pain interference score, the mean will be set to missing if fewer than 4 pain interference questions are answered.

5.3.4. Missing Toxicity Grade of Adverse Events

No data imputation for missing toxicity grades of adverse events.

5.3.5. Missing ECG Data

For ECG analyses, no values will be imputed for missing data.

If one or two of the triplicate measurements for an ECG parameter are missed, the average of the remaining two measurements or the single measurement can be used in the analyses. If all triplicate measurements are missing at a timepoint for an ECG parameter, no values will be imputed for this timepoint. If the triplicate needs to be repeated because of an artifact, then the

repeated triplicate will be reported on an unscheduled CRF page. Based on a review of the data these unscheduled assessments may be used in place of the assessments at the nominal time. Data review and consultation with the study team is required to flag these cases.

5.3.6. Missing Pharmacokinetic Data

Concentrations below the limit of quantification

For all calculations, figures, and estimation of individual pharmacokinetic parameters, all concentrations assayed as below the limit of quantification (BLQ) will be set to zero. PK parameter listings and summaries will be generated. In log-linear plots these values will not be represented. The BLQ values will be excluded from the estimation of individual pharmacokinetic parameters and from calculations of geometric means and their confidence intervals. A statement similar to 'All values reported as BLQ have been replaced with zero' should be included as a footnote to the appropriate tables and figures. In listings BLQ values will be reported as below limit of quantification ("<LLOQ"), where LLOQ will be replaced with the corresponding value from the analytical assay used.

Deviations, missing concentrations and anomalous values

In summary tables and plots of median profiles, concentrations will be set to missing if one of the following cases is true:

- A concentration has been reported as ND (ie, not done) or NS (ie, no sample);
- A deviation in sampling time is of sufficient concern or a concentration has been flagged as anomalous by the clinical pharmacologist.

Summary statistics will not be presented at a particular timepoint if more than 50% of the data are missing. For analysis of pharmacokinetic concentrations, no values will be imputed for missing data. If less than 3 evaluable concentrations or PK parameters at a given timepoint are available, only minimum and maximum will be presented.

Actual PK sampling times will be used in the derivation of PK parameters. If a PK parameter cannot be derived from the concentration data, the parameter will be coded as NC (ie, not calculated). NC values will not be generated beyond the day that a participant discontinues.

In summary tables of concentration-time profiles or PK parameters, statistics will be calculated by setting NC values to missing; and statistics will not be presented for a particular treatment if more than 50% of the data are not collected, not calculated, or below LLOQ. For statistical analyses (ie, analysis of variance), PK parameters coded as NC will also be set to missing.

If an individual participant has a known biased estimate of a PK parameter (due for example to a deviation for the assigned dose level), this will be footnoted in summary tables and will not be included in the calculation of summary statistics or statistical analyses.

6. ANALYSES AND SUMMARIES

FAS analysis set is the primary population for evaluating all efficacy endpoints as well as baseline characteristics. Safety analysis set is the primary population for evaluating treatment administration/compliance and safety.

6.1. Primary Endpoint – PFS

6.1.1. PFS

The primary endpoint is PFS which is defined as the time from the date of randomization to the date of the first documentation of objective PD assessed by BICR per RECIST v1.1, or death due to any cause, whichever occurs first.

6.1.1.1. Primary Analysis

Estimand strategy: *Hypothetical strategy* (Section 2.1.1.). The estimand is defined according to the primary objective and is in alignment with the primary endpoint PFS. The primary analysis will be based on BICR-assessed PFS in FAS analysis set for 1) all participants, and 2) participants with ESR1 mutation..

- **Population-level summary**: A log-rank (stratified by ESR1 mutaion status, Visceral Disease status) test will be used to compare PFS between the two treatment arms. HR for PFS with the corresponding 2-sided 95% CI will be calculated based on Cox's Proportional Hazard model. PFS time associated with each treatment arm will be summarized using the Kaplan-Meier method and median PFS time with the corresponding 2-sided 95% CI will be reported (Section 5.2.1.)
    - The PFS rates summarized using Kaplan-Meier method and corresponding 2-sided 95% CIs will be reported for each treatment arm at 3-month, 6-month, 9-month, and 12-month
    - Frequency (number and percentage) of participants with each PFS event type and censoring reasons will be presented by treatment arm. The PFS time or censoring time and the reasons for censoring will also be presented in a participator listing
    - Duration of follow-up for PFS in the treatment arms will be summarized using the reverse Kaplan-Meier method including the median time of follow-up for PFS and associated 2-sided 95% CIs

6.1.1.2. Sensitivity Analysis

The following sensitivity analyses will be performed to explore the robustness of the primary analysis results with the same *hypothetical* estimand strategy (Section 2.1.1.). All sensitivity analyses will be based on BICR-assessed PFS (except investigator assessed PFS analysis) in FAS analysis set.

- **Unstratified analysis** (PFS Sensitivity Analysis 1) – An unstratified log-rank test will be used to compare PFS between the two treatment arms. HR for PFS with the corresponding 2-sided 95% CI will be calculated based on unstratified Cox's Proportional Hazard model. PFS time associated with each treatment arm will be summarized using the Kaplan-Meier method and median PFS time with the corresponding 2-sided 95% CI will be reported (Section 5.2.1.)
- **Symptomatic deterioration is considered as an event for PFS** (PFS Sensitivity Analysis 2) – A log-rank (stratified) test will be used to compare PFS between the two treatment arms. HR for PFS with the corresponding 2-sided 95% CI will be calculated based on Cox's Proportional Hazard model. PFS time associated with each treatment arm will be summarized using the Kaplan-Meier method and median PFS time with the corresponding 2-sided 95% CI will be reported (Section 5.2.1.)
    - In addition to objective progression of disease (PD**)** or death due to any cause in the absence of documented PD, symptomatic deterioration is also considered as a PFS event. Symptomatic deterioration is defined as discontinuation of <u>any</u> study drug for reason of "global deterioration of health" or for reason of "progressive disease". The same censoring rules for the primary analysis apply to non-event data.
    - The PFS rates summarized using Kaplan-Meier method and corresponding 2-sided 95% CIs will be reported for each treatment arm at 3-month, 6-month, 9-month, and 12-month
    - Frequency (number and percentage) of participants with each PFS event type and censoring reasons will be presented by treatment arm. The PFS time or censoring time and the reasons for censoring will also be presented in a participator listing
- **Analysis of PFS determined by investigator** (PFS Sensitivity Analysis 3) – In this analysis, PD will be programmatically derived from local radiologist's tumor assessment (the target lesion measurements, non-target lesion status, and new lesions status recorded on the CRF) per RECIST v1.1. A log-rank (stratified) test will be used to compare PFS between the two treatment arms. HR for PFS with the corresponding 2-sided 95% CI will be calculated based on Cox's Proportional Hazard model. The PFS time associated with each treatment arm will be summarized using the Kaplan-Meier method and median PFS time with the corresponding 2-sided 95% CI will be reported (Section 5.2.1.)
    - In addition to the censoring rules defined for the primary analysis, the PFS data will also be censored on the date of the last adequate tumor assessment for participants due to the intercurrent event of "progression by BICR" but not by investigator
    - The PFS rates summarized using Kaplan-Meier method and corresponding 2-sided 95% CIs will be reported for each treatment arm at 3-month, 6-month, 9-month, and 12-month

- Frequency (number and percentage) of participants with each PFS event type and censoring reasons will be presented by treatment arm. The PFS time or censoring time and the reasons for censoring will also be presented in a participator listing.

