Protocol A5481023

MULTICENTER, RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED, PHASE 3 TRIAL OF FULVESTRANT (FASLODEX®) WITH OR WITHOUT PD-0332991 (PALBOCICLIB) ± GOSERELIN IN WOMEN WITH HORMONE RECEPTOR-POSITIVE, HER2-NEGATIVE METASTATIC BREAST CANCER WHOSE DISEASE PROGRESSED AFTER PRIOR ENDOCRINE THERAPY

Statistical Analysis Plan (SAP)

Version: 2.1.

Author: PPD

Date: January 10, 2018

TABLE OF CONTENTS

1. AMENDMENTS FROM PREVIOUS VERSIONS

1.1. Version 2.1

This version amends the statistical analysis plan dated December 17, 2014.

Since more patients have been randomized in the study than it was original planned, the previously determined final OS analysis based on 198 deaths (only 38% deaths out of 521 patients) becomes insufficient and inadequate to answer the overall survival question for the patient population. Therefore, the time of final OS analysis is shifted to when approximately 60% deaths are observed in the 521 randomized patients.

1.2. Version 2.0

This version amends the statistical analysis plan dated August 28, 2013. Changes have been made to incorporate relevant updates from protocol amendment 1 (dated April 4, 2014) and protocol amendment 2 (dated September 30, 2014).

Other major changes include the following items:

- the summary of early safety and PK unblinding plan was added;
- clarifications on statistical analysis models for drug-drug interaction were provided;
- the analysis plan for PK data was updated;
- the analysis plan for PRO data was updated;
- the plan of sensitivity analysis on PFS was updated;
- clarifications on subset analyses were provided;
- the analysis plan for ocular assessment data was added;
- the summary of key efficacy analyses table was updated.

2. INTRODUCTION

This is the Statistical Analysis Plan (SAP), Amendment 1, for study A5481023. This analysis plan is meant to supplement the study protocol. Any deviations from this analysis plan will be described in the Clinical Study Report.

2.1. Study Design

This is an international, multicenter, randomized, double-blind, placebo-controlled, parallel-group, Phase 3 clinical trial with the primary objective of demonstrating the superiority of palbociclib in combination with fulvestrant (Faslodex®) over fulvestrant alone

in prolonging investigator-assessed PFS in women with HR+/HER2-negative metastatic breast cancer whose disease has progressed after prior endocrine therapy. The safety between the two treatment arms will also be compared.

Eligible patients with:	**Stratification**:		**Blinded:**	**Treatment until**:	**All Patients should be followed for**:
-HR+/HER2-recurrent BC -failure of prior hormonal treatment - progressed ≤ 12 months from prior adjuvant therapy with AI*/Tam - progressed ≤ 1 month from prior advanced BC therapy with AI*/anti-endocrine - LHRH agonist for pre/peri menopausal women * for post menopausal women only	Documented sensitviity to prior hormonal therapy (yes vs. no) Menopausal status at study entry (pre/peri- vs. post menopausal) Presence of visceral metastatses (yes vs. no)	**R A N D O M I Z E (2:1)**	Palbociclib + Fulvestrant (N=278) Or Placebo + Fulvestrant (N=139)	Progressive disease per RECIST v1.1 Or Symptomatic deterioration Or Need for new additional anticancer therapy not specified in the protocol Or Unacceptable toxicities Or Investigator's conclusion that discontinuing therapy is in the patient's best interest Or Lost to follow-up Or Patient choice to withdraw from treatment Or Withdrawal of patient consent Or Death	Subsequent anti-cancer therapy (including type, start/stop dates, progression date, best response to the subsequent therapy) And Overall survival

Eligible patients must have histologically or cytologically proven diagnosis of adenocarcinoma of the breast with evidence of recurrent advanced (locally or metastatic) disease. Their breast cancer must have progressed during or within 12 months of completion of adjuvant therapy (with an AI, if postmenopausal, or with tamoxifen, if pre- or perimenopausal). Alternatively, it must have progressed while on, or within 1 month after the end of prior therapy for advanced/metastatic breast cancer (with an AI, if postmenopausal, or other prior endocrine treatment, if pre- or perimenopausal). One previous line of chemotherapy for metastatic disease is allowed. Patients must have measurable disease as per RECIST v.1.1 (Eisenhauer et al. 2009) or bone disease as their only site of disease. Tumor tissue is required for patient participation.

At least 417 patients will be randomized in a 2:1 ratio and stratified by documented sensitivity to prior hormonal therapy (yes vs. no), menopausal status at study entry (pre-/peri- vs. post menopausal), and presence of visceral metastases (yes vs. no).

Patients in Arm A (at least 278) will receive palbociclib 125 mg/day orally for 3 weeks followed by 1 week off plus fulvestrant 500 mg intramuscularly on Days 1 and 15 of Cycle 1, every 28 days (+/- 7 days) thereafter starting from Day 1 of Cycle 1.

Patients in Arm B (at least 139) will receive placebo orally daily for 3 weeks followed by 1 week off plus fulvestrant 500 mg intramuscularly on Days 1 and 15 of Cycle 1, every 28 days (+/- 7 days) thereafter starting from Day 1 of Cycle 1.

In both arms, pre- and peri-menopausal women will also receive the LHRH agonist goserelin (Zoladex® or generic).

Patients will undergo tumor assessment at Week 8, and then every 8 weeks for the first year, and then every 12 weeks, calculated from the date of randomization.

Patients will continue to receive assigned treatment until objective Progressive Disease (PD), symptomatic deterioration, unacceptable toxicity, death, or withdrawal of consent, whichever occurs first. Crossover between treatment arms will not be allowed. Patients showing RECIST-defined PD can continue with study treatment at the discretion of the investigator as long as that is considered to be in the best interest of the patient and no new anticancer treatment is initiated. In this case, the patient would continue with routine safety relevant assessments as per the Schedule of Activities for the active treatment period.

In addition, should palbociclib/placebo related toxicity mandate palbociclib/placebo discontinuation, patients can continue to receive fulvestrant alone. Patients discontinuing the active treatment phase (ie, discontinuing both palbociclib/placebo and fulvestrant) will enter a follow-up phase during which survival and new anti-cancer therapy information will be collected, initially every 3 months and then every 6 months.

An external data monitoring committee (E-DMC) will perform an early review of safety data from approximately the first 40 randomized patients with sufficient pharmacokinetic (PK) sampling to confirm safety and tolerability of the combination. The early safety review by the E-DMC will also include Pharmacokinetic (PK) data from these initial approximately 40 patients to explore potential Drug-Drug-Interactions (DDIs) between fulvestrant, goserelin (if applicable) and palbociclib. The study will continue while these analyses and review are ongoing.

In addition, blood samples will be collected from all patients to assess trough concentrations of palbociclib for exposure/response analysis for safety and efficacy findings.

Patient Reported Outcomes (PRO) will be collected to evaluate global quality of life and general health status.

The study will also include a molecular profiling component aimed at assessing the relationship between breast tumor sensitivity and resistance to palbociclib and the alteration of cell cycle pathway-related genes and proteins in tumor tissues.

2.2. Study Objectives

Primary Objective:

To demonstrate the superiority of palbociclib in combination with fulvestrant (with or without goserelin) over fulvestrant (with or without goserelin) alone in prolonging investigator-assessed PFS in women with HR+/HER2- metastatic breast cancer whose disease has progressed on prior endocrine therapy.

Secondary Objectives:

- To compare measures of tumor control (including PFS, DR, CBR) between the treatment arms;
- To compare safety and tolerability between the treatment arms;
- To evaluate trough concentrations of palbociclib when given in combination with fulvestrant or fulvestrant plus goserelin and compare them to historical data;
- To compare fulvestrant and goserelin trough concentrations when given in combination with palbociclib to those when given without palbociclib;
- To explore correlations between palbociclib exposures and efficacy/safety findings in this patient population;
- To compare Patient Reported Outcomes measures between treatment arms;
- To characterize alterations in genes, proteins, and RNAs relevant to the cell cycle, drug targets, tumor sensitivity and/or resistance;
- To conduct subgroup analysis for primary and secondary endpoints in stratified groups.

3. INTERIM ANALYSES, FINAL ANALYSES AND UNBLINDING

3.1. Progression Free Survival

The study is designed to have one interim analysis and the final analysis at 238 events based on the primary PFS endpoint with the investigator assessment. The interim analysis will be conducted to allow for early stopping of the study due to efficacy or to potentially re-estimate the sample size of the trial based upon the primary endpoint of PFS. The safety of the combination will also be assessed at the interim analysis. The analysis will be performed after approximately 143 investigator-assessed PFS events (documented progressive disease or death; approximately 60% of the total events expected). The information fraction for the interim analysis may be adjusted if needed.

- The Haybittle-Peto efficacy boundary (Haybittle 1971; Peto et al. 1976) is to be used at the IA. If the value of the test-statistic from the log-rank test for PFS exceeds the efficacy boundary ($z \geq 3$, $p \leq 0.00135$) the trial may be stopped for efficacy. It is assumed that a hazard ratio less than 1.0 is in favor of the palbociclib plus fulvestrant arm.

- As appropriate, the sample size of the study may be adjusted using the method outlined by Cui et al. (1999) and applied to the PFS endpoint. Using the Cui method guarantees that the overall type I error will still be preserved after a sample size increase. The technical details of sample size re-estimation approach will be pre-specified in a separate technical document that is to be shared only with the E-DMC before the interim analysis.

- If the results of the interim analysis indicate serious safety concerns, the sponsor will communicate with the Health Authorities regarding stopping the clinical trial.

- The final analysis of PFS will be performed after approximately 238 PFS events have been observed if the trial continues and the sample size is not adjusted after the IA .

Efficacy Stopping Boundary of PFS for Rejecting Null Hypothesis Expressed as Z Scales and p-values

Analysis	Number of Event (%)	Z Scale	p-value (1-sided)
Interim	143 (60%)	3	0.00135
Final	238(100%)	1.9644	0.02474

3.2. Overall Survival

Interim analyses of efficacy are also planned for the secondary OS endpoint and OS will be tested in a hierarchically approach. The analysis will be performed at the time of the interim or final PFS analyses if the primary analysis for PFS is positive. The nominal significance levels for the interim and final analyses of OS will be determined by the Lan-DeMet spending function based upon the O'Brien-Fleming boundary. The overall significance level for the efficacy analysis of OS will be preserved at 0.025 (one-sided test).

- Only one interim analysis of OS is planned. Although the first possible time for OS interim analysis could be at the time of the PFS IA, it is anticipated that the number of deaths could be low at PFS IA and yield non-robust analysis results. Therefore, the OS interim analysis will be planned at approximately 97 deaths (at the estimated time for planned PFS final analysis). If OS is not significant at the interim analysis, a final analysis will be performed after 198 deaths have been observed.

- If PFS is not significant at its final analysis, the formal conclusion on OS statistical significance will not be drawn.

Efficacy Stopping Boundary of OS for Rejecting Null Hypothesis Expressed as Z Scales and p-values

Analysis	Number of Event (%)	Z Scale	p-value (1-sided)
Interim	97 (49%)	3.0141	0.0013
Final	198 (100%)	1.9673	0.0246

3.3. Data Monitoring Committee

The study will use an External Data Monitoring Committee (E-DMC). The E-DMC membership and governance are outlined in a separate charter.The E-DMC will be responsible for ongoing monitoring of the efficacy and safety data from patients in the study according to the Charter. The E-DMC will make recommendation as to whether or not the trial should continue based on ongoing reviews of safety data. In addition, the E-DMC will also evaluate interim efficacy data and make a recommendation regarding study continuation based on observed results of the study. The Sponsor will designate a biostatistician not affiliated with the project to prepare data for E-DMC review. Only if action or consultation with Health Authorities is required will other sponsor staff be involved in the data preparation. Clinical sites will be restricted from access to study results until the conclusion of the study. More details can be found in the E-DMC charter.

3.4. Blinding

At the initiation of the trial, the trial sites will be instructed on the method for breaking the blind. The method will be either a manual or electronic process. Blinding codes should only be broken in emergency situations to preserve the patient safety. Blinding codes may also be broken after a patient discontinues treatment due to disease progression, as determined by the treating investigator using RECIST v.1.1 criteria, only if deemed essential to allow the investigator to select the patient's next treatment regimen and after discussion and agreement with the sponsor. Codes should not be broken in the absence of emergency situations or progressive disease as per RECIST v.1.1 (eg, in case of clinical deterioration, increase in tumor markers or any other evidence suggestive of disease progression but in the absence of RECIST-defined disease progression). When the blinding code is broken, the date and reason for unblinding must be fully documented in source documents and entered on the case report form. However, every effort should be made by the site staff to ensure that the treatment arm in which the unblinded patient is assigned is not communicated to any sponsor personnel or designee involved in the conduct of the trial.

3.4.1. Early Safety and PK Unblinding

There will be an early review of safety and PK data from approximately the first 40 randomized patients with sufficient pharmacokinetic (PK) sampling. A separate PK Unblinding Plan explains how the study will remain blinded while PK data are explored for potential Drug-Drug-Interactions (DDIs) between palbociclib + fulvestrant ± goserelin versus placebo + fulvestrant ± goserelin compared to historical palbociclib data and whether dosing adjustments are necessary. The study will continue while these analyses and review are ongoing.

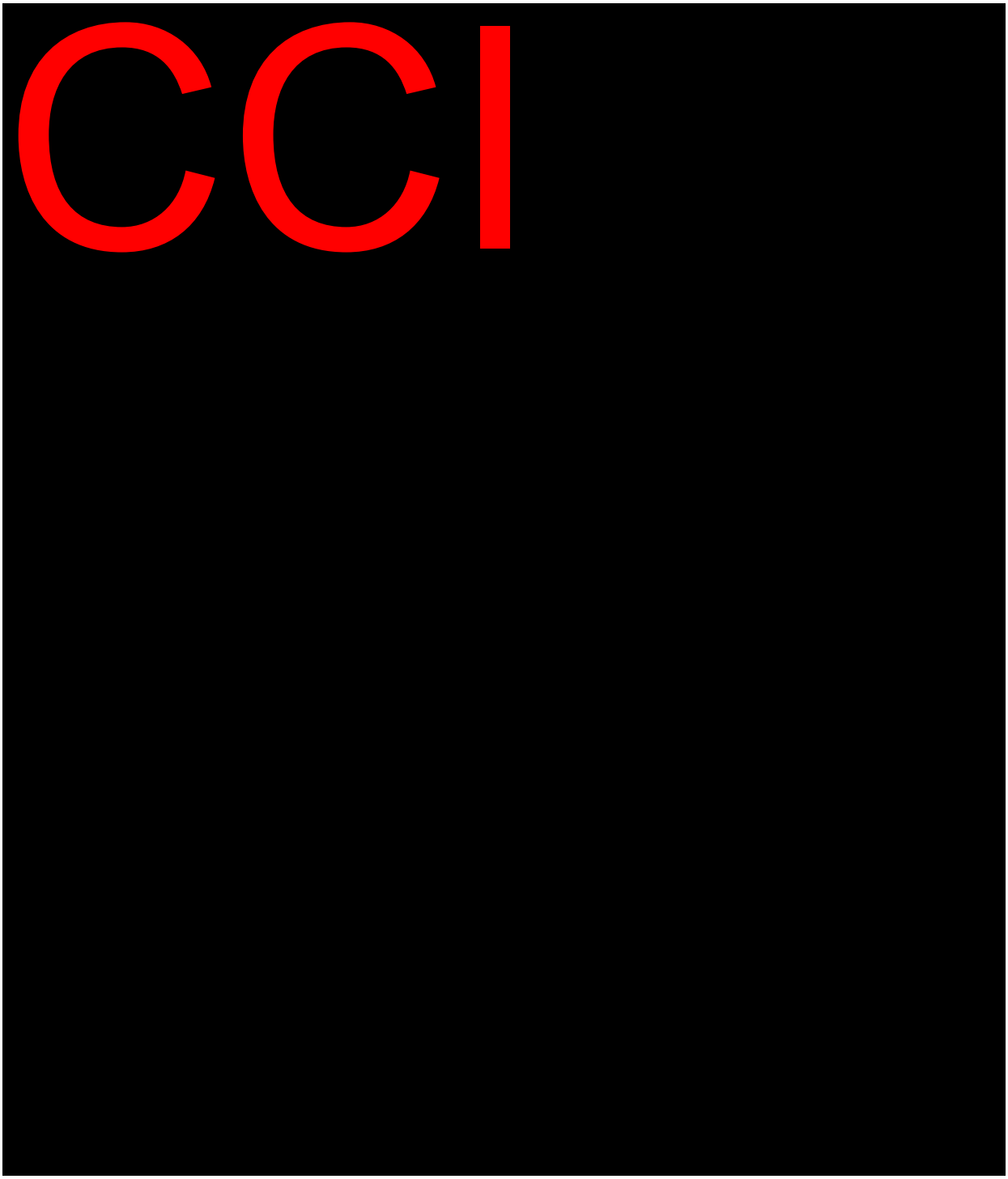
4. HYPOTHESES AND DECISION RULES

4.1. Statistical Hypotheses

The primary objective of this study is to demonstrate that the combination of palbociclib and fulvestrant is superior to the combination of placebo and fulvestrant in prolonging

investigator-assessed PFS in women with HR+/HER2-negative metastatic breast cancer that has progressed on prior endocrine therapy, and regardless of their menopausal status.

The study is designed to test the null hypothesis that the true PFS distributions for both palbociclib plus fulvestrant and placebo plus fulvestrant arms are the same with a median PFS 6.0 months versus the alternative hypothesis that the true PFS distribution has a median that is longer than 6.0 months for the palbociclib plus fuvestrant arm.



