STATISTICAL ANALYSIS PLAN

Protocol title:	**A randomized, multicenter, double-blind Phase 3 study of amcenestrant (SAR439859) plus palbociclib versus letrozole plus palbociclib for the treatment of patients with ER (+), HER2 (-) breast cancer who have not received prior systemic anti-cancer treatment for advanced disease**
Protocol number:	**EFC15935**
Compound number (INN/Trademark):	**SAR439859** **(Amcenestrant/not applicable)**
Study phase:	**Phase 3**
Short title:	**Amcenestrant (SAR439859) plus palbociclib as first line therapy for patients with ER (+) HER2 (-) advanced breast cancer** **(AMEERA-5)**
Statistician:	[redacted]
Statistical project leader:	[redacted]
Date of issue:	**11-May-2021**

Regulatory agency identifier number(s):

IND:	133204
EudraCT:	2020-001824-33
NCT:	NCT04478266
WHO:	U1111-1233-0486

Total number of pages: 74

TABLE OF CONTENTS

LIST OF TABLES

VERSION HISTORY

This statistical analysis plan (SAP) for study EFC15935 is based on the protocol amendment 03 dated 13-Dec-2020. The first participant was randomized on 19-Nov-2020. This SAP is approved before the first interim analysis is conducted.

Table 1 - Major changes in statistical analysis plan

SAP Version	Approval Date	Changes	Rationale
1	11-May-2021	Not Applicable	Original version

1 INTRODUCTION

1.1 STUDY DESIGN

This is a prospective multicenter, international, randomized, double-blind, double-dummy, Phase 3 trial comparing the efficacy and safety of amcenestrant in combination with palbociclib versus letrozole in combination with palbociclib in men, pre/peri-menopausal women (with goserelin), and postmenopausal women, all with ER(+)/HER2(-) breast cancer who have not received prior systemic treatment for advanced disease.

All eligible participants will be randomly assigned using an Interactive Response Technology (IRT) to either amcenestrant plus palbociclib (experimental) arm or letrozole plus palbociclib (control) arm in a 1:1 ratio.

The population will be stratified by:

- De-novo metastatic disease (Yes or No).
- Postmenopausal women (Yes or No).
- Visceral metastasis defined by at least 1 liver, lung, brain metastasis, pleural, or peritoneal involvement (Yes or No).

1066 participants will be randomly assigned to study intervention with a balanced randomization ratio (533 participants randomized per treatment arm).

1.2 OBJECTIVE AND ENDPOINTS

Table 2 - Objectives and endpoints

Objectives	Endpoints
Primary	
• To determine whether amcenestrant in combination with palbociclib improves progression free survival (PFS) when compared with letrozole in combination with palbociclib in participants with ER+, HER2- advanced breast cancer who have not received prior systemic anticancer therapies for advanced disease	• Progression-free survival is defined as the time interval from the date of randomization to the date of first documented tumor progression as per Response Evaluation Criteria in Solid Tumors (RECIST 1.1) assessed by local radiologist/investigator or death (due to any cause), whichever comes first.
Secondary	
Key secondary objective:	**Key secondary endpoint:**
• To compare the overall survival (OS) in both treatment arms	• Overall survival is defined as the time interval from the date of randomization to the date of documented death (due to any cause).
Other Secondary objectives:	**Other Secondary endpoints:**
• To evaluate the objective response rate (ORR) in both treatment arms	• Objective response rate is defined as the proportion of participants who have a complete response (CR) or partial response (PR), as best overall response determined as per RECIST 1.1, from the date of randomization until disease progression, death, cutoff date, initiation of post-treatment anti-cancer therapy, whichever occurs first.

Objectives	Endpoints
• To evaluate the duration of response (DOR) in both treatment arms	• Duration of response is defined as the time from first documented evidence of CR or PR until progressive disease (PD) as determined as per RECIST 1.1 or death from any cause, whichever occurs first.
• To evaluate the clinical benefit rate (CBR) in both treatment arms	• Clinical benefit rate is defined as the proportion of participants who have a confirmed CR, PR, or stable disease (SD) for at least 24 weeks determined as per RECIST 1.1, from the date of randomization until disease progression, death, cutoff date, initiation of post-treatment anti-cancer therapy, whichever occurs first.
• To evaluate progression-free survival on next line of therapy (PFS2)	• The PFS2 is defined as the time from the date of randomization to the date of first documentation of PD on the next systemic anti-cancer therapy according to investigator or death due to any cause in the absence of documented PD on the next systemic anti-cancer therapy, whichever occurs first.
• To evaluate the pharmacokinetics (PK) of amcenestrant, and palbociclib	• Plasma concentrations of amcenestrant, palbociclib
• To evaluate health-related quality of life (HRQL) in both treatment arms	• Symptoms and function related to HRQL as measured by EORTC QLQ-C30, breast cancer specific module (QLQ-BR23/BR45) and EQ-5D-5L Disease-specific and generic HRQL, disease and treatment-related symptoms, the impact of symptoms and treatment, health state utility, and health status will be evaluated using the European Organization for Research and Treatment of Cancer (EORTC) core quality of life questionnaire (QLQ-C30), the EORTC-QLQ breast cancer specific module (QLQ-BR23/BR45) and the EuroQoL questionnaire with 5 dimensions and 5 levels per dimension (EQ-5D-5L), from Cycle 1 Day 1 until 90 days after last dose of study treatment
• To evaluate the time to first chemotherapy in both treatment arms	• Time to chemotherapy is defined as the time interval from the date of randomization to the start date of the first chemotherapy after study treatment discontinuation
• To evaluate safety in both treatment arms	• Adverse events (AEs)/serious adverse events (SAEs) and laboratory abnormalities
Tertiary/exploratory	
• To evaluate in participants tumor biomarkers over time such as estrogen receptor (ER), Ki67, Bcl-2, and progesterone receptor (PgR) protein, and ribonucleic acid (RNA) gene expression profiles (for participants with tumor sites accessible for biopsy).	• Tumor ER, Ki67, Bcl-2, and PgR protein, and RNA gene expression profiles in paired biopsies at Cycle 1 Day 1 (pre-treatment) and optional at end of treatment for participants who discontinued treatment due to disease progression
• To evaluate in participants the gene mutation profile of the tumor over time (baseline and end of treatment) by cell-free deoxyribonucleic acid (cfDNA) analysis.	• The gene mutation profile of the tumor by cfDNA analysis over time (Cycle 1 Day 1 [pre-treatment], at Cycle 4 Day 1 predose (tumor assessment), and upon disease progression on treatment or EOT whichever comes first)
• To evaluate exposure/response relationship of amcenestrant and palbociclib	• PK of amcenestrant will be correlated with safety and/or efficacy
• To evaluate the PK of goserelin	• Plasma concentrations of goserelin

Objectives	Endpoints
• To evaluate PFS in participants expressing PIK3CA and ESR1 mutation in both treatment arms	• Progression-free survival is defined as the time interval from the date of randomization to the date of first documented tumor progression as per Response Evaluation Criteria in Solid Tumors (RECIST 1.1) assessed by local radiologist/investigator or death (due to any cause), whichever comes first.

1.2.1 Estimands

Primary estimands defined for the primary endpoint and key secondary endpoint are summarized in below Table 3. More details are provided in Section 4. For both estimands, the comparison of interest will be the comparison of amcenestrant in combination with palbociclib versus letrozole in combination with palbociclib. The analysis population corresponds to all participants from the intent-to-treat population.

Table 3 - Summary of primary estimand for main endpoints

Endpoint Category (estimand)	Estimands			
	Endpoint	**Population**	**Intercurrent event(s) handling strategy**	**Population-level summary (Analysis and missing data handling)**
Primary objective: to determine whether amcenestrant in combination with palbociclib improves the progression free survival (PFS) when compared to letrozole in combination with palbociclib.				
Primary endpoint (Estimand 1)	PFS	ITT	- Interruption/discontinuation of study intervention: PFS will be analyzed based on events irrespective of study intervention interruption/discontinuation (treatment policy strategy). - Start of new anti-cancer therapy prior to PFS event: PFS will be censored at the last evaluable tumor assessment prior to a new anti-cancer therapy is initiated (hypothetical strategy). - PFS event documented after two or more non-evaluable tumor assessments: PFS will be censored at the last evaluable tumor assessment prior to the event (hypothetical strategy).	One-sided log-rank test stratified by randomization stratification factors, as entered in the IRT system. Hazard ratio (HR) and corresponding 95% CI estimated using stratified Cox proportional hazard model. The Kaplan Meier estimate of PFS at specified time points, PFS quartiles and corresponding 95% CI from Kaplan Meier method.
Primary endpoint Sensitivity analysis #1 (Estimand 1 – SA#1)	PFS	ITT	- Interruption/discontinuation of study intervention: PFS will be analyzed based on events irrespective of study intervention interruption/discontinuation (treatment policy strategy). - Start of new anti-cancer therapy prior to PFS event: PFS will be analyzed based on events irrespective of start of new anti-cancer therapy (treatment policy strategy). - PFS event documented after two or more non-evaluable tumor assessments: PFS will be analyzed based on events irrespective of time interval between an event and the last evaluable tumor assessment prior to the event (treatment policy strategy)	One-sided log-rank test stratified by randomization stratification factors, as entered in the IRT system. Hazard ratio (HR) and corresponding 95% CI estimated using stratified Cox proportional hazard model. The Kaplan Meier estimate of PFS at specified time points, PFS quartiles and corresponding 95% CI from Kaplan Meier method.

Endpoint Category (estimand)	Estimands			
	Endpoint	**Population**	**Intercurrent event(s) handling strategy**	**Population-level summary (Analysis and missing data handling)**
Key secondary objective: to compare overall survival between two treatment arms.				
Key secondary endpoint (Estimand 2)	OS	ITT	- Interruption/discontinuation of study intervention: OS will be analyzed based on events irrespective of study intervention interruption/discontinuation (treatment policy strategy). - Start of new anti-cancer therapy prior to death: OS will be analyzed based on events irrespective of start of new anti-cancer therapy (treatment policy strategy).	One-sided log-rank test stratified by randomization stratification factors, as entered in the IRT system. Hazard ratio (HR) and corresponding 95% CI estimated using stratified Cox proportional hazard model. The Kaplan Meier estimate of OS at specified time points, OS quartiles and corresponding 95% CI from Kaplan Meier method.

2 SAMPLE SIZE DETERMINATION

Global population

For PFS, a total of 516 PFS events assessed by local radiologist/investigator will be needed to reject the null hypothesis using a logrank test at the one-sided level of 2.5% and a 90% power under the assumption of a HR of 0.75. Assuming proportional hazards under an exponential model and based on an anticipated median PFS of 24.8 months in the letrozole + palbociclib arm, this is expected to correspond to a median PFS of 33.1 months in the SAR439859 + palbociclib arm. Based on an expected accrual duration of 18 months (15% of total accrual in the first 4.5 months and 40% of total accrual in the first 9 months), a PFS analysis cut-off date (COD) 40 months after the first participant randomized and an annual dropout rate of 5%, a total of 1066 participants are expected to be randomized in a 1:1 ratio into the SAR439859 + palbociclib and letrozole + palbociclib arms. The power calculation accounts for two interim analyses at 40% and 70% of the planned number of events.

For OS, a total of 632 deaths will be needed to reject the null hypothesis using a logrank test at the one-sided level of 2.5% and a 80% power under the assumption of a HR of 0.80. Assuming proportional hazards under an exponential model and based on an anticipated median OS of 46 months in the letrozole + palbociclib arm, this is expected to correspond to a median OS of 57.5 months in the SAR439859 + palbociclib arm. OS COD is expected 80 months after the first participant randomized (assuming an annual dropout rate of 1%). The power calculation accounts for two interim analyses at the time of the PFS primary analysis and at 75% of the planned number of OS events.

Calculations were made using East 6.5 software.

3 ANALYSIS POPULATIONS

The following populations for analyses are defined:

Table 4 - Populations for analyses

Population	Description
Screened	All participants who signed the ICF.
Randomized	All participants from screened population who have been allocated to a randomized intervention by IRT regardless of whether the intervention was received or not.
Intent-to-treat (ITT)	All randomized participants. Participants will be analyzed according to the treatment arm assigned at randomization.
Safety	All randomized participants and who took at least 1 dose of study intervention. Participants will be analyzed according to the treatment arm they actually received.
Pharmacokinetic (amcenestrant)	All participants from the safety population who receive at least 1 dose of amcenestrant and with at least 1 evaluable plasma concentration of amcenestrant posttreatment.
Pharmacokinetic (palbociclib)	All participants from the safety population who receive at least 1 dose of palbociclib and with at least 1 evaluable plasma concentration of palbociclib posttreatment.
Pharmacokinetic (goserelin)	All participants from the safety population who receive at least 1 dose of goserelin and with at least 1 evaluable plasma concentration of goserelin posttreatment.
Population without trial impact (disruption) due to Covid-19	All randomized participants: • without any critical or major deviation related to Covid-19 • and who did not permanently discontinue full treatment due to Covid-19 • and who did not permanently discontinue study due to Covid-19.

Participants exposed to study intervention before or without being randomized will not be considered randomized and will not be included in any analysis population. The safety experience of these participants will be reported separately.

Randomized participants for whom it is unclear whether they took the study intervention will be considered as exposed and will be included in the safety population as randomized.

For any participant randomized more than once, only the data associated with the first randomization will be used in any analysis population. The safety experience associated with any later randomization will be reported separately.

For participants receiving more than one study intervention during the study, the intervention group for as-treated analyses will be the as-randomized intervention group if the participant has received at least one administration of the as-randomized intervention.

4 STATISTICAL ANALYSES

4.1 GENERAL CONSIDERATIONS

In general, continuous data will be summarized using the number of observations available, mean, standard deviation (SD), median, minimum, and maximum. Categorical and ordinal data will be summarized using the count and percentage of participants.

The baseline value is defined as the last value or measurement taken up to the date of randomization. This definition applies for all variables unless otherwise specified.

Unless otherwise specified, analyses will be performed by intervention group (and overall for baseline and demographics characteristics).

Observation period

The observation period will be divided into 3 segments:

- The **pre-treatment period** is defined as the period up to first IMP administration.
- The **on-treatment period** (ie, treatment-emergent (TE) period) is defined as the period from the first IMP administration to the last IMP administration + 30 days.
- The **post-treatment period** is defined as the period from the end of the on-treatment period.

4.2 PARTICIPANT DISPOSITIONS

The number (%) of participants included in each of the analysis populations listed in Table 4 will be summarized. Reasons for exclusion from the population without trial impact (disruption) due to Covid-19 will be summarized.

Screen failures are defined as participants who consent to participate in the study but are not subsequently randomized. The number (%) of screen failures and reasons for screen failures will be provided in the screened population.

Regarding intervention discontinuation, the following definitions will be used:

- Permanent **partial** intervention discontinuation of palbociclib is defined as the discontinuation of palbociclib but amcenestrant/letrozole is continued.
- Permanent **full** intervention discontinuation is defined as the discontinuation of all the study drugs.

The number (%) of participants in the following categories will be provided:

- Randomized participants.
- Randomized but not exposed participants.
- Randomized and exposed participants.

- Participants who discontinued all the study drugs and main reason for permanent full intervention discontinuation.
- Participants who discontinued palbociclib but amcenestrant/letrozole is continued, and main reason for permanent partial discontinuation of palbociclib.
- Participants still on treatment.
- Status at the cut-off date (Alive/Death).
- Participants with date of last contact obtained before the cutoff date and duration from last contact to cut-off date (0-4 weeks, 4-8 weeks, 8-12 weeks, >12 weeks).

Reasons for permanent study intervention "adverse event" and "other reasons" will be split as related versus not related to Covid-19, if applicable.

The number (%) of exposed and not randomized participants will also be summarized.

In addition, the number (%) of participants screened, screened-failed, randomized, with permanent full intervention discontinuation will be provided by country and site/geographical region.