  Discordance between BICR and Investigator assessments will be evaluated. The EDR and LDR will be calculated for each treatment arm and the differential discordance around each measure will be summarized as the rate on the experimental arm minus the rate on the control arm (Section 5.2.6.)

- **Analysis of PFS using CRF data for stratification** (PFS Sensitivity Analysis 4) – To evaluate potential impact of mis-stratification of the two stratification factors in IRT, the value from CRF (instead of IRT) will be used for the two stratification factors in this analysis. The censoring rules will follow the primary PFS analysis.

  HR for PFS with the corresponding 2-sided 95% CI will be calculated based on stratified Cox's Proportional Hazard model. PFS time associated with each treatment arm will be summarized using the Kaplan-Meier method and median PFS time with the corresponding 2-sided 95% CI will be reported (Section 5.2.1.)

- **Analysis of PFS in participants with ESR1 mutation negative** (PFS Sensitivity Analysis 5) – This sensitivity analysis of PFS will be performed among the participants whose ESR1 mutation status was negative and exclude participants whose ESR1 mutation status was unknown (biomarker test results were missing or not done). The censoring rules will follow the primary PFS analysis.

  HR for PFS with the corresponding 2-sided 95% CI will be calculated based on stratified Cox's Proportional Hazard model. PFS time associated with each treatment arm will be summarized using the Kaplan-Meier method and median PFS time with the corresponding 2-sided 95% CI will be reported (Section 5.2.1.)

6.1.1.3. Supplementary Analyses

The following supplementary analyses will be performed to provide additional insights into the understanding of the primary analysis results with the different estimand strategies. All supplementary analyses will be based on BICR-assessed PFS in FAS analysis set.

- **Multivariate analysis (covariate-adjusted)** (PFS Supplementary Analysis 1) – HR for PFS with the corresponding 2-sided 95% CI will be calculated based on Cox's Proportional Hazard model with stratification factors (based on CRF) and baseline characteristics as covariates (baseline characteristics described in section 6.4.1.1).
- **No censoring at start of anti-cancer therapies (*Treatment policy* estimand)** (PFS Supplementary Analysis 2) – A log-rank (stratified) test will be used to compare PFS between the two treatment arms. HR for PFS with the corresponding 2-sided 95% CI will be calculated based on Cox's Proportional Hazard model. PFS time associated

with each treatment arm will be summarized using the Kaplan-Meier method and median PFS time with the corresponding 2-sided 95% CI will be reported (Section 5.2.1.)

- PFS data will not be censored due to the start of new anti-cancer therapies and the event time is the actual time of assessment or death
- The PFS rates summarized using Kaplan-Meier method and corresponding 2-sided 95% CIs will be reported for each treatment arm at 3-month, 6-month, 9-month, and 12-month
- Frequency (number and percentage) of participants with each PFS event type and censoring reasons will be presented by treatment arm. The PFS time or censoring time and the reasons for censoring will also be presented in a participator listing.

- **Administration of new anti-cancer therapies is treated as part of event (*Composite* estimand)** (PFS Supplementary Analysis 3) – A log-rank (stratified) test will be used to compare PFS between the two treatment arms. HR for PFS with the corresponding 2-sided 95% CI will be calculated based on Cox's Proportional Hazard model. PFS time associated with each treatment arm will be summarized using the Kaplan-Meier method and median PFS time with the corresponding 2-sided 95% CI will be reported (Section 5.2.1.)

  - Impute as event at the date of start of new anti-cancer therapies
  - The PFS rates summarized using Kaplan-Meier method and corresponding 2-sided 95% CIs will be reported for each treatment arm at 3-month, 6-month, 9-month, and 12-month
  - Frequency (number and percentage) of participants with each PFS event type and censoring reasons will be presented by treatment arm. The PFS time or censoring time and the reasons for censoring will also be presented in a participator listing.

- **Include all events after 2 or more consecutive missing or inadequate assessments (*Treatment policy* estimand)** (PFS Supplementary Analysis 4) – A log-rank (stratified) test will be used to compare PFS between the two treatment arms. HR for PFS with the corresponding 2-sided 95% CI will be calculated based on Cox's Proportional Hazard model. PFS time associated with each treatment arm will be summarized using the Kaplan-Meier method and median PFS time with the corresponding 2-sided 95% CI will be reported (Section 5.2.1.)

  - PFS data will not be censored due to 2 or more missing or inadequate assessments and the event time is the actual time of assessment
  - The PFS rates summarized using Kaplan-Meier method and corresponding 2-sided 95% CIs will be reported for each treatment arm at 3-month, 6-month, 9-month, and 12-month
    Frequency (number and percentage) of participants with each PFS event type and censoring reasons will be presented by treatment arm. The PFS time or censoring time and the reasons for censoring will also be presented in a participator listing.

- **Time to Treatment Failure (TTF)** (Supplemental Analysis 5) – TTF is defined as the time from randomization to discontinuation of study treatment for any reason or

censoring, including but not limited to disease progression, treatment toxicity, or death. Participants who are ongoing on treatment will be censored at the analysis cutoff date.

- **Include all events regardless of violations, discontinuation of study drug or change of therapy (*Treatment policy* estimand)** (PFS Supplementary Analysis 6) – A log-rank (stratified) test will be used to compare PFS between the two treatment arms. HR for PFS with the corresponding 2-sided 95% CI will be calculated based on Cox's Proportional Hazard model. PFS time associated with each treatment arm will be summarized using the Kaplan-Meier method and median PFS time with the corresponding 2-sided 95% CI will be reported (Section 5.2.1.)
    - PFS data will not be censored due to study treatment discontinuation, 2 or more missing or inadequate assessments, receiving new anti-cancer therapy.
    - The PFS rates summarized using Kaplan-Meier method and corresponding 2-sided 95% CIs will be reported for each treatment arm at 3-month, 6-month, 9-month, and 12-month
      Frequency (number and percentage) of participants with each PFS event type and censoring reasons will be presented by treatment arm. The PFS time or censoring time and the reasons for censoring will also be presented in a participator listing

6.2. Secondary Endpoint(s)

6.2.1. Key Secondary Endpoint - OS

6.2.1.1. Main Analysis

Estimand strategy: *Treatment Policy Strategy* (Section 2.1.2.). The estimand is defined according to the secondary objective for OS and is in alignment with the secondary endpoint OS. The main analysis will be based on OS in FAS analysis set <u>for 1) all participants, and 2) participants with ESR1 mutation</u>

- **Population-level summary**: A log-rank (stratified by ESR1 mutaion status, Visceral Disease status) test will be used to compare OS between the two treatment arms. HR for OS with the corresponding 2-sided 95% CI will be calculated based on Cox's Proportional Hazard model. In order to account for the group sequential design for OS, the repeated CI (RCI) method will also be used to construct the 2-sided RCIs for the HR. OS time associated with each treatment arm will be summarized using the Kaplan-Meier method and median OS time with the corresponding 2-sided 95% CI, if estimable, will be reported (Section 5.2.1.)
    - The interim and final analyses on OS will follow section 5.1.3 and section 7 for the details of timing and alpha spending control.
    - The Lan-DeMets/O'Brien-Flemming efficacy boundary will be used for the interim and final analyses.