CCI

5. ANALYSIS SETS

5.1. Intent-to-Treat Population (Full Analysis Set)

The intent-to-treat (ITT) population or full analysis set will include all patients who are randomized, with study drug assignment designated according to initial randomization, regardless of whether patients receive study drug or receive a different drug from that to which they were randomized. The ITT population will be the primary population for evaluating all efficacy endpoints and patient characteristics.

5.2. As-Treated (AT) Population (Safety Analysis Set)

The as-treated (AT) population or safety analysis set will include all patients who receive at least 1 dose of study medication, with treatment assignments designated according to actual study treatment received. The AT population will be the primary population for evaluating treatment administration/compliance and safety. Efficacy and clinical benefit endpoints may be assessed in this population as well.

5.3. Other Analysis Sets

5.3.1. Pharmacokinetic Analysis Sets

There will be at least two PK analysis sets:

1. Early Safety Review Set: The 40 patients participating in the early safety review, who are treated with palbociclib + fulvestrant ± goserelin or placebo + fulvestrant ± goserelin and have at least one measured plasma drug concentration.

2. Palbociclib PK Analysis Set: All patients (including those in the Early Safety Review Set) who have PK blood samples collected for palbociclib and have at least one measured plasma drug concentration.

PK/Adverse Event Analysis Set (if applicable): patients experiencing unexpected or serious adverse events, or adverse events that lead to discontinuation who have PK blood samples collected for palbociclib or fulvestrant (or goserelin if applicable) and have at least one measured plasma drug concentration.

5.3.2. Biomarker Analysis Set

A subset of AT patients, who have both baseline and at least one follow-up values for at least one biomarkers.

5.3.3. Patient Reported Outcome (PRO) Evaluable Population (PRO Analysis Set)

The PRO –evaluable population is defined as a subset of ITT patients, who have completed a baseline and at least one post –baseline PRO assessment prior to end of study treatment.

5.4. Treatment Misallocations

- If patients were randomized but not treated, then they will be reported under their randomized treatment group for efficacy analyses. However, they are by definition excluded from the safety analyses.
- If patients were randomized but took incorrect treatment, then they will be reported under their randomized treatment group for efficacy analyses, but will be reported under the treatment they actually received at the time of onset of toxicity for all safety analyses.

5.5. Protocol Deviations

All deviations will be described when they appear and relate to the statistical analyses or populations.

5.5.1. Protocol Deviations Assessed Prior to Randomization

Deviations prior to randomization are not allowed (ie, no waivers are allowed). There is no distinction between acceptable and unacceptable, minor or major deviations. Selected eligibility violations may be described in a listing,but no patient will be excluded from the primary analyses based on eligibility

5.5.2. Protocol Deviations Assessed Post Randomization

Treatment not started. Patients who do not start treatment will not be included in the safety analysis. However, those patients will still be included in the primary efficacy analyses.

Randomized treatment on the wrong arm. In such cases efficacy is analyzed according to the assigned treatment arm, not the arm received (intent-to-treat principle). For safety, results are analyzed as treated at the time of the onset of the toxicity.

6. ENDPOINTS AND COVARIATES

6.1. Efficacy Endpoints

6.1.1. Primary Endpoint

Progression Free Survival (PFS) is defined as the time from the date of randomization to the date of the first documentation of objective progression of disease (PD) or death due to any cause in the absence of documented PD, whichever occurs first. If tumor progression data include more than 1 date, the first date will be used. Documentation of progression must be objective disease assessment. Objective disease assessments are based on Response Evaluation Criteria in Solid Tumor (RECIST) v1.1 guidelines.

The length of PFS will be calculated as
PFS time (months) = [progression/death date (censor date) – randomization date + 1]/30.4.

Censorship: Patients last known to be 1) alive 2) not to have started new (non-protocol) anti-cancer treatment and 3) progression-free, and who have a baseline and at least one disease assessment after dosing, are censored at the date of the last objective disease assessment that verified lack of disease progression. Patients with inadequate baseline disease assessment are censored at the date of randomization.

- Patients with no disease assessments after dosing are censored at the date of randomization unless death occurred prior to the first planned assessment (in which case the death is an event).

- Patients starting new anti-cancer treatment prior to progression are censored at the date of last objective disease assessment documenting no progression prior to the new treatment.

- If patients are removed from the study (withdrew the consent, lost to follow up, etc.) prior to progression and death, then censorship is at the date of the last objective disease assessment that verified lack of disease progression.

- Patients with documentation of progression or death after an unacceptably long interval (ie, 2 or more incomplete or indeterminate assessments) since the last tumor assessment will be censored at the time of last objective assessment documenting no progression.

Other scenarios for defining event/censoring may be investigated in supportive analyses (eg, if a substantial number of patients has questionable failure or censorship dates).

6.1.2. Secondary Endpoints

- **Overall Survival (OS)** is defined as the time from date of randomization to date of death due to any cause. In the absence of confirmation of death, survival time will be censored to last date the patient is known to be alive. For patients lacking survival data beyond the date of their last follow-up, the OS time will be censored on the last date they were known to be alive. Patients lacking survival data beyond randomization will have their OS times be censored at randomization. The length of OS will be calculated as OS time (months) = [death date (censor date) – randomization date + 1]/30.4.

- **One-, Two- or Three-year Survival Probability** is defined as the probability of survival 1 year, 2 or 3 years after the date of randomization based on the Kaplan-Meier estimate.

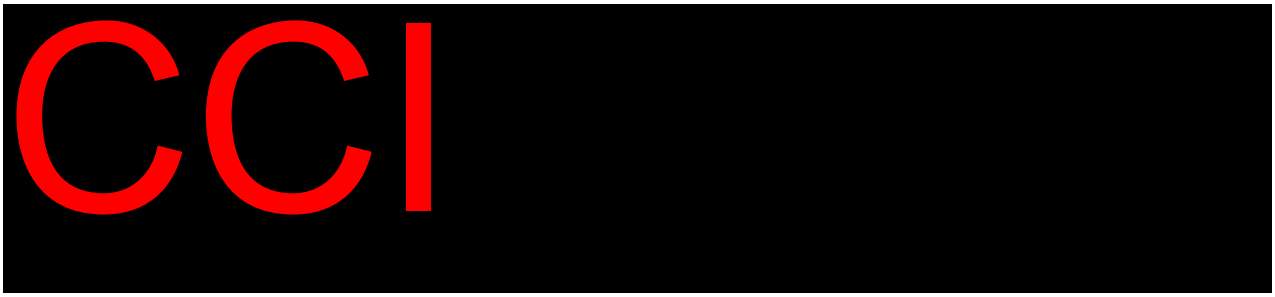
- **Objective Response (OR)** is defined as the overall complete response (CR) or partial response (PR) according to the Response Evaluation Criteria in Solid Tumors (RECIST version 1.1) **Objective Response Rate (ORR)** is defined as the proportion of patients with CR or PR relative to (1) all randomized patients and (2) randomized patients with measurable disease at baseline. Designation of best response of stable disease (SD) requires the criteria to be met at least 8 weeks after randomization. Patients who do not have on-study radiographic tumor re-evaluation, who receive anti-tumor treatment other than the study medication prior to reaching a CR or PR, or who die, progress, or drop out

for any reason prior to reaching a CR or PR will be counted as non-responders in the assessment of ORR.

- **Clinical Benefit Response (CBR)** is defined as the overall complete response (CR), partial response (PR) , or stable disease (SD) ≥24 weeks according to the Response Evaluation Criteria in Solid Tumors (RECIST version 1.1). **Clinical Benefit Response Rate** is defined as the proportion of patients with CR, PR, or SD ≥24 weeks relative to (1) all randomized patients and (2) randomized patients with measurable disease at baseline. Designation of best response of SD ≥24 weeks requires the criteria to be met at least 24 weeks after randomization. Patients who do not have on-study radiographic tumor re-evaluation, who receive anti-tumor treatment other than the study medication prior to reaching a CR or PR, a best response of SD ≥24 weeks, or who die, progress, or drop out for any reason prior to reaching a CR or PR and a best response of SD ≥24 weeks will be counted as non-responders in the assessment of CBR.
- **Duration of Response (DR)** is defined as the time from the first documentation of objective tumor response (CR or PR) to the first documentation of disease progression or to death due to any cause, whichever occurs first. If tumor progression data include more than 1 date, the first date will be used. DR will be calculated as [the date response ended (ie, date of PD or death) – first CR or PR date + 1)]/30.4. DR will only be calculated for the subgroup of patients with an objective tumor response.

Responders last known to be 1) alive 2) not to have started new (non-protocol) anti-cancer treatment and 3) progression-free are censored at the date of the last objective disease assessment that verified lack of disease progression.

- Patients starting new anti-cancer treatment prior to progression are censored at the date of last objective disease assessment documenting no progression prior to the new treatment.
- If patients are removed from the study (withdrew the consent, lost to follow up, etc.) prior to progression and death, then censorship is at the date of the last objective disease assessment that verified lack of disease progression.
- Patients with documentation of progression or death after an unacceptably long interval (ie, 2 or more incomplete or indeterminate assessments) since the last tumor assessment will be censored at the time of last objective assessment documenting no progression.




CCI

6.3. Safety Data

Overall safety profile as characterized by type, frequency, severity of adverse events as graded by NCI Common Toxicity Criteria for Adverse Events version 4 (NCI CTCAE v.4.0), timing and relationship to treatment on each arm, and laboratory abnormalities observed.

Serious and Non-serious Adverse events (AEs), hematology, blood chemistry, ECG and vital signs will be assessed as described in the Schedule of Activities of the protocol.

Adverse events will be classified using the MedDRA classification system. The severity of the toxicities will be graded according to the NCI CTCAE version 4. For labs without CTCAE grade definitions, results are summarized as normal or abnormal (per Pfizer Data Standards (PDS)) or not done. For other AEs without specific CTCAE definitions, results are identified according to CTCAE "other" categories. The "Schedule of Activities" in the protocol lists all safety parameters to be collected.

Adverse events leading to discontinuation of trial treatment, events classified using NCI CTCAE v.4.0 as Grade 3 or higher, trial drug related events, and serious adverse events will be considered with special attention.

Laboratory results will be graded according to the NCI CTCAE v.4.0 severity grade. For parameters for which an NCI CTCAE v.4.0 scale does not exist, the frequency of patients with values below, within, and above the normal range for the local lab will be summarized. Each patient will be summarized by the worst severity grade observed for a particular laboratory parameter. This will be provided for all cycles as well as by cycles. Patients who start treatment are assessed for toxicities up to 28 days after the final dose of treatment or start of new treatment (whichever comes first). Toxicities observed beyond 28 days and recorded in the database per Sponsor's agreement will be included in the summaries.

A 3-tier approach will be used to summarize AEs. Under this approach, AEs are classified into 1 of 3 tiers. Different analyses will be performed for different tiers (See Section 8.2.6).

6.3.1. Treatment Emergent Adverse Event

An adverse event is considered treatment emergent if:

- The event occurs for the first time after the start of study treatment and up to 28 days after final dose of study treatment and was not seen prior to the start of treatment or
- The event was seen prior to the start of treatment but increased in NCI CTCAE v.4.0 grade during study treatment.
- Disease progression is not considered a treatment emergent adverse event unless the patient dies of disease prior to 28 days after discontinuation of treatment.

6.3.2. Treatment Related Adverse Event

Adverse events defined as treatment emergent adverse events with cause possibly, probably or definitely related to treatment as judged by the investigator are defined as treatment related adverse events. Events that are continuation of baseline abnormalities are not considered treatment related unless there is an increase in grade, or if there is an increase following a decrease, and the increase is judged by the investigator to be caused by the treatment.

6.3.3. Laboratory Safety Assessments

Laboratory assessment will be assigned to cycles based on the collection date of the sample relative to the start dates of cycles from the study drug administration as described in the Schedule of Activities table in Appendix 10.1.

Baseline evaluations for laboratory are those collected

- Within 28 days prior to or on first day of study drug and
- If there is more than one baseline evaluation, closest to but any time prior to the 1st dosing on the first day of study treatment.

6.3.4. Electrocardiogram (ECG)

All Electrocardiograms (ECGs) will be performed using a 12-lead (with a 10-second rhythm strip) tracing. ECG measurements will include PR interval, QT interval, RR interval, and QRS complex. It is preferable that the machine used has a capacity to calculate the standard intervals automatically.

ECG interval readings by the ECG recorder's algorithm will be read and interpreted at the investigational site for eligibility determination and patient safety monitoring and documentation stored in the source documents.

Additional ECGs may be performed as clinically indicated at any time.

For the purpose of the study, triplicate ECGs are defined as three consecutive ECGs performed approximately 2 minutes apart at the protocol specified timepoints (see Schedule of Activities table) to determine the mean QTc interval.

Refer to the Dose Modification Section for details on dose management in case of abnormal ECGs considered related to investigational product.

6.3.5. Other Safety Assessment

A full physical examination including an examination of all major body systems and breasts, height (at screening only), weight, blood pressure and pulse rate which may be performed by a physician, registered nurse or other qualified health care provider, will be required at screening, and Day 1 of every cycle and at End of Treatment.

Symptom directed physical examinations, blood pressure and pulse rate will be performed at subsequent visits.

Performance Status: The ECOG performance status scale will be used.

6.4. Other Data

6.4.1. Pharmacokinetic Data

For the first approximately 40 patients participating in the early safety review, PK samples will be drawn at pre-dose on Day 1 and Day 15 of Cycle 1 and Cycle 2, and Day 1 of Cycle 3 for DDI assessments. Two (combinations without goserelin) or three (combinations with goserelin)

3-mL samples of venous blood will be collected in appropriately labeled collection tubes for assessment of palbociclib, fulvestrant, and/or goserelin levels at the protocol-specified times.

In all other palbociclib patients (not participating in the early safety review), plasma concentrations will be drawn on Day 15 of Cycle 1 and Cycle 2 for palbociclib determination. For these patients, one 3 mL sample of venous blood will be collected in appropriately labeled collection tubes for assessment of palbociclib.

All palbociclib concentrations will be used for exposure/response analysis for safety and efficacy findings.

The endpoint for these analyses is palbociclib steady state trough plasma concentration (Ctrough) which will be compared as follows: palbociclib and fulvestrant with/without goserelin versus placebo and fulvestrant with/without goserelin compared to historical palbociclib data from studies A5481001, A5481002, and A548003. The historical palbociclib data will be A5481001 (epharm artifact ID 8146749), A5481002 (epharm artifact ID 8145916) and A5481003 (epharm artifact ID 8149001). A5481001 was conducted in all comer patients and resulted in lower exposures for palbociclib, A5481002 was a pilot open-labeled clinical study in 17 patients with previously treated mantle cell lymphoma and A5481003 was conducted in breast cancer patients and resulted in higher exposures for palbociclib. These studies provide the “normal” palbociclib exposure ranges.

Palbociclib steady state trough plasma concentration (Ctrough) is defined as pre-dose plasma concentration following at least 8 consecutive days of 125 mg daily dose without dosing interruption. The time window for the PK collection should be between 22 and 26 hrs after the dose (the day prior to PK collection) and no more than 1 hr postdose on the day of PKcollection.

Covariates for this PK per PMAR-EQDD-A548b-DP4-269 are: age and baseline body weight.

Additional PK blood samples for palbociclib and fulvestrant (and goserelin if applicable) may be collected from patients experiencing unexpected or serious adverse events, or adverse events that lead to discontinuation.

CCI

Samples will be analyzed using validated analytical methods in compliance with Pfizer standard operating procedures.

All efforts will be made to obtain the pharmacokinetic samples at the scheduled nominal time relative to dosing. However, samples obtained within 10% of the nominal time AND collected prior to administration of the investigational products on that day will be considered protocol compliant.

6.4.2. Biomarkers

Tumor tissues are required from all patients for study participation.

HR (ER and PR) and HER2 status for eligibility will be based on local results utilizing an assay consistent with local standards for entry into the trial and for reporting of the efficacy endpoints. Central assessment of ER, PR, and HER2 on suitable samples will be performed retrospectively at a qualified central laboratory.

A biomarker will be nominated as the primary marker for analysis using data external to the current study. The interaction between this primary marker (eg, pRb or other from a set of biomarkers to be analyzed) and benefit from palbociclib in terms of prolongation of PFS will be examined. In addition, other tumor tissue biomarkers, including DNA, RNA and protein analytes, will be analyzed to investigate possible associations with resistance/sensitivity to treatment with palbociclib. Biomarkers that will be analyzed will be selected based on their known relevance to mechanisms involved in cell cycle regulation. Examples of such biomarkers include, but are not limited to, CCND1 and CDKN2A gene copy number, cdk4 and cdk6 RNA expression, and Ki67, pRb, cyclin E and p16 protein expression. The relationship between centrally reported ER/PR and HER2 status and resistance/sensitivity to treatment with palbociclib will also be assessed.

Plasma samples will be collected for analyses of circulating free DNA or RNA, and the relationship with resistance/sensitivity to treatment with Palbociclib/placebo.

If archived FFPE tissue from a recurrent tumor or distant metastasis is unavailable, a de novo biopsy is required for patient participation, except those with bone metastasis only, when, in the investigator's judgment, such biopsy is feasible and can be safely performed. Original diagnostic tumor tissue will be used for biomarker analyses in the event that patients have bone disease only. Patients, who relapsed while receiving adjuvant therapy and had surgery

within the last 3 years, may provide tissue from that surgery. Provision of de novo metastatic tissue in these cases is strongly encouraged but not mandated.