Protocol deviations

Critical and major protocol deviations (automatic or manual) will be summarized in the randomized population as well as displayed separately as related versus not related to Covid-19 if applicable.

4.3 PRIMARY ENDPOINT(S) ANALYSIS

4.3.1 Definition of endpoint(s)

The primary endpoint Progression-Free Survival (PFS) is defined as the time interval from the date of randomization to the date of first documented tumor progression as per Response Evaluation Criteria in Solid Tumors (RECIST 1.1) assessed by local radiologist/investigator or death (due to any cause), whichever comes first.

4.3.2 Main analytical approach

Primary efficacy analysis will consist of PFS according to local radiologist's/investigator's assessment comparison between the amcenestrant + palbociclib arm and the letrozole + palbociclib arm through a logrank test procedure stratified by the stratification factors as entered in the IRT (ie, de-novo metastatic disease, postmenopausal women and visceral metastasis defined by at least 1 liver, lung, brain metastasis, pleural, or peritoneal involvement). A onesided Type I error rate of 2.5% will be used for statistical testing. Following hypotheses will be tested:

- The null hypothesis that the survival distribution functions (SDF) for the PFS of the experimental arm is lower than or equal to the SDF of the control arm.

- H_0: SDF (Experimental) ≤ SDF (Control)

versus

- The alternative hypothesis that the SDF for the PFS of the experimental arm is superior to the SDF of the control arm.
  - H_1: SDF (Experimental) > SDF (Control)

The cut-off date for the analysis of PFS is the actual date when the 516 PFS events (first occurrence of either documented progression assessed by local radiologist/investigator or death due to any cause) have been observed according to the primary PFS analysis. It is approximately estimated 40 months after first randomized participant assuming an accrual duration of 18 months (15% of total accrual in the first 4.5 months and 40% of total accrual in the first 9 months), an annual dropout rate of 5%, and a median PFS of 24.8 months in the control arm and a hazard ratio of 0.75.

The primary efficacy analyses on PFS will be performed on the ITT population, based on local radiologist's/Investigator's assessment of tumor burden. Analysis based on blinded Independent Review Committee (BIRC) assessment will also be performed but considered as supportive analyses.

PFS will be analyzed with Estimand 1 introduced in Section 1.2.1 and defined according to the following attributes:

- Intercurrent events:
  - The IMP discontinuation IE will be handled with the **treatment policy** strategy: PFS will be assessed based on tumor assessments and death irrespective of IMP discontinuation,
  - The new anticancer therapy IE will be handled with the **hypothetical** strategy: PFS will be censored at last evaluable tumor assessments prior to a new anti-cancer therapy,
  - Events documented after two or more non-evaluable tumor assessment will be handled with the **hypothetical** strategy: PFS will be censored at last evaluable tumor assessment if the event is documented after two or more non-evaluable tumor assessment.

The following estimates will be provided:

- PFS data will be analyzed using the Kaplan-Meier method by treatment group:
  - Kaplan-Meier estimates of the 25th, 50th and 75th percentiles and their associated 95% confidence interval (CI) will be provided. The 95% CIs will be constructed using a log-log transformation of the survival function and the methods of Brookmeyer and Crowley,
  - Number of patients at risk as well as the probabilities of being event-free at different timepoints (for example, at least 3, 6, 9, 12, 15, 18, 21, 24, 30, 36 and 39 months) with 95% CIs will be estimated for each treatment group using the Kaplan-Meier method

and a log-log approach based on a normal approximation following the Greenwood's formula,

- Kaplan-Meier curves will be plotted. These plots will include the number of patients at risk at key time points by treatment group.

- The hazard ratio (HR) and its 95% CIs will be estimated using the Cox proportional hazards model stratified by the same stratification factors as those used for the log-rank test described above. Underlying assumptions of the Cox Proportional hazards model will be assessed by graphical methods (ie, log-log graphical methods).
- For patients with events, the type of event (documented progression or death) will be summarized by treatment group using counts and percentages.
- For patients who died without documented progression, the time from the last evaluable disease assessment to the death will be summarized by treatment group using descriptive statistics.
- The number (%) of censored patients, the reason and timing of their censoring (ie, censored at randomization, censored at the last evaluable tumor assessment before the initiation of further anti-cancer therapy, or censored at last evaluable tumor assessment before the cut-off date), and the time from the last evaluable disease assessment to the cut-off date will be summarized by treatment group. For each censoring reason, when applicable, distinction will be made between cases where no event was observed and cases where an event was observed after the censoring.
- Median follow-up time (months) for the overall population will be estimated by treatment group using the reverse Kaplan-Meier method, where censored data are treated as events and events are treated as censored data.
- Time to tumor assessment may be summarized for the two treatment arms using descriptive statistics.

4.3.3 Sensitivity analysis

Different censoring and events rules

The same statistical methods used in the primary analysis will be applied using different censoring and event rules as defined below.

The sensitivity analyses will include the following censoring rules:

- Ignoring further anti-cancer therapy and considering events (documented progression or death) occurring after two or more non-evaluable tumor assessment as event.
- Considering events (documented progression or death) occurring after exactly one or two or more non-evaluable tumor assessment as event and back-dated to the next scheduled assessment.

Additional details are provided in Section 5.9.

Sensitivity analysis #1 (PFS considering events occurring after two or more non-evaluable tumor assessment as event and ignoring further anti-cancer therapy)

PFS endpoint will be analyzed based on local radiologist's/Investigator's assessment, with Estimand 1-SA#1 introduced in Section 1.2.1. Events (documented progression or death) occurring after two or more non-evaluable tumor assessment and events (documented progression or death) occurring after the start of any further ani-cancer therapy will be included. The date of progression (or death) will be used for date of outcome.

Sensitivity analysis #2 (PFS considering events occurring after two or more non-evaluable tumor assessment as event and back-dating at the next scheduled assessment)

If more than 10% patients have two or more consecutive non-evaluable assessments prior to PFS event, PFS endpoint will be analyzed based on local radiologist's/Investigator's assessment, including events (documented progression or death) occurring after two or more non-evaluable tumor assessment as event. The date of the next scheduled assessment will be used for date of outcome.

Stratification factors

The same statistical methods used in the primary analysis will be applied using different stratification rules as defined below.

Sensitivity analysis #3 (PFS analysis using stratification factors derived from eCRF data)

If more than 10% patients have a discordance between the strata as entered in the IRT system and as derived from eCRF data, PFS endpoint will be analyzed based on local radiologist's/Investigator's assessment stratified by the stratification factors as derived from eCRF data.

Non-proportional hazard

As mentioned above, underlying assumptions of the Cox Proportional hazards model will be evaluated to assess the relevance of estimating the treatment effect from hazard ratio. In case the proportional hazards assumption is not valid, Restricted Mean Survival Time (RMST) method (1) may be conducted for PFS with the primary analysis censoring rule. The RMST methodology is valid under any distribution of the time to event in the treatment groups and provide an estimate of the expected progression-free time between randomization and a common timepoint denoted by τ. The timepoint τ should be limited to the largest uncensored time to event in the data:

- τ_{max} = minimum of (largest observed PFS event time for experimental arm, largest observed PFS event time for control arm).

The treatment effect will be estimated based on the difference in RMST between the two treatment arms. The associated 95% CI for the difference in means RMST as a function of τ will be generated and the associated treatment effect between the two treatment arms will be plotted against time τ.

4.3.4 Supplementary analyses

A supplementary analysis will be performed considering the clinical/non-radiological progression as event, with the date of clinical/non-radiological progression as date of outcome. In addition sensitivity analyses could be performed using different censoring and event rules or stratification rules as defined in Section 4.3.3, if relevant.

4.3.5 Evaluation of BIRC data

A random sample-audit BIRC approach will be used as an auditing tool to support the primary findings on PFS based on the local radiologist's/investigator's assessment (denoted by LE-PFS). Approximately 50% of the randomized participants will be randomly selected and the BIRC will assess the tumor progression of these participants based on the review of tumor assessments.

The audit size calculation approach will be based on the method proposed by Dodd, et al (2). Assuming local radiologist's/investigator's and BIRC assessments are similar and the estimated log of investigator-based HR is -0.2877 (ie, HR=0.75), the audit size of 50% will ensure that the upper bound of a one-sided 95% CI for BIRC-based log-hazard ratio has 83% probability of being below 0 (ie, HR <1) if the correlation between local radiologist's/investigator's assessment and BIRC assessment is 0.65. If the correlation based on the LE-PFS and BIRC-PFS is greater than our assumption, the power of the sample-audit will increase.

Two methods will be used to summarize the data from the sample-based BIRC assessment of PFS.

Robustness of LE-PFS treatment effect

Robustness of PFS treatment effect will be evaluated with the NCI (National Cancer Institute) method (2), which uses an auxiliary variable estimator of the log-hazard ratio that combines information from participant level local radiologist's/investigator's assessment from all participants and the BIRC assessment of participants randomly selected for central review. If the upper bound of the one-sided 95% CI of the auxiliary estimator is below 0, then the BIRC audit is considered to have confirmed findings based on local evaluations; otherwise, a full BIRC may be triggered.

The auxiliary variable estimates and its one-sided 95% CI will be calculated as follows:

- $\delta = \frac{m}{N}$, the ratio of sample-audit BIRC size (m) on the trial size (N).
- $\hat{\theta}_{LE}$; $\hat{V}_{LE}$, the estimate of the log-hazard ratio and associated variance of the LE-PFS based on all patients.
- $\hat{\theta}_{LE1}$; $\hat{V}_{LE1}$, the estimate of the log-hazard ratio and associated variance of the LE-PFS for the patients not selected based on the sample-audit BIRC.
- $\hat{\theta}_{LE2}$; $\hat{V}_{LE2}$, the estimate of the log-hazard ratio and associated variance of the LE-PFS for the patients selected based on the sample-audit BIRC.

- $\hat{\theta}_{BIRC}$; $\hat{V}_{BIRC}$, the estimate of the log-hazard ratio and associated variance of the BIRC-PFS for the patients selected based on the sample-audit BIRC.
- $\hat{\rho}$, a bootstrap estimator of the correlation between $\hat{\theta}_{BIRC}$ and $\hat{\theta}_{LE2}$.

Specifically, within the randomly selected sample (of size *m*) for BICR audit, a bootstrap sample of *m* participants will be sampled with replacement. The two sample-based log-hazard ratio estimates (LE-based vs. BICR-based) will be estimated with the bootstrap sample. The procedure will be repeated 10000 times, and 10000 pairs of sample estimates of log (HR) (LE-based versus BICR-based) will be calculated. The sample correlation coefficient between the LE-based log(HR) estimates and BICR-based log(HR) estimates denoted $\hat{\rho}$ will be obtained based on the Pearson correlation coefficient, which will be used in the calculation of the auxiliary estimator and the corresponding one-sided 95% CI.

The auxiliary estimator, $\widetilde{\theta}$, will be calculated as

$$\widetilde{\theta} = \hat{\theta}_{BIRC} + \hat{\rho}\sqrt{\delta}(1-\delta)\sqrt{\frac{\hat{V}_{BIRC}}{\hat{V}_{LE}}}(\hat{\theta}_{LE2} - \hat{\theta}_{LE1})$$

The estimate of variance of $\widetilde{\theta}$ will be calculated as

$$\widetilde{V} = \hat{V}_{BIRC}\{1 - \hat{\rho}^2(1-\delta)\}$$

Assuming asymptotic normality of $\widetilde{\theta}$, the upper bound of one-sided 95% CI of this estimator will be calculated as $\widetilde{\theta} + Z_{1-0.05}\sqrt{\widetilde{V}}$, where $Z_{1-0.05}$ is the 95% quantile of a standard normal distribution. Then the associated upper bound of the confidence interval for the BIRC-PFS HR is obtained with $e^{\widetilde{\theta} + Z_{1-0.05}\sqrt{\widetilde{V}}}$.

The criteria for determining the robustness of the LE-PFS treatment effect will be based on the comparison of this upper bound of the 95% one-sided confidence interval with a clinically irrelevant threshold corresponding to no treatment effect, ie, an HR of 1. If the upper bound of the 95% one-sided confidence interval is below 1, robustness of the LE-PFS treatment effect will be considered as a similar outcome is observed between the BIRC and LE assessment. Indeed, in this situation, the upper bound of the confidence interval associated with the estimate of the true LE-PFS and the true BIRC-PFS treatment effects both exclude the threshold of no treatment effect. Otherwise, if the upper bound of the 95% one-sided confidence interval is greater than 1, one cannot conclude that the sample-audit BIRC is supportive of the LE findings and a full BIRC may be triggered to support the primary findings based on the LE-PFS treatment effect.

Differential discordance on PD events

The differential discordance between the BIRC and the investigator assessment of documented progression will be assessed using the Pharmaceutical Research and Manufacturers of America (PhRMA) method (3). The early discrepancy rate (EDR) and late discrepancy rate (LDR) differences between the two treatment arms will be calculated as follows:

Table 5 - Cross-tabulation of ICR and investigator assessments of documented PD

Investigator/local radiologist	BIRC	
	Documented PD	No documented PD
Documented PD	$a = a_1 + a_2 + a_3$	b
No documented PD	c	d

Only documented PD component of PFS is considered; if death occurs without prior PD, a subject is counted under 'No documented PD'
a_1: number of agreements on timing and occurrence of documented PD
a_2: number of times investigator declares documented PD later than BIRC
a_3: number of times investigator declares documented PD earlier than BIRC

If full BIRC is not triggered, the calculation will be based on participants who were randomly selected for central review. Otherwise, the calculation will be based on all participants.

The timing of investigator/local radiologist and BIRC (for participants with agreement on documented PD) will be considered to agree if they occur within ±7 days of each other, aligned with the protocol-specified window for tumor assessments.

The EDR is defined as:

$$EDR = \frac{b + a_3}{a + b}$$

The EDR quantifies the frequency with which the investigator declares progression early relative to ICR as a proportion of the total number of investigator-assessed PDs, within each arm.

The LDR is defined as:

$$LDR = \frac{c + a_2}{b + c + a_2 + a_3}$$

The LDR quantifies the frequency with which the investigator declares progression later than ICR as a proportion of the total number of discrepancies, within each arm.

If the distribution of discrepancies is similar between the arms, then this 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 (DD) for each measure will be summarized as the rate on the SAR439859 + palbociclib arm minus the rate on the letrozole + palbociclib 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.

Concordance of PFS outcome

A comparison of PFS outcome (ie, "Documented progression", "Death", "Censored") and timing (ie, "Same time", "ICR after INV", "ICR before INV") between the ICR and the investigator assessments will be summarized for each treatment group.

If full BIRC is not triggered, the calculation will be based on participants who were randomly selected for central review. Otherwise, the calculation will be based on all participants.

Table 6 - Cross-tabulation of ICR and investigator assessments of PFS outcome

Investigator/local radiologist	ICR		
	Documented PD	Death	Censored
Documented PD	n_{11}	n_{12}	n_{13}
Death	n_{21}	n_{22}	n_{23}
Censored	n_{31}	n_{32}	n_{33}

The PFS Outcome Discrepancy Rate (PFS ODR) will be calculated for each treatment arm as follow.

$$\text{PFS ODR} = \frac{n_{13} + n_{23} + n_{31} + n_{32}}{\sum_{ij} n_{ij}}$$

4.3.6 Subgroup analyses

Evaluation of treatment effect consistency

The consistency of the results from the primary analysis will be evaluated across pre-defined subgroups in patients available in the subgroup of consideration. The definition of each subgroup is defined in Table 7. Depending upon the study results, additional subgroups may be examined, and subgroups with small sample sizes may be pooled to create a larger meaningful subgroup. For each subgroup, Kaplan-Meier estimates of the 25th, 50th and 75th percentiles and their associated 95% CIs will be provided for each treatment arm along with the HR and its 95% CI estimated using the unstratified Cox proportional hazards model. A forest plot summarizing the results for each subgroup will be provided.