- The OS rates summarized using Kaplan-Meier method and corresponding 2-sided 95% CIs will be reported for each treatment arm at 6-month, 12-month, 18-month, and 24-month
- Frequency (number and percentage) of participants with each OS event type and censoring reasons will be presented by treatment arm. The OS time or censoring time and the reasons for censoring will also be presented in a participator listing
- Duration of follow-up for OS in the treatment arms will be summarized using the reverse Kaplan-Meier method including the median time of follow-up for OS and associated 2-sided 95% CIs

6.2.1.2. Sensitivity Analysis

The following sensitivity analysis will be performed to explore the robustness of the main analysis results with the same *treatment policy* estimand strategy (Section 2.1.2.). The sensitivity analysis will be based on OS in FAS analysis set.

Unstratified analysis (OS Sensitivity Analysis 1) – An unstratified log-rank test will be used to compare OS between the two treatment arms. HR for OS with the corresponding 2-sided 95% CI will be calculated based on unstratified Cox's Proportional Hazard model. PFS time associated with each treatment arm will be summarized using the Kaplan-Meier method and median OS time with the corresponding 2-sided 95% CI will be reported (Section 5.2.1.)

6.2.1.3. Supplementary Analysis

The following supplementary analyses will be performed to provide additional insights into the understanding of the main analysis results with the different estimand strategies. All supplementary analyses will be based on OS in FAS analysis set.

- **Multivariate analysis (covariate-adjusted)** (OS Supplementary Analysis 1) – HR for OS with the corresponding 2-sided 95% CI will be calculated based on Cox's Proportional Hazard model with stratification factors (based on CRF) and baseline characteristics as covariates (baseline characteristics described in section 6.4.1.1).
- **Had crossover to SERD inhibitor treatment (treatment switch) not occurred** (OS Supplementary Analysis 2)
  - **Estimand strategy**: *Treatment Policy* (Section 2.1.2.2.) and *Hypothetical*
  - **Intercurrent events**: All data will be used regardless the occurrence of intercurrent events (*discontinuation of study treatment*, *use of subsequent treatments*, *discontinuation of study*, *etc*.) except for the use of SERD inhibitor(s) in the control arm as a subsequent treatment. *Had treatment switch to SERD treatment not occurred*
  - **Population-level summary**: HR for OS based on Cox's Proportional Hazard model with counterfactual survival times for crossover participants using RPSFT method. The corresponding 2-sided 95% CI will be estimated based on bootstrap samples. OS time associated with each treatment arm will be summarized using

the Kaplan-Meier method and median OS time with the corresponding 2-sided 95% CI, if estimable, will be reported (Section 5.2.1.)

The OS rates summarized using Kaplan-Meier method and corresponding 2-sided 95% CIs will be reported for each treatment arm at 6-month, 12-month, 18-month, and 24-month.

6.2.2. Other Secondary Endpoints

6.2.2.1. ORR /DOR /CBR

ORR

The best overall response is the best response recorded from randomization until disease progression, or death due to any cause. If any new anti-cancer therapies is taken during the study, only tumor assessments performed on or before the start date of the new anti-cancer therapy will be considered in the assessment of best overall response. Confirmation of response is required ≥4 weeks after initial response is observed.

The best overall response on confirmed responses:

- Complete response (CR): At least two objective statuses of CR ≥4 weeks apart before first documentation of PD and/or start of new anticancer therapy. A CR without confirmation is considered as an unconfirmed CR.

- Partial response (PR): At least two objective statuses of PR or better (PR followed by PR or PR followed by CR) performed ≥4 weeks apart before first documentation of PD (and not qualifying for a CR) and/or start of new anticancer therapy. Two PRs separated by one or more SD or indeterminate assessments can be considered a confirmed PRs if the two PR responses are observed ≥4 weeks apart. A PR without confirmation is considered as an unconfirmed PR

- Stable disease (SD) (applicable only to participants with measurable disease at baseline): At least one objective statuses of SD (or better) documented after randomization and before first documentation of PD (and not qualifying for CR or PR) and/or start of new anticancer therapy.

- Non-CR/non-PD (applicable only to participants with non-measurable disease at baseline): at least one non-CR/non-PD assessment (or better) documented after randomization and before first documentation of PD (and not qualifying for CR or PR) and have not received new anticancer therapy.

- Progressive disease (PD): Documentation of objective status of PD prior to the second planned assessment after randomization (and not qualifying for CR, PR, SD, or Non-CR/non-PD).

- Not Evaluable (NE): All other cases that do not apply in above categories.

An objective status of PR or SD cannot be after a CR. SD can follow PR only in the rare case, e.g., per RECIST v.1.1 that tumor increases by less than 20% from the nadir, but enough that a previously documented 30% decrease from baseline no longer holds. If this occurs, the sequence PR-SD-PR is considered a confirmed PR. A sequence of PR-SD-SD-PD would be a best response of SD if the window for SD definition has been met.

Participants will be considered non-responders until proven otherwise. Participants who:

- Do not have a confirmed CR or PR while on study
- Do not have a post-baseline radiographic tumor assessment
- Receive anticancer therapies other than the study treatments prior to reaching a confirmed CR or PR

will be counted as non-responders in the assessment of ORR.

In addition, the number and percentage of participants with a best overall response of CR, PR, SD, Non-CR/Non-PD, PD, and NE will be tabulated.

The number and proportion of participants achieving confirmed objective response (CR or PR) will be summarized in the FAS analysis set and the FAS analysis set with measurable disease at baseline along with the corresponding 2-sided 95% confidence interval calculated using the Wilson score method. The odds ratio will be estimated using Mantel-Haenszel method, and the corresponding 2-sided Exact 95% confidence interval will be calculated and p-value will be based on Cochran-Mantel-Haenszel test. OR analyses will be performed on both investigator assessments and BICR data.

As a sensitivity analysis, the number and proportion of participants achieving objective response (unconfirmed or confirmed) will be summarized in the FAS analysis set and the FAS analysis set with measurable disease at baseline along with the same statistics described above on both investigator assessments and BICR data.

DOR

DOR will be summarized using the Kaplan-Meier methods and displayed graphically where appropriate. DOR will be calculated for the subgroup of participants with confirmed objective tumor response. The median event time and 2-sided 95% confidence interval for the median will be provided. DOR analyses will be performed on both investigator assessments and BICR data.

DOR (in months) = (date of PD/Death – date of first documentation of

objective tumor response +1)/30.4375

DOR data will also be censored on the date of the last tumor assessment on study for participants who do not have objective tumor progression, and who do not die due to any cause while on study.