Detailed information about biomarker sample collection, preparation, storage, labeling, and shipment is indicated in the Study Reference Manual.

6.4.3. Patient Reported Outcome Endpoints

Patient reported outcomes of functioning, global quality of life and general health status will be assessed using the he EORTC QLQ-C30, EORTC QLQ-BR23 and EQ-5D.

Patients will complete each instrument at pre-dose on Day 1 of Cycles 1-4, then on Day 1 of every other subsequent Cycle starting with Cycle 6 (eg, Cycles 6, 8, 10, etc), and then at the end of treatment visit. Completed questionnaires are always considered source document and must be filed accordingly.

Patients must complete these instruments in clinic (cannot be taken home) and prior to having any tests and to any discussion of their progress with healthcare personnel at the site. Interviewer administration in clinic may be used under special circumstances (eg, patient forgot their glasses or feels too ill). The instruments will be given to the patient in the appropriate language for the site.

6.4.3.1. EORTC QLQ-C30

The EORTC-QLQ-C30 (see Protocol Appendix 6) is a 30-item questionnaire composed of five multi-item functional subscales (physical, role, emotional, cognitive , and social functioning), three multi-item symptom scales (fatigue, nausea/vomiting, and pain), a global quality of life (QOL) subscale, and six single item symptom scales assessing other cancer-related symptoms (dyspnea, sleep disturbance, appetite loss, constipation, diarrhea, , and the financial impact of cancer). The questionnaire employs 28 4-point Likert scales with responses from “not at all” to “very much” and two 7-point Likert scales for global health and overall QOL. Responses to all items are then converted to a 0 to 100 scale. For functional and global QOL scales, higher scores represent a better level of functioning/QOL. For symptom-oriented scales, a higher score represents more severe symptoms. A 10-point or higher change in scores from baseline is considered clinically significant.

6.4.3.2. EORTC QLQ-BR23

The EORTC-QLQ-BR23 (see Protocol Appendix 7) is a 23-item breast cancer-specific companion module to the EORTC-QLQ-C30 and consists of four functional scales (body image, sexual functioning, sexual enjoyment, future perspective) and four symptom scales (systemic side effects, breast symptoms, arm symptoms, upset by hair loss).

6.4.3.3. EQ-5D

The EuroQol-5D (EQ-5D) (version 3L) (see Protocol Appendix 5) is a brief self-administered, validated instrument consisting of 2 parts. The first part consists of 5 descriptors of current health state (mobility, self care, usual activities, pain/discomfort, and anxiety/depression); a patient is asked to rate each state on a three level scale (1=no problem,

2=some problem, and 3=extreme problem) with higher levels indicating greater severity/impairment Published weights are available that allow for the creation of a single summary score called the EQ-5D index, which basically ranges from 0 to 1 with low scores representing a higher level of dysfunction and 1 as perfect health. The second part consists of the EQ-5D general health status as measured by a visual analog scale (EQ-5D VAS). The EQ-5D VAS measures the patient's self-rated health status on a scale from 0 (worst imaginable health state) to 100 (best imaginable health state).

6.5. Covariates and Stratification Factors

6.5.1. Covariates

The potential influences of baseline patient characteristics such as age, ethnic origin, ECOG performance status, geographical region, selected biomarkers, and stratification factors on the primary PFS, OS, and OR endpoints may be evaluated.

6.5.2. Stratification Factors

Randomization will be stratified by documented sensitivity to prior hormonal therapy (yes vs. no), by menopausal status at study entry (pre-/peri- vs. post menopausal), and by the presence of visceral metastases (yes vs. no). Sensitivity to prior hormonal therapy is defined as either: 1) documented clinical benefit (complete response [CR], partial response [PR], stable disease [SD] ≥24 weeks) to at least 1 prior hormonal therapy in the metastatic setting, OR 2) at least 24 months of adjuvant hormonal therapy prior to recurrence. "Visceral" refers to lung, liver, brain, pleural, and peritoneal involvement.

7. HANDLING OF MISSING VALUES

7.1. Missing Dates

In compliance with Pfizer standards, if the day of the month is missing for any date used in a calculation, the 1st of the month will be used to replace the missing date unless the calculation results in a negative time duration (eg, date of onset cannot be prior to day one date). In this case, the date resulting in 1 day duration will be used. If the day of the month and the month are missing for any date used in a calculation, eg, January 1 will be used to replace the missing date.

Missing dates for adverse events will be imputed based on the similar principle.

If the start date is missing for an AE, the AE is considered to be treatment emergent unless the collection date is prior to the treatment start date, and provided the event is not a serious adverse event.

7.2. Missing Tumor Assessments

If baseline tumor assessment is inadequate the patient cannot be assessed for response.

Inadequate baseline assessment may include

- Not all required baseline assessments were done
- Assessments were done outside the required window
- Measurements were not provided for one or more target lesions
- One or more lesions designated as target were not measurable.

If measurements for one or more target lesions are missing for an evaluation and disease does not qualify as progression (or symptomatic deterioration if applicable), the objective status for that evaluation is Indeterminate.

If non-target disease was not assessed, then objective status cannot be a CR even if all target disease has disappeared. Otherwise, missing non-target disease assessments do not necessarily affect response determination. Such cases will be reviewed carefully.

If a lesion measurement is missing because it is documented as too small to measure, the value 5 mm will be assigned and objective status calculated accordingly.

In the assessment of OR, patients who do not have on study radiographic tumor re-evaluations will be counted as non-responders.

7.3. Missing Data in PFS Derivation

PFS cannot be assessed in patients with inadequate baseline tumor assessment. PFS cannot be assessed in patients who have no on-study assessments unless death occurs prior to the first planned assessment time.

If a substantial number of patients have questionable failure or censorship dates for either PFS definition (such as progression or death not documented until after multiple missing assessments) scenarios such as best case (failure at time of documentation) and worst case (progression at earliest possible planned assessment date) will be investigated.

For PFS analysis, no values will be imputed for missing data. For time to event endpoints, non-event observations will be censored as defined in Section 6.1.1.

7.4. Missing Values in Patient Reported Outcomes

For the QLQ-C30, QLQ-BR23 and EQ-5D, in cases where two answers are given to one item, the more severe answer will be counted. For QLQ-C30 and QLQ-BR23, if at least half of the constituent items for the multi-item functional or symptom scale have been answered, then the score for that scale may be pro-rated based on the non-missing items. For the EQ-5D, the index score is considered missing if the answer to any of the 5 items is missing. . For subsequent analysis purposes, missing items will be considered missing; they will not be imputed.

7.5. Missing Values in PK Data

7.5.1. Concentrations below the limit of quantification

In all data presentations (except listings), concentrations below the limit of quantification (BLQ) will be set to zero. In listings, BLQ values will be reported as "<LLQ", where LLQ will be replaced with the value for the lower limit of quantification.

7.5.2. Deviations, missing concentrations and anomalous values

In summary tables, statistics will be calculated having set concentrations to missing if 1 of the following cases is true:

1. A concentration has been noted as ND (ie, not done) or NS (ie, no sample);
2. A deviation in sampling time is of sufficient concern or a concentration has been flagged anomalous by the pharmacokineticist.
3. A predose concentration has no actual collection time relative to the dosing due to missing dosing time or missing sampling time.

7.6. Missing and/or Duplicate Values in Biomarker Data

Missing biomarker data will not be imputed. Duplicate biomarker data is not expected. If any is received, the study team will determine which record will be used or if an average of the duplicate records will be used, depending upon the type of data. A flag will be included in the biomarker data indicating the value used for biomarker analyses.

8. STATISTICAL METHODOLOGY AND STATISTICAL ANALYSES

8.1. Statistical Methods

8.1.1. Analyses for Time-to-Event Data

Time-to-event endpoints between the 2 treatment arms will be compared with a 1-sided stratified log-rank test adjusting for presence of visceral metastases and sensitivity to prior hormonal therapy (two of the baseline stratification factors as listed in Section 6.5.2) and/or a 1-sided unstratified log-rank test at the α=0.025 overall significance level. Hazard ratios and 2-sided 95% confidence intervals (subject to the multiplicity adjustment at the final analysis for PFS and OS) will be estimated using Cox proportional hazards regression.

Cox proportional hazard models will also be used to explore the potential influences of the baseline stratification factors (as listed in Section 6.5.2) on time-to-event endpoints. In addition, potential influences of baseline patient characteristics such as age, race, ethnic origin, ECOG performance status, geographical region, and selected biomarkers on the endpoints may be evaluated. A backward selection process (with treatment in the model) will be applied to these variables to identify the final set of relevant factors. Treatment-by-factor interactions will be explored only for the set of factors included in the

final model. The estimated hazard ratio and 2-sided 95% confidence interval will be provided. Additionally for each treatment arm, the median event time and a 2-sided 95% confidence interval will be provided for each level of stratification factors or baseline characteristics.

Time-to-event endpoints will be summarized using the Kaplan-Meier method and displayed graphically when appropriate. Median event times and 2-sided 95% confidence interval for each median will be provided.

The X-year survival probability will be estimated using the Kaplan-Meier method and a 2-sided 95% confidence interval for the log [-log(X-year survival probability)] will be calculated using a normal approximation and then back transformed to give a confidence interval for the X-year survival probability itself.

Since patients in both treatment arms may receive other available treatments after disease progression, the treatment effect on overall survival may not be able to estimate properly by above defined methods because of these confounding factors. Therefore, the proper testing statistics such as Wilcoxon test and methods like Rank-Preserving Structural Failure Time Model (RPSFTM) proposed by Robins and Tsiatis (1991) will be applied to the overall survival analysis.

8.1.2. Analyses for Binary Data

The odds ratio estimator (with the corresponding 2-sided 95%) and the 1-sided stratified exact test will be used to compare the rates of binary endpoints for the two treatments. In addition, point estimates of the rates for each treatment arm will be provided along with the corresponding exact 2-sided 95% confidence intervals using the exact method based on Clopper-Pearson method.

8.1.3. Analyses of Continuous Data

Descriptive statistics, including the mean, standard deviation, median, minimum, and maximum values, will be provided for continuous endpoints.

8.1.4. Analyses for Categorical Data

The number and percentage of patients in each category will be provided for categorical variables.

8.1.5. Evaluation of BICR Data

NCI Method:

The auxiliary estimator of the log-hazard ratio that combines the information from patient-level data by investigator assessment on all the cases and the retrospective random sample BICR audit cases will be calculated.

More specifically, here are some notations:

$\hat{\theta}_{CA}$ = an estimator of the log-hazard ratio (based on the proportional hazards model) using BCIR data of the audited subset;

$\hat{\theta}_{LA}$ = an estimator of the log-hazard ratio (based on the proportional hazards model) using investigator assessment data of audited subset;

$\hat{\theta}_{\overline{LA}}$ = an estimator of the log-hazard ratio (based on the proportional hazards model) using investigator assessment data of non-audited subset;

$\hat{V}_{CA}$ = an estimator of the variance of log-hazard ratio using BICR data of the audited subset;

$\hat{V}_{L}$ = an estimator of the variance of log-hazard ratio using investigator assessment data of the total population;

δ = proportion of study participants audited assuming the number of events in the audit is proportional to the sampled audit size;

$\hat{\rho}$ = a bootstrap estimator of the correlation between $\hat{\theta}_{LA}$ and $\hat{\theta}_{CA}$.

$\hat{\rho}$ will be estimated using a bootstrap approach. Within the randomly selected sample (of size m) for BICR audit, a bootstrap sample of m patients will be sampled with replacement. The two sample-based log-hazard ratio estimates (investigator-based vs. BICR –based) will be estimated for the bootstrap sample. The procedure will be repeated 1000 times to obtain 1000 bootstrap samples. Consequently, 1000 pairs of sample estimates of log(HR) (investigator-based vs. BICR-based) will be calculated. The sample correlation coefficient between the investigator-based log(HR) estimates and BICR-based log(HR) estimates will be used for $\hat{\rho}$ in calculating the auxiliary estimator and the corresponding one-sided 95% CI.

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

$$\widetilde{\theta}_{C} = \hat{\theta}_{CA} + \hat{\rho}\sqrt{\delta}(1-\delta)\sqrt{\frac{\hat{V}_{CA}}{\hat{V}_{L}}}(\hat{\theta}_{\overline{LA}} - \hat{\theta}_{LA})$$

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

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

Assuming asymptotic normality of $\widetilde{\theta}_{C}$, the upper bound of one-sided 95% CI of this estimator will consequently be calculated as $\widetilde{\theta}_{C} + Z_{1-0.05}\sqrt{\widetilde{V}_{C}}$, where $Z_{1-0.05}$ is the 95% quantile of a standard normal distribution.

The NCI method may be used to evaluate the BICR data only if the BICR approach is conducted for the randomly selected sample. In the circumstances where the BICR approach is conducted for the total population upon the request of the health authorities or the study sponsor, NCI method will not be implemented to evaluate the BICR data.

PhRMA Method

Differential discordance will be evaluated using two measures, the early discrepancy rate (EDR) and late discrepancy rate (LDR). The agreement between the investigator and the blinded independent central review within a treatment arm is represented in a tabular form below

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

a1: number of agreements on timing and occurrence of PD.
a2: number of times investigator declares PD later than BICR
a3: number of times investigator declares PD earlier than BICR

The early discrepancy rate (EDR) is defined as:

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

The EDR represents the positive predictive value of investigator assessment and quantifies the frequency with which the investigator declares progression early relative to blinded independent central review within each arm as a proportion of the total number of investigator assessed PD's.

The late discrepancy rate (LDR) is defined as:

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

The LDR quantifies the frequency that investigator declares progression later than blinded independent central review as a proportion of the total number of discrepancies within the 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 around each measure will be summarized as the rate on the experimental arm minus the rate on the control arm. A negative differential discordance for the EDR and/or positive differential discordance for the LDR are suggestive of a bias in the investigator favoring the experimental arm.

8.2. Statistical Analyses

All efficacy analyses will be based on intent-to-treat (ITT) population. Some efficacy sensitivity analyses will also be performed on AT populations. All analyses will be performed by using SAS® Version 9.2 or higher.

The primary and secondary analyses of endpoints dependent on disease assessments (PFS, OR, DR, and CBR) will be based on investigator assessments of disease response and progression.

The BICR review as described in section 6.2. will be implemented as an auditing tool with the objective being to corroborate the analysis results of the primary endpoint (ie, investigator-assessed PFS) and to assist in the evaluation of potential bias. The analyses based on the BICR data are considered as supportive. The BICR audit approach is not intended to provide an alternative means of definitive analysis.

All primary, secondary, and supportive analyses will be tested at a significance level of 0.025 (1-sided test). No adjustments are planned for multiple testing/comparisons in the secondary and supportive hypothesis tests except OS.

8.2.1. Primary Efficacy Analysis

PFS based on the assessment of investigator will be summarized in the ITT population using the Kaplan-Meier method and displayed graphically where appropriate. The median event time and corresponding 2-sided 95% confidence interval for the median will be provided for PFS. The hazard ratio and its 95% Confidence interval (subject to the multiplicity adjustment at the final analysis) will be estimated. A log-rank test (1-sided, α=0.025) stratified by the presence of visceral metastases and sensitivity to prior hormonal therapy (two of the baseline stratification factors as listed in Section 6.5.2) will be used to compare PFS between the two treatment arms.

8.2.2. Sensitivity Analyses for the Primary PFS Endpoint

The primary efficacy analysis on the primary PFS endpoint is based on well-documented and verifiable progression events and deaths due to any cause. Other data are censored at the time of the last tumor assessment documenting absence of progressive disease and death. In addition, the sensitivity analysis (SA) on the primary PFS endpoint will be performed in determining whether the result of the primary PFS analysis is robust.

Sensitivity Analysis 1: Influence of Analysis Population - A sensitivity analysis will be performed to investigate whether the analysis population influenced the outcome of the primary endpoint PFS. The stratified log-rank test (1-sided, α=0.025) and Cox regression model will be used to evaluate the primary efficacy endpoint, PFS, in the AT population.

Sensitivity Analysis 2: Influence of the Use of Stratified Statistical Methods - A sensitivity analysis will be performed to investigate whether the use of stratified statistical methods influenced the outcome of the primary endpoint PFS. The unstratified log-rank test

(1-sided, α=0.025) and Cox regression model will be used to evaluate the primary endpoint of PFS in the ITT population.

Sensitivity Analysis 3: Influence of Stratification Factors and Covariates - A sensitivity analysis will be performed to investigate whether the stratification factors and important covariates (Section 6.5.1) influenced the outcome of the primary endpoint PFS. The multivariate Cox regression model will be used to evaluate the primary endpoint of PFS in the ITT population.

Sensitivity Analysis 4: Influence of Disease Assessment Scheduling - A sensitivity analysis will be performed to investigate whether deviations in disease assessment scheduling influenced the outcome of the primary endpoint PFS. If disease progression is documented between 2 scheduled tumor assessments, then the date of progression will be assigned to the earlier scheduled tumor assessment. In the event of death, the date of the endpoint will not be adjusted. Handling of missed disease assessments will be similar to that in the primary analysis except that any missed assessment will result in censoring. The stratified log-rank test (1-sided, α=0.025) and Cox regression model will be used to evaluate the investigator-assessed PFS with the modified event/censoring definitions in the ITT population.