Table 7 - Subgroups analyses: covariates investigated

Subgroup	Description
De-novo metastatic disease as per IRT	Yes or No
Postmenopausal women as per IRT	Yes or No
Visceral metastasis as per IRT	Yes or No
ECOG	0 or >=1
Age	<65 years or ≥65 years
Race	Asian or White or Other
Geographical region	Europe or North America or Asia or Other
Prior chemotherapy	Yes or No
PgR status	PgR+ or PgR-
PIK3CA mutational status	Wild type or mutated

Subgroup	Description
ESR1 mutational status	Wild type or mutated
Number of metastatic sites	<3 or ≥3
Measurable disease status	Measurable or Non-measurable
Treatment-free interval	<2 years v.s. >=2 years

Note: In case of discrepancy between IRT and eCRF stratification factors, subgroup analyses will also be performed according to the stratification factors as per eCRF. Some subgroups may be pooled based on the number of patients within each subgroup.

Evaluation of interactions

For each pre-defined factor defined in Table 7, PFS will be analyzed using an unstratified Cox proportional hazards model with terms for the factor, treatment and their interaction. The p-value of the test of interaction will be provided.

Evaluation of confounding

Since the results from the primary analysis could be impacted by confounding factors, any potential issues will be examined and, if confirmed, exploratory analysis of the primary endpoint will be done accordingly. A multivariate Cox proportional hazards model will be used to identify prognostic factors among the demographic and baseline characteristics factors described in the Table 7 using a stepwise selection procedure with a 15% significance level for removing effects. For significant prognostic factors identified in the multivariate model, the balance between treatment groups will be assessed. If major confounding is identified through screening for treatment group imbalances in a prognostic factor at baseline, an exploratory analysis of PFS will be done after adjusting for the prognostic factors in the multivariate Cox proportional hazards model. Differences between the adjusted and unadjusted models will be discussed in the clinical study report.

4.4 SECONDARY ENDPOINT(S) ANALYSIS

The secondary endpoints detailed in this section are overall survival (OS), objective response rate (ORR), clinical benefit rate (CBR), duration of response (DOR), time to first chemotherapy (TT1C) and progression-free survival on the next systemic anti-cancer therapy (PFS2). Other secondary endpoints analyses are defined in Section 4.7.2 (AE, SAE), Section 4.7.3 (laboratory abnormalities), Section 4.8.1 (PK) and Section 4.8.2 (quality of life).

4.4.1 Key/Confirmatory secondary endpoint(s)

4.4.1.1 Definition of endpoint(s)

Overall survival is defined as the time from date of randomization to date of death due to any cause. In the absence of observation of death, survival time will be censored at the last date the participant is known to be alive or at the OS cut-off date, whichever occurs first.

4.4.1.2 Main analytical approach

Overall survival (OS) will be evaluated as a key secondary efficacy endpoint. A hierarchical testing strategy will be used to ensure a strong control of the overall Type I error rate at one-sided 2.5%. In other words, comparison between arms on the OS will be performed only if the primary analysis of the PFS is statistically significant.

In case of statistically significant PFS, OS will be compared between the amcenestrant + palbociclib arm and the letrozole + palbociclib arm through a logrank test procedure stratified by the stratification factors as entered in the IRT. Following hypotheses will be tested:

- The null hypothesis that the survival distribution functions (SDF) for the OS of the experimental arm is lower than or equal to the SDF of the control arm.
  - H_0: SDF (Experimental) ≤ SDF (Control)

versus

- The alternative hypothesis that the SDF for the OS of the experimental arm is superior to the SDF of the control arm.
  - H_1: SDF (Experimental) > SDF (Control)

Otherwise, descriptive statistics of OS will be provided at the time of final PFS analysis.

The COD for final OS analysis will be the date when 632 death events have been observed.

OS will be analyzed on the ITT population, with Estimand 2 introduced in Section 1.2.1. Intercurrent events will be handled with the **treatment policy** strategy: OS will be assessed based on overall survival data irrespective of IMP interruption/discontinuation and irrespective of start of new anti-cancer therapy. The same estimates will be provided as for PFS defined in Section 4.3.2, with the exception that the follow-up of OS will be defined as the time interval from the date of randomization to the date of last contact with the patient. Patients who have died will be censored on their date of death.

4.4.1.3 Sensitivity analysis

Sensitivity analyses adjusting OS for switch to subsequent anti-cancer treatment could be performed at interim and/or final analyses, eg, using inverse probability of censoring weighting (IPCW) method (4) and rank preserving structural failure time model (RPSFTM) (5).

4.4.2 Supportive secondary endpoint(s)

4.4.2.1 Definition of endpoint(s)

Objective response rate

The Best Overall Response observed from the date of randomization until documented disease progression, death, start of an anti-cancer therapy, or analysis cut-off date (COD), whichever occurs first, will be determined according to RECIST v1.1. Confirmation of responses (CR or PR) is necessary.

The ORR on each randomized treatment arm will be estimated by dividing the number of participants with objective response (confirmed CR or PR as BOR assessed by Investigator, according to RECIST 1.1) by the number of participants from the analysis population of the respective treatment arm.

In addition, in order to evaluate the tumor shrinkage of the target lesions, best relative change from baseline in tumor size will be assessed for each patient with measurable disease.

Duration of response

The DOR is defined as the time from the date of the first tumor assessment at which the overall response was recorded as CR or PR that is subsequently confirmed to the first date of tumor assessment at which the overall response was recorded as PD or death, whichever comes first. For participants with ongoing response at the time of the analysis, DOR will be censored at the date of the last valid disease assessment not showing documented progression performed before the initiation of a new anticancer treatment (if any) or COD, whichever comes first.

DOR is determined only for patients who have achieved a BOR of PR or better.

Clinical benefit rate

The CBR on each randomized treatment arm will be estimated by dividing the number of participants considered as clinical benefit responders based on investigator assessment, by the number of participants from the analysis population of the respective treatment arm. For patients with measurable disease at baseline, they will be considered as clinical benefit responders if they achieve a CR or PR as BOR, or SD with an overall response recorded as SD or better at 24-1=23 weeks or later from randomization, allowing for the ±7 days visit window for tumor assessment. For patients with non-measurable disease at baseline, they will be considered as clinical benefit responders if they achieve a CR as BOR or Non-CR/Non-PD with an overall response recorded as Non-CR/Non-PD or better at least 23 weeks after randomization. Only tumor assessments performed before documented disease progression, death, start of an anti-cancer therapy, or analysis cut-off date, whichever occurs first, will be used for the derivation of the CBR.

Progression-free survival on next line of therapy (PFS2)

The PFS2 is defined as the time from the date of randomization to the date of first documentation of PD on the next systemic anti-cancer therapy according to investigator or death due to any cause, whichever occurs first. Documentation of progression on the next systemic anti-cancer therapy is based on investigator assessment of PD as captured in the follow-up CRF page.

Time to first Chemotherapy (TT1C)

The time to first use of chemotherapy after disease progression is defined as the time from date of randomization to start of the first use of chemotherapy. In the absence of chemotherapy after study treatment discontinuation, TT1C will be censored at the last contact date or the cutoff date, whichever occurs first.

4.4.2.2 Main analytical approach

Analysis of response-based endpoints (ie, ORR, DOR and CBR) will be performed primarily on the ITT population and supported by the analyses based on the ITT population with measurable disease at study entry. The other secondary analyses will be performed on ITT population only.

Response-based endpoints

ORR and CBR according to investigator assessments will be summarized by treatment arm with descriptive statistics at the time of the primary analysis on PFS (based on data collected up to the PFS analysis cut-off date). In addition, 95% two-sided CIs will be computed using the Clopper-Pearson method.

Of note, the BOR for each participant will also be summarized according to Investigator assessments.

The DOR will only be summarized on the subgroup of participants who have achieved objective response (confirmed CR or PR as BOR). Duration of response by treatment arm will be summarized using Kaplan-Meier methods and displayed graphically, if appropriate. 25th, 50th and 75th percentiles of DOR and associated 95% CI will be provided. The 95% CIs will be estimated for each treatment group using the Kaplan-Meier method and a loglog approach based on a normal approximation following the Greenwood's formula.

Exploratory analyses of response-based endpoints may be provided according to subgroups of interest (Table 7).

Waterfall plots will be used to display the best relative change from baseline in tumor size observed for each patient. It will be sorted within each arm by decreasing order of best relative change from baseline in patients with measurable disease at baseline. BOR will be displayed. Only patients with post-baseline target lesions measurement will be included in the plot.

Progression-free survival on next line of therapy (PFS2)

The PFS2 will be assessed with the following censoring and event scheme:

- Alive patients who have not yet initiated a next systemic anticancer therapy will be censored at the last contact date, or COD, whichever comes first.
- Patients who have died prior to initiation of the next systemic anticancer therapy will be considered as having an event with death date as event date in the PFS2 analysis.
- Patients without PFS2 event who have initiated a first next systemic anticancer therapy but not a second next systemic anticancer therapy will be censored at the last contact date or COD, whichever comes first.
- Patients without PFS2 event who have initiated a first next systemic anticancer therapy and a second next systemic anticancer therapy without event will be censored at the start date of their second next systemic anticancer therapy or COD, whichever comes first.

The same statistical methods as for PFS will be used with the exception that no statistical test will be made.

Time to first Chemotherapy (TT1C)

The TT1C will be assessed with the following censoring scheme:

- Alive patients who have not yet initiated a new chemotherapy will be censored at the last contact date, or COD, whichever comes first.
- Patients who have died prior to initiation of a new chemotherapy will be censored at the death date, or COD, whichever comes first.

The same statistical methods as for PFS will be used with the exception that no statistical test will be made.

4.5 TERTIARY/EXPLORATORY ENDPOINT(S) ANALYSIS

Tertiary endpoints analyses are defined in Section 4.8.1 (PK), Section 4.8.3 (Biomarker) and Section 4.8.4 (Exposure/response relationship).

4.6 MULTIPLICITY ISSUES

Hypothesis testing of the key secondary efficacy endpoint will be carried out. In order to ensure a strong control of the overall Type I error rate at a one-sided 2.5%, a hierarchical testing strategy will be used. In other words, comparison between arms on the OS will be performed only if the primary analysis of the PFS is statistically significant.

4.7 SAFETY ANALYSES

All safety analyses will be performed on the safety population as defined in Section 3, unless otherwise specified, using the following common rules:

- The analysis of the safety variables will be essentially descriptive, and no testing is planned.
- Safety data in participants who do not belong to the safety population (eg, exposed but not randomized) will be provided.

4.7.1 Extent of exposure

The extent of IMP exposure will be assessed by the duration of IMP exposure and compliance and summarized within the safety population. In addition, summaries may be provided by trial impact (disruption) due to Covid-19.

Duration of IMP exposure

Duration of IMP exposure will be summarized quantitatively and categorically: Less than 24 weeks, 24 to 48 weeks, 48 to 72 weeks, 72 to 96 weeks, 96 to 120 weeks, 120 to 144 weeks and >144 weeks.

Treatment compliance

Percentage of treatment compliance for a participant will be evaluated by relative dose intensity for each IMP, as defined in Section 4.7.1.2 and Section 4.7.1.3.

4.7.1.1 Overall exposure

The dose information will be assessed by the following variables:

- Overall number of cycles started, defined by the number of cycles in which at least one dose of any study interventions is administered.
- Duration of IMP exposure (in weeks) is defined as (last day of exposure – first day of exposure + 1)/7.
- The first day of exposure is defined as the first administration date with non-zero dose for at least one of the IMP amcenestrant/letrozole, matching placebo, palbociclib. The last day of exposure is the last day of study treatment administration plus appropriate carryover days, defined as the maximum between:
  - The last administration date of amcenestrant/letrozole,
  - Minimum of (date of death, the last administration date of Palbociclib + 7.

The total number of cycles started, number of cycles started by participants will be summarized as a quantitative variable and by category (number (%) of participants receiving at least 4 cycles, at least 8 cycles, etc.). The duration of overall exposure will be summarized quantitatively.

The following variable will be computed to describe overall dose modification (cycle delay):

- Cycle delay: A cycle is deemed as delayed if the start date of the current cycle – start date of the previous cycle is >31 days. Cycle delay is not defined for the first cycle.

Cycle delay will be analyzed at the participant (with number of participants used as denominator) and cycle (with number of cycles used as denominator) levels, as follows:

- Number (%) of participants with a least 1 cycle delayed
  - Number (%) of participants with a cycle delayed between 4 and 7 days (using maximum delay across all cycles),
  - Number (%) of participants with a cycle delayed between 8 and 14 days (using maximum delay across all cycles),
  - Number (%) of participants with a cycle delayed >14 days (using maximum delay across all cycles).
- Number (%) of cycles delayed
  - Number (%) of cycles delayed between 4 and 7 days,
  - Number (%) of cycles delayed between 8 and 14 days,
  - Number (%) of cycles delayed >14 days.

4.7.1.2 Amcenestrant/Letrozole exposure

The dose information will be assessed by the following:

- Duration of amcenestrant/letrozole exposure (in weeks) is defined by (date of last administration of amcenestrant/letrozole – date of first administration of amcenestrant/letrozole+1) /7.
- Cumulative dose (mg): the cumulative dose is the sum of all actual doses of amcenestrant/letrozole, given from first to last administration.
- Actual dose intensity (ADI in mg/day) for amcenestrant/letrozole: defined as the cumulative dose divided by the duration of amcenestrant/letrozole exposure (in weeks×7)

$$\text{ADI}=\frac{\text{Cumulative Dose (mg)}}{\text{Duration of treatment in weeks} \times 7}$$

- Planned dose intensity (PDI in mg/day): planned dose of amcenestrant/letrozole on C1D1.
- Relative dose intensity (RDI, in %): $100 \times \frac{\text{ADI (mg/day)}}{\text{PDI (mg/day)}}$

The total number of doses, total number of cycles started, number of cycles started by participant will be summarized as a quantitative variable and by category (number [%] of participants receiving at least 1 cycle, at least 2 cycles, etc.). Duration of amcenestrant/letrozole exposure, cumulative dose, ADI and RDI will be summarized quantitatively. RDI will also be summarized according to the following categories: 0-80%, 80-100%, >100%.

The following variables will be derived to describe dose modifications:

- Dose reduction: The first administration will not be counted as a dose reduction. For the second and subsequent amcenestrant/letrozole administrations, a dose is deemed to have been reduced if the daily dose taken by a patient is lower than the daily dose taken on the previous day or the day before dose(s) omitted.
- Dose omission is defined as a dose not taken or equal to 0 mg/day but administered afterwards. Several consecutive dose omissions will be counted as one episode of omission.

Dose modifications will be analyzed by participant and dose as follows:

- **Participant** (number of participants treated will be used as denominator)
  - Number (%) of participants with at least 1 dose modification (dose reduction or dose omission),
  - Number (%) of participants with at least 1 dose reduction,
  - Number (%) of participants with at least 1 dose omission,
  - Number of participants with at least 7 consecutive days of dose omission,
  - Number of dose reductions by patient according to the following categories: 0, 1, >1,

- Number of dose omissions by patient according to the following categories: 0, 1-2, 3-4, 5-6, >6,
- Number of participants (respectively cycles) with at least one cycle delay.

- **Dose** (number of doses administered will be used as denominator)
  - Number (%) of dose reductions,
  - Number (%) of dose omissions.

In addition, time to first dose reduction and time to first episode of dose omission will be calculated and summarized. Time to first dose reduction is defined as the time interval from randomization to the date of first dose reduction. Time to first episode of dose omission is defined as the time interval from randomization to the first day of dose omission.