DOR data will be censored on the date of the last tumor assessment on study for participants who discontinue from the study treatment due to withdrawal of consent for all follow-up or for clinical follow-up only prior to an event. Additionally, participants who start a new anti-cancer therapy prior to an event will be censored at the date of the last tumor assessment prior to the start of the new therapy. Participants with documentation of PD, or death after an unacceptably long interval (i.e., 2 or more missed, incomplete or non-evaluable assessments) since the last tumor assessment will be censored at the time of last objective assessment that did not show PD.

If the number of participants with OR is small, the Kaplan-Meier method may not provide reliable estimates. In this case, only descriptive statistics or listings will be provided.

CBR

The number and proportion of participants achieving clinical benefit response (confirmed CR or PR at any time, or Non-CR/Non-PD (for participants with non-measurable disease) ≥24 weeks, or SD (for participants with measurable disease) ≥24 weeks will be summarized in the FAS population and the FAS population with measurable disease at baseline along with the corresponding 2-sided 95% confidence interval calculated using the Wilson score method. The odds ratio and the corresponding 2-sided 95% confidence interval using the Wilson score method will be calculated to contrast the treatment effects on response rates. A 1-sided unstratified exact test will also be used to compare CBR rate between the treatment arms. CBR analyses will be performed on both investigator assessments and BICR data.

As a sensitivity analysis, the number and proportion of participants achieving clinical benefit response (unconfirmed or confirmed CR or PR, SD ≥24 weeks, or Non-CR/Non-PD ≥24 weeks) will be summarized in the FAS analysis set and the FAS analysis set with measurable disease at baseline along with the same statistics described above on both investigator assessments and BICR data.

6.2.2.2. Pharmacokinetic Analysis

Plasma concentration data of ARV-471 and ARV-473 will be listed, summarized, and plotted for all participants in the PK concentration set. For summary statistics and mean/median plots by collection time/day, the nominal PK sample collection time/day will be used; for individual participant listings by time/day, the actual PK collection time/day will be used. PK sampling deviations from the protocol-specified collection time/day will be listed.

Plasma concentrations of ARV-471 and ARV-473 will be summarized descriptively (e.g., n, mean, standard deviation, coefficient of variation, median, minimum, maximum, geometric mean, and its associated coefficient of variation) by nominal collection time/day.

Median, mean (+/- SD), and individual plasma trough ARV-471 and ARV-473 concentrations vs. visit day will be plotted over treatment duration.

Steady state trough concentrations ($C_{trough,ss}$) for ARV-471 and ARV-473 will be summarized descriptively by collection day (e.g., n, arithmetic mean, standard deviation, median, minimum, maximum, geometric mean, and its associated coefficient of variation). $C_{trough, ss}$ will also be summarized descriptively for each individual (e.g. geometric mean and its associated coefficient of variation). $C_{trough, ss}$ will include participants in the PK Steady State Trough Concentration ($C_{trough, ss}$) Set who have pre-dose plasma concentrations collected between 22 and 26 hours post most recent dose and prior or equal to the in-clinic dosing time following at least 7 consecutive days of the 200 mg ARV-471 without dose interruption.

PK data from participants in the QTc Substudy Analysis Set will be analyzed separately according to the method described above for plasma concentration of ARV-471 and ARV-473 descriptively summarized by nominal collection time/day.

Population pharmacokinetic analysis of ARV-471 and ARV-473 (if data permit) in this study will be performed in accordance with the FDA guidance on Population Pharmacokinetics. The PK concentration data set from this study may be pooled with datasets from other ARV-471 clinical studies. In addition, population PK/PD modeling will be attempted to investigate any causal relationship between ARV-471 and/or ARV-473 exposure and biomarker, safety, efficacy, and/or laboratory data. These modeling analyses will be reported separately from the final Clinical Study Report.

6.2.2.3. Electrocardiogram Analysis for QTc Sub-study

The sub-study is designed to characterize the effect of ARV-471 on QTc and includes subjects from the QTc Substudy Analysis Set.

Evaluable subjects for the QTc Substudy Analysis Set are defined as those who:

- Receive at least 7 consecutive daily doses of ARV-471 200 mg prior to each post-baseline ECG measurement.

Centrally-read triplicate ECG measurements will be used for the data analysis and interpretation. Corrected QT interval by Fridericia's (QTcF) method will be used for the primary analysis. QTcF will be derived by programming using Fridericia's formula ($QTcF = QT / RR^{1/3}$). The triplicate data will be averaged, and all summary statistics and data presentations will use the triplicate averaged data. Any data obtained from ECGs repeated for safety reasons after the nominal time-points will not be averaged along with the preceding triplicates.

ECG data on Day 1 of Cycle 2 and Cycle 3 from baseline will be summarized using descriptive statistics and categorical analysis by each time point.

Central tendency of the effect of ARV-471 on the QTcF will be summarized by the mean (90% CI) change from baseline at each timepoint. In addition, absolute, baseline, and mean change from baseline values of QTcF interval, RR interval, PR interval, QT interval, QRS duration, and HR, will be summarized descriptively (e.g., n, mean, standard deviation, 90% CI, median, and range) by collection time/day.

Average change from baseline of QTcF will be plotted by collection time/day.

Categorical analysis of the ECG data will be conducted and summarized as number and percentage of subjects with the following measurements:

1. QTcF and QT interval maximum increase from baseline (≤30, >30 to ≤60, and >60 msec).

2. Maximum post-dose QTcF (≤450, 451-≤480, 481- ≤500, and ≥501 msec).

3. PR interval ≥ 220 msec.

4. QRS duration ≥ 120 msec.

5. HR - ≤ 50 bpm, ≥ 100 bpm

6. HR increase from baseline ≥ 20 bpm; HR decrease from baseline ≥ 20 bpm.

6.2.2.4. Pharmacokinetic/Pharmacodynamic Analysis of QTc

Concentration data of ARV-471 and ARV-473 will be listed by patient and by actual collection time and day.

Concentration-QTc analysis will be conducted using the PK and ECG data from the QTc Substudy Analysis set and the Safety Analysis Set. Linear, log-linear, and/or saturable models will be examined for the concentration-QTc relationship. Exploratory analyses (via graphical displays and/or model fitting) include accounting for a delayed effect and the justification for the choice of pharmacodynamic model. Diagnostic evaluation will be included to explore the adequacy of the model. The modeling analyses may be reported separately from the clinical study report.

6.2.2.5. PRO Data Analysis

EORTC QLQ-C30

This questionnaire contains 30 questions organized into 5 multi-item functional scales, 3 multi-item symptom scales, 6 single item symptom scales, and one global quality of life scale. For each of the 15 scales (functional, symptoms, and global quality of life), the results will be summarized using means (and standard deviation), 95% confidence interval, and medians (and range) for each treatment group at each time point. This will be done based on the observed values as well as changes from baseline where both within group and between

group differences will be displayed. For each of the 15 scales, a graphical display of means over time as well as mean changes from baseline over time will also be provided.

EORTC QLQ-BR23

This questionnaire contains 23 questions organized into 4 functional scales and 4 symptom scales. As with C30, the analysis of the BR23 scales for women participants only will consist of descriptive statistics on means and changes from baseline. Also, as with C30, graphical displays of means and changes from baseline over time will also be provided for each BR23 scale. BR23 questionnaire results for male participants will be listed.