Sensitivity Analysis 5: Influence of Deviations in Tumor Lesion Assessment - A sensitivity analysis will be performed to investigate whether deviations in tumor lesion assessment influenced the outcome of the primary endpoint PFS. If a lesion is classified as Indeterminate (IND) at Time Point X, and is adequately evaluated as PD at the next Time Point (X+1), then PD will be assigned to the Time Point X or earlier (the first date of the consecutive INDs) instead of the date of the next Time Point (X+1) as the primary analysis. The stratified log-rank test (1-sided, α=0.025) and Cox regression model will be used to evaluate the investigator-assessed PFS with the modified event/censoring definitions in the ITT population

Sensitivity Analysis 6: Influence of Bone-only Disease Patients - Three sensitivity analyses will be performed to investigate whether addition of bone-only disease patients influenced the outcome of the primary endpoint PFS.

1. For patients with bone-only disease, progression is defined as the appearance of ≥1 new lesions after the first on-study bone scan. In the following cases the patient will be censored at the date of prior tumor assessment with no PD in addition to the censoring rules defined in Section 6.1.1:

    - On-study fracture; or
    - On-study management of pain (palliative radiation therapy, palliative surgery); or
    - Clinical worsening not objectively confirmed (ECOG performance status increase from baseline by at least 2 points in 2 assessments); or

- On-study change of therapy

2. For patients with bone-only disease, progression is defined as the appearance of ≥1 new lesions after the first on-study bone scan. The following cases the patient will be considered as PD:

    - On-study fracture; or
    - On-study management of pain (palliative radiation therapy, palliative surgery);

      or

    - Clinical worsening not objectively confirmed (ECOG performance status increase from baseline by at least 2 points in 2 assessments); or
    - On-study change of therapy

3. Excluding bone-only disease patients from the analysis of the primary PFS endpoint.

Sensitivity Analysis 7: Influence of Missing Data - A sensitivity analysis will be performed to investigate whether the missing PFS data due to various reasons (eg, new anticancer treatment, lost to follow-up, and consent withdrawal, etc.) influenced the outcome of the primary endpoint of PFS. The following missing PFS data that may result in the censored PFS data in the primary analysis will be considered PFS events in addition to the documented PD and death in this sensitivity analysis:

- new anti-cancer treatment;
- lost to follow up;
- consent withdrawal (ie, no long willing to participate in study)
- Medication error without associated AE

The stratified log-rank test (1-sided, α=0.025) and Cox regression model will be used to evaluate the investigator-assessed PFS with the modified event/censoring definitions in the ITT population.

Sensitivity Analysis 8 (Supportive Analysis): Influence of Potential Investigator Bias - A sensitivity analysis will be performed to investigate whether the potential investigator (INV) bias on tumor assessment may influence the outcome of the primary endpoint of PFS. BICR and investigator PFS (event) data will be combined. For events identified by both BICR and INV, BICR data is to be used to determine event time. For patients who are censored by both BICR and INV, BICR (when applicable) data is to be used to determine the censoring time. The stratified log-rank test (1-sided, α=0.025) and Cox regression model will be used to evaluate the combined PFS data in the ITT population. *This analysis will be conducted only*

if the BICR is done for all the patients upon the request of the sponsor or the regulatory agencies.

8.2.3. Secondary Analyses

OS will be summarized in the ITT population using the Kaplan-Meier methods and displayed graphically where appropriate. The median event time and 2-sided 95% confidence interval for the median will be provided. A stratified (by the presence of visceral metastases and sensitivity to prior hormonal therapy) log-rank test will be used to compare OS between two treatment arms and the hazard ratio and its 95% confidence interval (subject to the multiplicity adjustment at the final analysis) will be estimated.

OS will be hierarchically tested for significance at the time of PFS analyses, provided the primary endpoint, PFS, is statistically significant at the interim and/or final PFS analyses. If OS does not yield a significant result at these analyses, OS will be tested at the final OS analysis. If PFS is not significant at the interim and/or final PFS analyses, the statistical significance of OS results will not be evaluated.

A naive hierarchical testing strategy is to test the secondary endpoint (OS) only if the primary endpoint (PFS) is significant, using the same level of significance α that was used to test PFS. However, as shown recently (Hung, Wang, and O'Neill, 2007) this hierarchical strategy when applied in a group sequential design does not control the overall type I error rate in the strong sense.

An alternative strategy is to apply the hierarchical group sequential testing with **separate error spending functions** at level 0.025 for PFS (primary) and OS (secondary) hypotheses proposed in this study to protect the family-wise error rate (FWER) strongly at level 0.025. Under this strategy, the secondary hypothesis for OS can be tested only when the primary hypothesis for PFS is rejected. That is, only if the primary hypothesis has been rejected (positive PFS) at the pre-planned interim analysis (or the final PFS analysis), the secondary hypothesis for OS can be tested on partial data at an interim analysis as well and again at the OS final analysis, if it is not significant before.

More specifically, here are some notations:

- let H_{0p} and H_{0o} denote the null hypotheses for testing PFS and OS, respectively;
- let $\alpha_p(t)$ and $\alpha_o(t)$ denote the alpha-spending functions for PFS and OS, respectively, at information fraction t;
- let $T_1 < T_2^* < T_3$ denote the time points for:

  T_1: pre-planned PFS interim analysis (driven by PFS events),

  T_2:planned final analysis of PFS and expected interim analysis of OS,

  T_3: planned final analysis of OS;

- let $t_p(T_1)$, $t_p(T_2)$ represent information fractions for PFS at time points T_1, T_2, respectively;
- let $t_o(T_2)$, $t_o(T_3)$ represent information fractions for OS at time points T_2 and T_3, respectively;
- let $u_p(t)$ and $u_o(t)$ are the efficacy stopping boundaries for PFS and OS, respectively, at information fraction t.

* In a special case when PFS is statistically significant at T_1, the timing of T_2 will no longer be driven by the targeted number of PFS events but rather by the number of OS events expected to be observed at time of planned final analysis of PFS. In other words, the timing of T_2 can change but the information fraction $t_o(T_2)$ should remain approximately the same.

Using the following testing algorithm or strategy, it is expected that the overall type-I error rate can be controlled at level 0.025 across both endpoints if the proper error spending approaches, each at level 0.025, are applied separately to both endpoints.

1. **Interim Analysis of PFS at T_1: Test PFS with $\alpha_p(t_p(T_1))$**

   If $u_p(t_p(T_1))$ is crossed, then H_{0p} is rejected and go to Bullet 3

   If $u_o(t_o(T_1))$ is not crossed, go to Bullet 2.

2. **Final Analysis of PFS at T_2 and Interim Analysis of OS at T_2: Test PFS at $\alpha_p(t_p(T_2))$ adjusting for $\alpha_p(t_p(T_1))$ and Test OS with $\alpha_o(t_o(T_2))$ at T_2***

   If $u_p(t_p(T_2))$ is crossed, then H_{0p} is rejected and IA of OS is conducted at T_2:

   if $u_o(t_o(T_2))$ is crossed, H_{0o} is rejected;

   else if $u_o(t_o(T_2))$ is not crossed, go to Bullet 4.

   If $u_p(t_p(T_2))$ is not crossed, the study does not meet the primary endpoint of PFS and no testing for OS statistical significance will be done.

3. **Interim Analysis of OS at T_2: Test OS with $\alpha_o(t_o(T_2))$ at T_2***

   If H_{0p} is rejected at IA, OS will be tested at T_2:

   if $u_o(t_o(T_2))$ is crossed, H_{0o} is rejected;

   else if $u_o(t_o(T_2))$ is not crossed, go to Bullet 4.

4. **Final Analysis of OS at T_3: Test OS with $\alpha_o(t_o(T_3))$ adjusting for $\alpha_o(t_o(T_2))$.**

   If $u_o(t_o(T_3))$ is crossed, H_{0o} is rejected;

   else if $u_o(t_o(T_3))$ is not crossed, statistical significance is not established for OS results.

* Although the formal statistical inference of OS result is not going to be drawn at T_1, a small alpha can be theoretically spent at T_1 for OS based on the pre-specified O'Brien-Fleming boundary generated by the Lan-DeMets spending function. It is expected that only approximately 47 OS events from both

arms will be observed at T_1; consequently, the theoretical alpha that can be spent for OS at T1 is highly close to 0 and therefore is negligible.

The 1-year, 2-year and 3-year survival probabilities will be provided with their 95% confidence intervals.

After the final results of Study A5481003 were communicated in April 2014, the enrollment rate of Study A5481023 accelerated significantly. Over 200 patients were enrolled into the study in the last two months of the enrollment period. Thus, 521 patients instead of the planned approximately 417 patients were randomized.

An interim analysis of OS was triggered by the significant PFS results observed at the planned PFS interim analysis. At the time of the analysis 19 (5.5%) patients had died in the palbociclib plus fulvestrant arm and 9 (5.2%) patients had died in the placebo plus fulvestrant arm. No conclusions on OS could be drawn due to immature OS data with such a small number of deaths.

The second interim OS analysis was performed based on 112 deaths and managed by the DMC. No detail interim analysis results have been shared with the sponsor and were not available to the public. The DMC only informed the sponsor that the analysis results did not cross the pre-specified O'Brien-Fleming efficacy boundary and therefore no conclusions regarding OS could be drawn at the time. As the investigator and patients are still blinded to the treatment, collection of OS data for the final OS analysis is being continued in the study,

Since more patients have been randomized in the study than it was originally planned, the previously determined final OS analysis based on 198 deaths becomes insufficient and inadequate in terms of the follow up time and data maturity. It would be difficult to answer the overall survival question for the patient population with 38% deaths (198 events out of 521 patients). Therefore, the time of final OS analysis is shifted from when the study reaches 198 deaths to approximately 60% deaths are observed in the 521 randomized patients. This adjustment will not only ensure more matured the survival data with longer follow-up time to be included in the final analysis, but also protect the study integrity and address the clinical question with much more robust results.

The nominal significance level for the final analyses of OS is determined by the Lan-DeMet spending function based on the O'Brien-Fleming boundary with the consideration of two interim analyses. The overall significance level for the efficacy analysis of OS will be preserved at 0.025 (one-sided test).

The number and proportion of patients achieving objective response (CR or PR) will be summarized in the ITT population and the ITT population with measurable disease at baseline along with the corresponding exact 2-sided 95% confidence interval calculated using a method based on Clopper-Pearson method. The odds ratio estimator and the corresponding 2-sided 95% will be calculated to contrast the treatment effects on response rates. A stratified 1-sided exact test will be used to compare ORR between two treatment arms.

The subgroup analyses will be conducted for the key efficacy endpoints that include PFS, OS, and ORR. For the time-to-event endpoints, the median (and other quartiles) of each treatment arm and the corresponding 2-sided 95% CIs will be provided. The HR estimate from the Cox model, the corresponding 2-sided 95% CI, and the p-value from the score test will also be provided for all the subgroups. The response rate of each treatment arm will also be summarized for each subgroup and presented along with the corresponding exact 2-sided 95% CI. The odds ratio estimator (with the corresponding 2-sided 95%) and the 1-sided stratified exact test may also be used to compare ORR between the treatment arms in each subgroup.

The following subgroups may be considered:

- subgroups defined by stratification factors and baseline patient characteristics (section 6.5.1)
- subgroups defined by ER mutation status (yes vs. no) using plasma DNA data when available
- subgroups defined by bone-only (yes vs. no) patients

The number and proportion of patients achieving clinical benefit response (CR or PR and SD≥24 weeks) will be summarized in the ITT population and the ITT population with measurable disease at baseline along with the corresponding exact 2-sided 95% confidence interval calculated using a method based on Clopper-Pearson method. The odds ratio estimator (with the corresponding 2-sided 95%) and the 1-sided stratified exact test may also be used to compare clinical benefit response rate between the treatment arms in each subgroup.

DR will be summarized using the Kaplan-Meier methods and displayed graphically where appropriate. DR will be calculated for the subgroup of patients with objective disease response. The median event time and 2-sided 95% confidence interval for the median will be provided.

8.2.4. Supportive Efficacy Analyses

The independent third-party core imaging laboratory assessment data will be evaluated as described in Section 8.1.5. Additional supportive analyses for the time-to event endpoints such as PFS and DR and ORR will be performed in the ITT population based on the independent third-party core imaging laboratory assessment as described in Sections 8.2.1 - 8.2.3. In addition, discordance rates between the investigator and independent Core imaging laboratory on assessing PFS data will be summarized by treatment arms. Concordance between the investigator and independent Core imaging laboratory assessments of tumor response will be summarized by treatment arms as well.

If sufficient data is available, outcome to subsequent treatments in terms of objective response rate and PFS after next line of treatment (PFS2) may be summarized by treatment arm in the ITT population. PFS2 is defined as the time from randomization to second

objective disease progression or death from any cause, whichever first. Patients alive and for whom a second objective disease progression has not been observed should be censored at the last time known to be alive and without second objective disease progression. Median PFS2 and the confidence interval for each arm will be estimated. If PFS2 cannot be reliably be determined, the outcome in terms of end-of-next-line treatment may be summarized by the treatment arm in the ITT population. For this analysis, an event is defined as end or discontinuation of next-line treatment, second objective disease progression, or death from any cause, whichever first. Time to end-of-next-line-treatment is defined as the time from randomization to the event defined in the previous sentence. Patients alive and for whom a second objective disease progression and end/discontinuation of next-line treatment have not been observed should be censored at the last time known to be alive and without second objective disease progression and end/discontinuation of next-line treatment. Median time to end-of-next-line-treatment and the confidence interval for each arm will be estimated.

In order to study the effect of the combination of palbociclib and an anti-hormonal agent (ie, fulvestrant or exemestane) on overall survival for women with HR+/HER2- advanced breast cancer, a supportive pooled OS analysis of individual patient data from this study and a collaborative IIR study (GEICAM /2013-02 “PEARL” Study) may be conducted. . The collaborative IIR study (subsequently referred to as PEARL study) is a Phase 3 study of exemestane plus palbociclib versus chemotherapy (i.e., capecitabine) in HR+/HER2- metastatic breast cancer patients with resistance to non-steroidal aromatase inhibitors. PFS is the primary endpoint and OS is one of the key secondary endpoints for PEARL study.

Although a qualitative interaction is not anticipated, no pooled analysis will be performed if the qualitative interaction does exist between the two studies. A qualitative interaction is defined as the heterogeneity between the two studies where the estimated hazard ratio of OS demonstrates a statistical significance in favor of palbociclib arm in one study and a statistical significance in favor of the control arm in the other study.

The same definition of OS, which is the time from date of randomization to date of death due to any cause, will be shared between this study and PEARL study and will be used in the pooled analysis. The pooled analysis will be conducted on an intent-to-treat (ITT) basis. The experimental group in the pooled analysis will include the patients in the ITT population of this study with the designated study drug assignment of palbociclib + fulvestrant and the patients in the ITT population of PEARL study with the designated study drug assignment of palbociclib + exemestane. The control group in the pooled analysis will include the patients in the ITT population of this study with the designated study drug assignment of fulvestrant and the patients in the ITT population of PEARL study with the designated study drug assignment of exemestane.

A log-rank test stratified by study will be used to compare the OS between the experimental and control groups based upon the pooled data and the two-sided p-value of the log-rank test will be calculated. Kaplan-Meier survival curves will be presented for the two groups and the median OS times and 95% CIs will be estimated using the pooled data. The 1-year survival probability will also be estimated using the Kaplan-Meier method based upon the pooled data and a two sided 95% CI for the log [-log(1 year survival probability)] will be calculated using a normal approximation with the Greenwood’s formula, and then back

transformed to give a CI for the 1-year survival probability itself (Brookmeyer and Crowley 1982). The 2-year, and 3-year survival probabilities will be estimated similarly based upon the pooled data.

A fixed effects meta-analysis proportional hazard model (Whitehead 2002), stratified by study, will be used to estimate the HR of experimental group to control group. Patient-level covariates including the important prognostic factors and baseline characteristics will be prospectively specified and consistently defined for both studies and be adjusted as fixed effects in the proportional hazard model.

The quantitative heterogeneity between the two studies will be assessed by fitting a proportional hazard model which included an additional study by treatment interaction term. The interaction term will be tested with a chi-squared distribution with 1 degree of freedom. The graphical techniques (eg, forest plot) may also be used to further explore the heterogeneity between the two studies.

It is expected that the interim OS data be analyzed and summarized separately when the final analysis of PFS in the individual study is conducted. Consequently, the planned pooled analysis may be performed on the interim OS data only if the final analysis of PFS has been completed and the interim OS results generated separately for both studies. The pooled analysis may also be performed on the final OS data when the results of final OS analyses are available for each individual study. No multiple testing adjustments will be made in the pooled OS analysis.

8.2.5. Standard Analyses

Descriptive statistics will be used to summarize study conduct and patient disposition, baseline characteristics, and treatment administration/compliance.