4.7.1.3 Palbociclib exposure

The dose information will be assessed by the following:

- Duration of palbociclib exposure (in weeks) is defined by [min(date of death, date of last cycle last palbociclib intake + 7) – date of first cycle first palbociclib intake + 1]/7.
- Cumulative dose (mg): the cumulative dose is the sum of all actual doses of palbociclib, given from first to last administration.
- Actual dose intensity (ADI in mg/week/cycle) for palbociclib: defined as the cumulative dose divided by the duration of palbociclib exposure (in weeks×7)

$$\text{ADI} = \frac{\text{Cumulative Dose (mg)}}{\text{Actual number of weeks on-treatment for palbociclib}}$$

- Planned dose intensity (PDI in mg/week/cycle) for palbociclib:

$$\text{PDI} = \frac{125\ x\ 7\ x\ 3}{4} = 656.25$$

- Relative dose intensity (RDI, in %): $100 \times \frac{\text{ADI (mg/week/cycle)}}{\text{PDI (mg/week/cycle)}}$

The total number of doses, total number of cycles started, number of cycles started by participant will be summarized as a quantitative variable and by category (number [%] of participants receiving at least 1 cycle, at least 2 cycles, etc.). Duration of palbociclib exposure, cumulative dose, ADI and RDI will be summarized quantitatively. RDI will also be summarized according to the following categories: 0-80%, 80-100%, >100%.

The following variables will be derived to describe dose modifications:

- Dose reduction: The first administration will not be counted as a dose reduction. For the second and subsequent palbociclib administrations, a dose is deemed to have been reduced if the daily dose taken by a patient is lower than the daily dose taken on the previous day or the day before dose(s) omitted.

- Dose omission is defined as a dose not taken or equal to 0 mg/day between two non-zero doses, except for the planned 7 days off according to the palbociclib regimen as per protocol. Several consecutive dose omissions will be counted as one episode of omission.

Dose modifications will be analyzed by participant, cycle and dose as follows:

- **Participant** (number of participants treated will be used as denominator)
  - Number (%) of participants with at least 1 dose modification (dose reduction or dose omission),
  - Number (%) of participants with at least 1 dose reduction,
  - Number (%) of patients with at least one dose reduction to 100 mg,
  - Number (%) of patients with at least one dose reduction to 75 mg,
  - Number (%) of participants with at least 1 dose omission,
  - Number of participants with at least 7 consecutive days of dose omission,
  - Number of dose reductions by patient according to the following categories: 0, 1, >1,
  - Number of dose omissions by patient according to the following categories: 0, 1, >1,
  - Number of participants (respectively cycles) with at least one cycle delay.
- **Dose** (number of doses administered will be used as denominator)
  - Number (%) of dose reductions (including reduced to 100mg, 75mg),
  - Number (%) of dose omissions.

In addition, time to first dose reduction and time to first episode of dose omission will be calculated and summarized. Time to first dose reduction is defined as the time interval from randomization to the date of first dose reduction. Time to first episode of dose omission is defined as the time interval from randomization to the first day of dose omission.

4.7.2 Adverse events

General common rules for adverse events

All AEs will be graded according to National cancer institute common terminology for adverse events (NCI-CTCAE version 5.0) and coded to a lower-level term (LLT), preferred term (PT), high-level term (HLT), high-level group term (HLGT), and associated primary system organ class (SOC) using the Medical Dictionary for Regulatory Activities (MedDRA) version currently in effect at Sanofi at the time of database lock.

The AEs will be analyzed in the following 3 categories:

- Pre-treatment AEs: AEs that developed, worsened or became serious during the pre-treatment period.
- Treatment-emergent adverse events (TEAE)s: AEs that developed, worsened or became serious during the treatment-emergent period.

- Post-treatment AEs: AEs that developed, worsened or became serious during the post-treatment period.

Similarly, the deaths will be analyzed in the pre-treatment, treatment-emergent and post-treatment periods.

The primary focus of AE reporting will be on TEAEs. Pre-treatment and post-treatment AEs will be described separately.

An AE with incomplete or missing date/time of onset (occurrence, worsening, or becoming serious) will be classified as a TEAE unless there is definitive information to determine it is a pre-treatment or a post-treatment AE.

If the assessment of the relationship to IMP is missing for an AE, this AE will be assumed as related to IMP. If the grade is missing for 1 of the treatment-emergent occurrences of an AE, the grade will be summarized with the maximal grade of the other occurrences. If the grade is missing for all the occurrences, the grade will be left as missing.

Multiple occurrences of the same event in the same participant will be counted only once in the tables within a treatment phase, using the maximum (worst) grade by treatment phase. Summaries will be provided for all grades combined and for grade ≥3 (including grade 5). Missing grades, if any, will be included in the "all grades" category.

The AE tables will be sorted as indicated in Table 8.

Table 8 - Sorting of AE tables

AE presentation	Sorting rules
SOC, HLGT, HLT and PT	By the internationally agreed SOC order and by alphabetic order of HLGTs, HLTs and PTs.
SOC, HLT and PT	By the internationally agreed SOC order and by alphabetic order of HLTs and PTs.
SOC and PT	By the internationally agreed SOC order and decreasing frequency of PTs[a, b]
PT	By decreasing frequency of PTs[a]

a Sorting will be based on the SAR439859 intervention group.
b The table of all TEAEs presented by SOC and PT will define the presentation order for all other tables (eg, treatment-emergent SAE) presented by SOC and PT, unless otherwise specified.

Analysis of all adverse events

The overview of TEAE with the details below will be generated:

- Any TEAE.
- Any grade ≥3 TEAE.
- Any treatment emergent SAE.
- Grade 5 TEAE (any TEAE with a fatal outcome during the treatment-emergent period).
- Any TEAE related to amcenestrant/letrozole.

- Any TEAE related to Palbociclib.
- Any grade ≥3 TEAE related to amcenestrant/letrozole.
- Any grade ≥3 TEAE related to palcociclib.
- Any treatment emergent SAE related to amcenestrant/letrozole.
- Any treatment emergent SAE related to Palbociclib.
- Any TEAE leading to permanent partial discontinuation of palbociclib.
- Any TEAE leading to permanent full intervention discontinuation.

The AE summaries of Table 9 will be generated with number (%) of participants experiencing at least one event. The analyses will be performed for all grades combined and for grades ≥3. The all TEAE summary by Primary SOC and PT (and other safety summaries (eg, SAEs, deaths), if deemed needed after TEAE evaluation) may be performed by trial impact (disruption) due to Covid-19.

Table 9 - Analyses of adverse events

Type of AE	MedDRA levels
TEAE	
All TEAE	Primary SOC, HGLT, HLT and PT
	Primary SOC and PT
Common TEAE (≥5% in any treatment arm)	Primary SOC and PT
TEAE related to amcenestrant/letrozole as per Investigator's judgment	Primary SOC and PT
TEAE related to palbociclib as per Investigator's judgment	Primary SOC and PT
SAE	
Treatment emergent SAE	Primary SOC, HGLT, HLT and PT
	Primary SOC and PT
Treatment emergent SAE related to amcenestrant/letrozole as per Investigator's judgment	Primary SOC and PT
Treatment emergent SAE related to palbociclib as per Investigator's judgment	Primary SOC and PT
TEAE leading to treatment discontinuation or modification	
TEAE leading to permanent full intervention discontinuation	Primary SOC and PT
TEAE leading to permanent partial discontinuation of palbociclib	Primary SOC and PT
TEAE leading to dose modification (including dose reduction and dose omission) of amcenestrant/letrozole	Primary SOC and PT
TEAE leading to dose modification (including dose reduction and dose omission) of palbociclib	Primary SOC and PT

Type of AE	MedDRA levels
Death	
TEAE leading to death[a]	Primary SOC and PT
AE leading to death[a]	Primary SOC and PT
• In context of disease progression[b]	
• In context other than disease progression[c]	

a death as an outcome of the AE as reported by the Investigator in the AE page
b death within 30 days from last study treatment administration and the cause of death is disease progression
c death within 30 days from last study treatment administration and for whom cause of death is not disease progression or the death occurred more than 30 days from last study treatment administration and the cause of death is adverse event

Analysis of deaths

In addition to the analyses of deaths included in Table 9, the number (%) of participants in the following categories will be provided:

- Deaths during the treatment-emergent and post-treatment periods by reason for death.
- Deaths in non-randomized participants or randomized but not treated participants.

An overview of Grade 5 AEs will be provided with the following categories:

- Grade 5 AE (TEAE and post-treatment).
- TEAE leading to death (regardless date of death/period).
  - Grade 5 TEAE with a fatal outcome during the treatment period,
  - Any Grade TEAE with a fatal outcome during the post-treatment period.
- Post-treatment Grade 5 AE (excluding a TEAE that worsened to Grade 5 during the post-treatment period).

In addition, a listing of all deaths during the study will be provided.

Analysis of adverse events of special interest (AESIs)

Adverse events of special interest (AESIs) will be selected for analyses as indicated in Table 10. Number (%) of participants experiencing at least one event will be provided for each event of interest. Tables will be sorted as indicated in Table 8.

Table 10 - Selections for AESIs

AESIs	Selection
Pregnancy of a female participant entered in a study as well as pregnancy occurring in a female partner of a male participant entered in a study with IMP/NIMP	e-CRF specific tick box on the AE page
Symptomatic overdose (serious or nonserious) with IMP/NIMP	e-CRF specific tick box on the AE page
Increase in alanine transaminase (ALT) ≥Grade 3	e-CRF specific tick box on the AE page
Photosensitivity	e-CRF specific tick box on the AE page

4.7.3 Laboratory variables, vital signs and electrocardiograms (ECGs)

The following laboratory variables, vital signs and electrocardiogram (ECG) variables will be analyzed. They will be converted into standard international units.

- Hematology:
  - Red blood cells and platelets and coagulation: hemoglobin, hematocrit, red blood cell count, platelet count, prothrombin time and international normalized ratio (INR),
  - White blood cells: white blood cell count, neutrophils, lymphocytes, monocytes, basophils, eosinophils.
- Clinical chemistry:
  - Metabolism: glucose, total cholesterol, high-density lipoprotein, low-density lipoprotein, triglycerides, albumin,
  - Electrolytes: sodium, potassium, chloride, calcium, phosphate, magnesium,
  - Renal function: creatinine, eGFR (as collected in the eCRF, blood urea nitrogen, urea,
  - Liver function: alanine aminotransferase (ALT), aspartate aminotransferase (AST), alkaline phosphatase, lactate dehydrogenase, total and conjugated bilirubin, gamma-glutamyltransferase (GGT).
- Urinalysis:
  - Urinalysis for qualitative analysis: pH, proteins, glucose, ketones, leukocytes, erythrocytes.
- Vital signs: heart rate, systolic and diastolic blood pressure, weight and ECOG PS (0, 1, 2, 3, 4).
- ECG variables: ECG assessments will be described as normal/missing or abnormal. ECG parameters collected at baseline will be summarized quantitatively.

For hematological parameters and some selected biochemistry parameters, Sanofi sponsor generic ranges (LLN, ULN) are defined and will be used for grading. For other biochemistry parameters, grading will be derived using local laboratory normal ranges.

Quantitative analyses

For blood pressure, a table will be provided with the baseline value and worst value during the on-treatment period.

Analyses according to PCSA and NCI grading

For laboratory variables, analyses according to NCI grading will be made based on NCI-CTCAE version as defined in the protocol. In addition, for laboratory variables for which NCI-CTCAE scale is not applicable and vital signs, PCSA analyses will be performed and PCSA values are defined as abnormal values considered medically important by the Sponsor according to predefined criteria/thresholds based on literature review (Section 5.5).

Analyses according to PCSA and NCI grading will be performed based on the worst value during the treatment-emergent period, using all measurements (either local or central, either scheduled, nonscheduled or repeated).

For laboratory variables, vital signs and ECG variables above, the incidence of participants with at least one PCSA during the treatment-emergent period will be summarized regardless of the baseline level and according to the following baseline status categories:

- Normal/missing.
- Abnormal according to PCSA criterion or criteria.

For hematology/clinical chemistry toxicities (except for liver function) and ECOG PS, shift tables showing the number of patients in each grade at baseline by worst grade during the on-treatment period will be provided.

For laboratory variables graded by NCI-CTCAE,

- The number (%) of participants with abnormal laboratory tests at baseline will be presented by grade.
- The number (%) of participants with abnormal laboratory tests during the treatment-emergent period will be summarized by grade. When appropriate, the number (%) of participants with abnormality of any grade and with Grade 3-4 abnormalities will be provided.

For ECG, the incidence of participants with at least one abnormal ECG during the treatment-emergent period will be summarized regardless of the baseline level and according to the following baseline status categories:

- Normal/missing.
- Abnormal.

4.8 OTHER ANALYSES

4.8.1 PK analyses

Plasma concentrations will be summarized for amcenestrant, palbociclib and goserelin by treatment arm for each time point using descriptive statistics (such as the number of observations, arithmetic and geometric mean, median, standard deviation (SD), standard error (SE), coefficient of variation (CV)%, minimum, and maximum). A summary of predose concentrations over time will be presented graphically by treatment arm. These analyses will be performed by specific subgroups (eg, gender, BMI, age) if appropriate.

Amcenestrant and palbociclib plasma concentrations will be included in the descriptive statistics if actual sampling occurs in the following time windows: [2-4h] for samples at 3h post dose on Day 1 Cycle 1 and Cycle 2; within 2 hours before administration for predose samples.

All concentration values below the lower limit of quantitation (LLOQ) will be treated as zero in all summary statistics excepted for the geometric mean and associated coefficient of variation for which they will be considered as missing.

In case of exposure response analysis (See Section 4.8.4) using PK estimates is conducted, corresponding descriptive statistics of PK parameters of interest will be presented.

4.8.2 Quality of life analyses

Health-related quality of life (HRQL)

Symptoms and function related to health-related quality of life (HRQL) will be assessed using:

- The European Organisation for Research and Treatment of Cancer Core Quality of Life Questionnaire with 30 items (EORTC QLQ-C30) (version 3.0) (6, 7, 8).
- The European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire – Breast cancer module with 23 items (EORTC QLQ-BR23) or with 45 items (EORTC QLQ-BR45) (9, 10).
- The EuroQol measure with 5-dimensions and 5-levels per dimension (EQ-5D-5L) (11).

EORTC-QLQ-C30

The EORTC QLQ-C30 (version 3.0) is a 30-item cancer-specific patient-reported outcome measure of health-related quality of life (HRQL) in patients with cancer (6, 7, 8). The QLQC30 assesses disease-related symptoms, treatment-related symptoms, and functioning relevant to patients with cancer.

The QLQ-C30 is composed of both multi-item scales and single-item scales, including: five functional scales (physical functioning (5 items), role functioning (2 items), emotional functioning (4 items), cognitive functioning (2 items), and social functioning (2 items)); eight symptom scales (fatigue (3 items), nausea and vomiting (2 items), pain (2 items), dyspnoea (1 item), insomnia (1 item), appetite loss (1 item), constipation (1 item), diarrhoea (1 item)); a financial difficulties scale (1 item); and a global health status (GHS)/quality of life (QOL) scale (2 items). Items in the functional, symptom, and financial difficulties scales have a 4-point verbal rating scale ranging from “not at all” (1) to “very much” (4). Items in the GHS/QOL scale each have a 7-point numeric rating scale ranging from “very poor” (1) to “excellent” (7). The recall period for each scale is one week.

Each QLQ-C30 scale is scored according to the validated scoring algorithms available in the QLQ-C30 scoring manual, which standardize and transform the raw scores to a 0-100 range (8). Each scale is scored such that higher scores indicate higher levels of each construct. Higher scores on the functional scales indicates higher levels of functioning. Higher scores on the GHS/QOL scale indicates higher levels of GHS/QOL. High scores on the symptom and financial difficulties scales indicate higher levels of symptoms and financial difficulties, respectively.

Missing data will be handled in accordance with the scale developer’s rules (Section 5.10), (8).

EORTC QLQ-BR23

The EORTC QLQ breast cancer module (QLQBR23) is a 23-item disease-specific patient-reported outcome measure of disease-related symptoms, treatment-related symptoms, and functioning relevant to patients with breast cancer in patients with breast cancer. The QLQ-BR23 is used in conjunction with the QLQC30. The QLQBR23 assesses.