EQ-5D Health Index

Analysis of the EQ-5D health index will consist of descriptive statistics on means and changes from baseline and graphical displays of means and changes from baseline over time, as with the EORTC scales. In addition, there will be a health status profile analysis consisting of a display of the number and percentage of participants in each of the 5 response levels for each of the 5 dimensions at each visit.

EQ-5D General Health Status (EQ-5D VAS)

Analysis of EQ-5D general health status will consist of descriptive statistics on means and changes from baseline, and graphical displays of means and changes from baseline over time, as with the EORTC scales.

BPI

Pain severity will be evaluated by rating pain on a scale from 0 (no pain) to 10 (pain as bad as you can imagine) for the ‘worst’, ‘least’, and ‘average’ pain experienced in the past 24 hours as well as the pain experienced ‘now’. Pain severity score will be generated by calculating the mean of the four items relative to pain severity.

Pain interference with daily functioning will be evaluated by rating interference on a scale from 0 (does not interfere) to 10 (completely interferes) for ‘general activity’, ‘mood’, ‘walking ability’, ‘normal work’, ‘relation with other people’, ‘sleep’, and ‘enjoyment of life’. The pain interference score will be generated by calculating the mean of the seven items relative to pain inference.

The score for both scales range from 0 to 10, where a higher score is associated to a worse level of pain. Analysis of the pain severity and pain interference will consist of descriptive statistics on means and changes from baseline and graphical displays of means and changes from baseline over time, as with the EORTC scales.

In the CSR, the following analyses will be conducted and reported.

Descriptive statistics (n, mean, SD, median, minimum, maximum) of absolute scores over time as well as change from baseline will be reported for each of the total and subscales of

EORTC QLQ-C30, EORTC QLQ-BR23 (for women participants only), EQ-5D-5L and BPI-SF. Scores of EORTC QLQ-BR23 for male participants will be listed.

Longitudinal analysis of change from baseline, time to definitive deterioration and other analyses will be described in a separate SAP specific to the PRO endpoints. The analyses results will be presented in the PRO report, separate from the CSR.

6.2.2.6. Biomarker and ctDNA Analysis

For baseline continuous biomarker data, descriptive statistics, including the mean, standard deviation, median, minimum, and maximum values, will be provided by treatment arm.

For baseline categorical biomarker data, the number and percentage of participants in each category will be provided by treatment arm.

The quantitative changes of ctDNA level C2D1 from baseline and their association with predictability of efficacy will be reported but may not be included in CSR pending on availability of biomarker data at the final analysis.

Appropriate statistical methods may be used to investigate any possible relationship of biomarker levels with vepdegestrant anti-tumor efficacy.

6.3. Subset Analyses

Subset analyses will be performed for primary and key secondary endpoints (i.e., PFS and OS) using univariate Cox regression analysis to calculate HR based on the FAS analysis set for the subgroups defined below to evaluate the consistency of the treatment effect.

- Randomization stratification factors
  - ESR1 mutation status (yes vs. no)
  - Visceral disease (yes vs. no)
- Baseline demographic and disease characteristics
  - Age at randomization (<65 vs. ≥65)
  - Gender (Male vs. Female)
  - Race (White, Black, Asian, Other)
  - Ethnicity (Hispanic or Latino or of Spanish origin vs. Not)
  - Country /Region (North America, Europe, Asia, Rest of the world)
  - Hormone receptor status (ER≥10%, ER<10%)

- HER2 status (HER2+, HER2-)
- ECOG performance status (0 vs. 1)
- Measurable disease by RECIST criteria (Yes vs. No)
- Bone only (Yes vs. No)
- Liver metastases (Yes or No)
- Menopausal status (Pre/perimenopausal vs. Postmenopausal)
- Duration of prior CDK4/6i treatment (≥12 months vs. <12 months)
- Line of prior therapy (1 prior or 2 prior lines)

- Other baseline disease characteristics (if appropriate)
  - Hormone receptor status
    - PGR+ (≥1%) vs PGR- (<1%)
    - ER ≥10% and PGR+
    - ER ≥ 10% and PGR-
  - Most recent TNM stage (IIB, IIIA, IIIB, etc.)
  - Histopathological Grade (1, 2, 3, X)
  - Histopathological Type
  - Prior treatments (surgery, radiation, chemotherapy, hormonal therapy, anti-HER2 therapy)
  - Other biomarker characteristics (PIK3A, AKT, PTEN, etc.)

The subset analyses for PFS and OS will use the same censoring rules as for the corresponding primary analysis. A 2-sided unstratified log-rank test will be used to compare treatment arms for each subgroup level. The unstratified HR and corresponding 2-sided 95% CI and median time with the corresponding 2-sided 95% CI will be provided per subgroup level.

All the subset analyses are exploratory. No adjustment for multiplicity will be performed. In the case of a low number of participants within a category (<5% of the randomized population), the categories will be pooled, if appropriate.

To assess the heterogeneity of treatment effects across the subgroup levels, a Cox regression model will be fitted with the time-to-event endpoints as the dependent variable and subgroup, treatment group, and treatment-by-subgroup interaction as explanatory variables. A p-value for the interaction test (Likelihood Ratio test) with the treatment-by-subgroup interaction will be provided as appropriate. The HR and corresponding 2-sided 95% CIs for all subgroups will be presented in a forest plot.

6.4. Baseline and Other Summaries and Analyses

6.4.1. Baseline Summaries

Baseline summaries will include (but not limited to) the following analyses based on the FAS analysis set by treatment arm.

6.4.1.1. Baseline Characteristics

Demographic and disease characteristics will be summarized by treatment arm using the following information from the 'Demographics', 'Disease Character", 'Eligibility', and 'Medical History' eCRF pages.

Demographic and baseline characteristics

- Gender: Male or Female
- Race: White, Black (Black or African American), Asian, Other (American Indian or Alaska Native, Native Hawaiian, or Other Pacific Islander), or Not Reported/Unknown.
- Hispanic / Latino, Not Hispanic / Latino, or Not Reported
- Country of registration / Region
- Age (years): summary statistics
- Age categories: < 65 years or ≥65 years
- Menopausal status: Pre/perimenopausal, man, or Postmenopausal
- ECOG Performance Status: 0 or 1
- Hormone receptor status (ER ≥ 10%, ER < 10%)
- HER2 status (HER2+, HER2-, unknown, equivocal, unknown)
- ECOG performance status (0 vs. 1)
- Measurable disease by RECIST criteria (Yes vs. No)
- Bone only (Yes vs. No)
- Liver metastases (Yes or No)
- Menopausal status (Pre/perimenopausal vs. Postmenopausal)

- Duration of prior CDK4/6i treatment (≥ 12 months vs. < 12 months)
- Line of prior therapy (1 prior or 2 prior lines)
- Prior treatments (surgery, radiation, chemotherapy, hormonal therapy, anti-HER2 therapy)

Randomization is stratified by the following factors, as reported in IRT system where randomization is performed and on the eCRF page:

- ESR1 mutation status (yes vs. no)
- Visceral disease (yes vs. no)

Stratification factors will be summarized by treatment arms, as well as the concordance between the data recorded in the IRT at randomization and CRF by treatment arms.