- **Study Conduct and Patient Disposition** - an accounting of the study patients will be tabulated including randomized (per stratification factors), treated, accrual by study center, assessed for AEs, laboratory data, biomarkers, PK, and QTc, etc. Patients not meeting the eligibility criteria will be identified. Patients not completing the study will be listed along with the reason for their premature discontinuation. Reasons for premature discontinuation will be summarized. Randomization errors and stratification errors will be described.
- **Baseline Characteristics** - patient characteristics such as patient age, height, weight, race, ethnicity, ECOG performance status, primary diagnosis, HR and HER2 status, prior therapy (radiotherapy, surgery, systemic therapy), baseline disease site, prior medication, medical history, and signs and symptoms at study entry will be summarized in frequency tables, and descriptive statistics will be provided for quantitative variables.
- **Treatment Administration and Compliance**
    - **Extent of Treatment**

      The extent of treatment will be summarized as follows:

        - The number and % of patients on treatment and off for each reason

- Treatment assigned vs. actual received
- The number and percent of patients beginning 1, 2, 3, 4, 5+ cycles of either study drug
- The number of cycles started (median, minimum, maximum) will be reported (overall and by study treatment).
- Duration of treatment (weeks) (overall and by study treatment)
- Cumulative dose and relative dose intensity (see Appendix 10.6 for details) (overall and by cycle; by study treatment)

- **Treatment Delays and Dose Modifications**

  Dose adjustment is permitted for palbociclib/placebo only. The fulvestrant dose cannot be adjusted, its dosing can only be delayed or interrupted. However, a single fulvestrant injection can be skipped in case of a fulvestrant-related toxicity. Treatment delays and dose modifications of study treatments will be summarized as follows including number and percent:

  - The number of patients with at least one palbociclib dose reduction and the number of patients with at least one palbociclib or fulvestrant dose omission at any time during drug administration will be reported.
  - The number of patients with at least one palbociclib dose reduction due to an adverse event will be reported.
  - The number of patients with at least one palbociclib dose delay (ie, start of following cycle is delayed) and percentage due to each reason for the delay will be reported

- **Concomitant medications and Non-drug treatments**

  Concomitant and non-drug treatments refer to all drug and non-drug treatments taken while on active treatment (during the effective duration of study treatment), whether or not they are recorded at baseline (ie, have stop day greater than or equal to day 1 relative to first dose of study drug). Concomitant medication will be summarized in frequency tables by treatment. All the collected data of goserelin (including the start and stop date for each dose of goserelin) will be listed in a separate table. In addition, the duration of goserelin will be summarized for pre/peri-menopausal patients by the general descriptive statistics including mean, median, standard deviation and range. The number of pre/peri-menopausal women who took goserelin prior to randomization and the number of pre/peri-menopausal women who took goserelin during the study will also be provided.

- **Follow-Up Therapy**

Follow-up cancer therapy will be summarized by treatment as patients with number of regimens (0, 1, 2, ≥3), and patients with particular agents.

8.2.6. Safety Analyses

Listings of AE, SAE, death, lab data, vital signs, and physical examinations will be provided according to reporting standard.

8.2.6.1. Adverse Events

All patients treated with at least one dose of study treatment (ie, palbociclib/Placebo or fulvestrant) will be included in all the safety analyses.

Adverse events will be classified using the medical dictionary for regulatory activities (MedDRA) classification system. The severity of the toxicities will be graded according to the NCI CTCAE v.4.0 whenever possible (http://evs.nci.nih.gov/ftp1/CTCAE/About.html).

It should be recognized that most studies are not designed to reliably demonstrate a causal relationship between the use of a pharmaceutical product and an adverse event or a group of adverse events. Except for select events in unique situations, studies do not employ formal adjudication procedures for the purpose of event classification. CCI [redacted] CCI [redacted] CCI [redacted] Adverse events will be summarized by treatment and by the frequency of patients experiencing treatment emergent adverse events corresponding to body systems and MedDRA preferred term. Adverse events will be graded by worst NCI CTCAE v.4.0 grade. Adverse events will be summarized by cycle and by relatedness to trial treatment. Detailed information collected for each AE will include a description of the event, duration, whether the AE was serious, intensity, relationship to study drug, action taken, and clinical outcome. Emphasis in the analysis will be placed on AEs classified as treatment emergent.

Adverse events leading to discontinuation of trial treatment, events classified as NCI CTCAE v.4.0 Grade 3 or higher, trial drug related events, and serious adverse events will be considered with special attention.

The percentage of patients with an event will be calculated using the number of patients in the as-treated population as the denominator. The denominator for summary tables for each laboratory parameter will be all patients in the as-treated population with at least one evaluable cycle for that parameter.

A 3-tier approach will be used to summarize AEs.

Tier-1 events: These are pre-specified events of clinical importance and are maintained in a list in the product's Safety Review Plan.

Tier-2 events: These are events that are not tier-1 but are "common". A MedDRA preferred term is defined as a tier-2 event if there are at least 10% for all a grades in any treatment

group. For grade 3/4/5 analysis, the events should be reported in at least 5% patients in any treatment group.

Tier-3 events: These are events that are neither tier-1 nor tier-2 events.

- For Tier-1 events, the MedDRA preferred term, treatment arm, n (%) for each MedDRA preferred term per arm, risk difference, 95% confidence interval and P-values for the risk difference will be provided. Graphical format may be presented as well. Presented in descending p-value order.

- For Tier-2 events, the MedDRA preferred term, treatment arm, n (%) for each MedDRA preferred term per treatment arm, risk difference and 95% confidence intervals for the risk difference will be provided in tabular format. Table by AE for All Grade and for Grade 3/4/5 will be provided. Graphical format may be presented as well. Presented in descending risk difference order.

- Tier-3 events will be presented by observed event proportions. The following will be provided:

    - Incidence and grades of treatment emergent (all causality, preferred term, and by System Organ Class) AEs for all cycles combined.

    - Incidence and grades of treatment emergent (all causality, preferred term) AEs for all cycles combined in descending frequency order.

    - Incidence and grades of treatment emergent (treatment related, preferred term and by System Organ Class) AEs for all cycles combined.

    - Incidence and grades of treatment emergent (treatment related, preferred term) AEs for all cycles combined in descending frequency order.

- Disease progression will not be included

- There will be no adjustment for multiplicity

The following summaries of treatment emergent adverse events will also be provided by arm:

- Discontinuations Due to Adverse Events including causality: all cause, treatment related, including relationship to specific study treatment of fulvestrant, palbociclib, and placebo

- Temporary Discontinuations or Dose Reductions Due to Adverse Events including causality and relationship to specific study treatment of fulvestrant, palbociclib, and placebo

- Treatment-Emergent Adverse Events (All Causality, and Treatment Related) including the number of patients evaluable for adverse events, total number of adverse events (counting each unique preferred term across all patients), number of

patients with serious adverse events, number of patients with Grades 3 and 4 adverse events, number of patients with Grade 5 adverse events, and number with dose reductions or temporary discontinuations due to adverse events

- Treatment-Emergent Adverse Events by MedDRA System Organ Class, Preferred Term and Maximum NCI CTCAE v.4.0 Grade (All Causality, and Treatment related)
- Treatment-Emergent Adverse Events by MedDRA Preferred Term sorted by Descending Order of AE Frequency (All Causality, and Treatment related)
- Treatment-Emergent Adverse Events by Preferred Term – Grade 3/4/5 events with number of patients experienced Grade 3-5 AEs and total number of Grade 3-5 AEs, sorted by Descending Order of AE Frequency (All Causality, and Treatment Related)

A summary of Serious Adverse Events and listing of deaths reported as serious adverse events will be provided.

8.2.6.2. Laboratory abnormalities

Hematologic, chemistry and urinalysis laboratory data will be summarized by cycle. The hematologic and chemistry laboratory results will be graded according to the NCI CTCAE v.4.0 severity grade. For parameters for which an NCI CTCAE v.4.0 scale does not exist, the frequency of patients with values below, within, and above the normal range for the local lab will be summarized. Each patient will be summarized by the worst severity grade observed for a particular laboratory parameter. This will be provided for all cycles as well as by cycles.

8.2.7. Patient Reported Outcomes Analyses

Analysis of the PRO endpoints will be based on the PRO evaluable population.

The PRO analysis endpoints will be based on the instruments QLQ-C30, QLQ-BR, and EQ-5D. . For each treatment group and at each time point, the number and percentage of patients who completed these instruments will be summarized, as will the reasons for non-completion of these measures. An instrument is considered completed if at least one item was answered by the patient.

EORTC QLQ-C30.

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

For each of the 15 scales, statistical comparison between the two treatment groups will be based on a longitudinal repeated measures analysis using a mixed effects model. The variables in the model will be treatment, time, treatment-by-time, with baseline used as a covariate. Parameter estimates will be based on a restricted maximum likelihood method and

an unstructured covariance matrix will be used. No adjustments for multiple comparisons will be made.

For the symptom scale of pain, a time to deterioration (TTD) analysis will be carried out using survival analysis methods. Deterioration will be defined increase in score of 10 points or greater from baseline. .

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

EORTC QLQ-BR23.

This questionnaire contains 23 questions organized into 4 functional scales and 4 symptom scales. As with C30, the analysis of the BR23 scales will consist of descriptive statistics on means and changes from baseline, and a between treatment comparison using a longitudinal mixed effects model. Also, as with C30, graphical displays of means and changes from baseline over time will also be provided for each BR23 scale. .

EQ-5D Health Index

Analysis of the EQ-5D health index will consist of descriptive statistics on means and changes from baseline, a between treatment comparison using a longitudinal mixed effects model, and graphical displays of means and changes from baseline over time, as with the EORTC scales. In addition, there will be a health status profile analysis consisting of a display of the number and percentage of patients in each of the 3 response levels for each of the 5 dimensions at each visit.

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

Analysis of EQ-5D general health status will consist of descriptive statistics on means and changes from baseline, a between treatment comparison using a longitudinal mixed effects model, and graphical displays of means and changes from baseline over time, as with the EORTC scales.

8.2.8. QTc Analyses

All ECGs obtained during the study will be evaluated for safety

Baseline will be defined as the average of the triplicate readings closest prior to dosing on the first day that study medication is administered

Baseline and changes from baseline in RR, PR, QT, QTc, QTcF, QRS and heart rate will be summarized by treatment and time post-dose.

ECG endpoints and changes from baseline (QTc, QTcF, PR, QRS RR and HR), over all measurements taken post-dose, will also be summarized descriptively by treatment using categories as defined in ICH E14 (ie, QTcF (ms): 450-480; 480-500; ≥500) Numbers and percentages of subjects meeting the categorical criteria will be provided and individual

values listed in the study report.Unplanned ECGs are captured when there are AE suggestive of arrhythmia, and such unplanned records will not be included as part of the summaries. All ECG recordings will be included in the listings

8.2.9. Analyses for Ocular Assessment Data

Ocular assessment data will be summarized only if sufficient data are collected from the study. All the collected ocular assessment data will be listed. The data may be combined with the ocular assessment data collected from other palbociclib studies to generate the pooled results that are not to be included in this clinical study report.

8.2.9.1. Snellen Best Corrected Visual Acuity and Refraction

Snellen visual acuity will be assessed by using a standard wall or projection chart before implementing any procedures that can affect vision (eg, pupil dilation, tonometry, and gonioscopy). The same optotype will be used throughout the study for a specific patient, and the right eye should be tested first. The refractive error will be determined at the Screening visit.

The line read with 2 or fewer errors will be recorded. If 3 of the 5 letters on a line are read correctly, the patient will be given credit for that line. For example, if the patient reads 20/25 +3, 20/20 will be recorded.

A decrease in best-corrected visual acuity of 3 lines or more from the Screening visit will be reported as an adverse event and included in the AE tables. An adverse event of visual acuity will be counted from the following lines: 20/20 or better, 20/25, 20/30, 20/40, 20/50, 20/60, 20/70, 20/80, 20/100, 20/125, 20/150, and 20/200. If the acuity at screening is better than 20/20, the decrease will be calculated from 20/20. The adverse event of visual acuity will be reported in the AE tables. If sufficient data are collected in the study, a table may be created to summarize the number of subjects (and the proportions of subjects) with visual acuity of improvement, no change, decrease by 1 line, 2 line and >2 lines by treatment groups and left/right eyes.

In the event of a decrease in visual acuity of 3 lines or more from screening, refraction will be rechecked at all subsequent study visits. A change in refraction power (spherical or cylindrical) of +/– 1.25 diopters compared with the screening examination will be reported as an adverse event and included in the AE tables.

8.2.9.2. Intraocular Pressure Measurement

Intraocular pressure (IOP) will be measured using a calibrated Goldmann applanation tonometer. Both eyes will be tested, with the right eye preceding the left eye. The operator will initially set the dial at 10 mm Hg, then look through the slit lamp and adjust the dial to take the reading, and then record the results, including the time assessment is made.

Any IOP increase of greater than 10 mmHg above baseline or any IOP that increases above 25 mm Hg will be reported as an adverse event and included in the AE tables. If sufficient data are collected in the study, a summary table that presents the descriptive statistics (eg,

mean, median, and range) of IOP and the maximum IOP change from baseline by treatment arm and left/right eyes may be created.

8.2.9.3. Slit-Lamp Biomicroscopy

Slit-lamp biomicroscopy with fluorescein will be performed. At each scheduled visit, deposition of pigment on the corneal endothelial layer or the lens capsule or any abnormalities of the lids, conjunctivae, cornea, anterior chamber, iris, or lens (see lens grading) will be graded as mild, moderate, or severe. Intraocular Inflammation Grading Scale for Biomicroscopy will be u sed. During the study, any new finding or deterioration from baseline findings should be reported as an adverse event and be included in the AE tables.

8.2.9.4. Lens Grading

The Wisconsin AREDS 2008 Clinical Lens Opacity Grading Procedure will be used. Each type of opacity will be graded in half steps from <1 to >3 (1=mild, 2=moderate and 3=severe). If sufficient data are collected in the study, a table that summarizes the lens grading results by treatment arm, left/right eyes, different types of opacity, and visit may be created.

8.2.9.5. Funduscopy (Ophthalmoscopy)

Funduscopy (Ophthalmoscopy) will be performed after dilation of the pupils to examine the vitreous body, retina, and optic nerve head. At screening, any abnormalities and pathologic findings will be graded as mild, moderate, or severe. Any new findings or deterioration from baseline findings will be reported as an adverse event and included in the AE tables.

8.2.10. Pharmacokinetic Analyses

For the pharmacokinetic analyses, four unique drug combinations will be studied: 1) fulvestrant + placebo, 2) fulvestrant + goserelin + placebo, 3) fulvestrant + palbociclib, and 4) fulvestrant + goserelin + palbociclib.

In the early safety review, the first approximately 40 patients will have the concentration data of palbociclib, fulvestrant and goserelin listed by analyte, patient, collection time and day and unique drug combination. Summary statistics will be provided for concentrations by analyte, collection time (Nominal Study Day), and unique drug combination. For fulvestrant and goserelin concentrations, graphical comparisons will be provided (box-plot by treatment) as described below.

For all patients on the palbociclib treatment arm (including those in the early safety review), palbociclib trough concentrations will be listed by patient, collection time (Nominal Study Day), and unique drug combination. Summary statistics will be provided by collection time (Nominal Study Day), and unique drug combination. If fulvestrant or goserelin do not have an effect on palbociclib PK, an average of the trough concentrations of palbociclib on Day 15 of Cycles 1 and 2 will be listed by patient for the corresponding unique drug combination, and summary statistics will be provided. If both fulvestrant and goserelin have no effect on palbociclib PK, summary statistics will be provided including the average of the trough

concentrations of palbociclib on Day 15 of Cycles 1 and 2 from all patients across all unique drug combinations.

The relationship between palbociclib trough concentrations (on Day 15 of Cycles 1 and 2) and potential covariates will be evaluated. If there is no treatment effect, data from all treatment groups will be combined for covariate analysis, and the mean of the Cycle 1 Day 15 and Cycle 2 Day 15 palbociclib troughs for each patient will be used in the analysis. Analysis of Covariance (ANCOVA) may be used to analyze the effect of baseline characteristics on pharmacokinetic parameters. The ANCOVA may include baseline characteristics (such as age, gender, race, renal impairment status, etc.) as covariates.

8.2.10.1. Baseline Characteristics Analyses

Patient characteristics such as age, gender, weight, height, BMI, ethnicity, diagnosis, extent of disease, ECOG performance status, smoking history and creatinine clearance will be summarized for the PK evaluable population per unique drug combination and study.

8.2.10.2. Drug-Drug Interaction Analyses

The drug combination effect on palbociclib, fulvestrant and goserelin will be determined by constructing 90% confidence intervals around the estimated difference between the Test and Reference treatments using the models specified below for analyzing natural log transformed Ctrough data. The statistical analyses models are pre-specified below, but not limited to these models. The validity of pre-specified analyses will be evaluated by the unblinded pharmacokineticist and statistician. Additional analyses may also be conducted if warranted by the data.

Three analytes will be compared: palbociclib, fulvestrant and goserelin. For palbociclib, Ctrough data from study A5481023 will be the Test while palbociclib Ctrough from historical data will be the Reference. For fulvestrant, fulvestrant± goserelin+palbociclib combination will be the test while fulvestrant±goserelin+ placebo combination will be the reference. For goserelin, fulvestrant+goserelin+palbociclib combination will be the test while fulvestrant+goserelin+ placebo combination will be the reference. Estimates of the adjusted mean differences (Test-Reference) and corresponding 90% confidence intervals (CI) obtained from the models will be exponentiated to provide estimates of the ratio of adjusted geometric means (Test/Reference) and 90% CIs for the ratios.

Residuals from the models will be examined for normality and the presence of outliers via visual inspection of plots of residuals versus predicted values and normal probability plots of residuals. A summary of findings will be described in the data summary from these analyses but will not be included in the report. If there are major deviations from normality then the effect of these on the conclusions will be investigated through alternative transformations and/or analyses excluding outliers. Justification for any alternatives to the planned analyses will be given in the report.

8.2.10.2.1. Palbociclib

- Base model: One-way analysis of variance (ANOVA) will be used to analyze the natural log transformed Ctrough with unique drug combination as a single factor.

The statistical summary table for the base model will provide the least square mean (lsm) for the test and reference, geometric mean ratio (Test/Reference) and its 90% CI.

- Full model: Analysis of covariance (ANCOVA) will be used to analyze the natural log transformed Ctrough with unique drug combination as a main effect along with age and baseline body weight as covariates. The statistical summary table for the full model will provide the lsm for the test and reference, adjusted geometric mean ratio (Test/Reference) and its 90% CI.