The QLQ-BR23 is composed of both multi-item scales and single-item scales, including: four functional scales (body image (4 items), sexual functioning (2 items), sexual enjoyment (1 item), and future perspective (1 item)); and four symptom scales (systemic therapy side effects (7 items), breast symptoms (4 items), arm symptoms (3 items), upset by hair loss (1 item)).

Each QLQ-BR23 scale is scored according to the validated scoring algorithms available in the QLQ-BR23 scoring manual, which standardize and transform the raw scores to a 0-100 range (8). Each scale is scored such that higher scores indicate higher levels of each construct. Higher scores on the functional scales indicates higher levels of functioning. High scores on the symptom scales indicate higher symptom levels.

Missing data will be handled in accordance with the scale developer's rules (Section 5.10), (8).

EORTC QLQ-BR45

The EORTC QLQ breast cancer module (QLQBR45) is a 45-item disease-specific patient-reported outcome measure of disease-related symptoms, treatment-related symptoms, and functioning relevant to patients with breast cancer (10). The QLQ-BR45 includes all 23 items from the QLQ-BR23 plus an additional 22 items assessing endocrine therapy symptoms (10 items), endocrine sexual symptoms (4 items), breast satisfaction (2 items), and skin/mucosis symptoms (6 items) (10). In countries where the translated and validated EORTC QLQ-BR45 questionnaire is not available, study participants can use EORTC QLQ-BR23 questionnaire until the availability of translated and validated EORTC QLQ-BR45 questionnaire.

Each QLQ-BR45 scale is scored according to the validated scoring algorithms available in the QLQ-BR45 scoring manual, which standardize and transform the raw scores to a 0-100 range (Section 5.10), (8, 10). Each scale is scored such that higher scores indicate higher levels of each construct. Higher scores on the functional scales indicates higher levels of functioning. High scores on the symptom scales indicate higher symptom levels.

Missing data will be handled in accordance with the scale developer's rules (Section 5.10), (8).

EQ-5D-5L

The EuroQoL questionnaire with 5 dimensions and 5 levels per dimension (EQ5D-5L) is a standardized measure of health status that provides a simple, generic measure of health utility, and consists of 2 sections: descriptive and visual analogue scale (VAS). The descriptive section consists of 5 dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension has 5 levels: no problems, slight problems, moderate problems, severe problems, and extreme problems. The VAS records the participant's self-rated health on a 20 cm vertical VAS with endpoints labeled 'the best health you can imagine' and 'the

worst health you can imagine'. This information can be used as a quantitative measure of health as judged by the individual participants.

The instrument is designed for selfcompletion by the participants. Response options are measured with a 5-point Likert scale with higher scores indicating better HRQL.

For EQ-5D-5L, VAS will be described as given by the participant and health utility index will be calculated according to EuroQoL-specific country algorithms/crosswalk developed by B. van Hout (11). In case a specific country algorithm is missing, the value sets based on the United Kingdom population will be used to generate health utility scores.

Statistical analyses

For each of the PRO/HRQL and health utility instruments (EORTC QLQ-C30, QLQ-BR45/QLQ-BR23, and EQ-5D-5L), patient reported outcomes endpoints will be analyzed in participants from the safety population who have completed the baseline and at least 1 postbaseline assessment.

For each questionnaire the compliance profile over time will be summarized on the safety population (number and percentage of forms received versus expected, and number and percentage of forms evaluable versus expected). A questionnaire is considered received if a least one item on the form is completed. A questionnaire is expected as defined in the SoA (Section 1.3 of the protocol) and based on the number of cycles started by the patient. A questionnaire is evaluable if at least one scale is complete. For multi-item scales, a scale is defined as complete when at least half of the scale items have responses (8). For single item scales, a scale is defined as complete when the item has a response (8). Reasons for non-completion will be summarized on the safety population.

For the QLQC30 (15 total scales), QLQ-BR45 (12 scales, 8 of which are QLQ-BR23 scales), and EQ-5D-5L (health index and visual analogue scale) instruments, descriptive statistics on the absolute value and changes from baseline will be done for each treatment arm at each time point, at EOT, and at the first post-treatment follow up visit.

For each scale within each instrument, statistical comparison between the two treatment arms will be based on a longitudinal repeated measures analysis using a mixed effects model on the change from baseline. The variables in the model will be treatment, time, treatment-by-time, with the baseline value and stratification factors used as covariates. Parameter estimates will be initially based on a restricted maximum likelihood method and an unstructured covariance matrix will be used. The structure of the correlation matrix will be investigated and simplified using likelihood ratio tests if appropriate. The differences in least square means between treatment and control group, and the corresponding 2-sided 95% CI at selected time points will be presented. No adjustments for multiple comparisons will be made. A graphical display of mean changes from baseline over time will also be provided for each scale of each instrument.

4.8.3 Biomarker analyses

4.8.3.1 Tumor biopsy biomarkers

Tumor biopsy tissue will be collected pre-treatment (or archival) and at EOT/progression. The expression of several proteins [eg, ER, progesterone receptor (PgR), B-Cell Lymphoma 2 (BCL2)] assessed by IHC will be described at baseline by considering the percentage of positive cells and/or the derived H-score. The relationship between baseline protein expression and efficacy endpoint (such as ORR, CBR or PFS) may be explored to assess potential prognostic and/or predictive effects.
The change from baseline to EOT/progression will be summarized per treatment arm.

Ribonucleic acid (RNA) may be also isolated from these tumor biopsies for additional exploratory analyses. Differential gene expression analyses may be performed, and some tumor gene signatures (such as the ER activation signature and other gene signature of interest) may be also derived and described.

4.8.3.2 cfDNA tumor and germline mutation analysis

cfDNA samples will be collected at baseline, Cycle 4 Day 1 predose and EOT/progression. From these samples, genomic aberrations from targeted gene panel will be identified (eg, single nucleotide variants, copy number variants, indels or fusion genes). For each gene, the prevalence of aberrations will be provided overall and possibly according to some baseline characteristics. For some genes of interest, the detail of the detected mutations will be provided. Descriptive statistics will be also used to describe the mutant allele frequency and concentration for single nucleotide variants and indels. The evolution of the genomic aberrations overtime (baseline, Cycle 4 Day 1 predose and EOT/progression) will be summarized per treatment arm.

Germline mutations from baseline normal tissue reference will be subtracted from mutations identified through cfDNA to identify somatic mutations.

The association between efficacy endpoints (such as ORR, CBR or PFS) and the mutation status at baseline for some key genes (such as ESR1 or PIK3CA, by comparing mutant versus wild-type) will be evaluated to assess potential prognostic and/or predictive effects.

4.8.3.3 Estradiol

Descriptive statistics of the circulating level of estradiol (such as the number of observations, arithmetic and geometric mean, median, standard deviation (SD), standard error (SE), coefficient of variation (CV)%, minimum, and maximum).will be provided at baseline and at Cycle 3 Day 1 for each treatment arm. The number of values below the limit of quantification (BLQ) will be summarized and imputed at the value LOQ/2 for descriptive statistics. The relationship between baseline circulating levels of estradiol and clinical response may be explored in the two treatment arms. On-treatment changes in circulating level of estradiol may be also associated with efficacy endpoints for exploratory purpose.

4.8.3.4 Genotyping analysis

The number of participants with available data for DMET genotyping will be reported. The frequencies of the different genotypes and phenotype will be tabulated by genes for UGT1A1 and UGT1A4.

4.8.4 Exposure response analysis

PK estimates may also be used to conduct the exploratory exposure-response analyses for safety (eg, incidence of AEs) and efficacy (eg, PFS, BOR, CBR, DOR). This analysis will be described in a separate analysis plan and reported in a stand-alone report.

4.8.5 Further therapy after discontinuation of investigational medicinal product administration during the study

A summary table will be provided for further therapies based on WHO-DD coding. Similar analysis will be performed for further radiotherapy and further surgery.

4.9 INTERIM ANALYSES

4.9.1 Interim analyses for PFS

Two interim analyses are planned based on the primary PFS endpoint at 40% (non-binding futility only) and 70% (efficacy only) of the planned total number of events expected. The stopping boundary for futility is based on the observed HR based on stratified Cox proportional hazard model, ie, an HR>1.1. The stopping boundary for efficacy will be derived based on the O'Brien and Fleming α-spending function and depend on the actual number of PFS events observed at the time of the interim analysis.

A summary of the PFS analyses is provided in Table 11. At IA 1 for futility, the observed HR based on Cox proportional hazard model stratified by the stratification factors as entered in the IRT system will be compared with the futility boundary. At IA2 for efficacy, the test statistic obtained from a logrank test procedure stratified by the stratification factors as entered in the IRT system will be compared with the efficacy boundary.

Demographics and baseline characteristics, prior or concomitant medication, extent of IMP exposure and compliance, AEs (TEAE, death, SAE, related TEAE, TEAE leading to discontinuation, TEAE leading to dose modification, AESI), and laboratory variables (abnormality of hematological and chemistry test) will be analyzed at interim analysis. If needed, more data will be analyzed to inform the futility decision including additional efficacy endpoints, PK, and biomarker.

Table 11 - PFS analyses

Analysis	Months after FPI (approx. under PFS HR=0.75)	Planned accrual	Number of events	Information fraction	Cumulative Power (under PFS HR=0.75)	Futility boundary	Efficacy boundary
PFS IA 1 (futility only)	19.5	1066	206	40%		HR >1.1	NA
PFS IA 2 (efficacy)	28	1066	361	70%	62%	NA	p ≤0.0074 (HR[a] ≤0.7736)
PFS Final analysis	40	1066	516	100%	90%	p >0.0228 (HR[a] >0.8386)	p ≤0.0228 (HR[a] ≤0.8386)

a HR is provided only for information purposes. The interim and final decisions will be based on p-values.
Note: number have been rounded. Calculations were made using East 6.5 software.
FPI = first participant in; HR = hazard ratio; IA = interim analysis; NA = not applicable; PFS = progression-free survival

In case of positive results at interim analysis, disease assessments data will continue to be collected according to the protocol until the final PFS analysis cut-off date (defined as the date when 516 PFS events assessed by radiologist/investigator are observed) and PFS results will be updated (non-inferential analysis only).

4.9.2 Interim analyses for OS

Comparison between arms on the OS will be performed only if the primary analysis of the PFS is statistically significant. Therefore, a maximum of three analyses are planned for OS, at the time of the primary analysis of PFS, at 75% of the planned number of OS events and at the final OS analysis.

A gamma error spending function (γ=-10) independent from the O'Brien and Fleming α-spending function for PFS will be used, along with the hierarchical testing strategy in order to strongly control the Family Wise Error Rate (FWER, overall Type I error rate). This guarantees the protection of the 2.5% FWER across hypotheses associated with PFS and OS and the repeated testing of the OS hypotheses at interim and the final analysis (12).

A summary of the OS analyses is provided in Table 12. The test statistic obtained from a logrank test procedure stratified by the stratification factors as entered in the IRT system will be compared with the futility boundary

Table 12 - OS analyses

Analysis	Months after FPI (approx.)	Planned accrual	Number of deaths (approx.)	Information fraction	Cumulative Power[a] (under HR=0.80)	Futility boundary	Efficacy boundary
Scenario 1: PFS is statistically significant at the PFS IA 2							
OS IA 1 (at PFS IA 2)	28	1066	228	36.1%	1.2%	NA	$p \leq 4.07\times10^{-5}$ (HR[b] ≤0.5934)
OS IA 2	55	1066	474	75%	32.9%	NA	p ≤0.0020 (HR[b] ≤0.7680)
Final analysis	80	1066	632	100%	80%	p >0.0248 (HR[b] >0.8554)	p ≤0.0248 (HR[b] ≤0.8554)
Scenario 2: PFS is statistically significant at the PFS final analysis							
OS IA 1 (at PFS final analysis)	40	1066	348	55.1%	8.5%	NA	$p \leq 2.78\times10^{-4}$ (HR[b] ≤0.6408)
OS IA 2	55	1066	474	75%	32.7%	NA	p ≤0.0020 (HR[b] ≤0.7673)
Final analysis	80	1066	632	100%	80%	p >0.0248 (HR[b] >0.8554)	p ≤0.0248 (HR[b] ≤0.8554)

a Marginal power conditional to statistically significance of PFS
b HR is provided only for information purposes. The interim and final decisions will be based on p-values.
Note: number have been rounded. Calculations were made using East 6.5 software. Assume an annual dropout rate of 1%.
FPI = first participant in; HR = hazard ratio; IA = interim analysis; NA = not applicable; OS = overall survival.

4.9.3 Data Monitoring Committee (DMC)

This study will use an independent DMC. The first DMC meeting will be set up to review early safety results (eg, after approximately 50 participants have completed at least 2 cycles, or after 6 months after first participant randomized), and then periodically. In addition to review of safety results, DMC will also evaluate efficacy at the interim analyses and make a recommendation regarding study continuation based on observed results of the study. Ad hoc DMC meetings may also be held if a significant safety issue or an issue deemed important for discussion arises on this or other amcenestrant studies. After each meeting, the DMC will make recommendations to the Sponsor's representatives regarding the continued safety of treating ongoing and future study participants, as well as the course of action regarding the conduct of the study. The DMC will also oversee the interim analyses on PFS detailed in Section 4.9.1.

5 SUPPORTING DOCUMENTATION

5.1 APPENDIX 1 LIST OF ABBREVIATIONS

ADI:	actual dose intensity
AE:	adverse event
AESIs:	adverse events of special interest
AI:	aromatase inhibitor
ALT:	alanine aminotransferase
AST:	aspartate aminotransferase
ATC:	anatomic category
BIRC:	blinded Independent Review Committee
BMI:	body mass index
BOR:	best overall response
CBR:	clinical benefit rate
cfDNA:	cell-free deoxyribonucleic acid
CI:	confidence interval
COD:	cut-off date
DOR:	duration of response
ECG:	electrocardiogram
ECOG:	Eastern Cooperative Oncology Group
eCRF:	electronic case report form
EDR:	early discrepancy rate
EORTC:	European Organization for Research and Treatment of Cancer
EOT:	end of treatment
EQ-5D-5L:	EuroQoL questionnaire with 5 dimensions and 5 levels per dimension
ER+:	estrogen receptor positive
ESR1:	estrogen receptor 1 gene
FWER:	family wise error rate
GGT:	gamma-glutamyltransferase
GHS:	global health status
HGLT:	high level group term
HR:	hazard ratio
IA:	interim analysis
ICF:	informed consent form
ICR:	independent comitee review
IHC:	immunohistochemistry
IMP:	investigational medicinal product
IRT:	interactive response technology
ITT:	intent-to-treat
LDR:	late discrepancy rate
LLT:	lower-level term
MedDRA:	medical dictionary for regulatory activities
NCI-CTCAE:	National cancer institute common terminology for adverse events

NIMP: noninvestigational medicinal product
ODR: outcome discrepancy rate
OS: overall survival
PCSA: potentially clinically significant abnormality
PD: progressive disease
PDI: planned dose intensity, planned dose intensity
PFS: progression free survival
PgR: progesterone receptor
PIK3CA: gene which encodes the p110alpha catalytic subunit of PI3K
PK: pharmacokinetic
PS: performance status
QLQ: quality of life questionnaire
RDI: relative dose intensity, relative dose intensity
RNA: ribonucleic acid
SAP: statistical analysis plan
SD: standard deviation
SDF: survival distribution function, survival distribution function
SERD: selective estrogen receptor down-regulator
SERM: selective estrogen receptor modulator
TE: treatment-emergent
TEAE: treatment-emergent adverse event
VAS: visual analogue scale
WHO-DD: World Health Organization-drug dictionary

5.2 APPENDIX 2 CHANGES TO PROTOCOL-PLANNED ANALYSES

This section summarizes major statistical changes in the protocol amendment(s).