6.4.1.2. Medical History and Prior Treatments

Medical history, cancer history, hormone receptor and HER2 status, and prior treatment will be summarized by treatment arm using the information from corresponding eCRF pages.

6.4.2. Study Conduct and Participant Disposition

Summaries will be based on the FAS analysis set by treatment arm. An accounting of the study participants will be tabulated including screened, screened but not randomized, randomized, randomized but not treated, treated (including discontinued, ongoing at cutoff date), assessed for safety. Participants not meeting the eligibility criteria will be identified. Participants not completing the study will be listed along with the reason for their early discontinuation. Reasons for premature discontinuation will be summarized. Randomization errors and stratification errors will be described, if any.

6.4.3. Protocol Deviations

Important protocol deviations will be described when they appear and relate to the statistical analyses or analysis populations. Deviations will be summarized separately from the database and may include, but are not limited to, eligibility violations, deviations from the conduct of the study after start of study treatment, or cases where study evaluations were not performed as required by the protocol.

6.4.4. Study Treatment Exposure

The following analyses will be based on the safety analysis set by treatment arm.

6.4.4.1. Extent of Exposure

Vepdegestrant:

The duration of vepdegestrant exposure is defined as last dose date – first dose date + 1 day, regardless of unplanned intermittent discontinuations.

The cumulative dose (mg) is defined as the sum of all doses taken from Cycle 1, Day 1 to the last dose in the last cycle, where the last cycle is based on the investigator's report in the CRF.

Relative dose intensity (RDI) is defined as the actual total dose intensity within the entire treatment duration by a participant relative to planned/intended dose intensity based on the planned schedule of the treatment duration. RDI will be calculated according to the following algorithm:

- Planned total dose (PTD) is the cumulative planned dose of the drug should have been taken over the entire treatment duration. PTD = 200 mg x (last dose date – first dose date +1)
- Actual total dose (ATD) is the cumulative dose of the drug that has been taken over the treatment duration.
- Relative Dose Intensity (RDI) is the ratio of ATD and PTD, expressed as a percentage.

Fulvestrant:

For fulvestrant, the duration of fulvestrant is defined as last dose date – first dose date + 28 day, regardless of delay or missing cycles.

- The planned total dose (PTD) is calculated as 500 mg times number of planned cycles (including missing cycles) plus the planned C1D15 dose.
- Actual total dose (ATD) is calculated as 500 mg times number of cycles with non-zero dose, plus C1D15 dose (if non-zero).
- Relative Dose Intensity (RDI) is the ratio of ATD and PTD, expressed as a percentage.

6.4.4.2. Treatment Modifications

According to the protocol dose modifications of ARV-471 may occur as dose interruption and dose reduction (from 200 mg to 100 mg daily). Incidence of dose interruptions, and dose reductions of ARV-471 as well as duration of dose interruptions will be reported for Arm A.

No dose reduction for fulvestrant is permitted. Incidence of dose delays of fulvestrant will be reported for arm B. Reasons for dose delays will be reported according to CRF.

A dose reduction for ARV-471 is defined as a decrease of the actual dose from 200 mg to 100 mg daily dose. Once a dose has been reduced for a given participant, all subsequent doses should be administered at that dose level, until further dose modification. The reduced

doses in those subsequent cycles are not defined as dose reductions. Dose re-escalation of ARV-471 from 100 mg to 200 mg is allowed for participants in Arm A.

6.4.5. Concomitant Medications and Nondrug Treatments

The analyses will be based on the safety analysis set by treatment arm.

Concomitant medications are medications, other than study drugs, which started prior to first dose date of study treatment and continued during the on-treatment period as well as those started during the on-treatment period. **Prior medications** are medications, other than study drugs and pre-medications for study drug, which are started before the first dose of study treatment.

Summary of concomitant medications will include the number and percentage of participants by Anatomical Therapeutic Chemical (ATC) Classification level 2 and preferred term. A participant will be counted only once within a given drug class and within a given drug name, even if he/she received the same medication at different times. If any concomitant medication is classified into multiple ATC classes, the medication will be summarized separately under each of these ATC classes. The summary tables will be sorted on decreasing frequency of drug class and decreasing frequency of drug name in a given drug class. In case of equal frequency regarding drug class (respectively drug name), alphabetical order will be used. In case any specific medication does not have ATC classification level 2 coded term, it will be summarized under ‘Unavailable ATC classification’ category.

6.4.6. Subsequent anti-cancer therapies

Anti-cancer drug treatment will be provided in a data listing with data retrieved from ‘Cancer Therapy’ eCRF page with category ‘Follow-up Cancer Therapy’.

Number and percentage of participants with any anti-cancer therapy after discontinuation will be tabulated overall and by type of therapy based on the data collected from the ‘Cancer Therapy’, ‘Radiation Treatment’ and ‘Cancer Surgery’ eCRF pages.

6.5. Safety Summaries and Analyses

The safety analysis set will be the primary population for safety evaluations. Summaries of AEs and other safety parameters during the on-treatment period will be based on the safety analysis set by treatment arm. In this study, safety summary and analyses will be performed in SAS for all participants. In addition, a sub-study is designed to characterize the effect of ARV-471 on QTc and the analyses will be performed in the QTc Sub-study Analysis Set.

6.5.1. Adverse Events

AEs will be classified using the latest version of MedDRA classification system. The severity of the AEs will be graded according to the NCI CTCAE v.5.0. Safety will be assessed on the basis of any grade 1-5 AEs and SAEs.

The active reporting period for all AEs (serious and non-serious) begins from the first dose of study drug and including 28 calendar days following the last dose of study treatment.

Adverse Event of Special Interest (AESIs) are examined as part of routine safety data review procedures throughout the clinical trials and as part of signal detection processes. Should an aggregate analysis indicate that these prespecified events occur more frequently than expected, eg, based on epidemiological data, literature, or other data, then this will be submitted and reported in accordance with Pfizer's safety reporting requirements.

Unless otherwise specified, AEs will be summarized by frequency and percentage of participants with the AE in the category of interest as described above, by treatment arm, primary SOC and PT in decreasing frequency based on the frequencies observed.

Each participant will be counted only once within each SOC or PT. If a participant experiences more than one AE within a SOC or PT for the same summary period, only the AE with the worst NCI CTCAE v.5.0 severity grade, as appropriate, will be included in the summaries for severity.

Detailed information collected for each AE will include a description of the event, severity, whether the AE was serious, relationship to study drug, and action taken.

AEs leading to death or discontinuation of study treatment, events classified as NCI CTCAE v.5.0 Grade 3 or higher, study drug related events, and SAEs will be summarized separately.

In case a participant has events with missing and non-missing grades, the maximum of the non-missing grades will be displayed. No imputation of missing grades will be performed.

The percentage of participants with an event will be calculated using the number of participants in the safety analysis set as the denominator.