- Final model: Regression analysis will be performed on natural log transformed Ctrough against unique drug combination and covariates identified by backward selection process from the full model at significance level of 10%. If unique drug combination is identified as one of the significant predictors, a statistical summary table for the final model will be provided with the lsm for the test and reference, adjusted geometric mean ratio (Test/Reference) and its 90% CI.

8.2.10.2.2. Fulvestrant

- Base model: One-way analysis of variance (ANOVA) will be used to analyze the natural log transformed Ctrough with unique drug combination as a single factor. The statistical summary table for the base model will provide the least square mean (lsm) for the test and reference, geometric mean ratio (Test/Reference) and its 90% CI.

8.2.10.2.3. Goserelin

- Base model: One-way analysis of variance (ANOVA) will be used to analyze the natural log transformed Ctrough with unique drug combination as a single factor. The statistical summary table for the base model will provide the least square mean (lsm) for the test and reference, geometric mean ratio (Test/Reference) and its 90% CI.

8.2.10.3. Exposure/Response Analysis

The relationship between exposure and efficacy/safety endpoints will be explored, as necessary, based on emerging efficacy and safety data. All patients treated with palbociclib (Palbociclib PK Analysis Set) and for whom drug plasma concentration results (from at least 1 visit) are available will be included in the analysis. The results of these modeling analyses may be reported separately from the clinical study report.

8.2.10.4. Adverse Events with PK Analyses

For patients experiencing unexpected or serious adverse events, or adverse events thatlead to discontinuation who have PK blood samples collected for palbociclib or fulvestrant (or goserelin, if applicable) and have at least one measured plasma drug concentration: Concentration data of palbociclib,fulvestrant and goserelin listed by analyte, patient, collection time and day and unique drug combination as well as the corresponding adverse

event analysis data. This will be done, if applicable, and separately for the early safety review data set and for the palbociclib PK analysis data set.

8.2.10.5. Pharmacokinetic Analyses Reporting

The study report will include, but not be limited to, the following listings, tables, and figures for PK data and analysis.

For the first approximately 40 patients participating in the early safety review:

1. Summary statistics of PK Subject Evaluation Groups and their Demographic Characteristics

2. Summary statistics of plasma predose concentration (ng/mL) for palbociclib, fulvestrant and goserelin, respectively, versus time (Nominal Study Day) by unique combination

3. Summary statistics of plasma mean predose concentrations (ng/mL) for Day 1 and Day 15 of Cycles 1, 2 and 3 visits for palbociclib by unique drug combination

4. Summary statistics of plasma mean predose concentrations (ng/mL) for palbociclib (125 mg dose only) historical control by study and all studies combined (A5481001-3).

5. Summary statistics of plasma mean predose concentrations (ng/mL) for Day 1 and Day 15 of Cycles 1, 2 and 3 visits for fulvestrant and goserelin by unique drug combination, respectively

6. Box-plot of the plasma predose concentration (ng/mL) versus study day for palbociclib, fulvestrant and goserelin, respectively, by unique combination - mean (+/- SD); median and individual. For palbociclib C_{trough} box-plot, historical control (combined data from studies A5481001-3 should be also plotted as reference.

7. Box-plot of the plasma predose concentration (ng/mL) versus unique drug combination for palbociclib, fulvestrant and goserelin, respectively, by unique study day - mean (+/- SD); median and individual

8. Box-plot of the mean plasma predose concentrations (ng/mL) of the Day 15 of Cycles 1 and 2 visits for palbociclib (ng/mL) by unique combination - mean (+/- SD); median and individual. Corresponding historical control data (combined from studies A5481001-3 should be also plotted as reference.

9. Box-plot of the mean plasma predose concentrations (ng/mL) of the Day 15 of Cycles 1 and 2 visits for palbociclib (ng/mL) by unique combination - mean (+/- SD); median and individual. Corresponding historical control data from studies A5481001-3 should each be plotted as reference.

10. Box-plot of the mean plasma predose concentrations (ng/mL) of the Day 15 of Cycle 1 for palbociclib (ng/mL) by unique combination - mean (+/- SD); median and individual. Corresponding historical control data from studies A5481001-3 should each be plotted as reference.

11. Box-plot of the mean plasma predose concentrations (ng/mL) of the Day 15 of Cycle 2 for palbociclib (ng/mL) by unique combination - mean (+/- SD); median and individual. Corresponding historical control data from studies A5481001-3 should each be plotted as reference.

12. Box-plot of the mean plasma predose concentrations (ng/mL) of Day 1 and Day 15 of Cycles 1, 2 and 3 visits for fulvestrant (ng/mL) by unique combination - mean (+/- SD); median and individual, respectively

13. Box-plot of the mean plasma predose concentrations (ng/mL) of Day 1 and Day 15 of Cycles 1, 2 and 3 visits for goserelin (ng/mL) by unique combination - mean (+/- SD); median and individual, respectively

14. Statistical comparison of adjusted geometric mean of palbociclib C_{trough} (ANCOVA and/or ANOVA analysis summary)

15. Statistical comparison of adjusted geometric mean of fulvestrant C_{trough} (ANOVA analysis summary)

16. Statistical comparison of adjusted geometric mean of goserelin C_{trough} (ANOVA analysis summary)

17. Listing of plasma predose concentration (ng/mL) of palbociclib, fulvestrant and goserelin versus time by visit

18. Listing and descriptive summary of plasma mean predose concentrations (ng/mL) of the Day 15 of Cycles 1 and 2 visits for palbociclib by unique drug combination

19. Listing and descriptive summary of plasma mean predose concentrations (ng/mL) of the Day 1 of Cycles 2 and 3 visits for fulvestrant and goserelin by unique drug combination, respectively

20. Listing of plasma sample time deviation and sample comments for palbociclib, fulvestrant and goserelin assay

21. Listing of Plasma Samples for Palbociclib (PD-0332991), Fulvestrant, and Goserelin Assays That Are Excluded From the Analysis

For all patients participating in the palbociclib PK analysis set:

1. Summary statistics of PK Subject Evaluation Groups and their Demographic Characteristics

2. Summary statistics of plasma predose concentration (ng/mL) for palbociclib, versus time (Nominal Study Day) by unique combination

3. Summary statistics of plasma mean predose concentrations (ng/mL) of the Day 15 of Cycles 1 and 2 visits for palbociclib by unique drug combination

4. Box-plot for the plasma predose concentration (ng/mL) versus study day for palbociclib by unique combination - mean (+/- SD); median and indi vidual

5. Box-plot for the plasma predose concentration (ng/mL) versus unique drug combination for palbociclib by unique study day - mean (+/- SD); median and individual

6. Box-plot of the individual patient mean plasma predose concentrations (ng/mL) of the Day 15 of Cycles 1 and 2 visits for palbociclib (ng/mL) by unique combination - mean (+/- SD); median and individual

7. Box-plot of the plasma predose concentration (ng/mL) for palbociclib by significant covariate(s) if applicable - mean (+/- SD); median and individual

8. Listing of plasma predose concentration (ng/mL) of palbociclib versus time by visit

9. Listing and descriptive summary of plasma mean predose concentrations (ng/mL) of the Day 15 of Cycles 1 and 2 visits for palbociclib by unique drug combination

10. Listing of plasma sample time deviation and sample comments for palbociclib assay (all patients)

11. Listing of Plasma Samples for Palbociclib (PD-0332991), Fulvestrant, and Goserelin Assays That Are Excluded From the Analysis

12. ANCOVA and/or ANOVA analysis summary table (if applicable)

For the PK/Adverse Event Analysis Set (if applicable, and separate for early safety review data set and palbociclib PK analysis set):

1. Listing of plasma predose concentration (ng/mL) of palbociclib, fulvestrant and goserelin by analyte, patient, collection time and day and unique drug combination as well as the corresponding adverse event data (suspect drug/dose, action taken, discontinuation day, discontinuation date or event onset date, event onset day, event stop day, AE term, investigator causality, clinical outcome/seriousness).

8.3. Summary of Key Efficacy Analyses

Type of Analysis	Endpoint	Analysis Set	Statistical Method
Primary Analysis	PFS	ITT Investigator assessment	• Stratified log-rank test (stratified by the presence of visceral metastases and sensitivity to prior hormonal therapy, 1-sided p-value), • K-M method (median and 95% CIs) • HR and 95% CIs from the stratified Cox model
Sensitivity analysis 1	PFS	AT Investigator assessment	• Stratified log-rank test (stratified by the presence of visceral metastases and sensitivity to prior hormonal therapy, 1-sided p-value), • K-M method (median and 95% CIs) • HR and 95% CIs from the stratified Cox model
Sensitivity analysis 2	PFS	ITT Investigator assessment	• Unstratified log-rank test (1-sided p-value), • K-M method (median and 95% CIs) • HR and 95% CIs from the Cox model
Sensitivity analyses 3	PFS	ITT Investigator assessment	• HR 95% CIs from the multivariate Cox model
Sensitivity analysis 4-7	PFS	ITT Investigator assessment	• Stratified log-rank test (stratified by the presence of visceral metastases and sensitivity to prior hormonal therapy, 1-sided p-value), • K-M method (median and 95% CIs) • HR and 95% CIs from the stratified Cox model
Supportive Analysis Sensitivity analysis 8	PFS	ITT Investigator assessment and BICR (only if 100% BICR is conducted)	• Stratified log-rank test (stratified by the presence of visceral metastases and sensitivity to prior hormonal therapy, 1-sided p-value), • K-M method (median and 95% CIs) • HR and 95% CIs from the stratified Cox model
Supportive Analysis	PFS	ITT BICR	• Stratified log-rank test (stratified by the presence of visceral metastases and sensitivity to prior hormonal therapy, 1-sided p-value). • K-M method (median and 95% CIs) • HR and 95% CIs from the stratified Cox model • Unstratified log-rank test (1-sided p-value). • HR and 95% CIs from the stratified Cox model
Supportive Analysis	PFS	ITT BICR	• PhRMA and NCI estimation results • Concordance/Discordance summary results
Secondary analyses	OS	ITT	• Stratified log-rank test (stratified by the presence of visceral metastases and sensitivity to prior hormonal therapy, 1-sided p-value) • K-M method (median and 95% CI) • HR and 95% CIs from the stratified Cox model • Wilcoxon test and RPSFT method (when appropriate)
	1, 2, and 3-year survival	ITT	• K-M method (95% CI)

Type of Analysis	Endpoint	Analysis Set	Statistical Method
Secondary analyses	OR	ITT Investigator assessment	• Stratified exact test (stratified by the presence of visceral metastases and sensitivity to prior hormonal therapy, 1-sided p-value), • Exact CI based on Clopper-Pearson method (95% CI) • Odds ratio and 95% CI
Supportive analyses	OR	ITT BICR	• Stratified exact test (stratified by the presence of visceral metastases and sensitivity to prior hormonal therapy, 1-sided p-value), • Exact CI based on Clopper-Pearson method (95% CI) • Odds ratio and 95% CI
Secondary analyses	OR	ITT Investigator assessment	• Stratified exact test (stratified by the presence of visceral metastases and sensitivity to prior hormonal therapy, 1-sided p-value), • Exact CI based on Clopper-Pearson method (95% CI) • Odds ratio and 95% CI
Supportive analyses	OR	ITT BICR	• Stratified exact test (stratified by the presence of visceral metastases and sensitivity to prior hormonal therapy, 1-sided p-value), • Exact CI based on Clopper-Pearson method (95% CI) • Odds ratio and 95% CI
Secondary analyses	DR	ITT patients with a CR or PR Investigator assessments	• K-M method (median and 95% CI)
Supportive analyses	DR	ITT patients with a CR or PR BICR	• K-M method (median and 95% CI)

Abbreviations: AT=as-treated; BICR=Blinded, Independent Central Review; CBR=clinical benefit rate; CR=complete response; CI=confidence interval; DR=duration of response; HR=hazard ratio; ITT=intent-to-treat; K-M=Kaplan-Meier; OR=objective response; OS=overall survival; PFS=progression-free survival; PR= partial response; SD=stable disease;

9. REFERENCES

1. Amit O, Mannino F, Stone AM, et al: "Blinded independent central review of progression in cancer clinical trials: results from a meta-analysis". European Journal of Cancer. 2011; 47:1772-1778.

2. Brookmeyer R, Crowley JJ. A confidence interval for the median survival time. Biometrics 1982; 38:29-41.

3. Cui L, Hung HMJ, Wang S, et al. Modification of sample size in group-sequential clinical trials. Biometrics. 1999; 55: 853-7.

4. Dodd LE, Korn EL, Freidlin B, et al. Blinded independent central review of progression-free survival in phase III clinical trials: important design element or unnecessary expense? J Clin Oncol 2008;26:3791-6.

5. Dodd E L, Korn L E, Freidlin B, Gray R, Bhattacharya S. An audit strategy for progression-free survival. Biometrics, Vol 67, Issue 3; 1092-99, 2011.

6. Dmitrienko A, Tamhane AC, Bretz F (2010). Multiple Testing Problems in Pharmaceutical Statistics. Chapman and Hall/CRC, Boca Raton.

7. Eisenhauer E, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, Dancey J, Arbuck S, Gwyther S, Mooney M, Rubinstein L, Shankar L, Dodd L, Kaplan R, Lacombe D, Verweij J. New response evaluation criteria in solid tumors: Revised RECIST guideline version 1.1. Eur J Can 45: 228-47, 2009.

8. Haybittle JL. Repeated assessment of results in clinical trials of cancer treatment. Brit. J. Radiology 1971; 44:793-797.

9. National Cancer Institute (NCI). Common Terminology Criteria For Adverse Events (CTCAE). http://evs.nci.nih.gov/ftp1/CTCAE/About.html

10. Peto, R., Pike, M. C., Armitage, P., Breslow, N.E., Cox, D.R., Howard, S.V., Mantel, N., McPherson, K., Peto, J., Smith, P.G. Design and Analysis of Randomized Clinical Trials Requiring Prolonged Observation of Each Patient: I. Introduction and Design. Brit. J. Cancer 1976; 34:585–612.

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

12. Whitehead, A. Meta-Analysis of Controlled Clinical Trials; John Wiley & Sons, Ltd: Chichester, England, 2002.

13. Wright GL, Blum J, Krekow LK, et al. Randomized Phase II Trial of Fulvestrant with or without Dasatinib in Postmenopausal Patients with Hormone Receptor-Positive Metastatic Breast Cancer Previously Treated with an Aromatase Inhibitor. Cancer Research: December 15, 2011; Volume 71, Issue 24, Supplement 3 doi: 10.1158/0008-5472.SABCS11-PD01-01

14. Zhang, J, Chen H, He K, et al. Evaluation of blinded independent central review of tumor progression in oncology clinical trials: a meta-analysis. Therapeutic Innovation & Regulatory Science. 2012 DOI: 10.1177/0092861512459733

10. APPENDICES

10.1. Schedule of Activities

The Schedule of Activities table provides an overview of the protocol visits and procedures. Refer to STUDY PROCEDURES and ASSESSMENTS sections of the protocol for detailed information on each procedure and assessment required for compliance with the protocol. The investigator may schedule visits (unplanned visits) in addition to those listed on the schedule of activities, in order to conduct evaluations or assessments required to protect the wellbeing of the patient.