Table 13 - Major statistical changes in protocol amendment(s)

Amendment Number	Approval Date	Changes	Rationale
01	27-Jul-2020	Pharmacodynamics analyses were removed, and genetics analyses were added	Pharmacodynamics analyses are not planned in study, and genetic analyses were missing.
02	30-Sep-2020	Updated the censoring rules for PFS analysis	The censoring rules have been detailed based on feedback from regulatory authorities.
02	30-Sep-2020	PFS assessed by the BIRC will be considered as supportive analysis of the primary PFS analysis even if a full BIRC is triggered	The strategy has been updated based on the feedback from regulatory authorities.

Amendment Number	Approval Date	Changes	Rationale
03	16-Dec-2020	Modification of statistical assumptions; modification of the study sample size from 708 to 1066 patients; addition of second interim analysis for OS; and change in study duration & estimated cut-off dates for PFS and OS analyses.	Revised PFS HR assumption used to power this endpoint, revised futility analysis strategy at interim analyses and revised dropout assumption due to change in PFS censoring scheme. Revised OS HR assumption used to power this endpoint.
03	16-Dec-2020	Addition of PFS2 as secondary endpoint and description of how the PFS2 will be assessed.	To assess the PFS after next line of therapy. PFS2 is a useful endpoint to assess any benefit beyond initial disease progression.
03	16-Dec-2020	Removed descriptions for China extension study including objective, population and sample size consideration.	In Nov 2020, an enrollment plan for China was adopted that targeted the completion of enrollment at the same time as global population. It is possible that an extension study is no longer needed. Therefore, relevant descriptions for sub-study are removed and flexible statement regarding the enrollment plan has been added. Consideration on patient number was also removed as all study participants from China may be treated within the global study. Specific sample size consideration will be provided in separate regulatory packages submitted to China National Medical Products Administration - Center for Drug Evaluation (NMPA-CDE).

5.3 APPENDIX 3 DEMOGRAPHICS AND BASELINE CHARACTERISTICS, PRIOR OR CONCOMITANT MEDICATIONS

Demographics, baseline characteristics, medical surgical history

The following demographics and baseline characteristics, medical and surgical history and disease characteristics at baseline will be summarized using descriptive statistics in the randomized population.

Demographic and baseline characteristics

- Age in years as quantitative variable and in categories (<65, 65 to <84, ≥85).
- Gender (Male, Female).
- Race (White, Black or African American, Asian, American Indian or Alaska Native, Native Hawaiian or other Pacific Island, Unknown).
- Ethnicity (Hispanic or Latino, not Hispanic or Latino, Unknown).
- Weight (kg).

- Eastern cooperative oncology group (ECOG) performance status.
- Menopausal status for females.

Baseline safety parameters (apart from those listed above) will be presented along with the safety summaries.

Medical (or surgical) history includes relevant history of previous or associated pathologies other than the tumor. Medical and surgical history will be coded to a LLT, PT, HLT, HLGT, and associated primary SOC using the MedDRA version currently in effect at Sanofi at the time of database lock.

Specific disease history at initial diagnosis includes time from initial diagnosis of breast cancer to randomization date (in years), histology, disease location and laterality, histopathology type and stage of the disease.

Specific disease status at study entry includes extent of the disease, number of organs involved, disease status, type of organ(s) involved, visceral metastasis or not, HER2 status, ER status, PgR status, disease status, ESR1 mutation status, PIK3CA mutation status.

Prior or concomitant medications

All medications will be coded using the World Health Organization-Drug Dictionary (WHO-DD) using the version currently in effect at Sanofi at the time of database lock.

- Prior medications are those the participant used prior to first IMP intake. Prior medications can be discontinued before first administration or can be ongoing during treatment period.
- Concomitant medications are any interventions received by the participant concomitantly to any IMP(s) from the first administration of IMP to the last IMP intake +30 days.
- Post-treatment medications are those the participant took in the period running from the end of the concomitant medications period up to the end of the study.
- A given medication can be classified as a prior medication and/or as a concomitant medication. If it cannot be determined whether a given medication was taken prior or concomitantly or post, it will be considered as prior and concomitant medication.

The prior and concomitant medications will be summarized for the randomized population, by anatomic and therapeutic level. The summaries will be sorted by decreasing frequency of anatomic category (ATC) based on overall incidence across treatment groups for prior medications, and incidence in the SAR439859 group for concomitant medications. In case of equal frequency, alphabetical order will be used. Participants will be counted once in each ATC category (anatomic or therapeutic) linked to the medication.

Anticancer therapies

Prior anticancer therapies include chemotherapy, hormonotherapy, immunotherapy, targeted therapy, surgery and radiotherapy.

The following variables will be collected/derived:

- Intent of prior anti-cancer therapy according to the following categories: neoadjuvant only, adjuvant only, both neoadjuvant and adjuvant, no prior treatment.
- Time from last relapse to randomization (in months).
- Intent of the last prior anti-cancer therapy according to the following categories: neoadjuvant only, adjuvant only, both neoadjuvant and adjuvant, no prior treatment.
- Type of prior anticancer therapy in neoadjuvant setting (targeted therapy, hormonotherapy, chemotherapy, immunotherapy or other as collected in eCRF).
- Type of prior anticancer therapy in adjuvant setting (targeted therapy, hormonotherapy, chemotherapy, immunotherapy or other as collected in eCRF).
- Time from start of adjuvant therapy to relapse in adjuvant setting (in years).
- Time from end of adjuvant therapy to relapse in adjuvant setting (in years).
- Duration of adjuvant therapy (in years).
- Among patients with prior adjuvant hormonotherapy
  - Number of patients with relapse <24 months after the start of adjuvant hormonotherapy,
  - Number of patients with relapse ≥24 months after the start and <12 months after the end of adjuvant hormonotherapy,
  - Number of patients with relapse ≥12 months after the end of adjuvant hormonotherapy.
- Among prior hormonotherapy:
  - Number of patients with intent: neoadjuvant only, adjuvant only, both neoadjuvant and adjuvant,
  - Type of prior hormonotherapy in neoadjuvant or adjuvant settings:
    - Aromatase inhibitors (AIs)
    - SERM (eg, Tamoxifen)
    - SERD (eg, Fulvestrant)
- Among prior chemotherapy:
  - Number of patients with intent: neoadjuvant only, adjuvant only, both neoadjuvant and adjuvant,
  - Type of prior chemotherapy in neoadjuvant or adjuvant settings:
    - Anthracyclins
    - Taxanes
    - Capecitabine
    - 5-fluorouracile

    - Other

- Among prior targeted therapy:
    - Number of patients with intent: neoadjuvant only, adjuvant only, both neoadjuvant and adjuvant,
    - Type of prior targeted therapy in neoadjuvant or adjuvant settings:
        - Anti-HER2
        - CDK4/6 inhibitors
        - Other
- Among prior immunotherapy:
    - Number of patients with intent: neoadjuvant only, adjuvant only, both neoadjuvant and adjuvant,
    - Type of prior immunotherapy in neoadjuvant or adjuvant settings:
        - Anti PD-1
        - Anti PD-L1
        - Other
    - Prior anti-cancer therapies in combination with endocrine therapy,
    - Type of prior endocrine-based combinations in neoadjuvant or adjuvant settings:
        - CDK4/6 inhibitors + AIs
        - AIs + SERM
        - Other
- Prior surgery: number (n, %) of patients with any prior surgery related to breast cancer, type of procedure (Preferred Term) and time from the last surgery to the randomization date (months).
- Prior radiotherapy: number (n, %) of patients with any prior external radiotherapy related to breast cancer, intent, intent of last prior external radiotherapy, time from the last external radiotherapy to the randomization date (months) overall and by intent (curative and palliative) and location of prior external radiation therapy by intent. Prior internal radiotherapy will be listed.

Further therapies after discontinuation of intervention will be summarized based on WHO-DD coding.

5.4 APPENDIX 4 DATA HANDLING CONVENTIONS

5.4.1 General conventions

The following formulas will be used for computation of parameters.

Time unit

A month length is 30.4375 days (365.25 / 12). If duration is to be reported in months, duration in days is divided by 30.4375. If duration is to be reported in years, duration in days will be divided by 365.25.

Duration

Unless otherwise specified, difference between two dates (Date A ≤ Date B) will be calculated as follows:

$$\text{Duration (days)} = \text{Date B} - \text{Date A} + 1$$

5.4.2 Data handling conventions for primary efficacy variables

The following formulas will be used for computation of PFS endpoint.

Date of tumor assessment

It is acknowledged that an assessment may include several methods of evaluation performed over a period of several days within a window of time around an expected assessment date. For each tumor assessment, a single date will be derived according to the overall response of that assessment.

The date of target lesion(s) assessment will be derived based on the date of the last assessment of target lesion(s).

When the overall response is different from PD, the date of tumor assessment is defined as the date of the last evaluation included in the series of evaluations performed within that time point (ie, target lesion(s), non-target lesions(s) and new lesion(s) assessments).

When the overall response is PD, the date of tumor assessment is the date when progression was first demonstrated according to the target lesion(s), non-target lesion(s) and new lesion(s), as specified below:

- For progression based on new lesion(s) the date of progression is the earliest date a new lesion has been detected.
- For progression based on non-target lesion(s), the date of progression is the earliest date a non-target lesion was considered as PD.
- For progression based on target lesion(s), the date of progression is the date of the last assessment of target lesion(s).

If progression is based on several events within the same tumor assessment (eg, new lesion(s) seen along with target lesion(s) progression), the earliest date of progression, according to the rules listed above, will be the date of assessment.

Evaluable tumor assessment

An evaluable tumor assessment is defined as a tumor assessment with an overall response different of non-evaluable (NE).

Date of documented progression

The date of documented progression is defined as the first date of tumor assessment at which the overall response was recorded as progressive disease.

Date of death

The date of death is defined as the date of death recorded in the eCRF.

Date of next scheduled assessment

The date of next scheduled assessment is the date of next theoretical tumor assessment per protocol after the last evaluable tumor assessment. This date may be used if back-dating is considered when the event occurred beyond the acceptable time window for the next tumor assessment (eg, 1 or 2 non-evaluable tumor assessment).

Date of last evaluable tumor assessment

The date of the last evaluable tumor assessment is defined as the last date of tumor assessment at which the overall response was recorded as CR, PR, SD or Non-CR/Non-PD before a censoring reason occurred.

5.4.3 Data handling conventions for secondary efficacy variables

The following formulas will be used for computation of secondary endpoint.

Best relative change from baseline in tumor size

Tumor size is defined as the sum of the longest diameters of the target lesions as per RECIST 1.1. It can be calculated only for measurable patients.

Relative change from baseline in tumor size at tumor assessment *t* will be calculated as follows:

> Relative change (%) from baseline in tumor size (*t*) = 100*(Tumor Size(*t*) – Tumor Size(*baseline*))/ Tumor Size(*baseline*)

Best relative change from baseline in tumor size will be the smallest relative change (up to documented progression, death, further anticancer therapy or COD, whichever occurs first) from baseline in tumor size.

Date of first documented response

The date of the first documented response is defined as the date of the first tumor assessment at which the overall response was recorded as CR or PR that is subsequently confirmed.

Date of last contact

The last contact date is derived for patients not known to have died at the analysis cut-off date based on the latest date among the following:

- Date of visits.
- Assessment dates (eg, laboratory, vital signs, ECOG performance status, ECG, tumor assessment, PK assessment, EOT completion etc).
- Medication and procedures dates including study medication, concomitant medications, surgical and medical procedures, further anti-cancer therapies administered after treatment discontinuation.
- Adverse event start and end dates.
- "Date of Last Available Information" collected on the "Subject status" page.
- Study treatment start/end date.
- Randomization date.

The last contact date is defined as the latest date from the above list or the cut-off date, whichever comes first. The last contact date could be used for censoring of patients in time to event analysis.

5.4.4 Missing data

The analyses and summaries of continuous and categorical variables will be based on observed data only. Percentages will be calculated using as denominator the number of patients with non-missing observation in the considered population. When relevant, the number of patients with missing data is presented.

Handling of disease characteristics missing/partial dates

- If the day is missing, it will be imputed by 1.
- If the month is missing, it will be imputed by 1 (only for medical history variables).
- If the year is missing, no imputation will be performed.

Incomplete date of cancer diagnosis:

- If the day of the cancer diagnosis is missing, the date will be imputed to the first day of the month.
- If day and month of the cancer diagnosis are missing, no imputation will be done.

Handling of medication missing/partial dates

No imputation of medication (other than anti-cancer therapies) start/end dates or times will be performed. If a medication date or time is missing or partially missing and it cannot be determined whether it was taken prior or concomitantly, it will be considered a prior, concomitant, and post-treatment medication.

For prior anti-cancer therapies, following rules will be applied:

- Incomplete start date of prior anti-cancer therapy: if the day of the start date of the prior regimen is missing, the date will be imputed to the first day of the month; if the month is missing, the date will be imputed to the first month of the year.
- Incomplete end date of prior anti-cancer therapy: if the day of the end date of last prior regimen is missing, the date will be imputed to the last day of the month; if the month is missing, the date will be imputed to the last month of the year.

Imputation of incomplete date for post anti-cancer treatment start date

For further anti-cancer treatments, if the medication start date is missing, it will be imputed as follows:

- If the medication start day and month are missing and the medication start year is the same as treatment end year, the medication start date will be set equal to treatment end date + 1.
- If the medication start day and month are missing and the medication start year is after the treatment end year, the medication start day and month will each be set to 01.
- If the medication start day is missing and medication start year and month is the same as the treatment end year and month, the medication start date will be set to the treatment end date + 1.
- If the medication start day is missing and medication start month is before the treatment end month and the medication start year is the same as treatment end year, the medication start day will be set to 01.
- If the medication start day is missing and the medication start month is after the treatment end month and the medication start year is the same as treatment end year, the medication start day will be set to 01.
- If the medication start day is missing and the medication start month is not missing and the medication start year is after the treatment end year, the medication start day will be set to 01.
- If the medication start day, start month and start year is missing, the medication start date will be set equal to the treatment end date + 1.

Handling of other missing dates

Incomplete date of progression for the last prior regimen:

- If the day of the progression for the last prior regimen is missing, the date will be imputed to the first day of the month.
- If day and month of the progression for the last prior regimen are missing, no imputation will be done.

Incomplete date of prior surgery:

- If the day of the prior surgery is missing, the date will be imputed to the end day of the month.

- If day and month of the prior surgery are missing, no imputation will be done.

Incomplete end date of prior radiotherapy:

- If the day of the end date of the prior radiotherapy is missing, the date will be imputed to the end day of the month.
- If the day and month of the end date of the prior radiotherapy are missing, no imputation will be done.

Handling of adverse events when date and time of first investigational medicinal product administration is missing

When the date and time of the first IMP administration is missing, all adverse events that occurred on or after the day of randomization should be considered as treatment-emergent adverse events.

Missing grade

If the grade is missing for one of the treatment emergent occurrences of an AE, the maximal severity on the remaining occurrences will be considered. If the grade is missing for all the occurrences, no imputation will be done and missing grades will be summarized in the "all grades" category.

Handling of missing assessment of relationship of adverse events to investigational medicinal product

If the assessment of the relationship to IMP is missing, then the relationship to IMP has to be assumed and the adverse event considered as such in the frequency tables of possibly related adverse events, but no imputation should be done at the data level.

Handling of potentially clinically significant abnormalities

If a patient has a missing baseline, he/she will be grouped in the category "normal/missing at baseline."

For PCSAs with 2 conditions, one based on a change from baseline value or a normal range and the other on a threshold value, with the first condition being missing, the PCSA will be based only on the second condition.

For a PCSA defined on a threshold and/or a normal range, this PCSA will be derived using this threshold if the normal range is missing; eg, for eosinophils the PCSA is >0.5 GIGA/L or >ULN if ULN ≥0.5 GIGA/L. When ULN is missing, the value 0.5 should be used.

Measurements flagged as invalid by the laboratory will not be summarized or taken into account in the computation of PCSA values.

Handling of Estradiol data below quantification limit

For Estradiol data with values below the quantification limit, the data will be imputed to half of the quantification limit.