The following summaries of TEAEs will also be provided by treatment arm:

- TEAEs (All Causality): Number of participants evaluable for AEs, total number of AEs (counting each unique PT across all participants), number of participants with SAEs, number of participants with Grades 3 and 4 AEs, number of participants with Grade 5 AEs, number of participants who discontinued study treatment due to AEs, and number of participants with dose reductions or dose interruption due to AEs.
- TEAEs by MedDRA SOC, PT and Maximum NCI CTCAE v.5.0 Grade (All Causality, and Treatment Related)

- TEAEs by MedDRA PT (including Clusters if available) and Maximum NCI CTCAE v.5.0 Grade, sorted by Descending Order of AE Frequency (All Causality, and Treatment Related).
- Adverse Events Leading to Dose Modification (All Causality, Treatment Related): AEs leading to dose reduction of study treatment (as recorded on the AE eCRF page, Action taken with study treatment = Dose reduced).
- Adverse Events Leading to Dose Interruption of Study Treatment (All Causality, Treatment Related): AEs leading to interruption of study treatment (as recorded on the AE eCRF page, Action taken with study treatment = Drug interrupted).
- Adverse Events Leading to Permanent Treatment Discontinuation (All Causality, Treatment Related): AEs leading to permanent discontinuation of study treatment (as recorded on the AE eCRF page, Action taken with study treatment = Drug withdrawn).
- Adverse Events Leading to Death: AEs with fatal outcome (as recorded on the AE eCRF page, Outcome = Fatal, as well as AEs of Grade 5).

A summary of SAEs and deaths reported as SAEs will be provided. Hospitalization solely due to signs and symptoms of recurrence or disease progression should not be reported as an SAE. The number and percentage of participants with each of the following will be presented for SAEs by treatment arm:

- SAEs by MedDRA SOC and PT.
- SAEs by MedDRA PT (including Clusters and PT).

6.5.2. Deaths

Summary of death will be presented with numbers and percentages as following:

- Deaths from start of treatment to and including 28 days after last dose of study treatment as well as primary cause of death
- Deaths during follow-up period occurring greater than 28 days after last dose of study treatment as well as primary cause of death
- Cause of death
    - Disease under study
    - Study treatment toxicity
    - Unknown
    - Other

Deaths during the active treatment period of the study and up to 28 days after last dose of study treatment or in the follow-up period will be listed together with the treatment arm, cycle/follow-up number, and primary cause of death.

6.5.3. Laboratory Data

The lab baseline value is defined as the last available value prior to the first dose of study treatment.

Laboratory data including hematology, and chemistry will be summarized for all cycles. The hematologic and chemistry laboratory results will be graded according to the NCI CTCAE v.5.0 severity grade.

The number and percentage of participants who experienced laboratory test abnormalities will be summarized according to worst toxicity grade observed for each laboratory assay. The analyses will summarize laboratory tests during the entire study treatment period. The denominator with respect to percentages for summary tables for each laboratory parameter will be all participants in the safety analysis set with at least one evaluable cycle for that parameter.

For laboratory tests without CTCAE grade definitions, results will be categorized as missing, low, normal, or high.

6.5.4. Summary of ECG Data

All participants in the safety analysis set with baseline and on-treatment ECG data will be included in the ECG summary and analysis. The triplicate data will be averaged, and all summary statistics and data presentations will use the triplicate averaged data. Any data obtained from ECGs repeated for safety reasons after the nominal time-points will not be averaged along with the preceding triplicates.

Central tendency of the effect of ARV-471 on the QTcF will be summarized by the mean (90% CI) change from baseline at each timepoint. In addition, absolute and change from baseline values of QTcF interval, RR interval, PR interval, QT interval, QRS duration, and HR, will be summarized descriptively (e.g., n, mean, standard deviation, 90% CI, median, and range) by collection time/day. Average change from baseline of QTcF will be plotted by collection time/day.

Categorical analysis of the ECG data will be conducted and summarized as number and percentage of subjects with the following measurements:

1. QT interval and QTcF maximum increase from baseline (≤30, >30 to ≤60, and >60 msec).

2. Maximum post-dose QTcF (≤450, 451-≤480, 481- ≤500, and ≥501 msec).

3. PR interval ≥ 220 msec.

4. QRS duration ≥ 120 msec.

5. HR - ≤ 50 bpm, ≥ 100 bpm

6. HR increase from baseline ≥ 20 bpm; HR decrease from baseline ≥ 20 bpm.

The analyses described above will be repeated separately for the QTc Substudy Analysis Set using centrally-read ECG data. See Section 6.2.2.3 for details.

7. INTERIM ANALYSES

No interim analysis is planned for the primary endpoint PFS.

Two interim analyses for efficacy are planned for the key secondary endpoint of OS.

The first OS interim analysis (IA1) will be performed at the time of the final PFS analysis for both the ESR1-mutation positive subgroup and all participants (ITT) populations. This is anticipated to occur approximately 18 months after the first participant is randomized into the study with approximately 49 and 107 deaths in ESR1 mutation positive subgroup and all participants population, respectively. OS will be tested in ESR1 mutation positive subgroup first with the allocated alpha (plus passed alpha from PFS testing per testing procedure if applicable) under group sequential trial design consideration with Lan-DeMets/O'Brien-Fleming boundary (O'Brien & Fleming, 1979). OS will be formally tested in all participants population only if OS testing in ESR1 mutation positive subgroup is positive. Otherwise, an administrative alpha spending (0.0000001) will be spent for OS IA1 analysis among all participants due to unblinded descriptive analyses and the study will proceed to the second OS interim analysis.

The second OS interim analysis (IA2) will be performed at 130 deaths in ESR1 mutation subgroup (approximately 67% of total 194 events required for the final analysis). This is anticipated to occur approximately 38 months after the first participant is randomized into the study. Following the same alpha spending principle as described in IA1, if OS IA2 is positive for ESR1 mutation positive subgroup, OS will be formally tested among all participants population per pre-specified testing procedure with available deaths events at the time (for example, approximately 265 deaths). Otherwise, OS for all participants will not be formally tested at IA2 and the study proceeds to the OS final analysis (396 events).

The significance level for OS family depends on the results from the PFS family. If the PFS tests are positive in both testing populations, the 3/4 α used for PFS testing will be reallocated to OS endpoint family and OS will be tested with full α (0.025). If PFS test is positive in ESR1 mutation positive population but negative in ITT population, the original allocated 1/4 α (0.00625) will be used for OS tests. A Lan-DeMets alpha-spending function that approximates O'Brien-Fleming stopping boundaries will be used to control the overall type I error rate at 0.00625/0.025 level (1-sided).

Table 7 summarized the operation characteristics and efficacy boundaries for the OS analysis in the ESR1-mutation positive population and in the ITT population. The exact efficacy boundaries and the actual alpha spending at the interim analyses will be updated at the time of the analyses based on the actual number of events observed and the information fraction.

Table 7. Summary of Planned Efficacy Boundaries

Population	Analysis	Information Fraction [a]	Cumulative OS Events	Efficacy Boundaries on P-value scale (1-sided)
Scenario 1: If the PFS tests are positive in both ESR1-mut and ITT populations (full α)				
ESR1-mut	IA 1	25%	49	<0.00001
	IA 2	67%	130	0.00598
	Final	100%	194	0.02315
ITT	IA 1	27%	107	0.000016
	IA 2	67%	265	0.006043
	Final	100%	396	0.023136
Scenario 2: If PFS test is positive in ESR1 mutation positive population but negative in ITT population (1/4 α)				
ESR1-mut	IA 1	25%	49	<0.000001
	IA 2	67%	130	0.000824
	Final	100%	194	0.005983
ITT	IA 1	27%	107	<0.000001
	IA 2	67%	265	0.000811
	Final	100%	396	0.005987

The efficacy boundaries are determined using Lan-DeMets alpha spending function that approximates O'Brien-Fleming stopping boundaries.
a. The exact efficacy boundaries and the actual alpha spent on each analysis will be updated based on the actual number of events observed.