Protocol Activity	**Screening**	**Active Treatment Phase [a] - One Cycle = 28 days**			**End of Treatment/Withdrawal [c]**	**Post-Treatment Follow-Up [d]**
		Cycles 1 and 2		**Cycles ≥3**		
Study Day	**Within 28 days prior to randomization unless specified otherwise**	**Day 1 [b]**	**Day 15**	**Day 1**		
Visit Window		**±-2 days**	**±2 days**	**±7 days [a]**		**±7 days**
Informed Consent [e]	X					
Medical/Oncological History [f]	X					
Baseline Signs/Symptoms [g]		X [g]				
Physical Examination/Vital Signs [h]	X	X [b]		X	X	
Ophthalmic Examination[i]	X			X [i]	X	
ECOG Performance Status	X	X		X	X	
Laboratory Studies						
Hematology [j]	X	X [b]	X	X	X	
Blood Chemistry [j]	X	X [b]	X	X	X	
Pregnancy test, serum estradiol and FSH (if applicable) [j]	X					
12-Lead ECG (in triplicate)	X				X	
Disease Assessment						
CT/MRI Scans of Chest, Abdomen, Pelvis, any clinically indicated sites of disease, and of bone lesions; Clinical evaluation of superficial disease [k]	X	◄--► [k] Performed every 8 weeks (±7 days) for the first year, and then every 12 weeks (±7 days) from the date of randomization (See tumor assessment requirements flowchart)			X	X
Radionuclide Bone Scan, Whole Body [l]	X	As clinically indicated or to confirm complete response. (See tumor assessment requirements flowchart) [l]			X	X

Protocol Activity	**Screening**	**Active Treatment Phase [a] - One Cycle = 28 days**			**End of Treatment/Withdrawal [c]**	**Post-Treatment Follow-Up [d]**
		Cycles 1 and 2		**Cycles ≥3**		
Study Day	**Within 28 days prior to randomization unless specified otherwise**	**Day 1 [b]**	**Day 15**	**Day 1**		
Visit Window		**±-2 days**	**±2 days**	**±7 days [a]**		**±7 days**
Other Clinical Assessments						
Adverse Event Reporting [m]	X	X	X	X	X	X
Concomitant Medications/Treatments	◄--► Recorded from 28 days prior to the start of study treatment up to 28 days after the last dose of study treatment					
Pharmacokinetics (PK) [n]		First 40 patients: Sampling at pre-dose on Day 1 and Day 15 of Cycles 1 and 2, and Day 1 of Cycle 3; all other patients: Pre-dose sampling on Day 15 of Cycles 1 and 2				
CCI						
EuroQol-5D (EQ-5D) [r]		Pre-dose on Day 1 of Cycles 1, 2, 3, 4 and Day 1 of every other cycle thereafter starting with Cycle 6 (ie, Cycle 6,8, 10, etc)			X	
European Organization for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC-QLQ-C30) [r]					X	
European Organisation for Research and Treatment of Cancer Breast Cancer Module (EORTC-QLQ-BR23) [r]					X	
Survival Follow-up [s]						X
Study Treatment						
Randomization		X				
Fulvestrant (both treatment arms) [t]		◄--►[t] IM administration on Days 1 and 15 of Cycle 1, every 28 days (±7 days) thereafter starting from Day 1 of Cycle 1				
Palbociclib or placebo (Arm A only) [u]		◄--►[u] Orally once daily on Days 1 to 21of each Cycle followed by 7 days off treatment				
For pre-/peri-menopausal patients only: Goserelin (both treatment arms, if applicable) [v]	SC administration at least 4 weeks before study treatment start [v]	◄--►[v] SC administration every 28 days				

a. **Active Treatment Phase:** Assessments should be performed prior to dosing on the visit day unless otherwise indicated. Acceptable time windows for performing each assessment are described in the column headers. One cycle consists of 28 days. A cycle could be longer than 28 days if persistent toxicity delays initiation of the subsequent cycle. **Day 1 of any cycle visit should coincide with the day the palbociclib/placebo treatment begins.** If there are delays due to toxicity, then the start of the next cycle visit will be delayed until the patient has recovered and can begin study treatment again. Fulvestrant injection will be given every 28 days (+/- 7 days) with the exception of Cycle 1 during which it will be administered on Days 1 and 15 (±2 days allowed according to the protocol visit time windows). Goserelin will be administered every 28 days (+/- number of days allowed according to the protocol visit time windows). The active treatment phase is ongoing as long as the patient is receiving both study drugs (ie, palbociclib/placebo and fulvestrant) or fulvestrant alone.
b. **Cycle 1/Day 1:** Blood chemistry, hematology, and physical examination not required if acceptable screening assessment is performed within 7 days prior to randomization.
c. **End of Treatment/Withdrawal:** Visit to be performed as soon as possible but no later than 4 weeks from the last dose of investigational products and prior to initiation of any new anticancer therapy. Obtain assessments if not completed during the previous 4 weeks on study (or within the previous 8 weeks or 12 weeks [as applicable] for disease assessments).
d. **Post Treatment Follow-up:** Patients who discontinue study treatment should be contacted 28 calendar days (±7 days) after discontinuation of study treatment (palbociclib/placebo or fulvestrant) to assess if there have been any new adverse events and/or any change to any previously reported adverse events. Telephone contact is acceptable. Patients who discontinue active study treatment for any reason other than objective disease progression or death will continue to have tumor assessments performed every 8 weeks (±7 days) for the first year, and then after 1 year every 12 weeks (±7 days) (calculated from the date of randomization) until documented progression or onset of new anticancer therapy. See Tumor Assessment Requirements Flowchart for details. For patients who discontinue study treatment due to objective disease progression, see table footnote s (Survival Follow-up) below.
e. **Informed Consent:** Informed consent must be obtained prior to any protocol required assessments being performed (with the exception of certain imaging assessments if meeting the criteria defined in the Screening Section of the study protocol amendment 2).
f. **Medical/Oncological History:** To include information on prior anticancer treatments.
g. **Baseline Signs/Symptoms:** Baseline tumor related signs and symptoms will be recorded at the Cycle 1 Day 1 visit prior to initiating treatment and then reported as adverse events during the trial if they worsen in severity or increase in frequency.
h. **Physical Examination/Vital signs:** A full physical examination including an examination of all major body systems and breasts, height (at screening only), weight, blood pressure and pulse rate, which may be performed by a physician, registered nurse or other qualified health care provider. Physical examinations will be carried out at Screening, Day 1 of every cycle and the End of Treatment/Withdrawal visit.

i. **Ophthalmology Examinations:** Upon approval of Amendment 1, newly enrolled lens grading evaluable patients will undergo an ophthalmic examination by an ophthalmologist at screening, during study treatment on Cycle 4 Day 1, on Cycle 7 Day 1, on Cycle 13 Day 1 (ie, after 3, 6 and 12 months), every 12 months thereafter (ie, Days 1 of Cycles 25, 37, etc.) and at the End of Treatment/Withdrawal visit. Additional ophthalmic examinations may be performed as clinically indicated. It is expected that a minimum of 100 evaluable patients will participate in these examinations. Sites will be informed once these examinations are no longer required for patients newly enrolled in this study. Refer to the Ocular Safety Assessment Section of the study protocol amendment 2 for further details.

j. **Laboratory tests:** Hematology includes hemoglobin, WBC, absolute neutrophil count, platelet count. Blood chemistry includes AST/ALT, alkaline phosphatase, sodium, potassium, magnesium, total calcium, total bilirubin, blood urea nitrogen (BUN) (or urea), serum creatinine, and albumin. Additional hematology/chemistries panels may be performed as clinically indicated. Upon approval of Amendment 2, hemoglobin A1c will be measured in all patients every 3 months from the date of randomization (ie, C4D1, C7D1, C10D1, etc), and at the End of Treatment/Withdrawal visit. Pregnancy test (serum) at screening only for women of childbearing potential. Test may be repeated as per request of IRB/IECs or if required by local regulations. Serum estradiol and Follicle stimulating hormone (FSH) levels are analysed at screening to confirm postmenopausal status of women <60 years old and who have been amenorrheic for at least 12 consecutive months.

k. **CT/MRI Scans of Chest, Abdomen, Pelvis:** Refer to the tumor assessment requirement flowchart for details and timing of procedures.

l. **Radionuclide Bone Scan, Whole Body:** Refer to the tumor assessment requirement flowchart for all details and timing of procedures.

m. **Adverse Events (AEs):** Serious Adverse events (SAEs) must be reported from the time the patient provides informed consent through and including 28 calendar days after the last administration of the study drug. SAEs occurring after the active reporting period has ended should be reported if the investigator becomes aware of them; at a minimum, all SAEs that the investigator believes have at least a reasonable possibility of being related to study drug are to be reported to the Sponsor. All AEs (serious and non serious) should be recorded on the CRF from the first dose of study treatment through last patient visit. It is expected that telephone contact with the patient will be made in order to assess SAEs and AEs 28 calendar days (+/- 7 days) after the last administration of the study drug.

n. **Pharmacokinetics (PK)**: In approximately the first 40 patients randomized in the study, plasma PK samples will be drawn pre-dose on Day 1 and Day 15 of Cycles 1 and 2, and Day 1 of Cycle 3 for DDI assessment for palbociclib and fulvestrant (and goserelin if applicable). In all other patients, plasma concentrations will be drawn on Day 15 of Cycle 1 and Cycle 2 for palbociclib only. Additional PK blood samples may be collected from patients experiencing unexpected or serious adverse events, or adverse events that lead to discontinuation.

r. **Patient Reported Outcomes Assessments:** All self-assessment questionnaires must be completed by the patients while in the clinic and cannot be taken home. Interviewer administration in clinic may be used under special circumstances.

s. **Survival Follow-Up:** For patients who discontinue study treatment due to objective disease progression, survival data (ie, patient status along with start, stop and type of new anticancer therapy) will be collected every 3 months for the first 9 months (Month 3, 6, and 9, ±14 days), then every 6 months starting at Month 15 (±14 days), calculated from the last dose of study treatment. Telephone contact is acceptable.

t. **Fulvestrant:** To be administered on-site according to the local Summary of Product Characteristics for fulvestrant (Faslodex®). Fulvestrant injection will be given every 28 days (+/- 7 days) with the exception of Cycle 1 during which it will be administered on Days 1 and 15 (±2 days allowed according to the protocol visit time windows).

u. **Palbociclib or Placebo:** Patients will be required to return all bottles of palbociclib/placebo as well as the completed patient diary on Day 1 of each cycle for drug accountability.

v. **Goserelin (if applicable):** Goserelin will be administered every 28 days (+/- number of days allowed according to the protocol visit time windows). Treatment with goserelin (Zoladex® or generic) as per local practice for all women who are pre- or peri-menopausal at study entry. Patients must have commenced treatment with goserelin or an alternative luteinizing hormone-releasing hormone (LHRH) agonist at least 4 weeks prior to randomization. If patients have not received goserelin as their LHRH agonist prior to study entry, they must switch to goserelin for the duration of the trial. It is recommended to administer goserelin (given every 28 days) on-site when monthly fulvestrant is given. If goserelin is administered at home by the patient, a patient diary will be implemented.

Tumor Assessment Requirements Flowchart

Method	Screening [a]	Treatment Period[b]	End of Treatment Visit c
CT d or MRI of chest, abdomen, and pelvis	Required [e]	Required	Required
CT d or MRI of any other site of disease, as clinically indicated	Required [e,f]	Required for sites of disease identified at screening	Required for sites of disease identified at screening, unless disease progression has been confirmed elsewhere
Radionuclide bone scan (whole body) and confirmatory imaging in case of any hot spots (CT, MRI or X-ray)	Required [g,h]	As clinically indicated or to confirm complete response [i]	Required for sites of disease identified at screening, unless disease progression has been confirmed elsewhere
Photographs of all superficial lesions as applicable[j]	Required	Required for sites of disease identified at screening	Required for sites of disease identified at screening, unless disease progression has been confirmed elsewhere

a. Screening scans must occur within 4 weeks (ie, 28 days) prior to randomization unless otherwise specified.

b. Tumor assessment must be done every 8 weeks (±7 days) for the first year during the treatment period, then after 1 year (6 assessments) every 12 weeks (±7 days) from the date of randomization until radiographically and/or clinically (ie, for photographed or palpable lesions) documented PD as per the Response Evaluation Criteria in Solid Tumors (RECIST) v.1.1, (Protocol Appendix 4) study treatment discontinuation (for patients continuing treatment beyond RECIST-defined disease progression), initiation of new anticancer therapy or discontinuation of patient from overall study participation (eg, death, patient's request, lost to follow-up), whichever occurs first. The schedule of assessments should be fixed according to the calendar, regardless of treatment delays/interruptions. Imaging assessments are to be scheduled using the randomization date as the reference date for all time-points and are NOT to be scheduled based on the date of the previous imaging time-point. Imaging assessment delay to conform to treatment delay is not permitted. The same tumor assessment technique MUST be used throughout the study for a given lesion/patient.

c. Patients who have already demonstrated objective disease progression as per RECIST v.1.1 do not need to have scans repeated at the end of treatment visit or during the post-treatment follow-up. For patients who do not have documented objective disease progression at time of study treatment discontinuation, tumor assessment will continue to be performed every 8 weeks (±7 days) for the first year, then after 1 year (6 assessments) every 12 weeks (±7 days) , until radiographically and/or clinically confirmed objective disease progression, initiation of new anticancer therapy, or discontinuation of patient from overall study participation (eg, death, patient's request, lost to follow-up).

d. The CT scans, including brain CT scan if applicable, should be performed with contrast agents unless contraindicated for medical reasons. If IV contrast is medically contraindicated, the imaging modality to be used to follow the disease (either CT without contrast or MRI) should be the modality which best evaluates the disease, and the choice should be determined by the investigator in conjunction with the local radiologist. MRI of the abdomen and pelvis

can be substituted for CT if MRI adequately depicts the disease. However, MRI of the chest should not be substituted for CT of chest even if IV contrast is contraindicated. In such case CT will be performed without contrast. If MRI is used to follow-up bone lesion(s) it must be performed a few days before any treatment that may affect bone-marrow cellularity (eg, G-CSF).

e. Radiographic assessments obtained per the patient's standard of care prior to randomization into the study do not need to be repeated and are acceptable to use as baseline evaluations, if (1) obtained within 28 days before randomization, (2) they were performed using the method requirements outlined in RECIST v.1.1 (3) (Protocol Appendix 4) the same technique/modality can be used to follow identified lesions throughout the trial for a given patient, and (4) appropriate documentation indicating that these radiographic tumor assessments were performed as standard of care is available in the patient's source notes.

f. Baseline brain scans are only required if signs and symptoms suggest presence of metastatic brain disease. Post-baseline repeat brain scans will only be required only if metastases are suspected.

g. Bone scans will be carried out at baseline for all patients within 12 weeks prior to randomization in order to detect bony sites of disease. Bone scans performed before the signing of informed consent as routine procedures (but within 12 weeks before randomization) do not need to be repeated and may be used as baseline assessments as long as (1) tests were performed using the method requirements outlined in RECIST v.1.1 (2) appropriate documentation indicating that these radiographic tumor assessments were performed as standard of care is available in the patient's source notes.

h. Any suspicious abnormalities (ie, hotspots) identified on the bone scans at baseline and on subsequent bone scans MUST be confirmed by X-ray, CT scan with bone windows or MRI. The same modality must be used throughout the trial for confirmation for a given lesion/patient. Bone lesions identified at baseline will be followed up according to the same assessment schedule (ie, Week 8 ±7 days from randomization) as for all other lesions. Areas that have received palliative radiotherapy cannot be used to assess response to study treatment.

i. Bone lesions identified at baseline as the only sites of a patient's disease will be followed by X-ray/CT scan/MRI every 8 weeks (±7 days) for the first year, and then after 1 year every 12 weeks (±7 days) during the active treatment phase (calculated from the date of randomization) and at the time of CR confirmation. If no bone lesions were identified at baseline, or if bone is not the only site of a patient's disease, bone scans will only be repeated during the active treatment phase as clinically indicated (ie, patient describes new or worsening bone pain, or has increasing alkaline phosphatase level, or other signs and symptoms of new/progressing bone metastases) but are required at the time of CR confirmation. New abnormalities found on subsequent bone scans must also be confirmed by X-ray, CT scan with bone windows or MRI.

j. Clinical assessment of superficial disease must be carried out on the same date as the imaging studies and will include photographs of all superficial metastatic lesions. All lesion measurements must be recorded in the CRF.

Notes:

Radiographic tumor assessments may be done at any time if there is clinical suspicion of disease progression at the discretion of the investigator. If progressive disease is confirmed per RECIST v.1.1, patients are expected to discontinue study therapy and begin the follow-up phase of the trial. However, patients may continue treatment as assigned at randomization beyond the time of RECIST-defined PD at the discretion of the investigator if that is considered to be in the best interest of the patient and as long as no new anticancer treatment is initiated.

10.2. RECIST (Response Evaluation Criteria in Solid Tumors) version 1.1 Guidelines

Adapted from *E.A. Eisenhauer, et al: New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). European Journal of Cancer 45 (2009) 228–247*

CATEGORIZING LESIONS AT BASELINE

Measurable Lesions

- Lesions that can be accurately measured in at least one dimension.
- Lesions with longest diameter twice the slice thickness and at least 10 mm or greater when assessed by CT or MRI (slice thickness 5-8 mm).
- Lesions with longest diameter at least 20 mm when assessed by Chest X-ray.
- Superficial lesions with longest diameter 10 mm or greater when assessed by caliper.
- Malignant lymph nodes with the short axis 15 mm or greater when assessed by CT.

NOTE: The shortest axis is used as the diameter for malignant lymph nodes, longest axis for all other measurable lesions.

Non-measurable disease

Non-measurable disease includes lesions too small to be considered measurable (including nodes with short axis between 10 and 14.9 mm) and truly non-measurable disease such as pleural or pericardial effusions, ascites, inflammatory breast disease, leptomeningeal disease, lymphangitic involvement of skin or lung, clinical lesions that cannot be accurately measured with calipers, abdominal masses identified by physical exam that are not measurable by reproducible imaging techniques.

- Bone disease: Bone disease is non-measurable with the exception of soft tissue components that can be evaluated by CT or MRI and meet the definition of measurability at baseline.
- Previous local treatment: A previously irradiated lesion (or lesion subjected to other local treatment) is non-measurable unless it has progressed since completion of treatment.

Normal sites

- Cystic lesions: Simple cysts should not be considered as malignant lesions and should not be recorded either as target or non-target disease. Cystic lesions thought to represent cystic metastases can be measurable lesions, if they meet the specific definition above. If non-cystic lesions are also present, these are preferred as target lesions.

- Normal nodes: Nodes with short axis <10 mm are considered normal and should not be recorded or followed either as measurable or non-measurable disease.

Recording Tumor Assessments

All sites of disease must be assessed at baseline. Baseline assessments should be done as close as possible prior to study start. For an adequate baseline assessment, all required scans must be done within 28 days prior to randomization and all disease must be documented appropriately. If baseline assessment is inadequate, subsequent statuses generally should be indeterminate.

Target Lesions

All measurable lesions up to a maximum of 2 lesions per organ, 5 lesions in total, representative of all involved organs, should be identified as target lesions at baseline. Target lesions should be selected on the basis of size (longest lesions) and suitability for accurate repeated measurements. Record the longest diameter for each lesion, except in the case of pathological lymph nodes for which the short axis should be recorded. The sum of the diameters (longest for non-nodal lesions, short axis for nodal lesions) for all target lesions at baseline will be the basis for comparison to assessments performed on study.

- If two target lesions coalesce the measurement of the coalesced mass is used. If a large target lesion splits, the sum of the parts is used.