5.4.5 Unscheduled visits

Unscheduled visit measurements of laboratory data, vital signs, and ECG will be used for computation of baseline and worst values and/or grades.

5.4.6 Pooling of centers for statistical analyses

Data from all sites will be pooled together for analyses.

5.4.7 Statistical technical issues

Not applicable.

5.5 APPENDIX 5 POTENTIALLY CLINICALLY SIGNIFICANT ABNORMALITIES CRITERIA

Parameter	PCSA	Comments
Vital signs		
HR	≤50 bpm and decrease from baseline ≥20 bpm ≥120 bpm and increase from baseline≥20 bpm	To be applied for all positions (including missing) except STANDING.
SBP	≤95 mmHg and decrease from baseline ≥20 mmHg ≥160 mmHg and increase from baseline ≥20 mmHg	To be applied for all positions (including missing) except STANDING.
DBP	≤45 mmHg and decrease from baseline ≥10 mmHg ≥110 mmHg and increase from baseline ≥10 mmHg	To be applied for all positions (including missing) except STANDING.
Weight	≥5% increase from baseline ≥5% decrease from baseline	FDA Feb 2007.
Laboratory		
Blood Urea Nitrogen	≥17 mmol/L	
Chloride	<80 mmol/L >115 mmol/L	
Basophils	>0.1 Giga/L	
Monocytes	>0.7 Giga/L	
RBC	≥6 Tera/L	
Hematocrit	≤0.37 v/v (Male); ≤0.32 v/v (Female) ≥0.55 v/v (Male); ≥0.5 v/v (Female)	

5.6 APPENDIX 6 SUMMARY OF STATISTICAL ANALYSES

EFFICACY ANALYSIS

Endpoint	Analysis population	Primary analysis	Sensitivity analysis	Supportive analysis	Subgroup analysis	Other analyses
Primary endpoint						
PFS according to investigator	ITT	Stratified Log-rank test Kaplan-Meier Cox proportional hazards model	#1 PFS considering events occurring after two or more non-evaluable tumor assessment as event and ignoring further anti-cancer therapy #2 PFS considering events occurring after two or more non-evaluable tumor assessment as event and back-dating at the next scheduled assessment #3 PFS analysis using stratification factors derived from eCRF data	- Robustness of LE-PFS treatment effect, evaluated with the NCI (National Cancer Institute) method with BIRC data - PFS according to BIRC assessments - PFS considering the clinical/non-radiological progression as event, and same rules as the sensitivity analyses if relevant) - Concordance of PFS outcome (PFS ODR) - Differential discordance on PD events (EDR and LDR) - RMST if proportional hazard assumption is not fulfilled	Yes	- Multivariate Cox proportional hazards model - PFS considering clinical/non-radiological progression as event
Secondary endpoints						
OS	ITT	Stratified Log-rank test Kaplan-Meier Cox proportional hazards model	No	IPCW and RPSFTM to account for informative censoring due to further anti-cancer therapy, if relevant	Yes	Multivariate Cox proportional hazards model
ORR according to investigator	ITT ITT measurable	Descriptive statistics Clopper-Pearson 95% CI	No		Yes if relevant based on subgroups of primary endpoint	- Best Overall Response - Best relative change from baseline in tumor size (waterfall plot)

Endpoint	Analysis population	Primary analysis	Sensitivity analysis	Supportive analysis	Subgroup analysis	Other analyses
DOR according to investigator	ITT (responders) ITT measurable (responders)	Kaplan-Meier	No		Yes if relevant based on subgroups of primary endpoint	
CBR according to investigator	ITT ITT measurable	Descriptive statistics Clopper-Pearson 95% CI	No		Yes if relevant based on subgroups of primary endpoint	
PFS2 according to investigator	ITT	Kaplan-Meier Cox proportional hazards model	No		Yes if relevant based on subgroups of primary endpoint	
Pharmacokinetics of SAR439859 and Palbociclib	Pharmacokinetic-evaluable population	Descriptive statistics	No	No	Yes if relevant	Safety/efficacy endpoints according to PK parameters
Time to first use of chemotherapy	ITT	Kaplan-Meier Cox proportional hazards model	No	No	No	
Tertiary/exploratory endpoints						
Tumor biomarkers	ITT	Descriptive statistics				
Mutation profile	ITT	Descriptive statistics				Relationship between the mutation status and clinical response
PFS according to PIK3CA and ESR1 mutation status at baseline based on investigator assessment	ITT	Kaplan-Meier Cox proportional hazards model	No	PFS according to PIK3CA and ESR1 mutation status at baseline based on BIRC assessment, if full BIRC is triggered	No	

BOR: Best overall response, CI: Confidence interval, CR: Complete response, EDR: Early discrepancy rate, ITT: intent-to-treat, LDR: Late discrepancy rate, ODR: Outcome discrepancy rate, ORR: Overall response rate, OS: overall survival, PFS: Progression free survival, PR: Partial response

Stratified analyses are performed with the stratification factors as entered in the IRT ie, presence of visceral metastasis, prior treatment with CDK4/6 inhibitors and ECOG.

SAFETY ANALYSES

Endpoint	***Analysis population***	***Primary analysis***	***Supportive analysis***	***Subgroup analysis***
Adverse events	Safety	Descriptive statistics by treatment group	Descriptive statistics by treatment arm	No
Deaths	Safety	Descriptive statistics by treatment group	Descriptive statistics by treatment arm	No
Laboratory	Safety	Descriptive statistics by treatment group	Descriptive statistics by treatment arm	No
Vital signs	Safety	Descriptive statistics by treatment group	Descriptive statistics by treatment arm	No
ECG	Safety	Descriptive statistics by treatment group	Descriptive statistics by treatment arm	No
Health-related quality of life (HRQL): EORTC QLQ-C30, EORTC QLQ-BR23/EORTC-BR45, EQ-5D-5L	Safety population who have completed the baseline and at least 1 postbaseline assessment	Descriptive statistics Mixed effects model on the change from baseline for each of the scales of each instrument	No	No

5.7 APPENDIX 7 INTERNATIONALLY AGREED SOC ORDER

The internationally agreed order (Guideline on summary of product characteristics, December 1999, European commission) for SOC:

1. Infections and infestations
2. Neoplasms benign and malignant (including cysts and polyps)
3. Blood and the lymphatic system disorders
4. Immune system disorders
5. Endocrine disorders
6. Metabolism and nutrition disorders
7. Psychiatric disorders
8. Nervous system disorders
9. Eye disorders
10. Ear and labyrinth disorders
11. Cardiac disorders
12. Vascular disorders
13. Respiratory, thoracic and mediastinal disorders
14. Gastrointestinal disorders
15. Hepato-biliary disorders
16. Skin and subcutaneous tissue disorders
17. Musculoskeletal, connective tissue and bone disorders
18. Renal and urinary disorders
19. Pregnancy, puerperium and perinatal conditions
20. Reproductive system and breast disorders
21. Congenital and familial/genetic disorders
22. General disorders and administration site conditions
23. Investigations
24. Injury and poisoning
25. Surgical and medical procedures
26. Social circumstances
27. Product Issues

The other terms are sorted by dictionary code order.

5.8 APPENDIX 8 RESPONSE EVALUATION CRITERIA IN SOLID TUMORS (RECIST 1.1)

Details provided in bibliographic reference (13).

Definitions

At baseline, tumor lesions/lymph nodes will be categorized measurable or nonmeasurable as follows:

Measurable

- Tumor lesions: Must be accurately measured in at least one dimension (longest diameter in the plane of measurement is to be recorded) with a minimum size of:
  - Ten millimeters by computed tomography (CT) scan (CT scan slice thickness no greater than 5 mm),
  - Ten-millimeter caliper measurement by clinical examination (lesions which cannot be accurately measured with calipers should be recorded as nonmeasurable),
  - Twenty millimeters by chest X-ray.
- Malignant lymph nodes: To be considered pathologically enlarged and measurable, a lymph node must be ≥15 mm in short axis when assessed by CT scan (CT scan slice thickness recommended to be no greater than 5 mm). At baseline and in follow-up, only the short axis will be measured and followed (see Special Issue 15 [13]). See also notes below on “Baseline documentation of target and nontarget lesions” for information on lymph node measurement.

Non-measurable

- All other lesions, including small lesions (longest diameter <10 mm or pathological lymph nodes with 10 to <15 mm short axis) as well as truly nonmeasurable lesions. Lesions considered truly nonmeasurable include: leptomeningeal disease, ascites, pleural or pericardial effusion, inflammatory breast disease, lymphangitic involvement of skin or lung, abdominal masses/abdominal organomegaly identified by physical examination that is not measurable by reproducible imaging techniques.

Special considerations regarding lesion measurability

Bone lesions, cystic lesions, and lesions previously treated with local therapy require particular comment:

Bone lesions:

- Bone scan, positron emission tomography (PET) scan or plain films are not considered adequate imaging techniques to measure bone lesions. However, these techniques can be used to confirm the presence or disappearance of bone lesions.

- Lytic bone lesions or mixed lytic-blastic lesions, with identifiable soft tissue components, that can be evaluated by cross sectional imaging techniques such as CT or magnetic resonance imaging (MRI) can be considered as measurable lesions if the soft tissue component meets the definition of measurability described above.
- Blastic bone lesions are nonmeasurable.

Cystic lesions:

- Lesions that meet the criteria for radiographically defined simple cysts should not be considered as malignant lesions (neither measurable nor nonmeasurable) since they are, by definition, simple cysts.
- 'Cystic lesions' thought to represent cystic metastases can be considered as measurable lesions, if they meet the definition of measurability described above. However, if noncystic lesions are present in the same participant, these are preferred for selection as target lesions.

Lesions with prior local treatment:

- Tumor lesions situated in a previously irradiated area, or in an area subjected to other locoregional therapy, are usually not considered measurable unless there has been demonstrated progression in the lesion. Study protocols should detail the conditions under which such lesions would be considered measurable.

Methods of measurement

- Measurement of lesions:
  All measurements should be recorded in metric notation, using calipers if clinically assessed. All baseline evaluations should be performed as close as possible to the treatment start and never more than 4 weeks before the beginning of the treatment.
- Method of assessment:
  The same method of assessment and the same technique should be used to characterize each identified and reported lesion at baseline and during follow-up. Imaging based evaluation should always be done rather than clinical examination unless the lesion(s) being followed cannot be imaged but are assessable by clinical examination. Clinical lesions: Clinical lesions will only be considered measurable when they are superficial and 10 mm diameter as assessed using calipers (eg, skin nodules). For the case of skin lesions, documentation by color photography including a ruler to estimate the size of the lesion is suggested. As noted above, when lesions can be evaluated by both clinical examination and imaging, imaging evaluation should be undertaken since it is more objective and may also be reviewed at the end of the study.
- Chest X-ray: Chest CT is preferred over chest X-ray, particularly when progression is an important endpoint, since CT is more sensitive than X-ray, particularly in identifying new lesions. However, lesions on chest X-ray may be considered measurable if they are clearly defined and surrounded by aerated lung.

- Computed tomography, MRI: CT is the best currently available and reproducible method to measure lesions selected for response assessment. This guideline has defined measurability of lesions on CT scan based on the assumption that CT slice thickness is 5 mm or less. When CT scans have slice thickness greater than 5 mm, the minimum size for a measurable lesion should be twice the slice thickness. Magnetic resonance imaging is also acceptable in certain situations (eg, for body scans).
- Ultrasound: Ultrasound is not useful in assessment of lesion size and should not be used as a method of measurement.
- Endoscopy, laparoscopy: The utilization of these techniques for objective tumor evaluation is not advised. However, they can be useful to confirm complete pathological response when biopsies are obtained or to determine relapse in trials where recurrence following complete response or surgical resection is an endpoint.
- Tumor markers: Tumor markers alone cannot be used to assess objective tumor response. If markers are initially above the upper normal limit, however, they must normalize for a participant to be considered in complete response. Specific guidelines for both cancer antigen 125 response (in recurrent ovarian cancer) and prostate-specific antigen response (in recurrent prostate cancer), have been published. In addition, the Gynecologic Cancer Intergroup has developed cancer antigen 125 progression criteria which are to be integrated with objective tumor assessment for use in first-line trials in ovarian cancer.
- Cytology, histology: These techniques can be used to differentiate between partial response (PR) and complete response (CR) in rare cases if required by protocol (eg, residual lesions in tumor types such as germ cell tumors, where known residual benign tumors can remain). When effusions are known to be a potential adverse effect of treatment (eg, with certain taxane compounds or angiogenesis inhibitors), the cytological confirmation of the neoplastic origin of any effusion that appears or worsens during treatment can be considered if the measurable tumor has met criteria for response or stable disease (SD) in order to differentiate between response (or SD) and progressive disease.

Tumor response evaluation

To assess objective response or future progression, it is necessary to estimate the overall tumor burden at baseline and use this as a comparator for subsequent measurements. Only participants with measurable disease at baseline should be included in protocols where objective tumor response is the primary endpoint. Measurable disease is defined by the presence of at least 1 measurable lesion. In studies where the primary endpoint is tumor progression (either time to progression or proportion with progression at a fixed date), the protocol must specify if entry is restricted to those with measurable disease or whether participants having non measurable disease only are also eligible.

Response criteria

Table 14 - Response criteria, evaluation of target lesions

Complete Response (CR)	Disappearance of all target lesions. Any pathological lymph nodes (whether target or nontarget) must have reduction in short axis to <10 mm.
Partial Response (PR):	At least a 30% decrease in the sum of diameters of target lesions, taking as reference the baseline sum diameters.
Progressive Disease (PD):	At least a 20% increase in the sum of diameters of target lesions, taking as reference the smallest sum on study (this includes the baseline sum if that is the smallest on study). In addition to the relative increase of 20%, the sum must also demonstrate an absolute increase of at least 5 mm. (Note: the appearance of one or more new lesions is also considered progression).
Stable Disease (SD):	Neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD, taking as reference the smallest sum diameters while on study

Table 15 - Response criteria, evaluation of nontarget lesions

Complete Response (CR):	Disappearance of all nontarget lesions and normalization of tumor marker level. All lymph nodes must be nonpathological in size (<10 mm short axis).
Non-CR/Non-PD:	Persistence of one or more nontarget lesion(s) and/or maintenance of tumor marker level above the normal limits.
Progressive Disease (PD):	Unequivocal progression (see comments below)[a] of existing nontarget lesions. (Note: the appearance of one or more new lesions is also considered progression).

a Although a clear progression of "nontarget" lesions only is exceptional, in such circumstances, the opinion of the treating physician should prevail and the progression status should be confirmed later on by the review panel (or study chair).

Evaluation of best overall response

It is assumed that at each protocol specified time point, a response assessment occurs. The following table (Table 16) provides a summary of the overall response status calculation at each time point for participants who have measurable disease at baseline. When patients have non-measurable (therefore non-target) disease only, Table 17 is to be used.

Table 16 - A summary of overall response status for measurable disease

Target lesions	Non-target lesions	New lesions	Overall response
CR	CR	No	CR
CR	Non-CR/non-PD	No	PR
CR	Not evaluated	No	PR
PR	Non-PD or not all evaluated	No	PR
SD	Non-PD or not all evaluated	No	SD
Not all evaluated	Non-PD	No	NE
PD	Any	Yes or No	PD
Any	PD	Yes or No	PD
Any	Any	Yes	PD

CR = complete response; PD = progressive disease; PR = partial response; SD = stable disease.

Table 17 - A summary of overall response status for non-measurable disease

Non-target lesions	New lesions	Overall response
CR	No	CR
Non-CR/non-PD	No	Non-CR/non-PD[a]
Not all evaluated	No	NE
Unequivocal PD	Yes or No	PD
Any	Yes	PD

CR = complete response; PD = progressive disease; PR = partial response; SD = stable disease.

a 'Non-CR/non-PD' is preferred over 'stable disease' for non-target disease since SD is increasingly used as endpoint for assessment of efficacy in some trials so to assign this category when no lesions can be measured is not advised.