8. REFERENCES

1. Brookmeyer R and Crowley J. A confidence interval for the median survival time. Biometrics 38:29-41, 1982

2. Therneau TM, Grambsch PM. Modeling Survival Data: Extending the Cox Model. Statistics for Biology and Health 2000.

3. Jennison C, Turnbull BW. In: Group Sequential Methods with Applications to Clinical Trials. Chapman & Hall/CRC, Boca Raton, 2000

4. Robins JM, and Tsiatis AA. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. Communications in Statistics. - Theory and Methods. 1991; 20:2609-2631

5. Schemper M, Smith TL. A note on quantifying follow-up in studies of failure time. Controlled Clinical Trials, 17: 343-346, 1996

6. Zhang X. Comparison of restricted mean survival times between treatments based on a stratified Cox model. Bio-Algorithms and Med-Systems 2013; 9(4): 183–9

7. McShane LM, Rothmann MD, Fleming TR. Finding the (biomarker-defined) subgroup of patients who benefit from a novel therapy: No time for a game of hide and seek. Clinical Trials, 2023 August; 20(4): 341-350.

8. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. Biometrics. 1979 Sep;35(3):549-56.

9. APPENDICES

Appendix 1. Summary of Efficacy Analyses

Endpoint	Analysis Type	Population	Brief Description	Notes
PFS	Primary Analysis	FAS/ESR1 mut	Primary Analysis defined in section 6.1.1.1	
	Sensitivity Analysis 1	FAS/ESR1 mut	Same as the primary analysis without stratification	
	Sensitivity Analysis 2	FAS/ESR1 mut	Treatment discontinuation due to symptomatic deterioration considered as a PFS event	
	Sensitivity Analysis 3	FAS/ESR1 mut	PD assessed by investigator	
	Sensitivity Analysis 4	FAS/ESR1 mut	Same as the primary analysis using data from CRF instead of IRT for the stratification factors	
	Supplementary Analysis 1	FAS/ESR1 mut	Multivariate Analysis	
	Supplementary Analysis 2	FAS/ESR1 mut	No censoring for new anti-cancer therapy	
	Supplementary Analysis 3	FAS/ESR1 mut	New anti-cancer therapy considered as a PFS event	
	Supplementary Analysis 4	FAS/ESR1 mut	No censoring for the PFS events after 2 or more missing disease assessments	
	Supplementary Analysis 5	FAS/ESR1 mut	Time to Treatment Failure	

	Supplementary Analysis 6	FAS/ESR1 mut	Include all PFS events regardless of censoring rules	
	Supplementary Analysis 7	FAS/ESR1 mut	Using RMST method	Only when PH assumption is violated.
	Subgroup Analysis	FAS/ESR1 mut	Subgroups defined in section 6.3	
OS	Primary Analysis	FAS/ESR1 mut	Primary Analysis defined in section 6.2.1.1	
	Sensitivity Analysis 1	FAS/ESR1 mut	Same as the primary analysis without stratification	
	Supplementary Analysis 1	FAS/ESR1 mut	Multivariate Analysis	
	Supplementary Analysis 2	FAS/ESR1 mut	RPSFT to evaluate crossover from control arm to treatment arm	
	Supplementary Analysis 3	FAS/ESR1 mut	Using RMST method	When PH assumption is violated.
	Subgroup Analysis		Subgroups defined in section 6.3	
ORR (confirmed)		FAS/ESR1 mut with measurable disease at baseline		
		FAS/ESR1 mut		

ORR (unconfirmed)		FAS/ESR1 mut with measurable disease at baseline		
		FAS/ESR1 mut		
DOR		Participants with confirmed PR/CR)		
CBR		FAS/ESR1 mut		
PRO				

Note: Formal analyses will be performed in all participants and in the subgroup of participants with ESR1 mutation.

Appendix 2. List of Abbreviations

Abbreviation	Term
aBC	advanced breast cancer
AE	adverse event
AESI	adverse event of special interest
ALT	alanine aminotransferase
ANC	absolute neutrophil count
AST	aspartate aminotransferase
ATC	Anatomic Therapeutic Chemical
ATD	Actual Total Dose

BICR	blinded independent central review
BLQ	below the limit of quantification
BP	blood pressure
BPI	Brief pain inventory
C1D1	Cycle 1 Day 1
CBR	clinical benefit response
CI	confidence interval
CR	complete response
CRF	case report form
CSR	clinical study report
CT	computed tomography/clinical trial/chemotherapy
CTCAE	Common Terminology Criteria for Adverse Events
ctDNA	circulating tumor DNA
DMC	data monitoring committee
DOR	duration of response
ECG	electrocardiogram
eCRF	electronic case report form
ECOG PS	Eastern Cooperative Oncology Group Performance Status
E-DMC	external data monitoring committee
EOT	end of treatment
ER	estrogen receptor
ESR1	Estrogen Receptor 1 gene
EudraCT	European Union Drug Regulating Authorities Clinical Trials (European Clinical Trials Database)
EuroQol	EuroQol Group comprises an international network of multidisciplinary researchers
FA	Final Analysis
FAS	full analysis set
FDA	Food and Drug Administration (United States)

FWER	Family-wise Error Rate
HER2	human epidermal growth factor receptor 2
HR	hazard ratio/heart rate/hormone receptor
IA	interim analysis
IF	information fraction
IRT	Interactive Response Technology
ITT	Intent to Treat
IV	intravenous
L	liter
LHRH	Luteinizing Hormone-Releasing Hormone
LLOQ	lower limit of quantitation
MedDRA	Medical Dictionary for Regulatory Activities
mg	milligram
MRI	Magnetic resonance imaging
ms	millisecond
N/A	not applicable
NC	not calculated
NCI	National Cancer Institute
ND	not done
NE	not evaluable
NR	not reached
OR	objective response
ORR	objective response rate
OS	overall survival
PD	progressive disease
PE	physical examination
PFS	progression free survival
PK	pharmacokinetic(s)
PR	partial response

PRO	patient-reported outcome
PT	preferred term
PTD	Planned Total Dose
QD	once daily
QLQ-BR23	EORTC Quality of Life Questionnaire Breast 23
QLQ-C30	EORTC Quality of Life Cancer Questionnaire 30
QoL	quality of life
QRS	combination of the Q wave, R wave and S wave
QT	time interval from the start of the Q wave to the end of the T wave, ie, time taken for ventricular depolarization and repolarization
QTc	corrected QT interval
QTcF	QTc calculated using Fridericia method
RCI	repeated CI
RECIST	Response Evaluation Criteria in Solid Tumors
RMST	restricted mean survival time
SAE	serious adverse event
SAP	statistical analysis plan
SD	standard deviation/stable disease
TTF	Time to Treatment Failure