- Measurements for target lesions that become small should continue to be recorded. If a target lesion becomes too small to measure, 0 mm should be recorded if the lesion is considered to have disappeared; otherwise a default value of 5 mm should be recorded.

NOTE: When nodal lesions decrease to <10 mm (normal), the actual measurement should still be recorded.

Non-target Disease

All non-measurable disease is non-target. All measurable lesions not identified as target lesions are also included as non-target disease. Measurements are not required but rather assessments will be expressed as ABSENT, INDETERMINATE, PRESENT/NOT INCREASED, INCREASED. Multiple non-target lesions in one organ may be recorded as a single item on the case report form (eg, ‘multiple enlarged pelvic lymph nodes’ or ‘multiple liver metastases’).

OBJECTIVE RESPONSE STATUS AT EACH EVALUATION

Disease sites must be assessed using the same technique as baseline, including consistent administration of contrast and timing of scanning. If a change needs to be made the case must be discussed with the radiologist to determine if substitution is possible. If not, subsequent objective statuses are indeterminate.

Target Disease

- Complete Response (CR): Complete disappearance of all target lesions with the exception of nodal disease. All target nodes must decrease to normal size (short axis <10 mm). All target lesions must be assessed.

- Partial Response (PR): Greater than or equal to 30% decrease under baseline of the sum of diameters of all target measurable lesions. The short diameter is used in the sum for target nodes, while the longest diameter is used in the sum for all other target lesions. All target lesions must be assessed.

- Stable: Does not qualify for CR, PR or Progression. All target lesions must be assessed. Stable can follow PR only in the rare case that the sum increases by less than 20% from the nadir, but enough that a previously documented 30% decrease no longer holds.

- Objective Progression (PD): 20% increase in the sum of diameters of target measurable lesions above the smallest sum observed (over baseline if no decrease in the sum is observed during therapy), with a minimum absolute increase of 5 mm.

- Indeterminate. Progression has not been documented, and

    - One or more target measurable lesions have not been assessed; or

    - Assessment methods used were inconsistent with those used at baseline; or

    - One or more target lesions cannot be measured accurately (eg, poorly visible unless due to being too small to measure); or

    - One or more target lesions were excised or irradiated and have not reappeared or increased.

Non-target disease

- CR: Disappearance of all non-target lesions and normalization of tumor marker levels. All lymph nodes must be 'normal' in size (<10 mm short axis).

- Non-CR/Non-PD: Persistence of any non-target lesions and/or tumor marker level above the normal limits.

- PD: Unequivocal progression of pre-existing lesions. Generally the overall tumor burden must increase sufficiently to merit discontinuation of therapy. In the presence of SD or PR in target disease, progression due to unequivocal increase in non-target disease should be rare.

- Indeterminate: Progression has not been determined and one or more non-target sites were not assessed or assessment methods were inconsistent with those used at baseline.

New Lesions

The appearance of any new unequivocal malignant lesion indicates PD. If a new lesion is equivocal, for example due to its small size, continued assessment will clarify the etiology. If repeat assessments confirm the lesion, then progression should be recorded on the date of the initial assessment. A lesion identified in an area not previously scanned will be considered a new lesion.

Supplemental Investigations

- If CR determination depends on a residual lesion that decreased in size but did not disappear completely, it is recommended the residual lesion be investigated with biopsy or fine needle aspirate. If no disease is identified, objective status is CR.

- If progression determination depends on a lesion with an increase possibly due to necrosis, the lesion may be investigated with biopsy or fine needle aspirate to clarify status.

Subjective Progression

Patients requiring discontinuation of treatment without objective evidence of disease progression should not be reported as PD on tumor assessment CRFs. This should be indicated on the end of treatment CRF as off treatment due to Global Deterioration of Health Status. Every effort should be made to document objective progression even after discontinuation of treatment.

Table 1. Objective Response Status at each Evaluation

Target Lesions	Non-target Disease	New Lesions	Objective status
CR	CR	No	CR
CR	Non-CR/Non-PD	No	PR
CR	Indeterminate or Missing	No	PR
PR	Non-CR/Non-PD, Indeterminate, or Missing	No	PR
SD	Non-CR/Non-PD, Indeterminate, or Missing	No	Stable
Indeterminate or Missing	Non-PD	No	Indeterminate
PD	Any	Yes or No	PD
Any	PD	Yes or No	PD
Any	Any	Yes	PD

If the protocol allows enrollment of patients with only non-target disease, the following table will be used:

Table 2. Objective Response Status at each Evaluation for Patients with Non-Target Disease Only		
Non-target Disease	**New Lesions**	**Objective status**
CR	No	CR
Non-CR/Non-PD	No	Non-CR/Non-PD
Indeterminate	No	Indeterminate
Unequivocal progression	Yes or No	PD
Any	Yes	PD

Best Overall Response

The best overall response (BOR) is the best response recorded from the randomization until disease progression or death due to any cause. This is derived from the sequence of objective statuses. Objective statuses are not considered after objective progression is documented or after start of the first anticancer treatment post discontinuation of protocol treatment. BOR for each patient will be derived as one of the following categories.

- **Complete response (CR)**: At least one objective status of CR documented before progression.
- **Partial response (PR)**: At least one objective status of PR documented before progression.
- **Stable disease (SD)**: At least one objective status of stable documented at least 8 weeks after randomization date and before progression but not qualifying as CR, PR.
- **Progressive Disease (PD):** Objective status of progression within 16 weeks of randomization, not qualifying as CR, PR or SD.
- **Indeterminate (IND):** Progression not documented within 16 weeks after randomization and no other response category applies.

10.3. Rules for Determining PFS Status and Date

Situation	Date of Progression/Censoring[1]	Outcome
Inadequate baseline assessment	Randomization date (Day 1)	Censored
No on-study assessments	Randomization date (Day 1)	Censored
Alive and no Progression	Date of last objective tumor assessment documenting no progression	Censored
Progression Documented on or between scheduled tumor assessments	Date of first objective tumor assessment documenting objective progression	Progressed (Event)
Patients are removed from the study (withdrew the consent, lost to follow up, etc.) prior to progression or death	Date of last objective tumor assessment documenting no progression	Censored
New anticancer treatment prior to progression or death	Date of last objective tumor assessment documenting no progression prior to new anticancer treatment	Censored
Death prior to first planned tumor assessment	Date of death	Death (Event)
Death without objective progression prior to treatment discontinuation[2]	Date of death	Death (Event)
Death or progression after 2 or more missed tumor assessments	Date of last objective tumor assessment documenting no progression prior to the event	Censored

[1] For date of censorship, if a tumor assessment takes place over a number of days (eg, superficial lesions one day, scans another), the last date is used as the assessment date.

10.4. Data Derivation Details

Enrollment/Randomization	Date of assignment of the randomization number
Study Day 1	Randomization day
Treatment start	Day 1 of Cycle 1
Day 1 (cycle start date)	Day 1 of a cycle is every 28 days unless there is a dosing delay.
Cycle length (all but final cycle)	Cycle length is 28 days (previous cycle length may exceed planned length if there is a delay in study treatment administration).
Final cycle	For patients off treatment, from Day 1 of final cycle to 28 days after final dose or until start of new anticancer treatment (whichever comes first). For patients on treatment, from Day 1 of most recent cycle start to protocol specified cycle length.
Follow-up Period for AEs	From 28 days after final dose until start of new anticancer treatment (whichever comes first).
Baseline lab values	From date closest to, but prior to, start of study treatment.
Tumor assessment baseline values	From date closest but prior to first dose.
Measurable disease	Defined by RECIST
Adequate baseline tumor assessment	Within 35 (28 + 7) days prior to first dose or 28 days prior to the randomization date with the exception of bone scan (within 12 weeks prior to randomization). Maximum diameter reported for each target lesion listed. Each target lesion is measurable, unless bone only disease. All required pre-treatment scans done.
Cycle k treatment delayed.	If study treatment administration is delayed for cycle k then cycle k-1 is extended.

10.5. Study Treatment Modification and Compliance

10.5.1. Dose Modification

No dose adjustment for fulvestrant is permitted but dosing interruptions are allowed. Treatment interruption for fulvestrant-related toxicities will be performed as per the investigator's best medical judgment.

In the event of significant treatment-related toxicity, palbociclib/placebo dosing may be interrupted or delayed and/or reduced as described below.

- A **treatment delay** is defined as any delay of the cycle start date, based on the previous cycle's start date.
- A **dose reduction** is defined as a day when the actual dose taken is less than the initial prescribed dose for any reason with the exception that a day with total dose administered of 0mg is not considered a dose reduction.
- A **dose interruptions/missed dose** is defined as a planned dosing day with 0 mg administered.

10.5.2. Summarizing Relative Dose (RD) and Relative Dose Intensity (RDI)

The following types of summaries are proposed for administration of palbociclib (palbociclib) and fulvestrant with the proposed dosing schedule in the protocol, which is

Palbociclib 125 mg/day orally for 3 weeks followed by 1 week off plus fulvestrant 500 mg intramuscularly on Days 1 and 15 of Cycle 1, and then on Day 1 of each subsequent 28-day cycle:

- RDI for palbociclib: by Cycle and Overall
- RD for fulvestrant: by Cycle and Overall

Note: the denominator for tables summarizing "fulvestrant" will be all patients who took at least a dose of fulvestrant and for tables summarizing "palbociclib" will be all patients who took at least a dose of palbociclib

Examples for the summaries described in above are included in the tables below.

Conventions:

- The Intended Dose Intensity is the same for all cycles: the daily dose is fixed at the start of treatment rather than start of a cycle and the intended treatment duration is the same for the entire dosing period, including last cycle (eg, for a 3/1 dosing schedule, all cycles have an intended duration of 4 weeks);
- Actual Dose Intensity is calculated based on actual cycle length in all but last cycle where it is fixed at the intended length (eg, 4 weeks for a 3/1 dosing schedule)

Table 1

Treatment / Summary Type	Calculation of RD/RDI	Example
Cyclical palbociclib / Overall	$RD = \frac{Actual\ Total\ Dose}{Intended\ Total\ Dose} * 100\ \%$ *Actual Total Dose* = (Sum over all cycles of the "*Actual Total Dose per cycle"*) *Intended Total Dose* = (Intended Total Dose per cycle†) * (Total number of cycles per CRF) Note: Calculation of RD is optional † = Calculated based on prescribed dose at the beginning of the study	• palbociclib is to be dosed at 125 mg QD on a 3/1 Schedule • Actual palbociclib dosing: 3/1 in Cycle 1, 3/2 in Cycle 2 and 2/2 in Cycle 3 (with 7 days dose interruption from Day 8 to Day 14) Actual Total Dose = (125*7*3)*2 + (125*7*2) (same dosing in the first 2 cycles) = 7,000 mg Intended Total Dose= (125*7*3) * 3 (same dosing in all 3 cycles) =7,875 mg RD = (7,000 / 7,875) * 100% = 88.9%
	$RDI = \frac{Actual\ Overall\ Dose\ Intensity}{Intended\ Overall\ Dose\ Intensity} * 100\ \%$ Actual Overall Dose Intensity = (Sum over all cycles of the "*Actual Total Dose per cycle"*) / (Sum over all cycles of the *"Actual number of weeks in cycle"* ♣), where ♣ is calculated as presented in Table 2 *Intended Overall Dose Intensity* = Intended Dose Intensity (per week per cycle) - calculated as presented in Table 3	• palbociclib is to be dosed at 125 mg QD on a 3/1 Schedule • Actual palbociclib dosing: 3/1 in Cycle 1, 3/2 in Cycles 2 and 2/2 in Cycle 3 (with 7 days dose interruption from Day 8 to Day 14) Actual Overall Dose Intensity = (2,625 + 2,625 + 1,750) / (4 + 5 + 4) = 538.46 Intended Overall Dose Intensity=656.25 RDI = (538.46 / 656.25) *100% = 82.1%

Table 2

Treatment / Summary Type	Calculation of RD/RDI	Example
Cyclical palbociclib / By Cycle	$RDI = \frac{Actual\ Dose\ Intensity}{Intended\ Dose\ Intensity} * 100\%$ *Actual Dose Intensity (per week)* = (Actual Total Dose per cycle) / (Actual number of weeks in cycle) *Actual Number of Weeks in Cycle* = (Start date of next cycle – Start date of current cycle) / 7. *Intended Dose Intensity (per week)* = (Intended Total Dose per cycle) / (Intended number of weeks in cycle)	• palbociclib is to be dosed at 125 mg QD on a 3/1 Schedule • Actual palbociclib dosing: 3/1 in Cycle 1 and 3/2 in Cycle 2 and 2/2 in Cycle 3 (with 7 days dose interruption from Day 8 to Day 14) Intended Dose Intensity = (125*7*3) / 4 = 656.25 mg/wk <u>Cycle 1:</u> Actual Dose Intensity = (125*7*3) / 4 = 656.25 mg/wk RDI = (656.25/656.25) * 100% = 100% <u>Cycle 2:</u> Actual Dose Intensity = (125*7*3) / 5 = 525 mg/wk RDI = (525/656.25) * 100% = 80% <u>Cycle 3:</u> Actual Dose Intensity = (125*7*2) / 4 = 437.5 mg/wk RDI = (437.5/656.25) * 100% = 66.7%

Table 3

Treatment / Summary Type	Calculation of RD/RDI	Example
fulvestrant By Cycle & Overall	$RD(Cycle) = \frac{Actual\ Dose}{Intended\ Dose} * 100\%$ *Actual Dose* = Actual Dose received within a specific cycle *Intended Dose* = Intended Dose within a specific cycle $RD(Overall) = \frac{Actual\ Total\ Dose}{Intended\ Total\ Dose} * 100\%$ *Actual Total Dose* = Actual dose received over all cycles *Intended Total Dose = Intended dose (1000 mg for Cycle 1 and 500 mg for the rest cycles) overall cycles* IM administration on Days 1 and 15 of Cycles 1, and then on Day 1 of each subsequent Cycle	• Fulvestrant is to be IM administered on Days 1 and 15 of Cycles 1, and then on Day 1 of each subsequent Cycle • Actual fulvestrant dosing: D1 on Cycle 1; D1 in Cycle 2 • Intended Dose is 1000 mg in Cycle 1 • Actual Dose is 500 mg in Cycle 2 • Actual Total Dose is 1000 mg over 2 cycles. <u>Cycle 1:</u> Actual Dose = 500 mg Intended Dose = 1000 mg RD= (500 /1000) * 100% =50% <u>Cycle 2:</u> Actual Dose = 500 mg Intended Dose = 500 mg RD=(500/500)*100%=100% <u>Overall:</u> Actual Total Dose = 500 + 500 = 1000 mg Intended Total Dose = 1000 + 500 = 1500 mg RD= (1000 / 1500) *100%=66.7%

10.6. List of Abbreviation

AE	Adverse Event
ALT	Alanine Aminotransferases
AI	Aromatase Inhibitor
AST	Aspartate Aminotransferases
AT	As Treated
AUC	Area Under the Curve
BC	Breast Cancer
BUN	Blood Urea Nitrogen
CBR	Clinical Benefit Response
CCND1	Cyclin D1
CDK	Cyclin-Dependent Kinase
CDKN2A, $p16^{Ink4A}$	Cyclin-Dependent Kinase Inhibitor 2A
CI	Confidence Interval
C_{max}	Maximum Plasma Concentration
CMH	Cochran Mantel Haenszel
CR	Complete Response
CRF	Case Report Form
CSF	Colony-Stimulating Factors
CT	Computed Tomography
CTCAE	Common Terminology Criteria for Adverse Events
CYP	Cytochrome P-450
DNA	Deoxyribonucleic Acid
DR	Duration of Response
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
E-DMC	External Data Monitoring Committee
EDR	The early discrepancy rate
EQ-5D	Dimension Health State EuroQoL Score
ER	Estrogen Receptor
FFPE	Formalin Fixed Paraffin Embedded
G-CSF	Granulocyte Colony Stimulating Factor
HER	Human Epidermal Growth Factor Receptor
HR	Hazard Ratio
HR	Hormone Receptor
ICH	International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use
IEC	Independent Ethics Committee
IHC	Immunohistochemistry
IRB	Institutional Review Board
ITT	Intent-to-treat
LDR	The late discrepancy rate
MedDRA	Medical Dictionary for Regulatory Activities

MRI	Magnetic Resonance Imaging
NCI	National Cancer Institute
OR	Objective Response
ORR	Objective Response Rate
OS	Overall Survival
PD	Progressive Disease
PET	Positron Emission Tomography
PFS	Progression Free Survival
PK	Pharmacokinetic
PR	Partial Response or Progesterone Receptor (depending on context)
PR	The PR interval is measured from the beginning of the P wave to the beginning of the QRS complex.
PS	Performance Status
PRO	Patient Reported Outcome
PT	Prothrombin Time
QD	Quaque Die (once daily)
QRS	The QRS complex is a name for the combination of three of the graphical deflections seen on a typical electrocardiogram. The QRS complex reflects the rapid depolarization of the right and left ventricles.
QT	Time between the start of the Q wave and the end of the T wave in the heart's electrical cycle
QT_c	QT interval corrected for heart rate
QTcB	QT interval corrected for heart rate using Bazett's fomula
QTcF	QT interval corrected for heart rate using Fridericia's fomula
Rb	Retinoblastoma
RECIST	Response Evaluation Criteria in Solid Tumors
RNA	Ribonucleic Acid
RPSFTM	Rank-Preserving Structural Failure Time Model
RR	The interval between an R wave and the next R wave
SAE	Serious Adverse Event
SD	Stable Disease or Standard Deviation (depending on context)
WBC	White Blood Cell