Table 18 - Evaluation of best overall response

Overall response First time point	Overall response Subsequent time point	BEST overall response
CR	CR	CR
CR	PR	SD, PD or PR[a]
CR	SD	SD provided minimum criteria for SD duration met, otherwise, PD
CR	PD	SD provided minimum criteria for SD duration met, otherwise, PD
CR	NE	SD provided minimum criteria for SD duration met, otherwise NE
PR	CR	PR
PR	PR	PR
PR	SD	SD
PR	PD	SD provided minimum criteria for SD duration met, otherwise, PD
PR	NE	SD provided minimum criteria for SD duration met, otherwise NE
NE	NE	NE

a If a CR is truly met at first time point, then any disease seen at a subsequent time point, even disease meeting PR criteria relative to baseline, makes the disease PD at that point (since disease must have reappeared after CR). Best response would depend on whether minimum duration for SD was met. However, sometimes 'CR' may be claimed when subsequent scans suggest small lesions were likely still present and in fact the participant had PR, not CR at the first time point. Under these circumstances, the original CR should be changed to PR and the best response is PR.

CR = complete response; NE = not evaluable; PD = progressive disease; PR = partial response; SD = stable disease

The best overall response is determined once all the data for the participant is known.

When SD is believed to be best response, it must also meet the protocol specified minimum time from baseline (defined as 63 days). If the minimum time is not met when SD is otherwise the best time point response, the participant's best response depends on the subsequent assessments. For example, a participant who has SD at first assessment, PD at second and does not meet minimum duration for SD, will have a best response of PD. The same participant lost to follow-up after the first SD assessment would be considered inevaluable. Complete or partial responses may be claimed only if the criteria for each are met at a subsequent time point as specified in the protocol (generally 4 weeks later). In this circumstance, the best overall response can be interpreted as in Table 18.

5.9 APPENDIX 9 DESCRIPTION OF CENSORING AND EVENT RULES FOR PRIMARY AND SENSITIVITY ANALYSES OF PFS

Table 19 - PFS Primary analysis

Situation	Date of outcome	Outcome	Category
No baseline tumor assesment[b]	Date of randomization	Censored	No baseline tumor assessments
No evaluable post-baseline tumor assessments[a]	Date of randomization	Censored	No evaluable post-baseline tumor assessments
Alive and no documented progression	Date of the last evaluable tumor assessment documenting no progression	Censored	Alive without progression
Documented progression (or Death) at or between scheduled visits	Date of progression (or Date of death)	Event	Documented progression (or Death without documented progression)
Documented progression (or Death) occurring after exactly one non-evaluable tumor assessment	Date of progression (or Date of death)	Event	Documented progression (or Death without documented progression)
Documented progression (or Death) occurring after two or more non-evaluable tumor assessment[b]	Date of the last evaluable tumor assessment documenting no progression	Censored	Event occurred after two or more missed tumor assessment
Clinical/non-radiological progression	Ignored	Ignored	
Initiation of further anti-cancer therapy	Date of the last evaluable tumor assessment before start date of further anti-cancer therapy (if any), date of randomization otherwise	Censored	Initiation of further anti-cancer therapy

a Except if the patient dies within 25 weeks after the date of randomization in which case a death event outcome will be considered for the PFS with date of outcome corresponding to the date of death

b An event occurring at least 26 weeks (excluded) after the last evaluable tumor assessment, documenting no progression or randomization date, whichever is last. 26 weeks corresponds to twice the time between two disease assessments per protocol (every 2*12 weeks), plus the 7-day window before and after.

Table 20 - PFS sensitivity analysis #1 (PFS considering events occurring after two or more non-evaluable tumor assessment as event and ignoring further anticancer therapy)

Situation	Date of outcome	Outcome	Category
No baseline tumor assessments[a]	Date of randomization	Censored	No baseline tumor assessments
No evaluable post-baseline tumor assessments[a]	Date of randomization	Censored	No evaluable post-baseline tumor assessments
Alive and no documented progression	Date of the last evaluable tumor assessment documenting no progression	Censored	Alive without progression
Documented progression (or Death) at or between scheduled visits	Date of progression (or Date of death)	Event	Documented progression (or Death without documented progression)
Documented progression (or Death) occurring after exactly one non-evaluable tumor assessment	Date of progression (or Date of death)	Event	Documented progression (or Death without documented progression)
Documented progression (or Death) occurring after two or more non-evaluable tumor assessment[b]	Date of progression (or Date of death)	Event	Documented progression (or Death without documented progression)
Clinical/non-radiological progression	Ignored	Ignored	
Initiation of further anti-cancer therapy	Ignored	Ignored	

a Except if the patient dies in which case a death event outcome will be considered for the PFS with date of outcome corresponding to the date of death

b An event occurring at least 26 weeks (excluded) after the last evaluable tumor assessment, documenting no progression or randomization date, whichever is last. 26 weeks corresponds to twice the time between two disease assessments per protocol (every 2*12 weeks), plus the 7-day window before and after.

Table 21 - PFS sensitivity analysis #2 (PFS considering events occurring after two or more non-evaluable tumor assessment as event and back-dating at the next scheduled assessment)

Situation	Date of outcome	Outcome	Category
No baseline tumor assessments[a]	Date of randomization	Censored	No baseline tumor assessments
No evaluable post-baseline tumor assessments[a]	Date of randomization	Censored	No evaluable post-baseline tumor assessments
Alive and no documented progression	Date of the last evaluable tumor assessment documenting no progression	Censored	Alive without progression
Documented progression (or Death) at or between scheduled visits	Date of progression (or Date of death)	Event	Documented progression (or Death without documented progression)
Documented progression (or Death) occurring after exactly one non-evaluable tumor assessment	Date of the next scheduled assessment	Event	Documented progression (or Death without documented progression)
Documented progression (or Death) occurring after two or more non-evaluable tumor assessment[b]	Date of the next scheduled assessment	Event	Documented progression (or Death without documented progression)
Clinical/non-radiological progression	Ignored	Ignored	
Initiation of further anti-cancer therapy	Date of the last evaluable tumor assessment before start date of further anti-cancer therapy (if any), date of randomization otherwise	Censored	Initiation of further anti-cancer therapy

a Except if the patient dies in which case a death event outcome will be considered for the PFS with date of outcome corresponding to the date of death

b An event occurring at least 26 weeks (excluded) after the last evaluable tumor assessment, documenting no progression or randomization date, whichever is last. 26 weeks corresponds to twice the time between two disease assessments per protocol (every 2*12 weeks), plus the 7-day window before and after.

Table 22 - PFS supplementary analysis for investigator assessment (taking clinical/non-radiological progression as event)

Situation	Date of outcome	Outcome	Category
No baseline tumor assessments[a]	Date of randomization	Censored	No baseline tumor assessments
No evaluable post-baseline tumor assessments[a]	Date of randomization	Censored	No evaluable post-baseline tumor assessments
Alive and no documented progression	Date of the last evaluable tumor assessment documenting no progression	Censored	Alive without progression
Documented progression (or Death) at or between scheduled visits	Date of progression (or Date of death)	Event	Documented progression (or Death without documented progression)
Documented progression (or Death) occurring after exactly one non-evaluable tumor assessment	Date of progression (or Date of death)	Event	Documented progression (or Death without documented progression)
Documented progression (or Death) occurring after two or more non-evaluable tumor assessment[b]	Date of the last evaluable tumor assessment documenting no progression	Censored	Event occurred after two or more missed tumor assessment
Clinical/non-radiological progression	Date of clinical/non-radiological progression	Event	Non-documented progression
Initiation of further anti-cancer therapy	Date of the last evaluable tumor assessment before start date of further anti-cancer therapy (if any), date of randomization otherwise	Censored	Initiation of further anticancer therapy

a Except if the patient dies within 25 weeks after the date of randomization in which case a death event outcome will be considered for the PFS with date of outcome corresponding to the date of death

b An event occurring at least 26 weeks (excluded) after the last evaluable tumor assessment, documenting no progression or randomization date, whichever is last. 26 weeks corresponds to twice the time between two disease assessments per protocol (every 2*12 weeks), plus the 7-day window before and after.

5.10 APPENDIX 10 EORTC QLQ-C30, QLQ-BR23, AND QLQ-BR45 ITEMS, SCALES AND SCORES

For QLQ-C30:

	Scale	Number of items	Item range*	**Version 3.0** Item numbers	Function scales
Global health status / QoL					
Global health status/QoL (revised)†	QL2	2	6	29, 30	
Functional scales					
Physical functioning (revised)†	PF2	5	3	1 to 5	F
Role functioning (revised)†	RF2	2	3	6, 7	F
Emotional functioning	EF	4	3	21 to 24	F
Cognitive functioning	CF	2	3	20, 25	F
Social functioning	SF	2	3	26, 27	F
Symptom scales / items					
Fatigue	FA	3	3	10, 12, 18	
Nausea and vomiting	NV	2	3	14, 15	
Pain	PA	2	3	9, 19	
Dyspnoea	DY	1	3	8	
Insomnia	SL	1	3	11	
Appetite loss	AP	1	3	13	
Constipation	CO	1	3	16	
Diarrhoea	DI	1	3	17	
Financial difficulties	FI	1	3	28	

* *Item range* is the difference between the possible maximum and the minimum response to individual items; most items take values from 1 to 4, giving *range* = 3.
† (revised) scales are those that have been changed since version 1.0, and their short names are indicated in this manual by a suffix "2" – for example, PF2.

For all scales, the *RawScore*, *RS*, is the mean of the component items:

$$RawScore = RS = (I_1 + I_2 + ... + I_n)/n$$

Then for **Functional scales**:

$$Score = \left\{1 - \frac{(RS-1)}{range}\right\} \times 100$$

and for **Symptom scales / items** and **Global health status / QoL:**

$$Score = \{(RS-1)/range\} \times 100$$

For QLQ-BR23

The scoring approach for the QLQ-BR23 is identical in principle to that for the function and symptom scales / single items of the QLQ-C30†

	Scale	Number of items	Item range*	QLQ-BR23 Item numbers	†
Functional scales					
Body image	BRBI	4	3	9 – 12	F
Sexual functioning †	BRSEF	2	3	14,15	†
Sexual enjoyment †	BRSEE	1	3	16	†
Future perspective	BRFU	1	3	13	F
Symptom scales / items					
Systemic therapy side effects	BRST	7	3	1 – 4,6,7,8	
Breast symptoms	BRBS	4	3	20 – 23	
Arm symptoms	BRAS	3	3	17,18,19	
Upset by hair loss	BRHL	1	3	5	

* "Item range" is the difference between the possible maximum and the minimum response to individual items.

† Items for the scales marked † are scored positively (i.e. "very much" is best) and therefore use the same algebraic equation as for symptom scales; however, the Body Image scale uses the algebraic equation for functioning scales.

BRSEE, sexual enjoyment, is not applicable if item 15 is "not at all."
BRHL, upset by hair loss, is not applicable if item 4 is "not at all."

For QLQ-BR45

	Scale	Number of items (n)	Item range*	QLQ-[XX] item numbers ($I_1, I_2, \ldots, I_n$)	Reverse scoring items
Functional scales / items					
Body Image	BI	4	3	39 - 42	
Future Perspective	FU	1	3	43	
Sexual Functioning	SX	2	3	44, 45	44, 45
Sexual Enjoyment	SE	1	3	46	46
Breast Satisfaction	BS	2	3	74, 75	74, 75
Symptom scales / items					
Systemic Therapy Side Effects	SYS	7	3	31 -34, 36 - 38	
Upset by Hair Loss	HU	1	3	35	
Arm Symptoms	ARM	3	3	47 - 49	
Breast Symptoms	BR	4	3	50 - 53	
Endocrine Therapy Symptoms	ET	10	3	54 -56, 63 - 69	
Skin Mucosis Symptoms	SM	6	3	57 - 62	
Endocrine Sexual Symptoms	ES	4	3	70 - 73	

* "Item range" is the difference between the possible maximum and the minimum response to individual items. All items are scored 1 to 4, giving range = 3.

SE, sexual enjoyment, is not applicable if item 45 is "not at all".
HU, upset by hair loss, is not applicable if item 34 is "not at all".
Problems with weight gain (item 69) is not applicable if item 68 is "not at all".

Principle for Scoring:
Take into account that the scoring of questions 44, 45, 46, 74, and 75 must be reversed prior to statistical analysis.

1) **Raw score**

For each multi-item scale, calculate the average of the corresponding items.

$$Raw\ Score = RS = \left\{\frac{(I_1 + I_2 + \ldots + I_n)}{n}\right\}$$

2) **Linear Transformation**

To obtain the Score S, standardize the raw score to a 0 – 100 range following the appropriate transformation:

$$\text{Functional scales / items: } S = \left\{1 - \frac{(RS-1)}{range}\right\} \times 100$$

$$\text{Symptom scales / item: } S = \left\{\frac{(RS-1)}{range}\right\} \times 100$$

Handling of Missing items for QLQ-C30, QLQ-BR23, and QLQ-BR45:

- **Multi-item Scales**

  Have at least half of the items from a given scale been answered?

  - If Yes, use all the items that were completed, and apply the standard equations given above for calculating the scale scores; ignore any items with missing values when making the calculations.
  - If No, set scale score to missing.

- **Single-item Scales:**

  Set to missing if no response.

6 REFERENCES

1. Royston P, Parmar MK. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. BMC Med Res Methodol. 2013;13(1):152.

2. Dodd LE, Korn EL, Freidlin B, Gray R, Bhattacharya S. An Audit Strategy for Progression-Free Survival. Biometrics. 2011;67(3):1092-9.

3. Amit O, Mannino F, Stone AM, Bushnell W, Denne J, Helterbrand J, et al. Blinded independent central review of progression in cancer clinical trials: Results from a meta-analysis. European Journal of Cancer. 2011;47:1772-8.

4. Watkins C, Huang X, Latimer N, Tang Y, Wright EJ. Adjusting overall survival for treatment switches: commonly used methods and practical application. Pharm Stat. 2013;12(6):348-57

5. Robins JM, Tsiatis AA. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. Communications in statistics-Theory and Methods. 1991;20(8):2609-31.

6. Aaronson N, Ahmedzai S, Bullinger M, Crabeels D, Estapè J, Filiberti A, et al. The EORTC core quality of life questionnaire: interim results of an international field study. Effect of cancer on quality of life. 1991:185-203.

7. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organisation for Research and Treatment of Cancer QLQ-C30: A quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst. 1993;85:365-76.

8. Fayers PM, Aaronson NK, Bjordal K, Groenvold M, Curran D, Bottomley A, on behalf of the EORTC Quality of Life Group. The EORTC QLQ-C30 Scoring Manual (3rd Edition). Published by: European Organisation for Research and Treatment of Cancer, Brussels 2001.

9. Sprangers MA, Groenvold M, Arraras JI, Franklin J, te Velde A, Muller M, et al. The European Organization for Research and Treatment of Cancer breast cancer-specific quality-of-life questionnaire module: first results from a three-country field study. J clin oncol. 1996;14(10):2756-68.

10. Bjelic-Radisic V, Cardoso F, Cameron D, Brain E, Kuljanic K, da Costa RA, et al. An international update of the EORTC questionnaire for assessing quality of life in breast cancer patients: EORTC QLQ-BR45. Ann Oncol. 2020;31(2):283-8.

11. Van Hout B, Janssen MF, Feng YS, Kohlmann T, Busschbach J, Golicki D, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. Value health. 2012;15(5):708-15.

12. Glimm E, Maurer W, Bretz F. Hierarchical testing of multiple endpoints in group-sequential trials. Stat Med. 2010;29:219-28.

13. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). 2009;45(2):228-47.

Signature Page for VV-CLIN-0612905 v1.0
efc15935-16-1-9-sap

Approve & eSign	[redacted] Clinical [redacted]

Approve & eSign	[redacted] Clinical [redacted]